MEDINFO 2019

Health and Wellbeing e-Networks for All

Proceedings
of the
17th World Congress
on Medical and Health
Informatics

Editors:
Lucila Ohno-Machado
Brigitte Séroussi

IMIA
IOS Press

Combining and integrating cross-institutional data remains a challenge for both researchers and those involved in patient care. Patient-generated data can contribute precious information to healthcare professionals by enabling monitoring under normal life conditions and also helping patients play a more active role in their own care.

This book presents the proceedings of MEDINFO 2019, the 17th World Congress on Medical and Health Informatics, held in Lyon, France, from 25 to 30 August 2019. The theme of this year's conference was 'Health and Wellbeing: E-Networks for All', stressing the increasing importance of networks in healthcare on the one hand, and the patient-centered perspective on the other. Over 1100 manuscripts were submitted to the conference and, after a thorough review process by at least three reviewers and assessment by a scientific program committee member, 285 papers and 296 posters were accepted, together with 47 podium abstracts, 7 demonstrations, 45 panels, 21 workshops and 9 tutorials. All accepted paper and poster contributions are included in these proceedings. The papers are grouped under four thematic tracks: interpreting health and biomedical data, supporting care delivery, enabling precision medicine and public health, and the human element in medical informatics. The posters are divided into the same four groups. The book presents an overview of state-of-the-art informatics projects from multiple regions of the world; it will be of interest to anyone working in the field of medical informatics.

ISBN 978-1-64368-002-6 (print)
ISBN 978-1-64368-003-3 (online)
ISSN 0926-9630 (print)
ISSN 1879-8365 (online)

MEDINFO 2019: HEALTH AND WELLBEING E-NETWORKS FOR ALL

Studies in Health Technology and Informatics

International health informatics is driven by developments in biomedical technologies and medical informatics research that are advancing in parallel and form one integrated world of information and communication media and result in massive amounts of health data. These components include genomics and precision medicine, machine learning, translational informatics, intelligent systems for clinicians and patients, mobile health applications, data-driven telecommunication and rehabilitative technology, sensors, intelligent home technology, EHR and patient-controlled data, and Internet of Things.

Studies in Health Technology and Informatics (HTI) series was started in 1990 in collaboration with EU programmes that preceded the Horizon 2020 to promote biomedical and health informatics research. It has developed into a highly visible global platform for the dissemination of original research in this field, containing more than 250 volumes of high-quality works from all over the world.

The international Editorial Board selects publications with relevance and quality for the field. All contributions to the volumes in the series are peer reviewed.

The HTI series is indexed by MEDLINE/PubMed; Web of Science: Conference Proceedings Citation Index – Science (CPCI-S) and Book Citation Index – Science (BKCI-S); Google Scholar; Scopus; EMCare.

Volume 264

Recently published in this series

ISSN 0926-9630 (print)
ISSN 1879-8365 (online)

MEDINFO 2019: Health and Wellbeing e-Networks for All

Proceedings of the 17th World Congress on Medical and Health Informatics

Edited by

Lucila Ohno-Machado

University of California San Diego Health, Department of Biomedical Informatics, Veterans Administration San Diego Healthcare System, Division of Health Services Research & Development, USA

and

Brigitte Séroussi

Sorbonne University, Laboratory of Medical Informatics and Knowledge Engineering for eHealth & Assistance Publique – Hôpitaux de Paris, Paris, France

IOS Press

Amsterdam • Berlin • Tokyo • Washington, DC

ISBN 978-1-64368-002-6 (print)
ISBN 978-1-64368-003-3 (online)
Library of Congress Control Number: 2019947555
doi: 10.3233/SHTI264

Publisher
IOS Press BV
Nieuwe Hemweg 6B
1013 BG Amsterdam
Netherlands
fax: +31 20 687 0019
e-mail: order@iospress.nl

For book sales in the USA and Canada:
IOS Press, Inc.
6751 Tepper Drive
Clifton, VA 20124
USA
Tel.: +1 703 830 6300
Fax: +1 703 830 2300
sales@iospress.com

PRINTED IN THE NETHERLANDS

From the Editorial Committee Chairs of MedInfo2019

Lucila Ohno-Machado[a,b], Brigitte Seroussi[c,d]

[a]UCSD Health Department of Biomedical Informatics, University of California San Diego, La Jolla, California, USA,
[b]Division of Health Services Research & Development, Veterans Administration San Diego Healthcare System, La Jolla, California, USA,
[c]Sorbonne Université, Université Paris 13, Sorbonne Paris Cité, INSERM, UMR S_1142, LIMICS, Paris, France,
[d]AP-HP, Hôpital Tenon, Département de santé publique, Paris, France.

The Proceedings of MEDINFO 2019 (the 17th World Congress of Medical and Health Informatics located in Lyon, France) illustrate how informatics scholars from all over the world are pursuing projects related to the theme of *Health and Wellbeing e-Networks for All*. Throughout this publication, readers will find innovative approaches to the collection, organization, analysis, and sharing of data and knowledge related to health and wellbeing. The articles in these proceedings not only document the state-of-the-art in our field worldwide, but also remind readers of how they can build on past discoveries and get motivated to pursue new paths in the future.

Every two years the medical and health informatics community assembles to discuss the latest findings in informatics, to meet old friends, and to make new ones. New generations come together to showcase their research, hear from their peers as well as from those who contribute to their training, and start to pave their way to be our future leaders. Those of us who have been attending these meetings for a long time continue to be amazed by the energy of newcomers, the wisdom and determination of informatics pioneers, the drive of foundational and practice-oriented informaticians, and the incredible amount of work that it takes to organize this conference.

These Proceedings are a small but critical part of what MEDINFO is all about. They feature articles and abstracts describing research, training, and service functions that informaticians all over the world are designing, implementing, and evaluating. They represent the wide range of informatics development in different regions. The way these Proceedings are edited is emblematic of our field: the collaboration needed to produce a top-quality publication, even in the absence of a dedicated staff or a large budget, is only possible because reaching our destination is more important than the few barriers that present themselves along the way. The Proceedings are the output of a large group of volunteers who receive a small token of appreciation and are rewarded by knowing that, because of their contributions, authors disseminate their work beyond regional boundaries and readers within and outside our field can learn and reuse tools, data, and knowledge.

It was our great pleasure and honor to work with Associate Editors Todd Lingren and Scott McGrath to produce these Proceedings. They coordinated a large group of informatics trainees from multiple institutions (Table 1) in order to ensure that the contents were easy to read and formatted uniformly. We thank the whole editorial team as well as the Scientific Program team for making the editing role a fun and enjoyable one. And we thank our readers for understanding the difficulties in editing a large number of accepted papers for clarity, grammar, and style in such a short amount of time.

These Proceedings feature state-of-the-art informatics projects from multiple regions of the world. Enjoy learning how informatics is changing the way we approach health and wellbeing for all.

Lucila Ohno-Machado, MD, PhD
Brigitte Seroussi, MD, PhD

Co-Chairs, MedInfo2019 Editorial Committee

Table 1– Editorial Committee Members

Marcy Antonio	University of Arizona, USA
Jacqueline Brixey	University of Southern California, USA
Melissa M Das	MaineHealth, USA
Smruti Deoghare	University of Cincinnati, USA
Grace Gao	University of Minnesota, USA
Mattias Georgsson	Blekinge Institute of Technology, Sweden
Zhe He	Florida State University, USA
Renee Marie Hendricks	Veterans Health Administration, USA
Felix Holl	Ulm University, Germany
Kate Fultz Hollis	Oregon Health & Science University, USA
Hyunggu Jung	University of Washington, USA
Tian Kang	Columbia University, USA
Emily Kawaler	New York University, USA
David Leander	Dartmouth Geisel School of Medicine, USA
Tiffany I. Leung	Maastricht University, The Netherlands
Lacey Lewis	Veterans Health Administration, USA
Elizabeth A. Lindemann	University of Minnesota, USA
Todd Lingren	University of Cincinnati, USA
Satish M. Mahajan	Veterans Health Administration, USA
Vincent Major	New York University, USA
Scott McGrath	University of Nebraska Omaha, USA
Melanie Meyer	University of Massachusetts, USA
Sean Mikles	University of North Carolina Chapel Hill, USA
Elliot Mitchell	Columbia University, USA
Vickie Nguyen	University of Texas Health Science Center at Houston, USA
Florence Odekunle	Rutgers University, USA
Adebowale Ojo	Center for Disease Control and Prevention, USA
Lisiane Pruinelli	University of Minnesota, USA
Satyajeet Raje	IBM, USA
Doug Redd	George Washington University, USA
Lincoln Sheets	University of Missouri, USA
Bryan Steitz	Vanderbilt University, USA
Vignesh Subbian	University of Arizona, USA
Lina Sulieman	Vanderbilt University, USA
Andreas Triantafyllidis	Centre for Research and Technology Hellas (CERTH), Greece
Elizabeth Umberfield	University of Michigan, USA
Jacob P. VanHouten	Vanderbilt University Medical Center, USA
Lois Walters-Threat	American Nurses Credentialing Center, USA
Wei Wei	University of Pittsburgh Medical Center, USA
Rafeek Adeyemi Yusuf	University of Texas Health Science Center at Houston, USA
Ling Zheng	Monmouth University, USA

From the Scientific Program Committee Chairs of MedInfo2019

Olivier Bodenreider[a], Michael Marschollek[b]

[a]National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA
[b]Peter L. Reichertz Institute for Medical Informatics, TU Braunschweig and Hannover Medical School, Hannover, Germany

The 17th World Congress of Medical and Health Informatics, MedInfo2019, was held in Lyon, France, from August 25th to 30th, 2019. The conference was hosted by the International Medical Informatics Association (IMIA) and the French Association of Medical Informatics (AIM).

MedInfo2019

The overarching theme of MedInfo2019 was *Health and Wellbeing: E-Networks for all*, stressing the increasing importance of networks in healthcare on the one hand, and the patient-centered perspective on the other hand. Combining and integrating cross-institutional data remains a challenge for both patient care and research. Patient-generated data, e.g. originating from mobile apps, sensor-based wearables, and smart home environments, contribute precious information to health care professionals. Not only do these data enable rigorous patient monitoring under normal life conditions, but they also help patients play a more active role in the care process.

The MedInfo2019 Scientific Program Committee (SPC) called for submissions in four thematic tracks:

1. *Interpreting health and biomedical data*,
2. *Supporting care delivery*,
3. *Enabling precision medicine and public health*, and
4. *The human element in medical informatics*.

We received over 1100 submissions from 62 countries across all IMIA regions – the largest number of submissions in the more recent history of MedInfo conferences. With support from four track chairs, 51 track members, and 990 active reviewers, we have conducted a thorough review process. Virtually all submissions were reviewed by at least three reviewers and assessed by one SPC track member. Based on these recommendations, final decisions were made by the SPC track chairs and the SPC co-chairs during a three-day face-to-face meeting in Paris, France. Finally, 285 full papers/student papers, 47 podium abstracts, 296 posters, seven demonstrations, 45 panels, 21 workshops and nine tutorials were accepted. All IMIA regions were well represented in the final program. All accepted paper and poster contributions are included in these proceedings.

MedInfo2019 included five keynotes by internationally recognized experts in medical informatics:

- *Artificial Intelligence in Medicine* by Tze-Yun Leong from Singapore,
- *Capacity Building in Developing Countries* by Binyam Tilahun from Ethiopia,
- *Drawing Reproducible Conclusions from Observational Clinical Data with OHDSI* by George Hripcsak from the United States,
- *From Biological Genotype to Digital Phenotype* by Jacques Demongeot from France, and
- *European Vision of Medical Informatics: The Role of the European Institute of Innovation and Technology* by Jean-Marc Bourez from France.

We thank the SPC track chairs, the SPC track members, and the reviewers for their invaluable contribution to MedInfo2019.

International Medical Informatics Association

IMIA plays a major global role in the application of information science and technology in the fields of healthcare and research in medical, health, and bio-informatics. Established in 1967, it continues to advance and nurture international cooperation in these fields, to stimulate education, research and applications in close collaboration with the World Health Organization (WHO).

Sincerely,

Olivier Bodenreider and Michael Marschollek,
Co-Chairs, MedInfo2019 Scientific Program Committee

Address for correspondence

Olivier Bodenreider, obodenreider@mail.nih.gov
Michael Marschollek, michael.marschollek@plri.de

This page intentionally left blank

Contents

Part 2

PAPERS

This page intentionally left blank

I. Interpreting Health and Biomedical Data

This page intentionally left blank

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190172

The MeSH-Gram Neural Network Model: Extending Word Embedding Vectors with MeSH Concepts for Semantic Similarity

Saïd Abdeddaïm[a], Sylvestre Vimard[a], Lina F. Soualmia[a,b]

[a] Normandie Univ., UNIROUEN, UNIHAVRE, INSA Rouen, LITIS, F-76000, Rouen, France,
[b] Sorbonne Univ., Inserm, LIMICS, F-75006 Paris, France

Abstract

Eliciting semantic similarity between concepts remains a challenging task. Recent approaches founded on embedding vectors have gained in popularity as they have risen to efficiently capture semantic relationships. The underlying idea is that two words that have close meaning gather similar contexts. In this study, we propose a new neural network model, named MeSH-gram, which relies on a straightforward approach that extends the skip-gram neural network model by considering MeSH (Medical Subject Headings) descriptors instead of words. Trained on publicly available PubMed/MEDLINE corpus, MeSH-gram is evaluated on reference standards manually annotated for semantic similarity. MeSH-gram is first compared to skip-gram with vectors of size 300 and at several windows' contexts. A deeper comparison is performed with twenty existing models. All the obtained results with Spearman's rank correlations between human scores and computed similarities show that MeSH-gram (i) outperforms the skip-gram model and (ii) is comparable to the best methods that need more computation and external resources.

Keywords:

Medical subject headings, Neural network models, Unified medical language system.

Introduction

Eliciting semantic similarity and relatedness between concepts is a major issue in the biomedical domain. Different measures have been proposed the last decades [1]. Those measures quantify the degree to which two concepts are similar. They rely either on knowledge-based approaches using ontologies and terminologies, or on corpus-based approaches, which are founded on distributional statistics (e.g. literature-based drug discovery) [2-5]. Several clinical applications of importance rely on semantic similarity and relatedness [6], such as biomedical information extraction and retrieval, clinical decision support, or disease prediction. For instance, biomedical information extraction and retrieval are improved by including semantically related terms and concepts [7-10].

The recent approaches that have given better results in semantic similarity and relatedness measures are founded on word embedding vectors computed by neural networks. Indeed, such architectures implemented initially by word2vec [11], have gained in popularity in the biomedical domain as they risen to efficiently capture semantic similarity and relatedness relationships between words and concepts [12-17]. Word embeddings is based on neural network language modeling where words are mapped to fixed-dimension vectors of real numbers. The similarity between words can thus be measured by the (cosine) similarity between vectors that are constructed over a training corpus. All co-occurrences of a word and its neighbors (*i.e.* contexts) within a predefined window size are considered. The idea behind those representation learning approaches is that two words that have close meaning generally have similar contexts [18]. For example, the words *epilepsy* and *convulsion* will both have *brain* and *mind* as neighbors.

The word2vec developed by Mikolov et al. [11] is a neural network language model that learns word vectors that either maximizes the probability of a word given the surrounding context, referred to as the CBOW (Continuous Bag Of Words) approach, or to maximize the probability of the context given a word, referred to as the skip-gram approach.

In this study we propose a new method, named MeSH-gram, which relies on a straightforward approach: it computes the word vectors by only using the MeSH (Medical Subject Headings) descriptors that are already included in the PubMed/MEDLINE corpus. The MeSH-gram model extends the skip-gram neural network model used in word2vec [11] and fastText tools [19]. The fastText is a successful re-implementation of word2vec which is designed to compute the vector of each word using its neighbors. The extension we propose in the MeSH-gram model replaces the neighbors by the MeSH descriptors of the abstract where each word occurs.

Related Works

Several semantic similarity and relatedness measures have been proposed the last decades [17]. Many of them have been implemented in the UMLS::Similarity package [20] available in the UMLS (Unified Medical Language System). They differ on the method used: path-based, content-based, UMLS-based, corpus-based, and more recently, methods based on word vectors and concepts vectors. Path-based measures [7] use the hierarchical structure of a taxonomy to measure similarity: concepts close to each other are more similar. For instance, Sajadi et al. [21, 22] developed a ranking algorithm based on Wikipedia graph metrics and used it to compare biomedical concepts. Content-based information measures [23, 24] quantify the amount of information a concept provides: the more specific concepts have a greater amount of information content. Other approaches [25, 26] use the entire UMLS Metathesaurus® [27] in order to compare the context in the definition of the concept to quantify its relatedness.

Several methods are vector-based: the concepts are represented by vectors, and the relatedness is usually estimated using the cosine similarity between them. In [26], the authors proposed to compute gloss vectors based on second order co–occurrences trained on WordNet. In [28], the authors computed the cosine

of two Latent Semantic Indexing concept vectors based on Pointwise Mutual Information association measure matrix. Recent vector-based methods use neural networks in order to compute concept vectors. The word2vec tool [11] was trained on different corpora: OSHUMED by Sajadi et al. [22]; PubMed/MEDLINE by Chui et al. [13]; PubMed Central by Muneeb et al. [12], Chiu et al. [13], and Pakhomov et al. [14]; and CLINICAL-ALL by [14]. Following the approach used by De Vine et al. [29] on OSHUMED, Yu et al. [15] trained word2vec on PubMed/MEDLINE transformed into UMLS concepts using the MetaMap indexing tool [30].

Other recent methods rely on word vectors. In their previous work, Yu et al. [31] retrofitted word vectors obtained by word2vec with hierarchical information from the MeSH thesaurus. Recently, Henry et al. [16] compared different ways to combine word vectors in order to compute multi-word term vectors. The compared multi-word term aggregation method consists in the summation (averaging) of component word vectors, creating concept vectors using the MetaMap indexing tool [30], and creating multi-word term vectors using the compoundify tool based on the UMLS Specialist Lexicon as glossary [27]. More recently, Henry et al. [17] used association measures for estimating semantic similarity and relatedness between biomedical concepts on PubMed/MEDLINE transformed into UMLS concepts. The best performance results were obtained by [15-17]. Their respective approach relies either on MetaMap in order to transform the text corpus into UMLS concepts, or on additional external resources such as the Specialist Lexicon.

The MeSH-gram model we propose in this study relies on a straightforward approach: it computes the word vectors by only using the MeSH descriptors that are already included in the PubMed/MEDLINE corpus. The extension we propose in the MeSH-gram model replaces the neighbors by the MeSH descriptors of the abstract where each word occurs.

In order to evaluate MeSH-gram, we use publicly available manually annotated corpora: two subsets from Mayo Clinic (MiniMayoSRS) of the MayoSRS (Mayo Semantic Relatedness Set) developed by Pakhomov et al. [32], and two from UMNSRS (The University of Minnesota Semantic Relatedness Set) developed by Pakhomov et al. [33]. MeSH-gram results are first compared to skip-gram and are then compared to twenty existing solutions reported in [17], including the best ones [15-17]. The MeSH-gram model has several advantages: (i) it avoids considering uninformative and too frequent words; (ii) there are less MeSH descriptors than possible context words; and (iii) MeSH descriptors are manually assigned and curated, which assures the best quality of indexing.

Methods

Neural network language models learn word vectors by either maximizing the probability of a word given the context, referred to as the CBOW approach, or by maximizing the probability of the context given a word, referred to as the skip-gram approach.

Skip-gram Word Embedding Model

Given $w_1\ w_2\ ...\ w_n$ a text line of words w_i, the skip-gram model maximizes the following average log probability:

$$\frac{1}{2r}\sum_{i=1}^{2r}\sum_{-r\le j\le r,\, j\neq 0} log\, p\left(w_{t+j}\middle|w_t\right)$$

where w_t is the target word, w_{t+j} is the context, and r is the context window radius. The context words surrounding the target term are determined by the context window radius r.

The probability of a context word w_c given a target word w_t, is computed by:

$$p\left(w_c\middle|w_t\right)=\frac{exp\left(V_{w_c}^T V_{w_t}\right)}{\sum_{w=1}^{N} exp\left(V_w^T V_{w_t}\right)}$$

where N is the vocabulary size, and V_w represents the vector of the word w.

MeSH-gram word embedding model

The MeSH-gram word-embedding model proposed in this paper extends the skip-gram neural network model used in word2vec [11] and fastText [19] tools: it uses MeSH descriptors that are already included in the PubMed/MEDLINE corpus to compute the word vectors.

Given $w_1\ w_2\ ...\ w_n$ the words of a PubMed/MEDLINE abstract, and $m_1\ m_2\ ...\ m_k$ the MeSH descriptors associated to this abstract, the MeSH-gram model maximizes the following average log probability:

$$\frac{1}{k}\sum_{i=1}^{k} log\, p\left(m_i\middle|w_t\right)$$

where w_t is the target word and m_i is a MeSH descriptor.

The probability of a context MeSH descriptor m_c given a target word w_t, is computed by:

$$p\left(m_c\middle|w_t\right)=\frac{exp\left(V_{m_c}^T V_{w_t}\right)}{\sum_{m=1}^{M} exp\left(V_m^T V_{w_t}\right)}$$

where M is the number of MeSH descriptors, V_m represents the vector of the MeSH descriptor m, and V_w the vector of the word w.

We have adapted fastText [19] in order to feed the neural network with pairs of (word, MeSH descriptor). For each abstract included in PubMed/MEDLINE, every word occurrence in the abstract text is associated to each MeSH descriptor, which means that each word vector reflects all the MeSH descriptors seen by its word occurrences in all the PubMed/MEDLINE abstracts.

Vector Representation and Similarity Computation

Using our MeSH-gram model and skip-gram model for comparison, we built word vectors of dimension 300. For the skip-gram model, we computed the vectors considering several window sizes W of 2, 5, 10 and 25.

In order to quantify the relatedness of a pair of words, the cosine distance between the distributional context vectors of each word is used. In the case of a multi-word term, the vector is generated by computing the average of the component word vectors that compose the term. As an example, for the term *epilepsy attack*, the vector $V_{epilepsy\_attack}$ will be computed as $V_{epilepsy\_attack}=(V_{epilepsy}+V_{attack})/2$ where $V_{epilepsy}$ and V_{attack} represent the vector of each word *epilepsy* and *attack* respectively. Rather than combining word vectors after construction, multi-word term vectors may be constructed directly from a preprocessed training corpus in which multi-word terms have been identified [17]. Otherwise, this will involve huge cost in preprocessing and storage requirements.

Training Corpus

We used the PubMed/MEDLINE[1] corpus that contains the abstracts of each article and the associated MeSH descriptors. The corpus was parsed with *pubmed_parser*[2], a python XML parser for PubMed dataset. Each abstract was tokenized using *polyglot*[3].

As fastText needs all the data integrated into one file, we have concatenated all the tokenized PubMed/MEDLINE abstracts. Each line of the resulting file consists of an abstract with its MeSH descriptors. We have adapted fastText in order to feed the neural network with all pairs of (word, MeSH descriptor) of each file line.

Gold Standard

In order to compare the MeSH-gram word embedding model proposed in this study with other methods, we used two evaluation benchmarks: MiniMayoSRS [32] and UMNSRS [33]. MiniMayoSRS consists of 29 clinical term pairs. Two thirty pairs (66.67%) contain a multi-word term. The relatedness of each word pair is rated by medical coders and also by physicians. UMNSRS consists of 566 and 586 pairs of medical terms, for measuring similarity and relatedness respectively. The degree of association between terms in each dataset was rated by four medical residents from the University of Minnesota medical school. As suggested by Pakhomov et al. [33], we use a subset of the ratings consisting of 401 pairs for the similarity set and 430 pairs for the relatedness set. Twenty (4.99%) and seventeen (3.95%) of the term pairs contain multi-word terms for the similarity and relatedness subsets respectively. All these clinical terms correspond to UMLS concepts included in the Metathesaurus®.

The correlations between the generated relatedness scores and the human-assigned scores are calculated using Spearman's rank.

Results

Skip-gram Model versus MeSH-gram Model

The results of the experiments are in Table 1 in which a comparison is performed between the results obtained with skip-gram model and those obtained by the MeSH-gram model using our modified version of fastText according to the four gold standards: MiniMayoSRS rated by physicians (MiniMayoSRS phys.), MiniMayoSRS rated by medical coders (MiniMayoSRS cod.), UMNSRS for similarity (UMNSRS Sim.) and UMNSRS for relatedness (UMNSRS Rel.). The number of calculated term pairs (*n*) is lower than the number of pairs in the UMNSRS gold standards (*Sim.* and *Rel.*) because the embedding vectors could not be computed for low frequency terms and contexts on PubMed/MEDLINE using fastText. Using descriptors rather than words as context allows to computed more term pairs.

MeSH-gram Model compared to Previous Works

Table 2 gathers the results obtained by the MeSH-gram model we developed and twenty previous works' results. It allows a comparison between all the models and on the same gold standards (MiniMayoSRS and UMNSRS). Table 2 complete the Table 12 given by Henry et al. [17].

Table 1 – Spearman's rank correlations between human scores and computed similarities

	MiniMayo		UMNSRS	
	Phys. n=29	**Cod. n=29**	**Sim. n=380**	**Rel. n=397**
Skip-gram				
W=2	0.740	0.757	0.679	0.529
W=5	0.763	0.779	0.704	0.576
W=10	0.776	0.789	0.716	0.589
W=25	0.766	0.781	0.718	0.608
MeSH-gram	**0.811**	**0.855**	**0.724*** *n=387	**0.643**** **n=407

Note: W: window size; n: number of pairs

Discussion

As one can see in Table 1 for the skip-gram model, the more the window is extended, the more the results are improved on the UMNSRS gold standard. The best results of skip-gram are obtained with a window size *W*=10 for the MiniMayoSRS set. This suggests that word vectors are a better solution when we consider an important number of context words in the abstract. The best results are obtained with the MeSH-gram model that considers MeSH descriptors as context for each term, suggesting that MeSH descriptors catch the semantics of all the abstracts associated with it. We can conclude that taking MeSH descriptors instead of context words gives better results than considering a large window size: (i) bigger window size does not lead necessary to better results and (ii) MeSH descriptors are fewer than context words (50 context words for window size *W*=25) leading also a reduced computation time.

For example, the pair *synthroid* and *hypothyroidism* is misclassified by the skip-gram model but better treated by the MeSH-gram model. They are considered as very related in UMNSRS as *Synthroid*® (*levothyroxime sodium*) is used to treat *hypothyroidism*. This term pair has a human-assigned score of 1473, which corresponds to the seventh most related mono-term pair in UMNRRS. While the skip-gram model (with *W*=2) ranks this pair at the 160th position with a similarity score of 0.40, the MeSH-gram model puts it at the 64th position with a similarity score of 0.58, which means that the MeSH-gram model better captures the relatedness for this example. The apparent reason for this could be that *synthroid* occurs only 107 times in PubMed/MEDLINE corpus, which is insufficient to construct reliable vectors using context words in the skip-gram model, while the MeSH-gram model computes its vectors using the more informative MeSH descriptors. The absolute difference between the skip-gram rank (160) and the MeSH-gram rank (64) is 96 for this term pair, however it is lower in average (36.5 with a standard deviation of 33.5) when we consider the scores obtained by the two methods for all the UMNSRS pairs.

On the contrary, the pair of the terms *weakness* and *emaciation* (low frequency term equals to 1433) are better scored by the skip-gram model (84th position) than the MeSH-gram model (165th position), while they are considered closely related by human scores (64th position).

From our observations, the better results obtained by the MeSH-gram model are not due to significant improvements for any specific category of term pairs (e.g. less frequent terms), but to the overall improvement (moderate increase or decrease)

[1] ftp://ftp.ncbi.nlm.nih.gov/pubmed/ [accessed Apr 1st, 2019]
[2] https://github.com/titipata/pubmed_parser [accessed Apr 1st, 2019]
[3] https://github.com/aboSamoor/polyglot [accessed Apr 1st, 2019]

of the similarity scores for each pair. We can only conclude from this that MeSH descriptors are globally more informative than context words.

The comparison with twenty methods displayed in Table 2 confirms that the MeSH-gram model gives comparable results with best previous work methods on the four gold standard datasets. While the methods (1) and (2) rely on the translation of PubMed/MEDLINE text data into ULMS concepts, and methods (4) and (5) require additional steps or resources such as compoundify tool (4) and MetaMapped MEDLINE corpus (5), the MeSH-gram model uses only the raw text corpus as input. The best previous works' results are obtained by the method (2) and then the method (3). However, the method (2) is not recommended by the authors themselves as it uses concept expansion, which requires additional computation cost without significantly increasing the performances for any dataset [17]. MeSH-gram is comparable to method (3) with better results on three datasets. All those results allow us to conclude that UMLS information used by the methods (1) to (5) is already contained in the MeSH descriptors available in the PubMed/MEDLINE corpus and used by the MeSH-gram model. Using MeSH descriptors as context is a good solution for datasets founded on UMLS concepts. However, MeSH-gram should be evaluated on other types of similarities such as BioSimVerb and BioSimLex [34].

Table 2 – Spearman's rank correlations between human scores and computed similarities using MeSH-gram and previous works' methods.

	MiniMayo		**UMNSRS**	
	Phys.	**Cod.**	**Sim.**	**Rel.**
MeSH-gram	0.81 (n=29)	**0.86** (n=29)	***0.72*** (n=387)	***0.64*** (n=407)
(1) Henry et al. [17]; recommended	**0.84** (n=29)	0.81 (n=29)	0.69 (n=392)	***0.64*** (n=418)
(2) Henry et al. [17]; not recommended	*0.85 (n=29)*	*0.84 (n=29)*	*0.73 (n=392)*	*0.66 (n=418)*
(3) Henry et al. [16]; CBOW words	***0.82*** (n=29)	0.82 (n=29)	0.69 (n=374)	0.61 (n=396)
(4) Henry et al. [16]; CBOW compounds	0.80 (n=29)	0.78 (n=28)	0.70 (n=373)	**0.65** (n=393)
(5) Henry et al. [16]; CBOW concepts	0.77 (n=29)	***0.83*** (n=29)	**0.73** (n=388)	0.60 (n=413)
(6) Yu et al. [15]; narrow +other relations	--	--	0.69 (n=526)	0.62 (n=543)
			0.68 (n=418)	0.63 (n=427)
(7) Yu et al. [15]; no lexicons	--	--	0.64 (n=526)	0.59 (n=543)
			0.63 (n=418)	0.59 (n=427)
(8) Yu et al. [31]	0.70 (n=25)	0.67 (n=25)	--	--
(9) Sajadi et al. [22]; HITS similarity	0.67 (n=29)	0.72 (n=29)	0.58 (n=566)	0.51 (n=587)
(10) Sajadi et al. [22]; (word2vec OSHUMED+UMLS)	--	--	0.39 (n=566)	0.39 (n=587)
(11) Sajadi et al. [22] (word2vec on OSHUMED)	--	--	0.26 (n=566)	0.29 (n=587)
(12) Chui et al. [13]	--	--	0.65 (n=n/a)	0.60 (n=n/a)
(13) Pakhomov et al. [14]	--	--	0.62 (n=449)	0.58 (n=458)
(14) Muneeb et al. [15]	--	--	0.52 (n=462)	0.45 (n=465)
(15) Workman et al. [22]	0.67 (n=29)	--	--	--
	0.69 (n=25)			
(16) Patawardhan and Pedersen [26]	0.59 (n=29)	0.58 (n=29)	0.58 (n=387)	0.45 (n=412)
(17) Lin [24]	0.42 (n=26)	0.53 (n=26)	0.49 (n=340)	0.29 (n=360)
(18) Resnik [23]	0.34 (n=26)	0.46 (n=26)	0.49 (n=340)	0.26 (n=360)
(19) Rada et al. [7]	0.35 (n=26)	0.44 (n=26)	0.53 (n=340)	0.29 (n=360)
(20) Lesk [25]	0.52 (n=29)	0.57 (n=29)	0.50 (n=387)	0.33 (n=412)

Note: n: number of pairs (inspired by [17]). The best result is highlighted in bold, the second best in italic and bold.

Conclusions

In this paper, we proposed a new method, MeSH-gram, to create distributional word vectors using MeSH descriptors as word context. We evaluated our results on four standard evaluation datasets, MiniMayoSRS Physicians, MiniMayoSRS Coders, UMNSRS tagged for relatedness, and UMNSRS tagged for similarity, and compared it against skip-gram model as a baseline and previous methods. All the obtained results of Spearman's rank correlations between human scores and computed similarities show that MeSH-gram (i) outperforms the skip-gram model and (ii) is comparable to the best recent methods, methods that need more computation and additional external resources.

We are trying different ways to combine the skip-gram model and the MeSH-gram model in order to improve the results. We also plan in our future works to include in MeSH-gram the MeSH qualifiers affiliated to the descriptors in order to have a more precise semantic meaning (e.g. the association *cancer/complications*, where *cancer* is a MeSH descriptor and *complications* is a MeSH qualifier, is more precise than *cancer* alone). A second step is to use fastText subwords and the evaluation of MeSH-gram for other kinds of similarities such as BioSimVerb and BioSimLex. MeSH-gram may also be used in other languages than English, for instance in French bibliographic corpora such as CISMeF [35], as well as in annotated electronic health records.

References

[1] C. Lofi, Measuring semantic similarity and relatedness with distributional and knowledge-based approaches, *Information and Media Technologies* **10** (3)(2015), 493-501.

[2] M. Yetisgen-Yildiz and W. Pratt, Using statistical and knowledge-based approaches for literature-based discovery. *J Biomed Inform* **39**(6)(2006), 600–611.

[3] P. Agarwal and D.B Searls, Can literature analysis identify innovation drivers in drug discovery? *Nat Rev Drug Discov* **8** (11)(2009), 865–878.
[4] F. Doshi-Velez et al. Graph-sparse LDA: a Topic Model with Structured Sparsity, In *Proc. 29th AAAI Conference on Artificial Intelligence* (2015), 2575–2581.
[5] M.R. Nelson, H. Tipney, J.L. Painter, J. Shen, P. Nicoletti, Y. Shen, et al., The support of human genetic evidence for approved drug indications, *Nat Genet* **47** (8)(2015), 856–860.
[6] B.T. McInnes and T. Pedersen, Evaluating semantic similarity and relatedness over the semantic grouping of clinical term pairs, *J Biomed Inform* **54** (2015), 329–336.
[7] R. Rada, H. Mili, E. Bicknell, and M. Blettner, Development and application of a metric on semantic nets, *IEEE SMC* **19** (1)(1989), 17–30.
[8] T. Cohen and D. Widdows, Empirical distributional semantics: methods and biomedical applications, *J Biomed Inform* **42** (2)(2009), 390–405.
[9] A. Henriksson, H. Moen, M. Skeppstedt, V. Daudaravičius, and M. Duneld, Synonym extraction and abbreviation expansion with ensembles of semantic spaces, *J Biomed Semantics* **5** (1)(2014), 6.
[10] C. Kurtz, C.F. Beaulieu, S. Napel, and D.L. Rubin, A hierarchical knowledge-based approach for retrieving similar medical images described with semantic annotations, *J Biomed Inform* **49** (2014), 227–244.
[11] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, and J. Dean, Distributed Representations of Words and Phrases and their Compositionality, In *Advances in Neural Information Processing Systems* (2013), 3111–3119.
[12] T.H. Muneeb, S.K. Sahu, and A. Anand, Evaluating distributed word representations for capturing semantics of biomedical concepts, In *ACL-International Joint Conference on Natural Language Processing* (2015), 158.
[13] B. Chiu, G. Crichton, A. Korhonen, and S. Pyysalo, How to train good word embeddings for biomedical NLP, In *Proc. of the 15th Workshop on BioNLP* (2016), 166–174.
[14] S.V. Pakhomov, G. Finley, R. McEwan, Y. Wang, and G.B. Melton, Corpus domain effects on distributional semantic modeling of medical terms, *Bioinformatics* **32** (23) (2016), 3635–3644.
[15] Z. Yu, B.C. Wallace, T. Johnson, and T. Cohen, Retrofitting concept vector representations of medical concepts to improve estimates of semantic similarity and relatedness, *Stud Health Technol Inform* **245** (2017), 657–661.
[16] S. Henry, C. Cuffy, and B.T. McInnes, Vector representations of multi-word terms for semantic relatedness, *J Biomed Inform* **77** (2018), 111–119.
[17] S. Henry, A. McQuilkin, and B.T. McInnes, Association measures for estimating semantic similarity and relatedness between biomedical concepts, *Artif Intell Med* **93** (2019), 1-10.
[18] Z.S. Harris, Distributional structure, *Word* **10** (2–3) (1954), 146–162.
[19] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, Enriching word vectors with subword information *TACL* **5** (2017), 135-146.
[20] B.T. McInnes, T. Pedersen, and S.V. Pakhomov, UMLS-Interface and UMLS-Similarity: open source software for measuring paths and semantic similarity, In *AMIA Annual Symposium Proc* (2009), 431.
[21] A. Sajadi, Graph-based domain-specific semantic relatedness from Wikipedia, In *Advances in Artificial Intelligence* (2014), 381–386.
[22] A. Sajadi, E. Milios, V. Kešelj, and J.C. Janssen, Domain-specific semantic relatedness from Wikipedia structure: a case study in biomedical text, In *CLILing* (2015), 347–360.
[23] P. Resnik, Using information content to evaluate semantic similarity in a taxonomy, In *14th IJCAI Proc-Vol 1* (1995), 448–453.
[24] D. Lin, Using syntactic dependency as local context to resolve word sense ambiguity, In *8th Conference on European Chapter of the ACL Proc* (1997), 64–71.
[25] M. Lesk, Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone, In *5th Annual ICIS Documentation Proc* (1986), 24–26.
[26] S. Patwardhan and T. Pedersen, Using WordNet-based context vectors to estimate the semantic relatedness of concepts, In *Workshop on Making Sense of Sense: Bringing Psycholinguistics and Comput Linguistics Together Proc* (2006).
[27] O. Bodenreider, The unified medical language system (UMLS): integrating biomedical terminology, *Nucleic acids research* **32** (suppl_1)(2004), D267-D270.
[28] T.E. Workman, G. Rosemblat, M. Fiszman, and T.C. Rindflesch, A literature-based assessment of concept pairs as a measure of semantic relatedness, In *AMIA Annual Symposium Proceedings* (2013), 1512.
[29] L. De Vine, G. Zuccon, B. Koopman, L. Sitbon, and P. Bruza, Medical semantic similarity with a neural language model, In *Proc. of the 23rd ACM CIKM* (2014), 1819-1822.
[30] A.R. Aronson and F.M. Lang, An overview of MetaMap: historical perspective and recent advances, *JAMIA* **17** (3)(2010), 229-236.
[31] Z. Yu, T. Cohen, B. Wallace, E. Bernstam, and T. Johnson, Retrofitting word vectors of MeSH terms to improve semantic similarity measures. In *7th LOUHI Proc* (2016), 43-51.
[32] S.V Pakhomov, T. Pedersen, B. McInnes, G.B. Melton, A. Ruggieri, and C.G. Chute, Towards a framework for developing semantic relatedness reference standards, *Journal of Biomedical Informatics* **44** (2) (2011), 251–265.
[33] S. Pakhomov, B. McInnes, T. Adam, Y. Liu, T. Pedersen, and G. Melton, Semantic similarity and relatedness between clinical terms: An experimental study, *In AMIA Annual Symposium Proc* (2010), 572–576.
[34] B. Chiu, S. Pyysalo, I. Vulić, and A. Korhonen, Bio-SimVerb and Bio-SimLex: wide-coverage evaluation sets of word similarity in biomedicine, *BMC bioinformatics* **19** (1)(2018), 33.
[35] L.F. Soualmia and S.J. Darmoni, Combining different standards and different approaches for health information retrieval in a quality-controlled gateway, *Int J Med Inform* **74** (2-4) (2005), 141-150.

Address for correspondence

Lina F. Soualmia, lina.soualmia@litislab.eu. Normandie Université, CURIB (LITIS), 25 Rue Lucien Tesnière, 76130 Mont-Saint-Aignan, FRANCE. +33 232 955 173

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190173

Graft Rejection Prediction Following Kidney Transplantation Using Machine Learning Techniques: A Systematic Review and Meta-Analysis

Aldilas Achmad Nursetyo[a,b], Shabbir Syed-Abdul[a,b], Mohy Uddin[c], Yu-Chuan (Jack) Li[a,b,d]

[a] *Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taiwan*
[b] *International Center for Health Information Technology, Taipei Medical University, Taiwan*
[c] *King Abdullah International Medical Research Center, King Saud bin Abdulaziz University for Health Sciences, Executive Office, King Abdulaziz Medical City, Ministry of National Guard Health Affairs, Riyadh, Kingdom of Saudi Arabia*
[d] *Research Center of Cancer Translational Medicine, Taipei Medical University, Taiwan*

Abstract

Kidney transplantation is recommended for patients with End-Stage Renal Disease (ESRD). However, complications, such as graft rejection are hard to predict due to donor and recipient variability. This study discusses the role of machine learning (ML) in predicting graft rejection following kidney transplantation, by reviewing the available related literature. PubMed, DBLP, and Scopus databases were searched to identify studies that utilized ML methods, in predicting outcome following kidney transplants. Fourteen studies were included. This study reviewed the deployment of ML in 109,317 kidney transplant patients from 14 studies. We extracted five different ML algorithms from reviewed studies. Decision Tree (DT) algorithms revealed slightly higher performance with overall mean Area Under the Curve (AUC) for DT (79.5% ± 0.06) was higher than Artificial Neural Network (ANN) (78.2% ± 0.08). For predicting graft rejection, ANN and DT were at the top among ML models that had higher accuracy and AUC.

Keywords:

Kidney Transplantation; Graft Rejection; Machine Learning

Introduction

Kidney transplantation provides high-quality life years to patients with ESRD. Outcomes following kidney transplantation are evaluated by renal function and graft rejection. Recipients' clinical status and outcomes after the transplant are influenced by recipients' ages, Human Leukocyte Antigen (HLA) matching, HLA immunization, ethnic background, time on dialysis, and cardiovascular comorbidities [10; 19].

Graft rejection is the most common problem for kidney transplant recipients. Antibody-mediated rejection requires a distinct therapy as compared to the therapy for usual T-cell-mediated acute rejection. Renal function, based on estimated Glomerular Filtration Rate (GFR) and/or proteinuria values, is a result of these factors. Renal function impairment, whether in a stable condition or as a progressing dysfunction, also has an impact. Confirmation of graft rejection needs renal biopsy, which in turn requires expenditure and time. Antibiotics and immunosuppressive drugs could decrease acute graft rejection incidence, but chronic graft rejection is still a major problem [7; 28]. Specific and nonspecific (diabetes, nephrotoxicity, infection and cancer) conditions could have significant negative long-term consequences [23].

Due to the increased availability clinical data and rapid development of computing technology, artificial inteligence (AI) has been successfully applied in the healthcare domain. AI uses advanced learning algorithms to analyze large volumes of healthcare data that facilitates decision making in clinical practice. Various forms of clinical data (such as, diagnosis, screening, and treatment assignment) can be used as training data before AI systems can be applied in daily healthcare settings. By doing so, the system can learn and apply AI to similar groups of subjects, associations between subject features, and outcomes of interest. Training data is not only limited to clinical data, but it also includes demographics, medical notes, electronic recordings from medical devices, physical examinations, and clinical laboratory and images [15].

Machine learning (ML) techniques are used for specific purposes. Each unique feature of ML can be used for a different function, therefore different results may be obtained with different ML models. In the literature review, we did not come across any meta-analysis study specifically evaluating the use of ML algorithms for predicting kidney transplant outcomes [17]. The objectives of this study are: (1) to review the role of ML in predicting graft rejection following kidney transplantation, and (2) to specifically identify ML algorithms that have a higher accuracy and performance for predicting graft rejection following kidney transplantation, reported in the literature.

Methods

This review paper followed the flowchart and checklist provided by Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [24].

Search Strategy and Data Sources

In order to get accurate results from the vast biomedical and health research databases—PubMed and Scopus, we used combined search keywords, such as "kidney transplants AND machine learning", "kidney transplants AND data mining", and "kidney transplants AND artificial intelligence". For covering the domain of computer science, we also used DBLP, which is a database for scientific journals in the field of computer science that are not yet indexed by PubMed, although the proposed published methods are applied on biomedical datasets [16]. In DBLP, we used the keyword "kidney transplants" to get all the related studies with this specific search criterion. We included all studies found in the

literature until July 9, 2018, and checked for duplicate findings.

Eligibility Criteria

The articles were included based on the following criteria:

1. Written in English.
2. Using ML techniques to predict graft rejection in patients who had undergone kidney transplantations.
3. Building model extracted on medical record (eg. renal registry).

Articles were excluded based on the following criteria:

1. Used features other than clinical features extracted from medical record (eg. –omics features, radiological imaging).
2. Using ML techniques other than graft rejection prediction purpose.
3. Did not mention the result of ML performance.

Data Extraction and Synthesis

The selected studies were summarized based on the author name, year of publication, number of patients, dataset details, types of input variables, used ML techniques, and also validation method. The aim of the summarizing process was to get a detailed description of each ML model generated by the studies. Performance of the ML models was also recorded by mentioning each performance metric (accuracy and AUC). Comparison of each ML techniques was analyzed based on a previous study done by Malhotra [22]. Performance was visualized for each ML techniques and performance metrics.

Results

By applying inclusion and exclusion criteria mentioned above, 14 articles were identified for detailed study analysis (Figure 1).

Datasets and Patients

From 14 selected studies, most studies used publicly available datasets, such as United States Renal Data System (USRDS), United Network for Organ Sharing (UNOS) and the Eurotransplant database. While other studies used self collected data from cohorts hosted by their organizations. Table 2 describes the datasets and number of patients used in each study. Number of patients used in studies ranged from 80 to 57,389. A total of 109,317 patients were described in this study.

USRDS is a national data system that collects, analyzes, and distributes information about ESRD and CKD in United States population [4]. It collects data on patient demographic characteristics, contact information, treatment, laboratory values, quality-of-life survey interviews and nutrition survey interviews for dialysis patients, and also facilitates data sharing by filling request form provided in their web page [9]. UNOS is a private, non-profit organization based in US focused on organ transplant procurement by maintaining contact with volunteers. UNOS also maintains Organ Procurement and Transplantation Network which contains pre-transplant data pertains to transplant candidates [3].

ANZDATA is a registry that holds records of the incidence, prevalence and outcome of dialysis and transplant treatment for patients with end stage renal failure in Australia and New Zealand population [1]. Like UNOS, Eurotransplant is also a non-profit organization that facilitates procurement organ transplant across Europe, especially post-mortem donor organs. It collects various data both from the donors' side and the recipients' side along with the outcomes of procedures [2]. The other included studies collected longitudinal data of kidney transplantation from local organization (e.g. teaching hospital) in order to conduct their research.

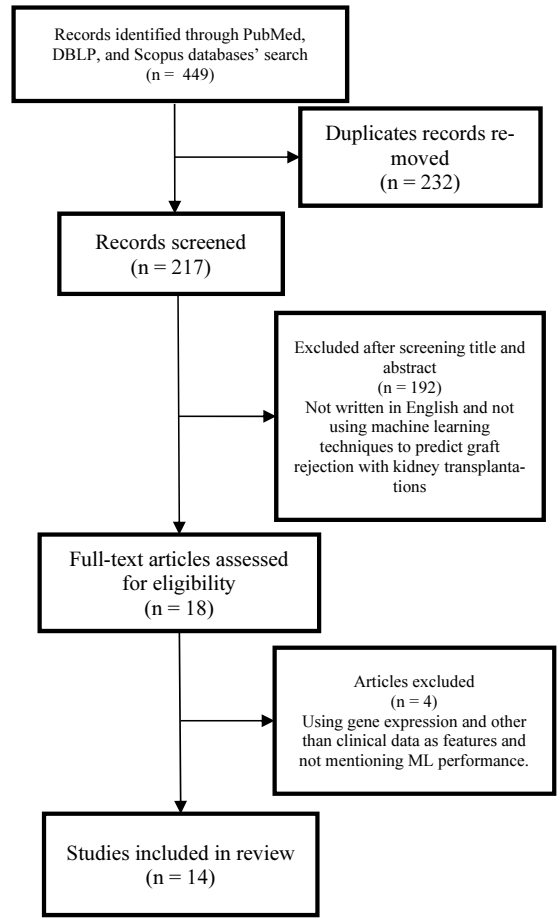


Figure 1. PRISMA Flow Diagram for Study Selection Process

ML Techniques and Overall Performance

Overall, 5 ML techniques that were used in the 14 included studies are: Artificial Neural Network (ANN), Support Vector Machine (SVM), Bayesian Belief Network (BBN), Decision Tree (DT) and an ensemble learning method called Random Forests. ANN was the topmost technique used in 7 studies. The second most commonly used technique was DT algorithm that appeared with different type such as C.50, Classification & Regression Tree (CART/C&RTree) and Random Forests.

Figure 2 describes the ML performances in the form of box plots. From the box plots, it is clear that DT and ANN mostly outperform all other techniques that had been used in studies. Tang et al [31] showed that ANN could perform better than statistical learning methods, such as Logistic Regression (LR). While Shaikhina et al. [27] showed that DT still can be the technique of choice even after being applied in ensemble methods, such as Random Forest. Study done by Esteban et al. [8] showed high performance by utilizing Recurrent Neural

Table 2. Dataset and number of patients included in studies.

Study No.	Author	Year	Dataset	Number of patients
1	Lin et al. [20]	2008	USRDS (2003) + UNOS	57389
2	Topuz et al. [33]	2017	UNOS (2004-2015)	31207
3	Brown et al. [6]	2012	USRDS (2004)	7348
4	Tang et al. [31]	2011	USRDS (2002)	4754
5	Yoo et al. [34]	2017	Misc.	3117
6	Esteban et al. [8]	2016	Misc.	2061
7	Shadabi et al. [25]	2004	ANZDATA Registry Database (2000)	1344
8	Lasserre et al. [18]	2012	Eurotransplants database (1998-2008)	707
9	Shahmoradi et al. [26]	2016	Misc.	513
10	Tapak et al. [32]	2017	Misc.	378
11	Greco et al. [13]	2010	Misc.	194
12	Hummel et al. [14]	2010	Misc.	145
13	Lofaro et al. [21]	2010	Misc.	80
14	Shaikhina et al. [27]	2017	Misc.	80

USRDS = United States Renal Data System; UNOS = United Network for Organ Sharing; ANZDATA = Australia & New Zealand Dialysis and Transplant Registry

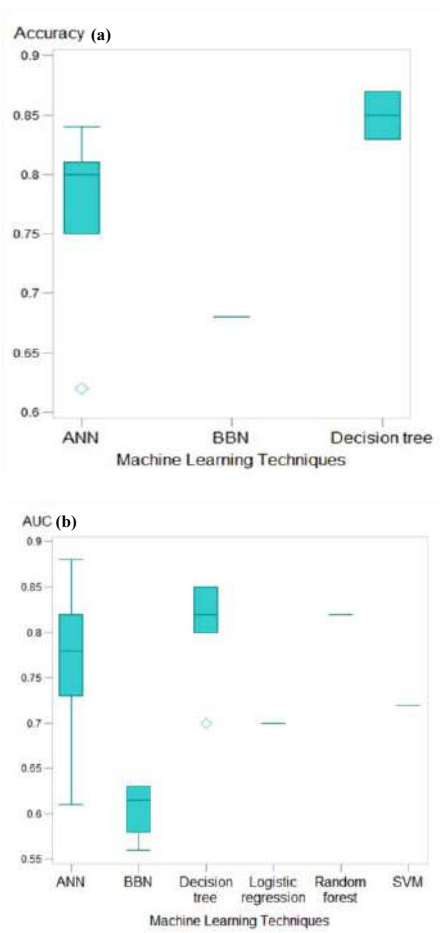


Figure 2 (a,b) Box plots showing ML performances of studies based on (a) Accuracy and (b) AUC. Outliers are shown by study number depicted in Table 2. ANN = Artificial Neural Network, BBN = Bayesian Belief Network, SVM = Support Vector Machine

Network (RNN). The model combined non-linear and linear features (medication perscriptions and laboratory results), along with with static features (gender, age, weight) showing 82% of AUC performance.

Discussion

This paper reviewed the role of applying AI techniques (ML methods) in predicting graft rejection following kidney transplantation, and described the algorithms used by critically reviewing their performances. It is important to be clear about the specific outcomes to be studied, before deploying ML methods.

Based on our results, ANN and DT were the most commonly used models. These techniques showed better performance than SVM, Random Forest and BBN. As DT has the robustness to noise, low computational cost, and ability to deal with redundant features; it has advantages over other learning algorithms. DT could be inducted in various ways, such as C5.0 and CART, but none have been shown to be superior to other methods [5].

ANN is an mathematical algorithm that represents the human neural architecture and resembling the function like learning and generalizing ability. Nowadays, these techniques are widely applied in various research fields because they can show good performance in finding relationship among unknown or complex variables, such as non-linear variables. ANN can be applied in various ways, the most used techniques are Multi-layer Perceptron (MLP), with 3 important layers: input layer, hidden layer, and output layer. This technique is described as being fully connected to every node in the next and previous layer. MLP are trained by selecting suitable connecting weights and transfer functions between the input and output vectors [11]. In this group, prediction model using RNN algorithm developed by Esteban et al. [8] are the most powerful in classification power. RNN are kind of neural networks that usually applied in sequential data such as voice recognition and natural language processing (NLP). The algorithm elaborate both dynamic and static data from medical record that are relevant to predict future outcomes [12].

While doing the literature search, we also found some other systematic reviews related to AI techniques and transplantation. Sousa et al. [30] has reviewed AI techniques used for analysing organ transplant databases from 2009 to 2010 from PubMed and Web of Knowledge. They inferred that the main techniques used were: ANN, LR, DT, Markov Models (MM), and Bayesian Networks (BN). ANN was most preferred for knowledge extraction. Singh et al. [29] provided a systematic review of clinical prediction models of patient and graft survival in kidney transplant recipient using Medline and EMBASE databases covering the time period from 1966 to 2013. They showed the model discrimination with 'C' statistics for patient survival models and graft survival models and reported calibration and external validation of the methods. They also deduced modest discriminatory ability in most clinical prediction models, variability in other measures of model performance, and inconsistency for external validation of models. While Sousa [30] focused on AI techniques that were applied to extract knowledge from transplantation databases, Singh [29] reviewed articles that developed clinical prediction models of patient and graft survival in kidney transplant recipients. In comparison to these articles, our review article specifically studied the role of AI techniques (ML methods) utilized in predicting outcomes following kidney transplantation, and also evaluated the performances of the algorithms used.

More studies are desirable to compare different models. Hybrid models could be used for prediction enhancements. Ensemble and deep learning methods could also be considered in the future.

Conclusions

Based on the PRISMA guidelines, this study evaluated the role of AI techniques (ML algorithms) in predicting treatment outcome following kidney transplantation by examining the available literature. From the literature and our results it was clear that there is no 'One size fits all' approach for applying ML methods. Selection of the right algorithm, provided input variables and volume, and accuracy of the training datasets are critical. Based on performance measured by sensitivity, specificity, accuracy, and AUC, we concluded that ANN and DT were the most suitable and prevalent methods to predict graft rejection following transplantation procedure. A new model built with features taken from both donors' and recipients' side is desirable. Comparison of various models, especially Ensemble method and Deep Learning is required for the future work.

Acknowledgements

This research is sponsored in part by Ministry of Science and Technology (MOST) under grant MOST 106-2221-E-038-005, 106-2923-E-038-001-MY2, 107-2923-E-038 -001 -MY2, Taipei Medical University under grant number 106-3805-004-111 and Wanfang hospital under grant number 106TMU-WFH-01-4.

Address of Correspondence

Shabbir Syed-Abdul

E-mail : drshabbir@tmu.edu.tw

Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, 15F., No. 172-1, Sec. 2, Keelung Rd., Da'an Dist., Taipei, Taiwan 10675

References

[1] ANZDATA Registry, in.
[2] Eurotransplant Database, in.
[3] United Network for Organ Sharing, in.
[4] United States Renal Data System (USRDS), in.
[5] R.C. Barros, M.P. Basgalupp, A.C.P.L.F. de Carvalho, and A.A. Freitas, A Survey of Evolutionary Algorithms for Decision-Tree Induction, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **42** (2012), 291-312.
[6] T.S. Brown, E.A. Elster, K. Stevens, J.C. Graybill, S. Gillern, S. Phinney, M.O. Salifu, and R.M. Jindal, Bayesian modeling of pretransplant variables accurately predicts kidney graft survival., *American Journal of Nephrology* **36** (2012), 561-569.
[7] R.B. Colvin, Antibody-mediated renal allograft rejection: diagnosis and pathogenesis, *J Am Soc Nephrol* **18** (2007), 1046-1056.
[8] C. Esteban, O. Staeck, S. Baier, Y. Yang, and V. Tresp, Predicting Clinical Events by Combining Static and Dynamic Information Using Recurrent Neural Networks, (2016), 93-101.
[9] R.N. Foley and A.J. Collins, The USRDS: what you need to know about what it can and can't tell us about ESRD, *Clin J Am Soc Nephrol* **8** (2013), 845-851.
[10] G. Garcia-Garcia, P. Harden, and J. Chapman, The global role of kidney transplantation, *Indian Journal of Nephrology* **22** (2012), 77-82.
[11] M.W. Gardner and S.R. Dorling, Artificial Neural Networks (The Multilayer Perceptron)—A Review Of Applications In The Atmospheric Sciences, *Atmospheric Environment* **32** (1998), 2627-2636.
[12] A. Graves, A.-r. Mohamed, and G. Hinton, Speech recognition with deep recurrent neural networks, in: *2013 IEEE international conference on acoustics, speech and signal processing*, IEEE, 2013, pp. 6645-6649.
[13] R. Greco, T. Papalia, D. Lofaro, S. Maestripieri, D. Mancuso, and R. Bonofiglio, Decisional trees in renal transplant follow-up, *Transplant Proc* **42** (2010), 1134-1136.
[14] A.D. Hummel, R.F. Maciel, R.G. Rodrigues, and I.T. Pisa, Application of artificial neural networks in renal transplantation: classification of nephrotoxicity and acute cellular rejection episodes, *Transplant Proc* **42** (2010), 471-472.
[15] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang, Artificial intelligence in healthcare: past, present and future, *Stroke Vasc Neurol* **2** (2017), 230-243.
[16] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, Machine Learning and Data Mining Methods in Diabetes Research, *Comput Struct Biotechnol J* **15** (2017), 104-116.
[17] K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, and D.I. Fotiadis, Machine learning applications in cancer prognosis and prediction, *Comput Struct Biotechnol J* **13** (2015), 8-17.
[18] J. Lasserre, S. Arnold, M. Vingron, P. Reinke, and C. Hinrichs, Predicting the outcome of renal transplantation., *Journal of the American Medical Informatics Association* **19** (2012), 255-262.
[19] C. Legendre, G. Canaud, and F. Martinez, Factors influencing long-term outcome after kidney transplantation, *Transplant International* **27** (2014), 19-27.
[20] R.S. Lin, S.D. Horn, J.F. Hurdle, and A.S. Goldfarb-Rumyantzev, Single and multiple time-point prediction models in kidney transplant outcomes, *Journal of Biomedical Informatics* **41** (2008), 944-952.

[21] D. Lofaro, S. Maestripieri, R. Greco, T. Papalia, D. Mancuso, D. Conforti, and R. Bonofiglio, Prediction of Chronic Allograft Nephropathy Using Classification Trees, *Transplantation Proceedings* **42** (2010), 1130-1133.
[22] R. Malhotra, A systematic review of machine learning techniques for software fault prediction, *Applied Soft Computing* **27** (2015), 504-518.
[23] N. Mohammadpour, S. Elyasi, N. Vahdati, A.H. Mohammadpour, and J. Shamsara, A Review on Therapeutic Drug Monitoring of Immunosuppressant Drugs, *Iranian Journal of Basic Medical Sciences* **14** (2011), 485-498.
[24] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, and P.G. The, Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement, *PLOS Medicine* **6** (2009), e1000097.
[25] F. Shadabi, R. Cox, D. Sharma, and N. Petrovsky, Use of Artificial Neural Networks in the Prediction of Kidney Transplant Outcomes, in: *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, 2004.
[26] L. Shahmoradi, M. Langarizadeh, G. Pourmand, Z.A. Fard, and A. Borhani, Comparing Three Data Mining Methods to Predict Kidney Transplant Survival, *Acta Inform Med* **24** (2016), 322-327.
[27] T. Shaikhina, D. Lowe, S. Daga, D. Briggs, R. Higgins, and N. Khovanova, Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation, *Biomedical Signal Processing and Control* (2017).
[28] B. Shrestha, J. Haylor, and A. Raftery, Historical perspectives in kidney transplantation: an updated review., *Progress in Transplantation* **25** (2015), 64-69, 76.
[29] S. Singh, D. Naimark, J. Victor, and S. Kim, Clinical Prediction Models of Patient and Graft Survival in Kidney Transplant Recipients: A Systematic Review.: Abstract# D2483, *Transplantation* **98** (2014), 631-632.
[30] F.S. Sousa, A.D. Hummel, R.F. Maciel, F.M. Cohrs, A.E. Falcao, F. Teixeira, R. Baptista, F. Mancini, T.M. da Costa, D. Alves, and I.T. Pisa, Application of the intelligent techniques in transplantation databases: a review of articles published in 2009 and 2010, *Transplant Proc* **43** (2011), 1340-1342.
[31] H. Tang, M.R. Poynton, J.F. Hurdle, B.C. Baird, J.K. Koford, and A.S. Goldfarb-Rumyantzev, Predicting Three-Year Kidney Graft Survival in Recipients with Systemic Lupus Erythematosus, *American Society of Artificial Internal Organs* (2011).
[32] L. Tapak, O. Hamidi, P. Amini, and J. Poorolajal, Prediction of Kidney Graft Rejection Using Artificial Neural Network., *Healthcare informatics research* **23** (2017), 277-284.
[33] K. Topuz, F.D. Zengul, A. Dag, A. Almehmi, and M.B. Yildirim, Predicting graft survival among kidney transplant recipients: A Bayesian decision support model, *Decision Support Systems* **106** (2017), 97-109.
[34] K.D. Yoo, J. Noh, H. Lee, D.K. Kim, C.S. Lim, Y.H. Kim, J.P. Lee, G. Kim, and Y.S. Kim, A Machine Learning Approach Using Survival Statistics to Predict Graft Survival in Kidney Transplant Recipients: A Multicenter Cohort Study, *Scientific Reports* **7** (2017), 8904.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190174

Automatic Methods to Extract Prescription Status Quality Measures from Unstructured Health Records

Patrick R Alba, MS[a,b], Olga V Patterson, PhD[a,b], Joshua S Richman, MD, PhD[c,d], Scott L DuVall, PhD[a,b]

[a] VA Salt Lake City Health Care System; [b] University of Utah, Salt Lake City, UT;
[c] Birmingham VAMC; [d] University of Alabama at Birmingham, Birmingham, AL

Abstract

Hospital systems frequently implement quality measures to quantify healthcare processes and patient outcomes. One such measure that has previously been used is the Surgical Care Improvement Project (SCIP) quality measure of perioperative beta blocker continuation, SCIP-Card-2. The SCIP-Card-2 measure requires resource-intensive medical chart abstraction, limiting its application to a small sample of eligible patients. This paper describes a natural language processing (NLP) system for automatic extraction of SCIP-Card-2 quality measures in clinical text notes.

Keywords:

Natural language processing; Quality Indicators; Health Resources

Introduction

Quality measures are tools that healthcare systems frequently implement to assess the effectiveness, safety, efficiency, and timely care of various medical procedures and processes [1]. This study focuses on the Surgical Care Improvement Project's (SCIP) quality measure to assess perioperative beta blocker continuation, SCIP-card-2. Studies have shown that mortality in patients who discontinued taking beta blockers pre- or post-surgery is significantly greater than mortality in patients who continued taking beta blockers [2]. Recording the SCIP-card-2 measure allowed health care organizations to take appropriate steps to ensure that patients receive proper care [3].

While the SCIP-card-2 measure was retired as of 2015, this measure provides an important use case for measures that are complicated to assess and require more than just structured data. The SCIP-card-2 measure included patients with current outpatient beta blocker therapy undergoing non-laparoscopic surgery. The quality measure evaluated two periods: (1) the perioperative period, defined as the day prior to surgery through discharge from the post-anesthesia care unit and (2) the postoperative period, consisting of the first two postoperative inpatient days. If patients received beta blockers during both periods, the surgical care was considered to adhere to the guideline. The SCIP-card-2 measure indicated guideline adherence failure if the patient received no beta blockers and no reason for intentional non-adherence was documented. Patients were excluded if no beta blockers were received but the rationale was documented.

Assessing this quality measure has been done using resource-intensive manual chart abstraction to identify and classify the prescription status of beta blockers [4]. Manual chart abstraction required numerous person hours, which limits any study population to a relatively small set of patients. By automatically classifying the prescription status of beta blockers, this study aims to show the potential benefits of using natural language processing (NLP) when assessing adherence to quality measures.

Numerous studies have focused on the automatic extraction of medication mentions and medication features from text. Xu et al developed the NLP system MedEx to extract a drug name, strength, route, and frequency with F-measures above 90% for each concept [5]. The 2009 i2b2 NLP challenge resulted in the successful creation of several NLP systems that also focused on the identification of medication information including: dosage, frequency, treatment duration, mode of administration, and the reason for administration [6]. Much work has been put into automatically extracting drug information, while less work has been done to extract the prescription status of a medication. However, it has been shown to be feasible. Pakhomov et al. used NLP to identify outpatient aspirin usage and contraindications to aspirin; using a sample of 499 diabetic patients, their system successfully identified patients taking aspirin and those not taking aspirin with a sensitivity of 99% and a specificity of 91% [7]. Sohn et al contributed further by testing both machine learning methods and rule-based methods to extract prescription status, finding that rule-based methods could be improved by better discerning the difference between automatically generated medication lists and narrative text sections [8]. Meystre et al identified Angiotensin Converting Enzyme Inhibitors, and successfully classified the medications as being active, discontinued, or negated with an overall accuracy of 95.49%, finding minimal improvements using machine learning methods over rule-based methods for this specific task [9].

The identification and classification of beta blocker administration, specifically using data from the Department of Veterans Affairs (VA) electronic health record (EHR), has not been previously explored. The subsequent sections describe the methods used within the scope of a larger study to create a comprehensive dictionary of terms reflecting beta blocker medications, a custom sectionizer to discern automatically generated medication lists from narrative text, and a rule-based NLP algorithm; all used together to detect and classify mentions of beta blockers in clinical text.

Methods

Cohort definition

All work for this study was completed using VA Informatics and Computing Infrastructure (VINCI) resources. Using structured administrative data from the VA Corporate Data Warehouse (CDW), 8679 patients met the criteria for this study as having current outpatient prescriptions for beta blockers and having undergone non-laparoscopic surgery within the study time period. Additionally, this sample of patients was further limited to those who did not have pharmacy data indicating beta blocker administration at all periods to meet measure requirements, providing a set of potentially ambiguous cases requiring manual review or NLP. All documents for the day prior to an operation and two days post-operation were used to create a corpus of 400,534 documents. This set of documents was randomly partitioned into a 75% training and 25% testing set.

Dictionary definition

An initial dictionary of medication terms was created by identifying all known beta blockers, their synonyms, and varied trade names in RxNorm, a thesaurus, nomenclature, and coding system for clinical drugs [10]. The dictionary was further expanded with additional terms and abbreviations using an iterative manual bootstrapping process on documents in the training set [11]. Finally, clinical experts with knowledge relevant to the surgical subdomain reviewed and contributed to the initial dictionary, which contained 59 seed terms. (Appendix A)

Due to the frequency of misspellings and variation of terms that exist in clinical documentation, using a simple dictionary of keywords to capture instances of a medication did not provide an acceptable level of term coverage. Previous studies have shown the effectiveness of using edit distance algorithms in a clinical setting to improve the term coverage of custom dictionaries [9]. The Levenshtein edit distance is a well-established fuzzy matching algorithm that was shown to be an effective method for identifying medication terms. Additional work has shown that combining another popular method, the Jaro-Winkler distance metric, could further enhance the accuracy of a fuzzy matching algorithm [12]. Edit distance algorithms, however, can be extremely computationally expensive as each term must be compared to every other string in the dictionary along with the millions of possible strings that exist in a dataset. In an effort to create a high-throughput system that could be applied to future, increasingly larger corpora, a comprehensive dictionary was created. Having a comprehensive dictionary eliminates the need to include edit distance algorithms as part of the NLP pipeline. Using the VA CDW full text index, each individual token ever used in the VA EHR was ranked with a total count of the number of times each token has been used. Terms found in the indexed table were then mapped to terms found in our initial dictionary of beta blocker terms. After manual review of the candidate dictionary of terms, a threshold was then chosen for the indexed tokens that matched beta blocker terms with a Levenshtein edit distance of <=2 and a Jaro-Winkler edit distance of >90%. Any term that fell within this threshold was considered a candidate match and then manually reviewed for validity.

By applying the edit distance algorithms to the full text index, rather than to each individual document in our set, we created a comprehensive dictionary that could be used to identify mentions of beta blockers in VA clinical documents using simple string match, which is a significantly faster processing algorithm than applying more complex document processing for beta blocker term identification. Using both distance metrics and the additional manual review, our expanded dictionary was evaluated on the training documents and exhibited an initial precision and recall of 100%. Thus, this comprehensive dictionary was determined to be fully sufficient to find every possible mention of any beta blocker medication term in the complete VA EHR.

Annotation

To create a set of prescription status examples, all documents in the training set were manually annotated using terms from the expanded dictionary. The training set contained 2345 instances, which were annotated at the sentence level using the VINCI developed annotation tool Chex [13]. Chex is a browser-based application that allows rapid annotation at a mention level with a user-defined context window that can be expanded as needed.

The context for the mapped terms was then classified as being one of the following types:

1. Active Medication: Any medication found in a narrative format, and described as being actively taken, or given on the day of the note.
2. Medication List: Any instance of a medication term found within an automatically generated medication list.
3. Negated Medication: A medication found in the narrative section of a document that was not given, e.g. “Metoprolol not given”.
4. Discontinued Medication: A medication that was prescribed to a patient but discontinued on the day of the note.
5. Other beta blocker mention: Any mention of a medication without an explicit mention of the prescription status, which appears in medication warnings or discussion of the medication or other similar contexts, e.g. “Metoprolol is frequently associated with…”

The annotators were also given an option for irrelevant terms being mapped as a beta blocker, however, no instances of an incorrect or irrelevant term was found, which confirms the accuracy of the utilized dictionary. Table 1 displays the frequency of each classification type that was found in the training set.

Table 1 Training set Classification Labels

Classification Label	Frequency
Active Medication	1018
Medication List	990
Negated Medication	57
Discontinued Medication	104
Other Beta Blocker Mention	176
Irrelevant Term	0

Rule-based medication and phrase detection

Previous studies have shown that automatically generated medication lists do not provide an accurate representation of a patient’s current prescription status [14,15]. However, these lists are frequently included as a semi-structured text element inside a document. To get an accurate representation of th

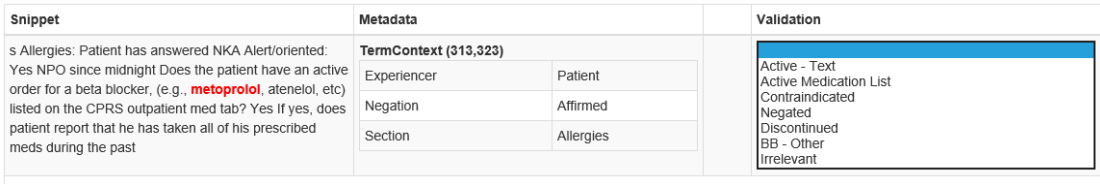


Figure 1. An example of the annotation interface using Chex that shows an instance of an "Active" beta blocker usage classification, it's span location within the document, and the options presented to an annotator.

prescription status on the day a note is recorded, the first step of the system distinguished these semi-structured lists from free text mentions found in narrative form. Medication mentions found in an automated list were detected using two methods. First, a custom sectionizer was created to locate the most frequently occurring section headers, many of which were specific to VA documents, e.g. "Veteran's Active Medication:".

After distinguishing medications in narrative format from those in medication lists, the NLP system focused on classifying those instances found in the narrative. To account for the diversity of terms, writing styles, document structures, and typographic errors that exist across individual clinicians, hospitals, and regions, our system relied on several sets of dictionaries used by separate modules throughout the NLP pipeline. These dictionaries contain numerous regular expressions compiled with manual knowledge engineering while also building upon the existing NLP systems and resources. Specifically, we utilized ConText [16], RxNorm [10], and MedEx [5]. The four concept dictionaries are:

- **Medication Term:** This dictionary consists of all known beta blocker terms as described in the dictionary definition section of this paper.
- **Modifier Terms:** These include terms that would commonly be associated with the noun phrase of a medication term such as: "*Oral* metoprolol", "*p.o* metoprolol" and "*home* atenolol". This list was created manually using examples found in the training set of documents and expanded using modified MedEx dictionaries.
- **Administration Terms:** This dictionary contains terms meant to describe the administration, discontinuation or negation of a medication such as: "Discontinued", "D/C", "held", "withheld", "stopped", "started", "began", etc. All terms in this dictionary were added using examples from the training corpus, their synonyms, and other known abbreviations and terms based on subject matter expert knowledge.
- **Context Terms:** This dictionary consists of terms found in the ConText algorithm to determine if a medication term is negated, hypothetical, historical, or experienced by someone other than the patient. The ConText dictionaries were also expanded using examples from our training set as well as other commonly occurring phrases specific to documentation in a VA healthcare setting.

After the terms from each dictionary were identified in a document, the system incorporated two pattern annotator modules [17]. The first pattern set combined instances of a beta blocker medication term and the modifying terms to create a medication phrase, e.g. "home metoprolol", "25mg oral atenolol". A second pattern set combined the medication phrases with administration terms to create a final administration phrase. Administration phrases were then normalized and labeled as being: Active, Active Medication List, Discontinued, Negated, or Other. One additional dictionary was included in the NLP system that captured terms describing the rationale for a discontinued medication e.g. "d/c atenolol due to" or "because of", however, there were very few instances of these patterns found in the training set. Due to the infrequency of examples, this dictionary was not included in the final validation. Figure 2 shows the basic structure used for combining all terms identified by different dictionaries into separate medication and administration patterns.

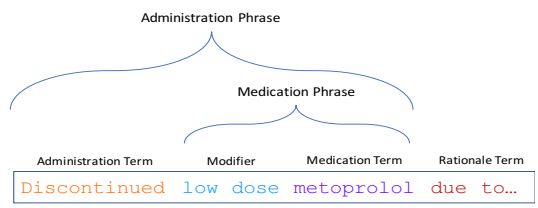


Figure 2. An example dictionary and pattern annotator structure for a discontinued administration phrase.

The rule-based NLP system to extract and classify the prescription status was implemented using Leo, a VINCI-developed set of services and libraries that aid in the development of NLP pipelines based on the Apache Unstructured Information Management Architecture Asynchronous Scaleout (UIMA AS) [18,19]. This system used simple regular expressions connecting the different concept dictionaries together into the relevant classification labels without the use of tokenizing, stemming, or additional parsing.

Results

System validation was performed using 200 random documents from the testing set that contained 274 mentions of beta blocker medication. The documents were processed by the final system and classification accuracy was evaluated.

Table 2. Confusion matrix of validation results

Actual labels Predicted labels	Active	Discontinued	Negated	Medication List	Other	Total predicted
Active	65	1	0	0	1	67
Discontinued	4	46	0	0	0	50
Negated	0	2	22	0	0	24
Medication List	13	0	0	59	4	76
No Status Found	33	1	3	10	10	57
Total actual	115	50	25	69	15	274

Table 3 Validation Results

	Precision	Recall	F-Measure
Active	0.970	0.565	0.714
Discontinued	0.920	0.920	0.920
Negated	0.917	0.880	0.897
Medication List	0.776	0.855	0.813

Error analysis revealed that a primary cause for error was due to patterns and terms present in the testing set that were previously unseen in the training set of documents, this is a known drawback when using rule-based NLP systems and is once again magnified in this work[20]. The "active" recall score was especially affected in this work as a result of the large variety of verbiage used by clinicians when notating documents in a narrative form.

Discussion

By using two edit distance methods and processing all terms ever used in the VA, we show the potential for creating comprehensive term dictionaries when a full-text index exists for the complete corpus. Incorporating these methods as a cached dictionary of terms prior to processing documents has a positive impact on the speed of processing. Through the use of existing and modified medication dictionaries, a light-weight targeted sectionizer, and a simple rule-based NLP pipeline, we were also able to show the potential for automatically extracting surgical quality measures.

Future work will be done to assess the feasibility of extracting the rationale for a discontinued medication as well as alternative methods of approaching the NLP pipeline to improve precision and recall. With a limited set of negated and discontinued examples in our training set, it was determined that in this specific study there would not be enough data to explore machine learning approaches as they generally require a larger example set for accurate training. Recent research has shown that this may still be improved using weakly labeled data[21], or alternative methods of dictionary creation[22], but those were not attempted for this work.

Conclusion

A comprehensive dictionary of terms associated with beta-blocker use in the VA clinical environment was developed and evaluated. A custom sectionizer to distinguish medication mentions in a clinical narrative from those created automatically using structured data was tested and applied to improve automatic prescription status extraction. This study showed that combining the dictionary creation methods with a custom sectionizer and a rule-based NLP pipeline can effectively classify the prescriptions status of the beta blockers, achieving a critical component of the SCIP-card-2 quality measure.

References

[1] L.H. Rosenberger, A.D. Politano, and R.G. Sawyer, The surgical care improvement project and prevention of post-operative infection, including surgical site infection., *Surg. Infect. (Larchmt).* **12** (2011) 163–8.

[2] J.B. Shammash, J.C. Trost, J.M. Gold, J.A. Berlin, M.A. Golden, and S.E. Kimmel, Perioperative beta-blocker withdrawal and mortality in vascular surgical patients, *Am. Heart J.* **141** (2001) 148–153.

[3] M.D. Kertai, M. Cooter, R.J. Pollard, W. Buhrman, S. Aronson, J.P. Mathew, and M. Stafford-Smith, Is Compliance With Surgical Care Improvement Project Cardiac (SCIP-Card-2) Measures for Perioperative β-Blockers Associated With Reduced Incidence of Mortality and Cardiovascular-Related Critical Quality Indicators After Noncardiac Surgery?, *Anesth. Analg.* (2017).

[4] T. Sutherland, J. Beloff, M. Lightowler, X. Liu, A. Kachalia, D.W. Bates, and R.D. Urman, Implementation of measures to improve SCIP perioperative beta-blocker compliance: Quality and financial implications, *Health Care Manag. (Frederick).* **34** (2015) 192–198.

[5] H. Xu, S.P. Stenner, S. Doan, K.B. Johnson, L.R. Waitman, and J.C. Denny, MedEx: A medication information extraction system for clinical narratives, *J. Am. Med. Informatics Assoc.* **17** (2010) 19–24.

[6] O. Uzuner, I. Solti, and E. Cadag, Extracting medication information from clinical text., *J. Am. Med. Inform. Assoc.* **17** (2010) 514–8.

[7] S.V. Pakhomov, N.D. Shah, P. Hanson, S.C. Balasubramaniam, and S.A. Smith, Automated processing of electronic medical records is a reliable method of determining aspirin use in populations at risk for cardiovascular events., *Inform. Prim. Care*. **18** (2010) 125–133.

[8] S. Sohn, S.P. Murphy, J.J. Masanz, J.-P.A. Kocher, and G.K. Savova, Classification of medication status change in clinical narratives., *AMIASymposium*. **2010** (2010) 762–766.

[9] S.M. Meystre, Y. Kim, J. Heavirland, J. Williams,

B.E. Bray, and J. Garvin, Heart Failure Medications Detection and Prescription Status Classification in Clinical Narrative Documents, in: Stud. Health Technol. Inform., 2015: pp. 609–613.

[10] S. Liu, W. Ma, R. Moore, V. Ganesan, and S. Nelson, RxNorm: Prescription for electronic drug information exchange, *IT Prof.* **7** (2005) 17–23.

[11] P.R. Alba, OV Patterson, B. Viernes, D. Denhalter, N. Bailey, A. Wilson, A.W. Kamauu, and S.L. DuVall, The Super Annotator: A Method of Semi-Automated Rare Event Identification for Large Clinical Data Sets, in: AMIA Annu Symp Proc, 2016.

[12] R.A. Baeza-Yates, and C.H. Perleberg, Fast and practical approximate string matching, *Inf. Process. Lett.* **59** (1996) 21–27.

[13] O.. Cornia R., DuVall S.L., Forbush T., Halls C., Patterson, Check it with Chex: A Validation Tool for Iterative NLP Development, in: AMIA Annu Symp Proc, 2014.

[14] P.L. Cornish, S.R. Knowles, R. Marchesano, V. Tam, S. Shadowitz, D.N. Juurlink, and E.E. Etchells, Unintended Medication Discrepancies at the Time of Hospital Admission, *Arch. Intern. Med.* **165** (2005) 424..

[15] E.M. Breydo, J.T. Chu, and A. Turchin, Identification of inactive medications in narrative medical text, *AMIA Annu. Symp. Proc.* (2008) 66–70.

[16] H. Harkema, J.N. Dowling, T. Thornblade, and W.W. Chapman, ConText: an algorithm for determining negation, experiencer, and temporal status from clinical reports., *J. Biomed. Inform.* **42** (2009) 839–51.

[17] B.R. Soundrarajan, T. Ginter, and S.L. DuVall, A Novel Approach to Building Relationships Between Annotations in Pipeline Natural Language Processing Systems, in: Int. Conf. Mach. Learn. Work. Learn. from Unstructured Clin. Text, Bellevue, WA, USA, 2011.

[18] R. Cornia, O. V Patterson, T. Ginter, and S.L. Duvall, Rapid NLP Development with Leo, in: AMIA Annu Symp Proc, Washington, DC, USA, 2014: p. 1356.

[19] D. FERRUCCI, and A. LALLY, UIMA: an architectural approach to unstructured information processing in the corporate research environment, *Nat. Lang. Eng.* **10** (2004) 327–348.

[20] B. Waltl, G. Bonczek, and F. Matthes, Rule-Based Information Extraction: Advantages, Limitations, and Perspectives, (2018).

[21] Y. Guo, A. Korhonen, I. Silins, and U. Stenius, Weakly supervised learning of information structure of scientific abstracts-is it accurate enough to benefit real-world tasks in biomedicine?, *Bioinformatics*. (2011).

[22] I. Banerjee, M.C. Chen, M.P. Lungren, and D.L. Rubin, Radiology report annotation using intelligent word embeddings: Applied to multi-institutional chest CT cohort, *J. Biomed. Inform.* (2018).

Address for Correspondence

Olga Patterson

olga.patterson@utah.edu

Appendix A: Beta blocker initial dictionary terms.

Generic_Name	Trade_Names
Acebutolol	Sectral
Alprenolol	
Atenolol	Tenormin,Tenoretic
Betaxolol	Kerlone,Betoptic
Bevantolol	
Bisoprolol	Zebeta,Ziac
Bopindolol	
Bupranolol	
Carazolol	
Carteolol	Cartrol
Carvedilol	Coreg
Celiprolol	
Cloranolol	
Epanolol	
Labetalol	Trandate, Normodyne,Trandate
Landiolol	
Mepindolol	
Metoprolol	Lopressor,Toprol
Nadolol	Corgard,Corzide
Nebivolol	
Oxprenolol	
Penbutolol	
Pindolol	Visken,Iodopindolol
Practolol	
Propranolol	Inderal,Innopran,Hemangeol,Inderide
S-atenolol	
Sotalol	Betapace,Sorine
Talinolol	
Tertatolol	
Timolol	Blocadren,Timoptic,Betimol,Istalol

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190175

mHealth4Afrika - Implementing HL7 FHIR Based Interoperability

Mert Baskaya[a], Mustafa Yuksel[a], Gokce Banu Laleci Erturkmen[a], Miriam Cunningham[b] and Paul Cunningham[b]

[a] SRDC Software Research Development & Consultancy Corp, Ankara, Turkey
[b] IIMC / IST-Africa Institute, Dublin, Ireland

Abstract

Supported by the European Commission under Horizon 2020, mHealth4Afrika is co-designing and validating a modular, multilingual, state-of-the-art health information system addressing primary healthcare requirements in resource constrained environments. mHealth4Afrika has co-designed a comprehensive range of functionality and medical programs in partnership with Ministries of Health, district health officers, clinic managers and primary healthcare workers from urban, rural and deep rural health facilities in Ethiopia, Kenya, Malawi and South Africa. This paper provides insights into how mHealth4Afrika is leveraging HL7 FHIR to support standards-based data exchange and interoperability between Electronic Medical Records and DHIS2. This work is currently being validated in the field.

Keywords:

Health Information Interoperability, Electronic Health Records, HL7 FHIR

Introduction

While electronic patient records are gradually being introduced into larger hospitals in Ethiopia, Kenya, Malawi and South Africa (current mHealth4Afrika beneficiary countries), paper-based registries [1, 2, 5-7] remain the default data capture method in resource constrained urban, rural and deep rural health facilities. mHealth4Afrika supports the objectives of UN Sustainable Development Goal 3 (SDG3) by co-designing a comprehensive, patient-centric health platform that is adaptable and extensible, modular and multilingual [1 - 6]. It integrates Electronic Medical Record (EMR) and Electronic Health Record (EHR) functionality, with the use of medical sensors and data visualization tools at the point of care [5, 6]. It supports the automatic counting of aggregate program indicator data required by Ministries of Health, SMS appointment notifications and lab system integration.

In some resource constrained environments, including Africa, donors have adopted a silo-based application approach, addressing requirements for specific programs they fund, including ART (HIV/AIDS) and Tuberculosis (TB) [6, 8].

Standards and interoperability are key enabling environment components of the WHO and ITU National eHealth Strategy Toolkit [9]. As the number of patient centric technology-enabled health applications (eHealth and mHealth) grow, the importance of interoperability becomes ever more critical to avoid unnecessary duplication of effort and fragmentation of electronic health record data. mHealth4Afrika has taken a standards-based approach to data interoperability to support data exchange between applications and systems.

The South African National Department of Health published the National Health Normative Standards Framework for Interoperability in eHealth in South Africa (HNSF) in 2014 [10]. This aims to provide a framework supporting the development of interoperable health systems. It proposed that Health Level Seven (HL7) is used as the messaging standard for the exchange and integration of electronic clinical healthcare information between systems.

In April 2017 the Ministry of Health of Kenya published the Kenya Standards and Guidelines for mHealth Systems [11] to support data and information sharing across multiple health systems. It is even more specific than South Africa, requiring that eHealth/mHealth solutions used in Kenya should leverage HL7 FHIR (Fast Healthcare Interoperability Resources) APIs (Application Programming Interfaces) [12].

mHealth4Afrika research objectives include to co-design a comprehensive, patient-centric health platform leveraging some of the functionality of District Health Information System 2.0 (DHIS2) [6] and support standards-based data exchange to transfer medical sensor readings and lab requests to and from the patient medical record, and import and export of patient records between EMRs. DHIS2 and the Tracker Capture application have several limitations. As a result, mHealth4Afrika has designed a comprehensive and extensible patient-centric, multi-program custom platform and user interface for use in medical facilities which interacts with the mHealth4Afrika data model set up in DHIS2 via the native DHIS2 WebAPI [6].

mHealth4Afrika has co-designed a comprehensive range of medical programs supporting easy and systematic data capture, storage and searching of patient centric data. Based on priorities of initial intervention countries (Ethiopia, Kenya, Malawi and South Africa) [1, 2, 5, 6], medical programs currently implemented and validated include: medical history, maternal health, family planning, cervical cancer screening, child under 5, tuberculosis, antiretroviral therapy, diabetes, general and specialist outpatient department (OPD). mHealth4Afrika supports single registration of a patient at a health facility and subsequent enrolment in a range of different programs based on their health conditions over time [1, 6].

mHealth4Afrika has applied a standards-based approach to data interoperability, taking account of state-of-the-art standards and international good practices, while respecting national policies and legislation in intervention countries. In the context of introducing medical sensors at the point of care, mHealth4Afrika developed a HL7 FHIR Service in 2017 to authenticate users and support data transfer of readings from BLE medical sensors to the appropriate patient record in the

mHealth4Afrika data model stored in the DHIS2 server via the DHIS2 API. We are not aware of any other project to date that has used HL7 FHIR for data transfer of readings from medical sensors (e.g. for blood pressure, blood glucose, SpO2, heart rate, weight, temperature) with a data model set up in DHIS2 or similar implementations designed for use in primary healthcare facilities in resource constrained environments.

During 2018 the mHealth4Afrika HL7 FHIR service was extended to support standards-based data exchange of lab requests to a laboratory system and the transfer of lab results into the appropriate mHealth4Afrika patient medical record.

A design requirement for mHealth4frika was to support standards-based data exchange of patient clinical data between health information systems. The target use case is to support a hospital referral or a patient moving temporarily or permanently from one health facility to another. Currently DHIS2 does not support exporting or importing the health record of a specific patient. mHealth4Afrika has extended its HL7 FHIR service to support import and export of a patient's health records between mHealth4Afrika platform instances and to support HL7 FHIR data exchange of patient's records with other EMR and EHR being used nationally.

This paper provides an overview of the design of the standards-based approach to support the import and export of patient health records by extending the mHealth4Afrika HL7 FHIR Service. The methods section provides insights into mHealth4Afrika compliance with Continua Design Guidelines (CDG) by Personal Connected Health Alliance (PCHA), the two-way data mapping between the mHealth4Afrika data model set up in DHIS2 and HL7 FHIR STU3 resources. The following section summarizes and discusses early results, while the last section presents the conclusion.

Methods

The mHealth4Afrika FHIR Service is designed as an independent service to achieve full HL7 FHIR based interoperability without disrupting how the mHealth4Afrika browser-based custom application communicates via the native DHIS2 WebAPI with the mHealth4Afrika data model set up in the DHIS2 server. The service supports exporting and importing data. The mHealth4Afrika application has been extended to include a user interface to support import/export functionality and interaction with the FHIR Service.

Data flow is explained below in Figure 1.

For authentication between the service and DHIS2 server, the mHealth4Afrika application passes the 'Cookie' header in each call made to the Import / Export service. The appropriate 'Cookie' header is obtained through the user login process.

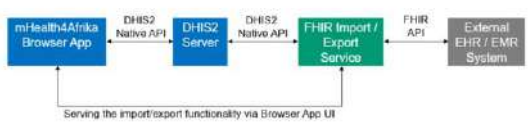


Figure 1 – mHealth4Afrika Import / Export Service Data Flow

Continua Design Guidelines Compliance

The mHealth4Afrika FHIR Service has been implemented with the purpose of achieving standards-based interoperability between the DHIS2 server and other EHR/EMR systems. Figure 2 briefly summarizes mHealth4Afrika compliance with the Continua Design Guidelines (CDG) by Personal Connected Health Alliance (PCHA) [13].

The Personal Health Devices Interface already implemented between medical devices supported by mHealth4Afrika and the mHealth4Afrika application, and the Services Interface already implemented between mHealth4Afrika applications and DHIS2 server, are beyond the scope of this paper.

This paper focuses on the mHealth4Afrika HL7 FHIR Service, which acts as a Healthcare Information Service (HIS) Interface between DHIS2 server and EHR / EMR systems.

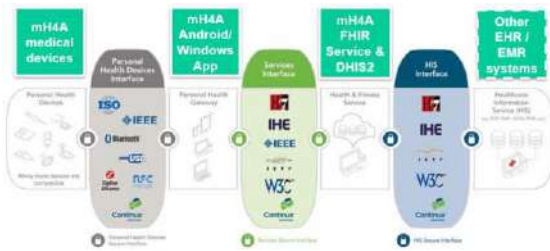


Figure 2 – mHealth4Afrika Compliance with Continua Design Guidelines

DHIS2 Data Model

Two-way data mapping between the DHIS2 data model and HL7 FHIR STU3 resources is required to support import and export capabilities based on FHIR. In this context the DHIS2 data model has been analyzed and data model elements which must be mapped to FHIR resources have been identified:

- DHIS2 OrganisationUnit represents healthcare facilities.
- DHIS2 TrackedEntityInstance represents patients.
- DHIS2 User represents practitioners or clinic managers.
- DHIS2 ProgramStage defines the content of the program specific forms, i.e. which DataElements should be presented in each mHealth4Afrika program stage (e.g. In Maternal Health program expected date of delivery should exist in Antenatal Care visits but not in Postnatal Care visits).
- DHIS2 Event represents patient visits corresponding to a program stage, i.e. filled-in program data.
- DHIS2 DataElement represents the leaf-level patient data definitions, such as vital sign measurements (e.g. blood glucose reading) as well as observational questions (e.g. Does the patient have edema?), which are all used in program stage definitions.
- DHIS2 Enrollment represents patient enrollments to mHealth4Afrika programs (Maternal Health, Tuberculosis, etc.). A patient may be enrolled in multiple programs.

Mapping of Resources

TrackedEntityInstance and Event are the two main DHIS2 resources directly related to patient healthcare records. TrackedEntityInstance contains demographic information about patients, while Event contains medical information captured during clinical (or program stage) visits by patients.

FHIR Patient resource is used to represent DHIS2 TrackedEntityInstances. TrackedEntityInstance attributes are mapped to respective FHIR Patient attributes. Standard FHIR attributes are used where possible and some extensions are defined for mHealth4Afrika specific attributes, e.g. patient type (adult or child). Figure 3 provides an example of some of the mHealth4Afrika TrackedEntityInstance set up in DHIS2 with the corresponding FHIR Patient resources presented in Figure 4. To date, a total of 61 TrackedEntityInstance attributes have been mapped to FHIR Patient attributes.


```
{
    "trackedEntityInstance": "t4TA3mk3hT6",
    "orgUnit": "YimTrMUaHPT",
    "attributes": [

      {"displayName": "Date of Birth",
       "valueType": "DATE",
       "attribute": "gCafpJzwByC",
       "value": "1995-11-02" },

      {"displayName": "Last Name/ Surname",
       "valueType": "TEXT",
       "attribute": "PUt8IjUQ5s6",
       "value": "Johns" },

      {"displayName": "First Name",
       "valueType": "TEXT",
       "attribute": "sQG4zPPzP5x",
       "value": "Mike" },

      {"displayName": "Patient Type",
       "valueType": "TEXT",
       "attribute": "DqsQdsl08uG",
       "value": "adult" },

      {"displayName": "Medical Record Number",
       "valueType": "TEXT",
       "attribute": "r9wKBvoh4WT",
       "value": "34708253912" },

      {"displayName": "Gender",
       "valueType": "TEXT",
       "attribute": "Tgz9w5PrpJM",
       "value": "male" },

      {"displayName": "Marital Status",
       "valueType": "TEXT",
       "attribute": "rncpcIYS0Jn",
       "value": "single" }
    ]}
```


Figure 3 – mHealth4Afrika Tracked Entity Instance

FHIR QuestionnaireResponse resource is used to represent DHIS2 Events. Events contain multiple dataValues that are similar to the QuestionnaireResponse attribute "item". As mentioned before, Events represent program stage visits and each dataValue contains a dataElement attribute which references the respective dataElement in the ProgramStage resource. In other words, ProgramStage contains a template of what the visits should contain (i.e. metadata) and Event contains actual values that is gathered in a visit (i.e. instances). The structure is similar in FHIR Questionnaire and QuestionnaireResponse. Questionnaire resource contains questions and, in some cases, possible values and option sets of the questions and QuestionnaireResponse resource contains answers to those questions defined in the Questionnaire. Having a very similar structure, Questionnaire resources are used to represent ProgramStages and QuestionnaireResponse resources are used for Events. Mapping of attributes, and the relations between ProgramStage, Event, Questionnaire and QuestionnaireResponse resources can be seen in Figure 5.

Event attributes are mapped to respective FHIRQuestionnaireResponse attributes. Similar to TrackedEntityInstances, standard FHIR attributes are used where possible and some custom extensions are defined for mHealth4Afrika specific values such as program. ProgramStage attributes are mapped to respective FHIR Questionnaire attributes.

FHIR Organization and EpisodeOfCare resources are used to represent OrganisationUnits and Enrollments for the purpose of having a complete and consistent FHIR Bundle. As outlined in Figure 4, FHIR Patient resource contains a reference to the managing organization with its id. Without including an Organization resource with the same id in the FHIR Bundle, the bundle will fail referential integrity. Hence, respective FHIR Organization resources are created corresponding to the OrganisationUnits in DHIS2. It is similar with Enrollments and Events, which contain enrollment references.


```
{
"resourceType": "Patient",
"id": "LoKI7VAJX8k",
"extension": [{
  "url": "http://www.mhealth4afrika.eu/fhir/PatientType"
  "valueCode": "adult"
}],
"identifier": [{
 "system":"http://www.mhealth4afrika.eu/fhir/MedicalRecordNo",
 "value": "34708253912"
}],
"name": [{
 "family": "Johns",
 "given": [ "Mike" ]
}],
"gender": "male",
"birthDate": "1995-11-02",
"maritalStatus": {
 "coding": [{
  "system": "http://hl7.org/fhir/v3/MaritalStatus",
  "code": "S",
  "display": "Never Married"
 }]
},
"managingOrganization": {
 "reference": "Organization/YimTrMUaHPT"
}
}
```


Figure 4 – HL7 FHIR Patient Resource

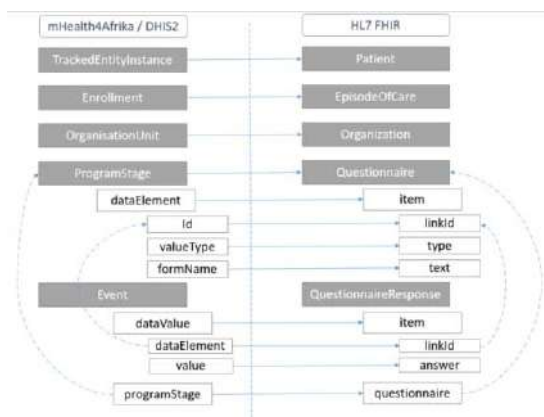


Figure 5 – DHIS2 Data Model to FHIR Resource mapping

FHIR Export

The service provides an endpoint that is used to export HL7 FHIR bundles containing information about the demographics and medical visits of the patients. The program, organizationUnit and the individual patient is specified at the request call to customize the output. The service uses several methods from the DHIS2 Native API to gather relevant data from DHIS2, transforms the gathered data to FHIR resources based on the mapping explained in the previous section, and finally returns a HL7 FHIR Bundle containing HL7 FHIR resources. The export can include all patient specific medical records or only a specific program, depending on the permission provided by the patient and why the electronic medical record is being provided to a different health facility.

The bundle contains several types of FHIR resources.

- Patient: This resource contains demographic information about the patient.
- QuestionnaireResponse: This resource contains actual data about the program stage visits of the patient.
- Questionnaire: This resource contains metadata about the program stage visits.
- Organization: This resource contains information about related Organisational Units.
- EpisodeOfCare: This resource contains information about program enrollments of the patients.

FHIR Import

The service provides an endpoint that is used to import HL7 FHIR bundles containing information about the complete or partial medical history of the patients. It enables patients to receive medical care in different health facilities as required and request that their electronic medical record is transferred between facilities and kept up to date.

The output of the FHIR Import, i.e. FHIR bundle, is parsed and relevant resources are mapped to the mHealth4Afrika DHIS2 data model. Several DHIS2 API calls are used sequentially to preserve referential integrity, e.g. patient demographic information must be imported before a lab result.

Initial Results and Discussion

The mHealth4Afrika platform has been co-designed and validated with Ministries of Health, District Health Offices, Clinic Managers and nurses in Ethiopia, Kenya, Malawi and South Africa over the past three years. It is currently being used in a mix of primary health care facilities and hospitals.

The initial focus of this extension to the mHealth4Afrika FHIR service commenced with a mapping of patient centric data attributes and data elements to FHIR resources. As outlined above, it has been necessary to use some extensions based on specific mHealth4Afrika program requirements. When the export and functionality was implemented, this required a modification of the mHealth4Afrika platform to engage with the extended FHIR service.

The standards-based approach being implemented builds on an existing HL7 FHIR service that mHealth4Afrika developed in 2017 to support data exchange of medical sensors readings and lab requests and lab results to and from the patient's electronic medical record. It takes account of national policy and legislative requirements in intervention countries, while leveraging the most up to date international standards.

In the current mHealth4Afrika beneficiary countries, individual patient medical records are only stored within the health facility that provides the healthcare services.

This interoperability work will enable mHealth4Afrika to both export and import complete individual patient records (as well as specific programs from individual patient records) from an instance in one health facility to an instance in another health facility. This ensures that existing electronic medical records can be transferred as required between health facilities, based on patient consent. This is a practical requirement that is not currently supported within DHIS2 standard tools.

It is normal in many countries that a pregnant mother may move from her place of normal residence to another location closer to family members towards the end of her term. In this context, it is important that the healthcare facility where she is currently receiving care has access, with her consent, to her full medical record - medical history, details of medical data captured during antenatal visits etc. The mHealth4Afrika platform allows the health facility managers to print off reports of data captured across antenatal visits. However, if the healthcare facility providing delivery and postnatal care are using the mHealth4Afrika platform or another EMR that supports data exchange using HL7 FHIR, it is more efficient if the electronic health record can be imported, updated and then exported as a FHIR bundle. This will allow the patient's health record to be kept fully up to date when they return to the health facility where they are normally resident.

This functionality will also support patients with medical conditions, such as TB, HIV, Diabetes, and Chronic Hypertension, to move their electronic health records for the purposes of referral or when transferring to another facility.

This import / export functionality is exposed to the clinic manager role via the mHealth4Afrika custom interface. This functionality is currently field tested in the intervention countries and will be adapted as necessary going forward.

Related Work

Healthcare system interoperability is an essential topic in the domain as modern medical care is inherently distributed. Standards-based transfer of medical information is researched and developed over decades and more recently HL7 FHIR based interoperability solutions are being implemented. OpenMRS FHIR Module is such an example [14].

When the FHIR HL7 service work commenced in mHealth4Afrika during 2017 and the import / export work described in this paper commenced during early 2018, there was no implementation of such a FHIR based interoperability service in DHIS2. In September 2018, DHIS2 started work on a use case focused on importing TrackedEntityInstances (patient data) with Observations and Immunisations as Events into a DHIS2 server using HL7 FHIR. This will allow EMRs and applications to connect to a FHIR repository to upload patient data. An adapter can then connect to the FHIR repository and enroll patients in the DHIS2 server. It is proposed mapping is done through a transformation engine.

While this use case was only a sub-set of our current requirements, mHealth4Afrika had meetings with DHIS2 to learn more about this proposed approach and share insight about the approach we have been implementing since 2017.

Recently, the DHIS2 team started to implement an interface based FHIR adapter, similar to the mHealth4Afrika

import/export service. This has the objective of eliminating the need of a FHIR repository to simplify integration of the interoperability functionality into non FHIR repository related use cases. This work is initially focused on import.

Conclusions

This paper provides insight into some of the research objectives of mHealth4Afrika, including designing a standards-based approach to support both import and export of patient health records by extending the existing mHealth4frika HL7 FHIR Service. It outlines two-way data mapping between the mHealth4Afrika data model set up in DHIS2 and HL7 FHIR STU3 resources.

Some of the mapping decisions are designed to be as generic as possible. For example, FHIR QuestionnaireResponse is used to represent program stage visits, as it allows multiple types of items that can cover all program stage data elements. While some of the data elements exactly fit to questionnaire questions, some of them represent specific laboratory tests or vital sign measurements which might be more suitable to be mapped to more specific FHIR medical resources such as Observation or DiagnosticReport. This has been implemented within mHealth4Afrika as an import mechanism, while acquiring vital sign measurements from patients via Bluetooth LE medical sensors, with the mHealth4Afrika Android Application acting as a sensor gateway. This initial import capability was extended within the context of the full program set, extending the FHIR service to support exporting individual patient records using specific FHIR resources.

The import/export functionality for individual patient's medical records was completed in February 2019 and is currently being field tested in the mHealth4Afrika intervention countries.

Acknowledgements

This research is co-funded by the European Commission under the Horizon 2020 Research and Innovation Framework Programme (mHealth4Afrika, Grant Agreement No. 688015).

References

[1] Cunningham M., Cunningham P., van Greunen D., Veldsman A., Kanjo C., Kweyu E. and Tilahun B. (2018) *mHealth4Afrika Beta v1 Validation in Rural and Deep Rural Clinics in Ethiopia, Kenya, Malawi and South Africa*, Proceedings of IEEE Global Humanitarian Technology Conference (GHTC) 2018, IEEE Xplore

[2] Cunningham, M., Cunningham, P. (2018) *mHealth4Afrika - Supporting Primary Healthcare Delivery in Resource Constrained Environment*s, Proceedings of 2018 International Conference on Sustainable Development (ICSD)

[3] UN Sustainable Development Goal 3 (Ensure healthy lives and promote well-being for all at all ages) www.un.org/sustainabledevelopment/health/

[4] Cunningham M., Cunningham P., van Greunen D., Veldsman A., Kanjo C., Kweyu E. and Tilahun B. (2017) *mHealth4Afrika Alpha Validation in Rural and Deep Rural Clinics in Ethiopia, Kenya, Malawi and South Africa*, Proceedings of IEEE Global Humanitarian Technology Conference (GHTC) 2017, IEEE Xplore, ISBN: 978-1-5090-6046-7, DOI: 10.1109/GHTC.2017.8239347

[5] Cunningham, P., Cunningham, M. (2018) *mHealth4Afrika – Challenges When Co-Designing a Cross-Border Primary Healthcare Solution*, 2018 IEEE International Symposium on Technology in Society (ISTAS) Proceedings, ISBN: 978-1-5386-9479-4, IEEE Xplore

[6] Cunningham, M., Cunningham, P., Van Greunen, D. (2018) *mHealth4Afrika - Co-designing an Integrated Solution for Resource Constrained Environments*, Proceedings of HELINA 2018, Journal of Health Informatics in Africa, 2018;5(2):1-9. DOI: 10.12856/JHIA-2018-v5-i2-198

[7] Cunningham, P., Cunningham, M., van Greunen, D., Veldsman, A., Kanjo, C., Kweyu, E. and Gebeyehu, A. (2016) *Implications of Baseline Study Findings from Rural and Deep Rural Clinics in Ethiopia, Kenya, Malawi and South Africa for the co-design of mHealth4Afrika*, Proceedings of IEEE Global Humanitarian Technology Conference (GHTC) 2016, IEEE Xplore, ISBN: 978-1-5090-2432-2, DOI: 10.1109/GHTC.2016.7857350

[8] Fraser HS, Bindich, P, Moodley D, Choi S, Mamlin B, Szolovits P. (2005) *Implementing electronic medical record systems in developing countries*, Journal of Innovation in Health Informatics, Vol13, No.2, ISSN 2058-4563, DOI: http://dx.doi.org/10.14236/jhi.v13i2.585

[9] World Health Organisation, International Telecommunication Union (2012), *National eHealth Strategy Toolkit*, Geneva, WHO, ISBN 978 92 4 154846 5

[10] South African National Department of Health (2014) *Health Normative Standards Framework for Interoperability in eHealth in South Africa* Version 2.0 2014 https://www.gov.za/documents/national-health-act-national-health-normative-standards-framework-interoperability-ehealth

[11] Kenya Ministry of Health (2017) *Kenya Standards and Guidelines for mHealth Systems*, April 2017 https://mhealthkenya.org/wp-content/uploads/2017/06/Kenya-Standards-and-Guidelines-for-mHealth-Systems-April-2017.pdf

[12] Health Level Seven *Fast Healthcare Interoperability Resources (FHIR®)*, https://www.hl7.org/fhir/summary.html

[13] Personal Health Connected Alliance (2018) *Continua Design Guidelines* https://www.pchalliance.org/continua-design-guidelines

[14] Kasthurirathne S.N., Mamlin B., Grieve G., Biondich P. (2015) *Towards Standardized Patient Data Exchange: Integrating a FHIR Based API for the Open Medical Record System*, Studies in health technology and informatics. 216. 932.

Address for correspondence

Mert Baskaya – baskaya@srdc.com.tr

Mustafa Yuksel – mustafa@srdc.com.tr

Gokce Banu Laleci Erturkmen – gokce@srdc.com.tr

Miriam Cunningham – miriam@iimg.com

Paul Cunningham – paul@iimg.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190176

Bridging the Gap Between Consumers' Medication Questions and Trusted Answers

Asma Ben Abacha, Yassine Mrabet, Mark Sharp,

Travis R. Goodwin, Sonya E. Shooshan and Dina Demner-Fushman

Lister Hill National Center of Biomedical Communications
National Library of Medicine, Bethesda, MD

Abstract

This paper addresses the task of answering consumer health questions about medications. To better understand the challenge and needs in terms of methods and resources, we first introduce a gold standard corpus for Medication Question Answering created using real consumer questions. The gold standard[1] consists of six hundred and seventy-four question-answer pairs with annotations of the question focus and type and the answer source. We first present the manual annotation and answering process. In the second part of this paper, we test the performance of recurrent and convolutional neural networks in question type identification and focus recognition. Finally, we discuss the research insights from both the dataset creation process and our experiments. This study provides new resources and experiments on answering consumers' medication questions and discusses the limitations and directions for future research efforts.

Keywords:

Health Informatics, Data Collection, Natural Language Processing

Introduction

Digital information and the World Wide Web make it possible to find information on demand, helping both patients and professionals to find relevant and valuable medical information. With these benefits come growing concerns about the potential misinformation of patients and non-expert consumers. Harmful consequences include self-diagnosis and unfounded anxiety about common symptoms after reviewing online health information (i.e., cyberchondria) [1]. The Pew Research Center reports that 72% of U.S. internet users (over 200 million) have gone online in the past year specifically for health-related information. Of that group, 77% say that their research started with search engines such as Google, while only 13% say they began at a specialized website such as WebMD[2].

Safe and efficient search alternatives are needed to protect non-expert consumers from misleading health information found online. Specialized Question Answering (QA) systems are a potential solution for the medical domain and the consumers' need for reliable health information [2]. Several research efforts tackled the problem of QA in the (bio)medical domain and highlighted the challenges related to question understanding, answer retrieval, and ranking [3-6]. To the best of our knowledge, no studies were previously dedicated to QA about medications.

Figure 1– Word Cloud Representing the Consumer Questions about Drugs that We Used to Create the Gold Standard Corpus.

While several previous studies included such questions as a part of larger and more heterogeneous QA tasks, we argue that questions about medications have specific characteristics that need to be analyzed and evaluated in a focused study. These aspects include (i) different question types (e.g. action, storage, usage time, stopping, and tapering), (ii) more conditional answers (e.g. ingredients and appearance depend often on the manufacturer), (iii) different answer sources having a specific data structure (e.g. websites recommended by the FDA such as DailyMed[3]), and (iv) potentially more difficult questions (e.g. drug alternatives, questions about not formalized/written medical knowledge such as usage time), and a higher sensitivity to error (e.g. drug interactions that cannot be confirmed online).

Several efforts focused on methods for medical question answering [7-10] as well as the creation of relevant datasets. For instance, Kilicoglu et al. [11] introduced a dataset of 2,614 medical questions annotated with the question focus, type and triggers of the question types. In that dataset, "Problems" are the most frequent named entities, while "treatment" and "information" are the most common question types. Ben Abacha et al. [12] organized a medical QA task at the TREC 2017 LiveQA track and published training and testing datasets including consumer health questions received by the NLM, annotations and reference answers retrieved by medical informatics experts. The coverage of questions about medications in the above collections, however, was insufficient to train and develop efficient systems to answer questions about

[1] https://github.com/abachaa/Medication_QA_MedInfo2019
[2] http://www.pewinternet.org/2013/01/15/health-online-2013
[3] https://dailymed.nlm.nih.gov/

medications. In this paper, we present a study of consumers' medication questions with the following contributions:

1. The development and publication of a manually annotated dataset of medication question-answer pairs based on real consumer questions submitted to MedlinePlus.
2. A synthesis of the manual annotation effort summarizing the insights obtained by a variety of experts from annotating and answering those questions.
3. New experiments using deep learning networks trained specifically for the tasks of identifying the main focus in the user's question, and the question type.

In what follows, we first describe our annotation methodology and the baseline approaches in the Methods section. We present the statistics and characteristics of the developed dataset and the empirical results in the Results section.

Methods

In this section, we describe the guidelines used in the manual annotation process and our first empirical evaluation methods based on the dataset.

Data Creation

Selecting Consumer Questions about Drugs. We selected anonymized consumer questions submitted to MedlinePlus[4]. We first performed Medical Entity Recognition using MetaMapLite [13]. We restricted the recognized entities to the following UMLS semantic types associated with medications: Antibiotic [antb], Clinical Drug [clnd], Neuroreactive Substance or Biogenic Amine [nsba], Pharmacologic Substance [phsu], Steroid [strd], and Vitamin [vita]. Finally, we manually selected the questions that (i) were deemed understandable and potentially answerable and (ii) have a drug name as focus. Figure 1 presents a word cloud of the most frequent terms in the selected consumer health questions.

Annotating the Questions. In a study conducted recently on consumer health question answering, Deardorff et al. [14] showed that for 62% of the questions, it was possible for librarians to find an answer in the top 5 search results in MedlinePlus using only the focus and question type. Given the importance of these two elements for QA, we, therefore, focused our efforts on manually annotating each question with a:

- Question focus (always a Drug name in this dataset),
- Question type (e.g. Dose, Interaction, Side effects).

Searching for Reference Answers. For each answerable question, annotators had to retrieve manually a correct and complete reference answer (with its URL and section title):

- **Correct** with regards to the question's explicit and implicit information (e.g. a question about a drug in the UK, a specific form or dose), and extracted from **reliable** websites or scientific papers.
- **Complete** with regards to all possible answers (e.g. all doses, ingredient lists from all manufacturers), and preferably written in **a consumer-friendly** language.

To select our answer sources, we followed the FDA recommendations suggesting[5] MedlinePlus-Drugs as a consumer-friendly website for consumer drug information and DailyMed for trustworthy information about FDA-approved and marketed drugs in the United States. Our final guideline was to search the following sources sequentially until an answer is retrieved:

1. MedlinePlus and DailyMed.
2. Other NIH or U.S. government websites.
3. Other trustworthy websites (e.g., the Mayo Clinic) or academic institutions' websites.
4. Other websites returned by a Google search.

Four annotators participated in the manual annotation and answering process. Then, a medical doctor and an expert in question answering reconciled the annotations of the question types and validated the retrieved answers.

Baseline Methods

Focus Recognition. We adapted, extended and evaluated Bi-directional Long Short-Term Memory (Bi-LSTM) networks[6] on the task of recognizing the question focus (i.e., main drug name) according to the state-of-the-art architecture proposed by Xuezhe and Hovy [15]. The network includes a first Bi-LSTM network to build character-level embeddings, and a second Bi-LSTM taking as input both word embeddings built from the UMLS and pre-trained embeddings built with GloVe [16] and the character-level embeddings built during training with the first Bi-LSTM layer. The token labels were generated with a final Conditional Random Fields (CRF) layer. Our UMLS embeddings consist of binary vectors, where each word is tagged as the Beginning, Inside, Outside, End, or Single token (BIOES) of each semantic concept in the UMLS.

Question Type Identification. We implemented and evaluated a Convolutional Neural Network (CNN) on the task of identifying the question type (e.g., dosage, usage, contraindications). The input embeddings include UMLS BIOES embeddings and a randomly initialized vector of 128 dimensions updated with back-propagation during training.

Answer Retrieval. We conduct a first qualitative study on answer retrieval using twenty questions randomly selected from our dataset and the CHiQA question answering system[7]. CHiQA is the first online medical QA system for consumer health questions from reliable sources such as NIH websites (e.g., MedlinePlus, GARD, NCI), Mayo Clinic and DailyMed. The system relies on different machine learning and knowledge-based methods to recognize the question's focus and type [17-18] and uses the extracted information to retrieve answers with the Lucene search engine and a question-entailment recognition approach [19-20]. Figures 2 and 3 present the first answers returned by CHiQA to two questions selected randomly from the gold standard.

Figure 2– First Answer Returned by CHiQA to the Question: "How to Stop Taking Bisoprolol?"

[4] https://medlineplus.gov
[5] https://www.fda.gov/drugs/resourcesforyou/consumers
[6] https://github.com/guillaumegenthial/sequence_tagging
[7] https://chiqa.nlm.nih.gov

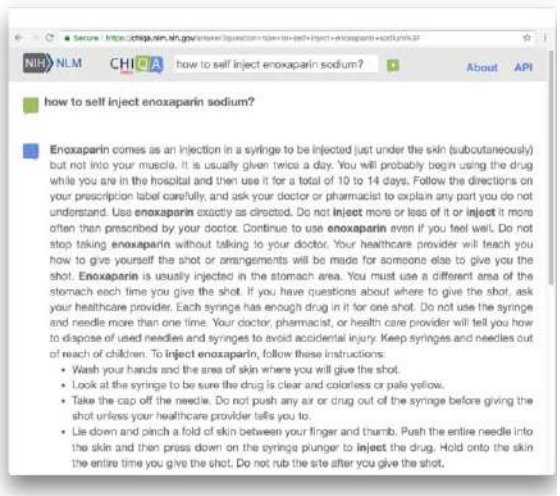


Figure 3– First Answer Returned by CHiQA to the Question: "How to Self Inject Enoxaparin Sodium?"

Results

Characteristics of the Dataset

The final gold standard contains **674** question-answer pairs with their associated annotations. These annotations include 25 question types, reported with examples in Table 1, and the answer sources, summarized in Figure 4. In particular, on the 674 answers, DailyMed was used to answer 290 questions, MedlinePlus for 128 questions, and other websites were used to answer 256 questions (e.g. cdc.gov, mayoclinic.org, health.harvard.edu, and PubMed abstracts and articles). Table 2 presents token-and-sentence-level statistics about the questions and the answers in the dataset.

Table 1 – Question Types in the Gold Standard

Question Type	#	Example
Information	112	what type of drug is amphetamine?
Dose	70	what is a daily amount of prednisolone eye drops to take?
Usage	61	how to self inject enoxaparin sodium?
Side Effects	60	does benazepril aggravate hepatitis?
Indication	55	why is pyridostigmine prescribed?
Interaction	51	can i drink cataflam when i drink medrol?
Action	39	how xarelto affects in the process of homeostasis?
Appearance	38	what color is 30mg prednisone?
Usage/time	36	when is the best time to take lotensin?
Stopping/tapering	31	how to come off citalopram?
Ingredient	28	what opioid is in the bupropion patch?
Action/time	23	how soon does losartan afffect blood pressure?
Storage and disposal	13	in how much temp bcg vaccine should store?
Comparison	11	why is losartin prescribed rather than a calcium channel blocker?
Contraindication	11	if i am allergic to sufa can i take glipizide?
Overdose	10	what happens if your child ate a tylenol tablet?
Alternatives	8	what medicine besides statins lower cholesterol?
Usage/duration	7	how long should i take dutasteride?
Time *(other time-related types)*	6	how long are you protected after taking the hep b vaccine?
Brand names	3	what is brand name of acetaminophen?
Combination	3	how to combine dapaliflozin with metformin?
Pronunciation	3	how do you pronounce humira?
Manufacturer	2	who makes this drug nitrofurantoin?
Availability	1	has lisinopril been taken off the market?
Long term consequences	1	what are the side effects and long term consequences of using nicotine?

Table 2 – Statistics about the Questions and Answers

Average number of tokens per question	**7.16**
Average number of tokens per answer	**69.06**
Average number of sentences per answer	**3.23**
Percentage of questions tokens present in answers	**34.72%**

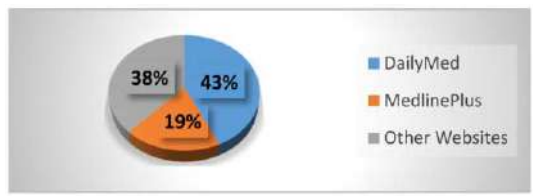


Figure 4 – Websites Used to Answer the Drug Questions

Evaluation Results

We report the average performance on 5 runs and the variation range for focus recognition and question type identification.

Focus Recognition. The Bi-LSTM-CRF network was implemented with Python 3.5 and Tensorflow 1.4. We trained the network with a random subset containing 80% of the data, using 10% for hyperparameter tuning and the remaining 10% for the final test. All data were represented in the BIOES token annotation format. The results on the test set are summarized in Table 3. We used a learning rate of 0.01, a dropout of 0.5, a batch size of 40, and the Adam optimizer [21] to minimize the CRF-based loss function.

Table 3 – Bi-LSTM-CRF Results for Focus Recognition

Results (%)	F1	P	R
Exact entity match	74.07 [+/-2.1]	78.12	70.42
Partial entity match	90.37 [+/- 3.4]	95.31	85.92

Question Type Identification. Our CNN network was also implemented using Python 3.5 and Tensorflow 1.4. Data were split according to same 80/10/10 percent subsets for training, development and test. To have enough training samples per class, we reduced the question types into the 14 most common types by aggregating the subtypes into their hypernym types: Information, Dose, Usage, Tapering, Interaction, Side effects, Indication, Action, Ingredient, Alternatives, Contraindication, Comparison, Manufacturing, and Appearance. We used a learning rate of 0.005, a dropout of 0.6, a batch size of 100, and the Adam optimizer to minimize the softmax-based loss function. The CNN network achieved an average accuracy of 75.7% on 5 runs with a variation in the [0, 2.5%] range. Despite using fixed random seeds for Tensorflow in both experiments, the relatively small size of the training data made them more

susceptible to the non-deterministic implementation of GPU reduction operations and the Adam optimizer in TensorFlow[8].

Answer Retrieval. In our qualitative study of the answers returned to 20 random questions from the dataset, CHiQA found the correct answer in the top four results in 35% of the cases, only related answers for 35% of them and irrelevant answers for the remaining 30%. While this limited evaluation must be taken with caution, our independent observations from the annotation process also hint that classical QA systems may not be the best fit for medication questions.

We discuss these insights and potential ways to improve answering questions about medications in the following section.

Table 4– Example Questions and Answers from the Dataset

ID	Question (Q) / Answer (A)
1a	(Q) "*what does prednisone do to the body?"*
1b	(Q) "*what would a normal dose be for valacyclovir?"*
1c	(Q) "*how much gravol to kill you?"*
1d	(Q) "*what time should take memantine?"*
2a	(Q) "*what color is phenytoin?"* (A) [depends on the manufacturer] • PINK • WHITE (/Light Lavender) • ORANGE
2b	(Q) "*why would my urine test be negative for benzodiazepines when i take Ativan?"* (A) "Common limitations exist for screening benzodiazepines when using traditional immunoassay (IA) tests. IA testing for benzodiazepines often targets nordiazepam and oxazepam (…)" **AND** "Some commonly prescribed drugs have limited cross-reactivity (…)"
2c	(Q) "*is it alright touse fluticasone when using oxygen?"* (A) "Pharmacological therapy can influence morbidity and mortality in severe chronic obstructive pulmonary disease (COPD). Long-term domiciliary oxygen therapy (LTOT) improves survival in COPD with chronic hypoxaemia. Oral steroid medication has been associated with improved survival in men and increased mortality in women, while inhaled steroid medication has been associated with a reduction in the exacerbation rate (…)."
2d	(Q) "*how long does vicodin stay in breast milk?"* (A) "Following a 10 mg oral dose of hydrocodone administered to five adult male subjects, the mean peak concentration was 23.6 ± 5.2 ng/mL. Maximum serum levels were achieved at 1.3 ± 0.3 hours and the half-life was determined to be 3.8 ± 0.3 hours."

Discussion

Difficulty and Ambiguity of Medication Questions

In our annotation and answering efforts, questions about medications showed traits common to consumer health questions such as **linguistic ambiguity** due to misspellings, or wrong grammar leading to unclear meaning or multiple interpretations. For instance, in Table 4, example 1a, it is unclear whether the question is about the side effects or the action of "prednisone". In addition, a distinct pattern emerged with many questions lacking essential information or context, making them too **underspecified** [22] to answer. For instance, in example 1b, finding a relevant answer about the right dose of "Valacyclovir" requires additional information about the patient and the condition or purpose of administration. **Circumlocution** [23] was another phenomenon that we encountered, i.e., the use of many words or general terms to express a medical term. For instance, in example 1c, the medical term "toxic dose" or "overdose" is replaced with the expression "how much X to kill you?". In several cases, we also encountered a **knowledge barrier** when the requested medical information/knowledge was not formalized or written online (e.g. Table 4, 1d).

Complexity of Manual Answer Retrieval

Our annotators reported **conditional answers** frequently. These include answers that depend on the manufacturer, on the disease, or on patient information. For instance, in Table 4, 2a, four different answers are possible according to the manufacturer of phenytoin. **Distributed answers** were also encountered in several cases. These are answers that can be formed only by combining different text snippets from different answers and/or sources (e.g., Table 4, 2b). Many answers also required an understanding of **expert terminology**. These include answers that need to be translated to consumer-friendly language and questions that required rephrasing to expert language in order to find relevant answers (e.g., Table 4, 2c). Other answers could not be found without **expert inference** based on background knowledge (e.g., Table 4, 2d). External resources like *eHealthMe* were needed for specific types of questions such as Interaction questions. Answering some of the questions was also time consuming even for medical experts. For example, an hour was necessary to answer the question "is it alright to use fluticasone when using oxygen?" from PubMed.

Challenges of Automating Medication QA

Prior to our study, the lack of gold standard datasets for medication QA was the major bottleneck in automatic QA for drug questions. This new dataset opens new opportunities for both qualitative studies and quantitative evaluation of QA systems. In addition, systems relying on big training data can use it both (i) as a development set for fine-tuning the hyperparameters and testing different architectures and (ii) as a test benchmark.

In question understanding, our baseline networks achieved an encouraging performance despite the limited training data, with (i) 74% F1 score in question focus recognition for exact span matching and 90% for partial span matching, and (ii) 75.7% accuracy in identifying the question type. Several improvements can be considered for future developments, such as a richer set of embeddings, or a relevant language model. A more fine-grained adaptation of the UMLS can also be applied by restricting to the list of relevant semantic types either through pre-filtering or through trainable masks.

In answer retrieval, insights from both our annotation process and our CHiQA-based evaluation provide additional guidance on the relevant solutions and ways of improvement for medication question answering. In particular:

- Medical text translation [24] and simplification [25] are often needed to find relevant answers and make the retrieved answers readable for non-expert users.
- More data and resources are needed to cover information about drug interactions and usage guidelines. Such information can be extracted from both the scientific literature and clinical sources [26].
- Conditional answers require different solutions such as providing a list of answers or interacting with the user in a dialogue-based approach.

[8] https://www.twosigma.com/insights/article/a-workaround-for-non-determinism-in-tensorflow

- Due to the frequency of these conditional answers, fully unsupervised approaches are less likely to succeed than approaches based on (advanced) inference methods that can rely on explicit context semantics.

Conclusions

We studied consumers' questions about medications. We created a new gold standard corpus for question answering about drugs that we shared in the scope of this paper[9]. We presented statistics, insights and conclusions based on the manual annotation process, deep learning experiments, and preliminary evaluation of automatic answer retrieval. We hope that this new benchmark and initial experiments will foster new approaches and additional community efforts in addressing the growing need for reliable information about medications online.

Acknowledgements

This work was supported by the intramural research program at the U.S. National Library of Medicine, NIH.

References

[1] R.W. White and E. Horvitz, Experiences with web search on medical concerns and self diagnosis, *AMIA Annu Symp Proc* **2009** (2009), 696-700.
[2] A.B. Abacha and D. Demner-Fushman, On the Role of Question Summarization and Information Source Restriction in Consumer Health Question Answering, in: *AMIA Informatics Summit*, 2019.
[3] P. Jacquemart and P. Zweigenbaum, Towards a medical question-answering system: a feasibility study, *Stud Health Technol Inform* **95** (2003), 463-468.
[4] D. Moll and J.L. Vicedo, Question Answering in Restricted Domains: An Overview, *Comput. Linguist.* **33** (2007), 41-61.
[5] S.J. Athenikos and H. Han, Biomedical question answering: a survey, *Comput Methods Programs Biomed* **99** (2010), 1-24.
[6] A. Ben Abacha and P. Zweigenbaum, MEANS: A medical question-answering system combining NLP techniques and semantic Web technologies, *Information Processing & Management* **51** (2015), 570-594.
[7] Y. Niu and G. Hirst, Analysis of semantic classes in medical text for question answering, in: *Workshop on Question Answeing in Restricted Domains, ACL*, 2004.
[8] J. Jeon, W.B. Croft, and J.H. Lee, Finding similar questions in large question and answer archives, in: *Proceedings of the 14th ACM international conference on Information and knowledge management*, ACM, Bremen, Germany, 2005, pp. 84-90.
[9] H. Yu, C. Sable, and H.R. Zhu, Classifying Medical Questions based on an Evidence, in: *Workshop on Question Answering in Restricted Domains, AAAI*, 2005.
[10] L.C. Llanos, S. Rosset, and P. Zweigenbaum, Automatic classification of doctor-patient questions for a virtual patient record query task, in: *BioNLP*, 2017.
[11] H. Kilicoglu, A. Ben Abacha, Y. Mrabet, S.E. Shooshan, L. Rodriguez, K. Masterton, and D. Demner-Fushman, Semantic annotation of consumer health questions, *BMC Bioinformatics* **19** (2018), 34.
[12] E.A. Asma Ben Abacha, Yuval Pinter, Dina Demner-Fushman, Overview of the Medical Question Answering Task at TREC 2017 LiveQA, in: *TREC*, 2017.
[13] D. Demner-Fushman, W.J. Rogers, and A.R. Aronson, MetaMap Lite: an evaluation of a new Java implementation of MetaMap, *J Am Med Inform Assoc* **24** (2017), 841-844.
[14] A. Deardorff, K. Masterton, K. Roberts, H. Kilicoglu, and D. Demner-Fushman, A protocol-driven approach to automatically finding authoritative answers to consumer health questions in online resources, *Journal of the Association for Information Science and Technology* **68** (2017), 1724-1736.
[15] X. Ma and E. Hovy, End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF, in: *ACl*, 2016.
[16] J. Pennington, Richard Socher, and Christopher Manning, Glove: Global vectors for word representation, in: *EMNLP*, 2014.
[17] Y. Mrabet, H. Kilicoglu, K. Roberts, and D. Demner-Fushman, Combining Open-domain and Biomedical Knowledge for Topic Recognition in Consumer Health Questions, *AMIA Annu Symp Proc* **2016** (2016), 914-923.
[18] K. Roberts, H. Kilicoglu, M. Fiszman, and D. Demner-Fushman, Automatically classifying question types for consumer health questions, *AMIA Annu Symp Proc* **2014** (2014), 1018-1027.
[19] A. Ben Abacha and D. Demner-Fushman, Recognizing Question Entailment for Medical Question Answering, *AMIA Annu Symp Proc* **2016** (2016), 310-318.
[20] A.B. Abacha and D. Demner-Fushman, A Question-Entailment Approach to Question Answering, *arXiv:1901.08079 [cs.CL]* (2019).
[21] D.P. Kingma and J.B. Adam, A Method for Stochastic Optimization, in: *ICLR*, 2015.
[22] D. Schlangen, A. Lascarides, and A. Copestake, Resolving underspecification using discourse information, in: *Perspectives on Dialogue in the New Millennium*, 2003.
[23] I. Stanton, S. Ieong, and N. Mishra, Circumlocution in diagnostic medical queries, in: *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, ACM, Gold Coast, Queensland, Australia, 2014, pp. 133-142.
[24] Q. Zeng-Treitler, S. Goryachev, H. Kim, A. Keselman, and D. Rosendale, Making texts in electronic health records comprehensible to consumers: a prototype translator, *AMIA Annu Symp Proc* (2007), 846-850.
[25] E. Abrahamsson, T. Forni, M. Skeppstedt, and M. Kvist, Medical text simplification using synonym replacement : adapting assessment of word difficulty to a compounding language, in: *14th Conference of the European Chapter of the Association for Computational Linguistics,April 27, 2014 Gothenburg, Sweden,* W. Sandra, ed., Association for Computational Linguistics, Stroudsburg, 2014, pp. 57-65.
[26] S. Vilar, C. Friedman, and G. Hripcsak, Detection of drug-drug interactions through data mining studies using clinical sources, scientific literature and social media, *Brief Bioinform* **19** (2018), 863-877.

Address for Correspondence

Asma Ben Abacha, PhD, asma.benabacha@nih.gov
Research Scientist
U.S. National Library of Medicine
8600 Rockville Pike, Bethesda, MD 20894, USA.

[9] https://github.com/abachaa/Medication_QA_MedInfo2019

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190177

Detecting Drug Non-Compliance in Internet Fora Using Information Retrieval and Machine Learning Approaches

Élise Bigeard[a], Frantz Thiessard[b], Natalia Grabar[b]

[a] CNRS, Univ Lille, UMR 8163 STL - Savoirs Textes Langage, F-59000 Lille, France,
[b] U Bordeaux, Inserm, Bordeaux Population Health Research Center, team ERIAS, UMR 1219, F-33000 Bordeaux, France

Abstract

Non-compliance situations happen when patients do not follow their prescriptions and take actions that lead to potentially harmful situations. Although such situations are dangerous, patients usually do not report them to their physicians. Hence, it is necessary to study other sources of information. We propose to study online health fora. The purpose of our work is to explore online health fora with supervised classification and information retrieval methods in order to identify messages that contain drug non-compliance. The supervised classification method permits detection of non-compliance with up to 0.824 F-measure, while the information retrieval method permits detection non-compliance with up to 0.529 F-measure. For some fine-grained categories and new data, it shows up to 0.65-0.70 Precision.

Keywords:

Patient Compliance, Information Storage and Retrieval, Machine Learning

Introduction

Drug non-compliance situations happen when patients do not follow instructions given by their doctors in the prescriptions. Among current situations, we can, for instance, mention modification of dosage, refusal to take prescribed drugs, use of drugs prescribed to other persons. The misuse of drugs, which is part of non-compliance, covers more precise situations, like use of drugs with different intents than those for which drugs are prescribed. We can thus mention recreational or suicidal use of drugs. Such situations are dangerous because they endanger patients and their health. Yet, patients do not inform their doctors or the authorities that they do not follow the instructions. Hence, it is necessary to study other sources of information to gain some insights into patients' lives. We propose to study social media, in which patients are producing large amounts of contents on various subjects [1], including the use of drugs.

Currently, social media have become an important source of information for various research areas, such as geo-localization, opinion mining, event extraction, translation, or automatic summation [2]. In the medical domain, social media have been efficiently used in information retrieval for epidemiological surveillance [3,4], in studying patient's quality of life [5], and drug adverse effects [6].

Yet, very few projects are focused on drug misuse and non-compliance. We can mention here non-supervised analysis of tweets about non-medical use of drugs [7], and creation of a semantic web platform on drug abuse [8]. Both of these projects are dedicated to one specific case of misuse, that is, drug abuse. In our work, we propose to study non-compliance and misuse situations, and more particularly to identify messages related to such situations in health fora. To reach these objectives, we propose two different methods: machine learning and information retrieval.

In what follows, we first detail the methods, then present and discuss the results. Finally, we conclude with directions for future research.

Methods

We propose to address the automatic detection of messages related to drug non-compliance as a categorization problem, and to use two methods: supervised machine learning and information retrieval. We first introduce the reference data used and then describe the two methods.

Reference and Test data

The reference and test data were built from corpora collected in several health websites in French. Several fora were collected from the Doctissimo[1] website (pregnancy, general questions on drugs, back pain, accidents in sport activities, diabetes). Doctissimo is indeed the most well-known and used health website and forum in French. We also used data from three other fora: AlloDocteur[2], masante.net[3], and Les diabétiques[4].

The contributors in all the fora were mainly diseased persons and their relatives, who joined the community to ask questions or provide accounts of their disorders, treatments, etc. Overall, these people may be affected by chronic or non-chronic disorders.

To build the reference data, we used two fora of Doctissimo (pregnancy and general questions on drugs). We collected messages written between 2010 and 2015, and kept only those messages that mentioned at least one drug. This gave a total of 119,562 messages (15,699,467 words). For the test data, we collected 145,012 messages from other corpora. In each message, the occurrence of drugs was detected with specific vocabulary containing French commercial drug names from several sources: base *CNHIM Thériaque*[5], *base publique du médicament*[6], and *base Medic'AM*[7] from Assurance Maladie.

Each drug name was associated with the corresponding ATC code [9]. For the manual annotation process of the reference

1 http://www.doctissimo.fr
2 http://www.allodocteur.fr
3 http://masantenet.com
4 http://www.lesdiabetiques.com
5 http://www.theriaque.org
6 http://base-donnees-publique.medicaments.gouv.fr
7 https://www.ameli.fr/l-assurance-maladie/statistiques-et-publications/donnees-statistiques/medicament/medic-am/medic-am-mensuel-2017.php

data, messages longer than 2,500 characters were excluded because they provided heterogeneous content difficult to categorize and process, both manually and automatically. Then, three annotators were asked to assign each message to one of the two categories:

- *Non-compliance* category contained messages that reported on drug non-compliance or misuse. When this category was selected, the annotators were also asked to shortly indicate what type of non-compliance was indicative (overuse, dosage change, brutal quitting). This indication is written as free text with no defined categories. For instance, the following example shows non-compliance situation due to the forgotten dose of medication: "*bon moi la miss boulette et la tete en l'air je devais commencer mon "utrogestran 200" a j16 bien sur j'ai oublier! donc je l'ai pris ce soir!!!!*" (*I blundered the missed dose and with the head in the clouds I had to start the "utrogestran 200" on d16 and I forgot of course! so I took it this evening!!!!*)
- *Compliance* category contains messages reporting normal drug use ("*Mais la question que je pose est 'est ce que c'est normal que le loxapac que je prends met des heures à agir ???*" (*Anyway the question I'm asking is whether it is normal that loxapac I'm taking needs hours to take effect???*)) and messages without use of drugs ("*ouf boo, repose toi surtout, il ne t'a pas prescris d'aspegic nourisson??*" (*ouch boo, above all give me a break, he didn't prescribe aspegic for the baby??*))

When annotators were unable to decide, they marked up the corresponding messages accordingly. The categorization of these messages as well as the categorization of annotation disagreements are discussed later. The three annotators involved in the process were: one medical expert in pharmacology, and two computer scientists familiar with medical texts and annotation tasks. Because this kind of annotation was a complicated task, especially concerning the decision on drug non-compliance, all messages annotated as non-compliant were additionally verified by one of the annotators.

The manual annotation process permitted to double-annotate 1,850 messages, among which we counted 1,717 messages in the compliance category and 133 messages in the non-compliance category. These numbers indicated the natural distribution of non-compliance messages (approximately 7%). Within the non-compliance category, we counted 16 types of non-compliance: they contained between 1 and 29 messages. For example, the *change of weight* type contained 2 messages, *recreational use of drugs* 2 messages, *suicide attempt* 2 messages, and *overuse* of drugs 20 messages. Concerning the annotation into *compliance* and *non-compliance* categories, the inter-annotator agreement [10] was moderate (0.46) [11]. It seemed that the task at hand was quite complicated and the sparsity of data for some types of non-compliance made the task even more difficult.

The corpus was pre-processed using Treetagger [12] to obtain its tokenization (typically, segmentation of words and punctuation), POS-tagging (assigning syntactic categories to words, such as *anxiétés/Nom (anxieties/Noun)*)), and lemmatization (normalization to canonical forms and removal of inflections for plurals, feminines, such as *anxiétés/anxiété (anxieties/anxiety)*). The corpus was used in three versions: (1) in the *forms* corpus, the messages were only tokenized and lowercased (eg, i'm taking 3 pills each day); (2) in the *lemmas* corpus, the messages were also lemmatized, the numbers were replaced by a unique placeholder, and diacritics were removed such as in a*nxiété/anxiete (anxiety)*; (3) in the *lexical lemmas* corpus, we kept only lemmas of the main lexical categories (verbs, nouns, adjectives, and adverbs) (eg, take pill day). Each message was also indexed with the three first characters of the ATC categories of drugs occurring in the message. We obtained 18355 distinct words, 12231 distinct lemmas, and 12096 distinct lemmas in the *forms* corpus, the *lemmas* corpus, and the *lexical lemma* corpus respectively.

Categorization with Supervised Machine Learning

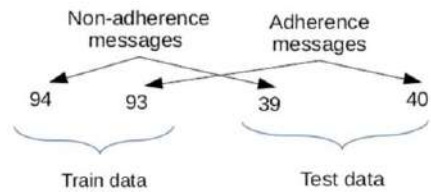


Figure 1 – Datasets Used for Machine Learning

With this method, the supervised machine learning algorithms learn a language model from manually annotated data, which can then be applied to new and unseen data. The categories aimed were drug *compliance* and *non-compliance*. The unit processed was the message. Several sets of features were exploited: the vectorized text of messages (forms, lemmas, and lexical lemmas) and the ATC indexing of drugs. The train set contained 94 non-compliant messages and 93 compliant messages. The test set contained 39 non-compliant messages and 40 compliant messages: we used two thirds of the reference data for training and one third for testing as shown in Figure 1.

We used the Weka [13] implementation of several supervised algorithms: NaiveBayes [14], Bayes Multinomial [15], J48, Random Forest, and Simple Logistic. These algorithms were used with their default parameters and with the string to word vector function.

Categorization with Information Retrieval

The information retrieval system was exploited to make the distinction between relevant and irrelevant messages with this method. This approach was unsupervised, although we took advantage of the reference data as well. We used the Indri information retrieval system in two ways:

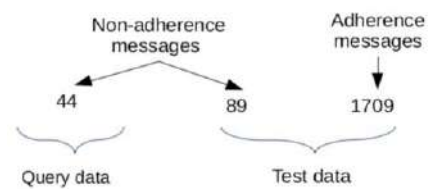


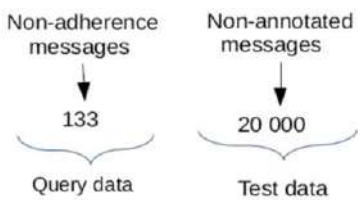
Figure 2 – Datasets Used for Global Information Retrieval

At the global level, we distinguished between drug compliant and non-compliant messages. The corpus was split into two sets: (1) 44 non-compliance messages (one third of the whole non-compliance category) were used for the creation of queries, and the query lexicon was weighted proportional to its frequency in the messages; (2) All compliance messages and 89 non-compliance messages (two thirds of the whole non-compliance category) were used for the evaluation as shown in Figure 2. The question we wanted to answer was whether the subset of non-compliance messages permited to retrieve other non-compliance messages. The evaluation was done automatically, computing Precision, Recall, and F-measure

with each version of the corpus (forms, lemmas, and lexical lemmas). This may give an idea of the performance of this method when searching similar information in new non-annotated data.

At the fine-grained level, we looked for various types of drug non-compliance. The question we wanted to answer was whether the messages already assigned to each non-compliance type can help in retrieving other similar messages and enrich the reference data through this unsupervised approach. This issue was particularly important for types with scare data availability: as we indicated above, some types contained only two messages. As previously, the messages from different non-compliance types were exploited to create queries. All of the annotated non-adherence data was used to build the queries. These queries were applied to a large corpus of 20,000 randomly selected messages that contained at least one mention of a drug. The results were evaluated manually computing the Precision, which mainly allows query optimize the queries. This is described in Figure 3.

Figure 3 – Datasets Used for Fine-grained Information Retrieval



Results and Discussion

Table 1– Machine Learning Results Obtained for the Categorization of Messages into the Non-compliance Category

	Precision	Recall	F-mesure
	NaiveBayes		
Forms	0.769	0.769	0.769
Lemma	0.786	0.846	0.815
Lexical lemmas	0.761	0.897	0.824
	NaiveBayesMultinomial		
Forms	0.732	0.769	0.750
Lemmas	0.795	0.795	0.795
Lexical lemmas	0.786	0.846	0.815

Categorization with Supervised Machine Learning

The results of categorization of messages into the *non-compliance* category, obtained with supervised machine learning algorithms, are presented in Table 1. We tested several algorithms but show only the results for the two best algorithms, Naive Bayes and Naive Bayes Multinomial. We can observe that the best results (up to 0.824 F-measure) are obtained on the lexical lemmas corpus. In all the experiments, Recall is higher than or equal to Precision.

Among the errors observed with NaiveBayes, 12 messages were wrongly categorized as *non-compliant* and 9 as *compliant*. Within these 12 messages, four contained terms associated with excess and negative effects (such as "*Je n'imaginais pas que c'était si grave*" (*I didn't imagine it was that bad*) or "*s'il vous plait ne faites pas n'importe quoi*" (*please don't make a mess*), usually specific to *non-compliance* messages.

Categorization with Information Retrieval

The results obtained with the information retrieval system Indri are presented in Table 2. The evaluation values are computed for the top 10, 20, 50 and 100 results. With lower cut-off (10, 20, 50), the Recall is limited by the efficiency of the system and also by the cut-off. As a matter of fact, the optimal Recall would be 0.112 at 10, 0.225 at 20 and 0.561 at 50, although it is hardly possible to reach such results. With this experiment, the best results (up to 0.529 F-measure) are obtained with the lexical lemmas corpus. Besides, the lemmatization shows an important improvement over the forms corpus, which means that lemmatization is important for information retrieval applications because it provides linguistic normalization of the corpus. As expected, the values of Recall and Precision are improved with the increase in the evaluated sample. Hence, there is more probability that the 89 relevant messages are found among the top 100 messages. With up to 0.5 of Precision, a human user is able to quickly find relevant messages, making this solution usable as an exploration tool. Overall, we can see that this information retrieval system can find non-compliant messages although the results are noisy.

At the fine-grained level, we tested several queries focusing on precise types of non-compliance and misuse of drugs. We will present queries and their results related to important drug misuse situations such as gain and loss of weight, recreational drug use, suicide attempts or ideas, and overdoses. The top 20 results are analyzed for each query. On the basis of messages that convey the expected contents similar to a given query and related to the aimed drug misuses, we can compute the Precision of the results.

Gaining/losing weight. The keywords used were *poids, kilo, grossir, maigrir* (*weight, kilo, gain weight, lose weight*), such as suggested by the manually built reference data. This query was applied to the lemmatized corpus. We expected to find mainly messages related to the use of drugs with the purpose to intentionally lose or gain weight and also some messages related to weight changes due to side effects of drugs. In reality, among the top 20 messages, 17 were about weight change as side effects of drugs, one message was about the use of drugs to lose weight intentionally, and two messages were about weight loss but with no relation to drugs. This means that, among the top 20 messages, only one new relevant message was found. It may correspond to the reality (misuse of drugs for weight changes was less frequent than weight change due to drug side effects) or to the corpus used (several messages were related to antidepressant drugs that have common side effect of weight change). This query gives 0.05 Precision.

Table 2 – Information Retrieval Results for the Categorization of Messages into the Non-compliance Category

	Precision	Recall	F-mesure
	Top 10 Results		
Forms	0.100	0.011	0.020
Lemma	0.400	0.045	0.081
Lexical lemmas	0.400	0.045	0.081
	Top 20 Results		
Forms	0.250	0.056	0.091
Lemmas	0.350	0.079	0.129
Lexical lemmas	0.300	0.067	0.109
	Top 50 Results		
Forms	0.340	0.191	0.244
Lemmas	0.400	0.045	0.081
Lexical lemmas	0.420	0.236	0.302

		Top 100 Results	
Forms	0.480	0.539	0.508
Lemmas	0.480	0.539	0.508
Lexical lemmas	**0.500**	**0.561**	**0.529**

Recreational drug use. The main purpose was to find messages in which prescription drugs were used with recreational objectives, such as looking for high sensations, hallucinations, sensations of happiness. We tried several queries:

- First, the keywords *drogue, droguer* (*non-medical drug, to take non-medical drugs*) were used. In French, the word *drogue* usually refers to street drugs, and not to prescription drugs. Yet, in the corpus, people use this word for neuroleptic medication particularly, in order to illustrate their feeling that these drugs open the way to addictions and have the same neuroleptic effects as the street drugs. Hence, we could find messages such as "*J'ai été drogué pendant 3 ans au xanax*" (*I was drugged with xanax for 3 years*) or "*Sa soulage mais ses une vrai drogue ce truc !!!*" (*It helps but this stuff is really a drug!!!*) These queries found interesting results (15 out of 20) but provide different insights than those expected;
- The keywords *hallu, allu, hallucination* (*hallucination*) were used. Among the top 20 messages, 2 messages were about intentionally seeking hallucination effects caused by some drugs, 7 messages were about people experiencing hallucinations but as unwanted side effects, 11 messages were about people suffering from hallucinations and taking drugs to reduce them;
- Finally, the keyword *planer* (*to be high from drugs*) was used. Among the top 20 messages, 19 were about the high effect of drugs, be it intentional (9 messages) or non-intentional (10 messages). For instance, "*J'ai déjà posté quelques sujets à propos de ce fléau qu'est le stilnox (...) je prends du stilnox, pour m'évader, pour planer*" (*I already posted a few topics about this plague that is stilnox (...) I take stilnox, to escape, to get high*).

Overall, these three queries related to recreational use of drugs gave 0.35 Precision on average.

Suicide. The keyword used was *suicide* (*suicide*). The query was applied to the lemmatized corpus. We expected to find messages in which people report on taking drugs (like antidepressants) or planning to do that with suicidal intentions. Among the top 20 results, 9 messages were about drugs and suicide with no particular relation between them, 5 other messages were about the fact that some drugs may increase the risk of suicide, 5 messages were critical about the fact that drugs may increase the risk of suicide, and one message reported on a real suicide attempt caused by drug withdrawal. Discussions on relation between drugs and suicide, and of course reporting on suicide attempt, may be important for our research because they represented the importance of these topics in the analyzed fora. This query gave 0.7 Precision.

Overuse. The keyword used is *boites* (*boxes*) because it often represented the quantity of drugs taken, in case of overuse, in the reference data. This query was applied to the *forms* corpus because it was important to preserve plural forms for this query. Among the top 20 messages, six messages were directly related to drug overuse, three messages were related to high dosage that corresponded to overuse, two messages related to the suicide attempts by ingestion of large amounts of drugs, two messages proposed to share unused prescription drugs, and seven messages were unrelated to overuse. This query gave 0.65 Precision.

Comparison of the Two Categorization Approaches

On one hand, supervised machine learning shows better results for the specific categorization task, but it heavily depends on the availability of manual annotations that are costly to produce. On the other hand, information retrieval methods provide lower precision. Yet, manual filtering of the information retrieval results may be less costly than manual annotation of the data. Besides, information retrieval provides results that are more easily understandable and usable by users. Hence, depending on users and availability of the reference data, either of the approaches may be preferred. The Machine Learning method reaches high Recall, suggesting that this method can detect up to 80% of non-adhering patients, provided that they talk about it in the forums.

In relation to small categories (types of misuses, such as those related to drug overuse or recreational use of drugs), supervised machine learning usually performs poorly, while information retrieval may achieve interesting results due to the definition of suitable queries: 0.45 average Precision, and up to 0.65-0.70 Precision for some queries. This may be useful for human users when they want to quickly find relevant messages. We can see that Precision numbers obtained for these small categories are higher than those obtained at the more global level (*compliance* and *non-compliance* categories) and presented in Table 2. This means that more precise and targeted information may be more reliable for the information retrieval process, and also for enriching categories for which there is little data available.

Overall, we assume that these two approaches are complementary: combination of their results may provide an efficient way to enrich the reference data. Besides, the approaches can also be combined: information retrieval queries can bootstrap the enrichment of categories and thus help supervised machine learning to perform better.

Limitations of the Current Work

The main limitation related to the machine learning categorization is the reduced size of the reference data. It contains indeed only 133 messages in the *non-compliance* category. Yet, these reference data allow the creation of quite efficient categorization models, which reach up to 0.824 F-measure. We assume that availability of larger reference data will improve the overall performance of the results. As explained above, one of the main motivations to exploit information retrieval methods is the possibility to enrich the reference data with this unsupervised approach.

Yet, the information retrieval approach requires manual analysis and evaluation of the retrieved messages. As we have seen, at the global level, the increase in size of the top sample improves overall results, but relevant messages are found together with irrelevant messages. When information retrieval is used at a fine-grained level, the results vary according to the types of non-compliance and the keywords used. We propose that information retrieval methods can be used for targeted enrichment of the corpus (for instance for some types of non-compliance), for manual exploration of corpora by health professionals (the results can be easily understood comparing to the results provided by machine learning algorithms), and for combining the results provided by this and other methods.

Another limitation of the work is that messages detected as cases of non-compliance are not currently fully analyzed by medical doctors, pharmacists, or for pharmacovigilance. This

method needs to be packaged into a software application easy to use by medical professionals in clinical settings. On one hand, the messages found further in our experiments permit to have clear insights in the real use of drugs, which is a very important issue that motivates our work. On the other hand, when methods are efficient enough, their results can be used by concerned experts (pharmaceutical industry, public health, general practitioners) to prepare and provide prevention and education actions to patients and their relatives. For example, packaging of drugs can be further adapted to their real use, dedicated brochures and discussions can be done with patients on known and possible drug side effects, and on necessary precautions.

Conclusions

This work presented exposition of two approaches for the detection of drug non-compliance situations in Internet fora. We used the French forum Doctissimo together with other fora to cover several disorders. The messages were first manually assigned to *compliance* and *non-compliance* categories. Automatic categorization with machine learning approach using NaiveBayes showed 0.824 F-measure, while with information retrieval approach using Indri, it showed 0.60 Precision at top 10 results and 0.34 at top 50 results. Information retrieval was also used for a more fine-grained categorization of messages at the level of individual types of non-compliance. Four topics were addressed with different queries suggested by messages available in the reference data. This provided 0.45 average Precision, and up to 0.65-0.70 Precision for some queries, as computed for the top 20 results. We also observed that with information retrieval, precise and targeted categories showed better Precision than the one obtained at a more global level of compliance and non-compliance messages. We considered this as encouraging point for the processing of categories with few messages available. We also proposed some thoughts on combination of two approaches (supervised machine learning and information retrieval) for enriching the reference data and for generating more efficient supervised models.

The main objective of the current work was to enrich the reference data and to work more closely with health professionals for their use.

Acknowledgements

This work has been performed as part of the DRUGSSAFE project funded by the ANSM, France and of the MIAM project funded by the ANR, France within the reference ANR-16-CE23-0012. We thank both programs for their funding. We would also like to thank the annotators who helped us with the manual annotation of misuses, Bruno Thiao Layel for extracting the corpus, Vianney Jouhet and Bruno Thiao Layel for building the list with drugs names, and The-Hien Dao for the set of disorders used. Finally, we thank the whole ERIAS team for the discussions.

References

[1] Daugherty T, Eastin M, Bright L. Exploring Consumer Motivations for Creating User-Generated Content. Journal of Interactive Advertising 2008;8.

[2] Louis A. Natural language processing for social media. Computational Linguistics 2016;42(4):833–6.

[3] Collier N. Towards cross-lingual alerting for bursty epidemic events. J Biomed Semantics 2011;2(5).

[4] Lejeune G, Brixtel R, Lecluze C, Doucet A, and Lucas N. Added-value of automatic multilingual text analysis for epidemic surveillance. In: Artificial Intelligence in Medicine (AIME), 2013.

[5] Tapi Nzali M. Analyse des médias sociaux de santé pour évaluer la qualité de vie des patientes atteintes d'un cancer du sein. Thèse de doctorat, Université de Montpellier, Montpellier, France, 2017.

[6] Morlane-Hondère F, Grouin C, and Zweigenbaum P. Identification of drug-related medical conditions in social media. In: LREC, 2016:1–7.

[7] Kalyanam J, Katsuki T, Lanckriet GRG, and Mackey TK. Exploring trends of nonmedical use of prescription drugs and polydrug abuse in the twitter- sphere using unsupervised machine learning. Addictive Behaviors 2017;65:289–95.

[8] Cameron D, Smith GA, Daniulaityte R, et al. PREDOSE: a semantic web platform for drug abuse epidemiology using social media. 2013;46(6):985–97.

[9] Skrbo A, Begović B, and Skrbo S. Classification of drugs using the atc system (anatomic, therapeutic, chemical classification) and the latest changes. Med Arh 2004;58(2):138–41.

[10] Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 1960;20(1):37–46.

[11] Landis J and Koch G. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.

[12] Schmid H. Probabilistic part-of-speech tagging using decision trees. In: ICNMLP, Manchester, UK. 1994:44–9.

[13] Witten I and Frank E. Data mining: Practical machine learning tools and techniques. Morgan Kaufmann, San Francisco, 2005.

[14] John GH and Langley P. Estimating continuous distributions in bayesian classifiers. In: Kaufmann M, ed, Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo. 1995:338–45.

[15] McCallum A and Nigam K. A comparison of event models for naive bayes text classification. In: AAAI workshop on Learning for Text Categorization, Madison, Wisconsin. 1998.

Address for correspondence

Elise Bigeard
CNRS, Univ Lille, UMR 8163 STL - Savoirs Textes Langage,
F-59000 Lille, France
e-mail: bigeard@limsi.fr.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190178

Carnival: A Graph-Based Data Integration and Query Tool to Support Patient Cohort Generation for Clinical Research

David Birtwell[a], Heather Williams[a], Reed Pyeritz[b], Scott Damrauer[c], Danielle L. Mowery[a,d]

[a] *Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA, USA,*
[b] *Department of Medicine, University of Pennsylvania, Philadelphia, PA, USA*
[c] *Department of Vascular Surgery, University of Pennsylvania, Philadelphia, PA, USA*
[d] *Department of Biostatistics, Epidemiology, & Informatics, University of Pennsylvania, Philadelphia, PA, USA*

Abstract

Clinical research studies often leverage various heterogeneous data sources including patient electronic health record, online survey, and genomic data. We introduce a graph-based, data integration and query tool called Carnival. We demonstrate its powerful ability to unify data from these disparate data sources to create datasets for two studies: prevalence and incidence case/control matches for coronary artery disease and controls for Marfan syndrome. We conclude with future directions for Carnival development.

Keywords:

biomedical research, cohort studies, information storage and retrieval

Introduction

In biomedicine, clinical research studies are conducted to understand how best to prevent, diagnose, or treat disease in patients. A fundamental step to conducting a clinical research study is aggregating data and abstracting facts (i.e., clinical variables and treatment outcomes pertinent for defining a patient population for study). Patient data containing clinical facts (e.g., administrative, clinical, and genomic data) are generated at various points of care, and subsequently are stored across disparate, siloed resources. For example, large academic medical centers may store patient data within clinical registries, electronic health records, document stores, survey tools, and biobanks. Once study data have been integrated, the patient data must then be modeled to accurately represent and classify the patient's clinical case according to each study arm (i.e., case/control). For example, in a matched case/control study design, a clinical researcher might match patients based on age, biological sex, race, and genes, then their clinical data (disease, treatments, and outcomes) might be queried to understand disease progression and/or the effectiveness of therapeutic interventions. More complex match criteria can include periods of time, geographical locations, and environmental exposures. These clinical facts are stored in structured (hospital billing codes, laboratory, and medications) and/or unstructured (clinical notes) data formats.

Integrating such diverse data and aggregating patient clinical facts in a traditional relational database is challenging for several reasons. First, complex relationships of the data cannot easily be modelled into a sufficiently expressive relational schema. Second, complex relationship queries for generating and matching patient cohorts using relational databases could be hindered by suboptimal query responses (e.g., stalled or unfulfilled requests due to multiple-join statements) [1]. Graph databases such as Neo4j have been shown to support both semantic integration of disparate data [2] while improving query times over traditional relational databases such as MySQL [1]. Although graph databases have been used to integrate and represent biological disease networks (protein-protein interactions and drug-target pairs) and represent genotype-phenotype associations, few have demonstrated how graph databases might be leveraged to query heterogeneous data, clinical and genomic, to generate patient cohorts for clinical research studies [3–6].

In this work, we present ***Carnival***, a data unification technology that takes a novel approach — a strictly-formatted property graph database with a data model inspired by the Open Biological and Biomedical Ontology (OBO) Foundry ontologies — towards the integration of disparate data into a unified graph data resource. Carnival leverages this model to support the execution of common investigatory tasks, i.e., patient cohort identification, automated case/control matching, and the production of data sets for scientific analysis. For this work, we aim to 1) provide an overview of Carnival's infrastructure, 2) review a menu of predefined operations that can be combined and stacked to query, integrate, and reason over clinical and genomic data for creating case/control study populations, 3) present two case/control cohorts generated by Carnival, and 4) preview how Carnival will support semantic interoperability and intelligent queries using ontologies, leverage textual variables using natural language processing, and improve its usability with a graphical user interface for wider adoption by clinical research partners.

Methods

We describe the infrastructure, functionality, and utility of Carnival for supporting case/control studies from clinical and genomic data collected from the University of Pennsylvania Health System (UPHS). UPHS includes the first university-owned teaching hospital (Hospital of the University of Pennsylvania est. 1874) and the first hospital in the United States (Pennsylvania Hospital est. 1751). As part of the Penn Medicine BioBank (PMBB), over 60,000 UPHS patients have been consented for their clinical (EHR) and genomic (blood and tissue samples) data to be studied for clinical research. These biological specimens have been whole-exome sequenced by Regeneron Genetics Center and the context of their collection is represented using the Ontology for BioBanking (OBIB) [7]. Carnival leverages data from UPHS sources (e.g., Penn Data Store, a clinical data warehouse) and PMBB to generate datasets for clinical research studies.

Carnival

We describe our graph-based data unification tool named Carnival, so named because Carnival is a party of information and inspired by the squash. Carnival is a multi-layered, Groovy-powered application that includes data source adapters (*vines*) that extract lightly-processed data from their sources, a set of relational database utilities, and caching functions for querying source data, graph data writers (*reapers*) that attach extracted data to the Carnival data store (*graph)*, logical rule engines (*reasoners*) that modify and validate data within the graph, sample stratifiers (*algorithms*) to perform common tasks such as case-control matching, and data writers (*sowers*) that write data to external resources (see Figure 1).

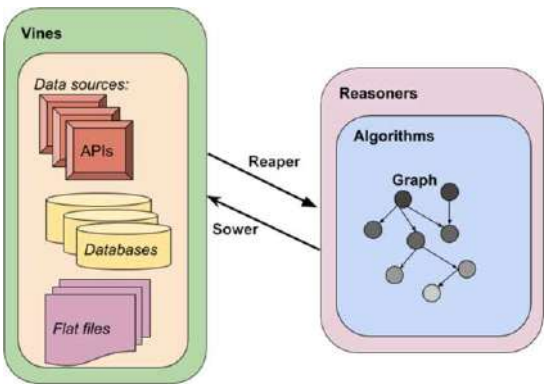


Figure 1– Carnival's Conceptual Framework

Vines

Data are pulled as needed from source systems (i.e., relational databases--Oracle, Microsoft SQL Server), application programming interfaces (API), or flat files (comma-separated value files). Typically, the vine contains methods (functions) that converse with the associated data source in its native language (e.g., SQL, Java, HTTP, etc.) to extract data. Each method produces a single matrix of data. Although there is nothing besides scale issues preventing the wholesale transfer of data from a source to the Carnival graph, vine functions are designed to extract the minimum data necessary for scientific study. For example, a vine for an EHR source might include an ICD Ever/Never method that accepts a set of ICD codes and patient identifiers then returns only the patient identifiers for patients whose medical record was assigned one or more of the ICD codes in the set.

Performing queries against large disparate data sources can be problematic for a variety of technical reasons: 1) connections can be dropped, 2) complex queries can take hours to return a result, and 3) queries that contain long lists of codes or identifiers can be cumbersome to compose. Carnival provides a suite of supportive classes and methods to address these difficulties, including SQL utilities for query composition, automated caching of vine method results, incremental caching to support restartable queries, and monitor threads that estimate time-to-completion of long-running queries.

Reapers

Reapers contain methods that extract data from source systems via vines and attach those data to the Carnival graph in a standardized way. A reaper method may be parameterized to limit its scope and call upon any number of vines to gather source data. The execution of the reaper method itself is recorded in the graph, including the inputs and outputs of the process. For example, an ICD Ever/Never reaper method that accepts a set of ICD codes and patient identifiers would execute a series of tasks: 1) create a vertex in the graph to represent the execution of the reaper method, 2) record the start time, finish time, and input ICD codes, 3) create links to the patient identifiers as inputs, 4) create vertices to represent the ICD Ever/Never status for each patient, 5) link those statuses to the appropriate patients, and 6) create links to those statuses from the execution vertex to mark them as outputs.

Graph

All data about a patient and metadata describing Carnival's processes are represented in Carnival's property graph structure. These highly-connected graphs contain meta information regarding the execution of the reapers, algorithms, and sowers queried and traversed using Tinkerpop-Gremlin and Cypher. We drew upon previous work in ontology modelling as inspiration for the Carnival graph data model: the Ontology for Biobanking (OBIB), the Ontology for Biomedical Investigations (OBI), and the Basic Formal Ontology (BFO) [7–9]. For example, BFO introduces the term Process, which denotes an event that occurs in a time and place. OBI introduces Planned Process, which extends Process to include a pre-defined plan, participants, inputs, and outputs. In the Carnival graph, healthcare encounters are modelled as planned processes, where participants include the patient and clinician and the outputs may be diagnoses and medications.

Reasoners

Reasoners apply logical rules to the graph to make modifications or additions to data. Reasoners execute their logic against the graph to validate that the graph is consistent with the reasoner logic. For example, a reasoner might assign a Boolean classification to patients as having or not having a disease state based on whether they have ever been assigned any of a set of ICD codes. The reasoner logic would check the ICD Ever/Never statuses of each patient and assign the appropriate Boolean disease state classification. The validation functionality might check that no patients have been assigned to both the have-disease and does-not-have-disease classes in the final dataset, which represents a logical inconsistency.

Operational Algorithms

Carnival leverages a graph-based, stratified-sampling case/control algorithm to match case patients with control patients for clinical research studies (i.e., case and control strata groups with equal numbers of males between the age of 35-40). Strata creation is managed by a strata manager class that takes as input the criteria for group partitioning, creates strata and strata group vertices in the graph, and assigns patient vertices to the appropriate groups.

First, the strata sampler groups patients by primary strata. In each stratum, patients are exhaustively grouped into disjoint sets. Strata can be defined along any singular value extracted from a patient: a *numeric* (e.g., current body mass index [BMI]), a *string* (race), a *Boolean* (e.g., FBN1 gene loss of function), or an *enumerated value set* (e.g., the existence of 2 or more ICD codes). Strata can be defined by a range (e.g., 20-29) for numerics and by multiple values grouped into the same strata group for enumerated values or strings. Each stratum also has a group for undefined or unknown values. Second, the cohort matcher then takes as input: patient cohorts for cases, candidate controls, and the primary strata that correspond to the desired matching criteria. The primary strata are combined to make a compound stratification that characterizes the cohorts for matching. For each strata group in the compound stratification, a number of patients in the candidate control cohort are selected as controls

corresponding to the number of patients in the case cohort from the same strata group. Additional criteria can also be provided to prioritize which of the controls within a group would be selected if there are more potential controls than necessary. For example, in a study where specimens associated with the patients will be expended, it may be advantageous to prioritize controls that have more specimens available to maximize the utility of PMBB specimens.

Sowers

Sowers extract data from the Carnival stratified case/control population graph and write it to external resources (e.g., databases, applications, and flat files). Data imported from disparate sources and linked in the Carnival graph may be queried producing data views useful for export to external systems. For example, an investigator may have an approved study that monitors patients who have a specific disease state, the data for which are tracked in a REDCap project. A Carnival sower can query the graph for disease states and write those data to the REDCap project via the REDCap API. In this way, Carnival supports extract, transform, and load (ETL) operations. This functionality can be leveraged to add new patient information (e.g., genomic facts into the EHR).

Clinical Data Elements for Case/Control Matching

Currently, case/control cohorts can be defined and analyzed using several clinical data elements shown in Table 1.

Table 1– FBN1 Controls

Clinical Data Element	Example
Demographics	current age, biological sex, race/ethnicity
Vital signs and risk factors	current BMI, weight, height, blood pressure, smoking status
Medications	administered at a given time or during time window
Specimen collection contexts	age or BMI at the time of specimen collection, medications administered before/after date of specimen collection
Hospital administration	ICD9/10, CPT procedures, DRG codes, and fee codes
Clinical status	death status
Genetic data	loss of function genes

Building a Cohort Using Operational Building Blocks

Carnival contains a number of operational algorithms or building blocks that can be combined to support cohort identification and case-control selection: *graph building, reasoners,* and *patient cohort algorithms*. *Graph building algorithms* aggregate data, via vines and reapers, and attach those data to the graph, examples include:

- Instantiating patients and encounters with identifiers
- Computing phecode assignments according to Phewas.org
- Computing BMI closest to the patient's PMBB recruitment date
- Obtaining the most recent and earliest healthcare encounter for each patient
- Calculating each patient's date of birth
- Gathering each patient's available specimens.

Reasoners operate over data in the graph. To support cohort building, Carnival contains a patient stratification algorithm that assigns patients to specific strata, which currently correspond to their current age and biological sex. A case-control matching algorithm creates a cohort of controls for a predefined set of cases based on a cohort of candidate controls and selected strata. *Patient cohort algorithms* facilitate the creation of patient cohorts and perform set operations on them. Current patient cohort algorithms include:

- Creating a cohort based on a set of identifiers, e.g., medical record number (MRN), enterprise master patient index (EMPI), or encounter identifier
- Creating a complement cohort that contains all patients not in an existing cohort
- Creating a cohort containing all patients who have been diagnosed with any of a set of ICD codes
- Creating a cohort containing all patients who have a loss of function mutation for a given gene.

We describe how Carnival uses clinical data elements and operational algorithms to support two case/control studies leveraging patient data from UPHS and PMBB.

Identifying Markers for Coronary Artery Disease

For the first case/control study, Carnival was leveraged to define the study populations for discovering biological markers for coronary artery disease. In the PMBB, blood specimens are collected at the time of enrollment. *Cases* are defined as patients without coronary artery disease (CAD) codes (410, I21) before enrollment into PMBB, and subsequently, have a myocardial infarction (MI), coronary revascularization, or another CAD event. *Controls* are defined as individuals without CAD codes who have had no MI or CAD events after PMBB enrollment.

Both cases and controls must have plasma and either buffy coat or DNA specimens available. A report of demographic and phenotype data relative to the date of enrollment was requested. Carnival was leveraged to generate the case/control list and report the following patient-specific information:

- Age at PMBB enrollment
- Current age
- Biological sex
- EHR race
- PMBB recruitment location
- Height, weight, and BMI measurement closest to PMBB enrollment
- Count, min, max, and median of lab results for: glucose, fasting glucose, hemoglobin A1C
- Count of distinct dates diabetes codes were assigned before PMBB enrollment
- Code and age of the patient for the first code matching the time filter for the most recent CAD codes before PMBB enrollment, most recent MI code before PMBB enrollment, first MI or CAD code after PMBB enrollment, and first revascularization code after PMBB enrollment.

The principal investigator (PI) manually reviewed the report and patient charts for the final case selection. When patient cases were chosen, Carnival generated two sets of controls. *Control group 1*: a 1:1 control instance matched to the case population on biological sex, current age +- 4 years, BMI closest to enrollment +- 6 points, and recruitment location. *Control group 2:* a frequency control matched to the case population on sex,

recruitment location, current age, and BMI closest to PMBB enrollment.

Defining Marfan Syndrome Study Controls

For this second case/control study, *Cases* were predefined by using the MRNs of patients provided by the PI. *Controls* were defined by Carnival. The Marfan Syndrome gene loss of function data value was critical for defining the control cohort. Candidate control matches were queried from PMBB based on age and biological sex. For each patient, Carnival determined whether a genetic assessment was conducted for the Marfan syndrome FBN1 gene and validated whether the outcome was negative to ensure that all patients within the control set were FBN1 negative.

Results

We leveraged Carnival to generate cohorts for two diseases: coronary artery disease and Marfan syndrome.

Identifying Markers for Coronary Artery Disease

Based on the CAD definitions for prevalence and incidence and specimen availability, Carnival identified 425 candidate case and 5,872 candidate control populations. The principal investigator conducted chart review to create the final 170 selected cases, then Carnival generated the final control groups 1 (n=170 cases) and 2 (n=60 cases) for these cases. Carnival identified a 1:1 case/control match (see Table 2).

Table 2– Coronary Artery Disease Cases and Controls

Sex	Age	Candi-date Cases	Selected Cases	Candi-date Controls	Control Group 1	Control Group 2
Female	<20	0	0	7	0	0
Female	20-29	0	0	486	0	1
Female	30-39	5	2	681	2	1
Female	40-49	8	7	599	6	2
Female	50-59	13	11	731	13	6
Female	60-70	39	20	814	23	3
Female	>70	68	31	450	27	10
Male	<20	0	0	1	0	0
Male	20-29	2	2	98	1	0
Male	30-39	3	2	156	2	2
Male	40-49	15	9	272	9	3
Male	50-59	40	24	529	23	5
Male	60-70	92	27	682	30	18
Male	>70	140	35	366	34	9
Totals		**425**	**170**	**5,872**	**170**	**60**

Defining Marfan Syndrome Study Controls

Using the exclusion criteria of patients who are not cases (case distribution not shown) and who do not have a loss of function mutation in the FBN1 gene, Carnival identified a 2:1 control/case match resulting in 146 selected controls (see Table 3).

Table 3– Marfan Syndrome Controls

Sex	Age	Selected Controls	Candidate Controls
Female	20-40	2	294
Female	40-50	12	300
Female	50-60	10	667
Female	60-70	18	1154
Female	>70	20	1718
Male	20-40	4	175
Male	40-50	12	343
Male	50-60	12	1004
Male	60-70	26	1892
Male	>70	30	2627
Totals		**146**	**10,174**

Discussion and Future Work

We demonstrated how Carnival's multi-layered framework, OBO foundry-inspired data model, and graph-based functionality can be leveraged to generate case/control matches and resulting datasets using PMBB and UPHS sources for two clinical research studies. We envision expanding Carnival's functionality 1) to improve semantic integration and intelligent query of data using ontologies, 2) to incorporate clinical data elements generated from clinical texts, and 3) to create a user-friendly interface to promote wider adoption of this tool by clinical research partners.

Improve Semantic Integration and Query using Ontologies

Carnival is a graph-based query tool that operates at the data level. We have partnered Carnival with TURBO technologies to provide richer semantic integration and reasoning among clinical data elements [10]. For example, patient diagnoses are currently defined using logical rules that operate over discrete diagnosis billing codes (e.g., ICD9 or ICD10). A single code rarely encapsulates a diagnosis for investigative purposes. Furthermore, compiling a list of diagnosis codes relevant to a particular disease or disease classification can be a daunting task. To accurately identify a diagnosis, TURBO has integrated ontologies (e.g., Monarch Disease Ontology) [11], to semantically link ICD codes to disease concepts. To provide semantic information services (i.e., returning a set of ICD diagnosis codes for a given disease classification), we have integrated Drivetrain [12], a TURBO technology that uses an RDF triple store and OBO Foundry ontologies [13]. This integration has permitted Carnival to operate at both the disease classification and diagnosis code levels.

Medication prescriptions have similarly benefitted from integration with Drivetrain. The Chemical Entities of Biological Interest (ChEBI) ontology [14,15] contains a rich semantic network of medication names, ingredients, and roles that Drivetrain links with medication order names in the UPHS EHR. Carnival operates over these elements, obviating the need for investigators to spend time deciphering individual medication order names. We aim to provide semantic integration over laboratory test results as well, when Drivetrain services for lab results become available.

Incorporate Clinical Data Elements from Clinical Notes

A wealth of clinical facts are locked within clinical free-text notes (e.g., discharge summaries, progress notes, radiology exams, and surgical pathology reports) [16]. We aim to integrate outputs from natural language processing tools including symptoms, signs, treatments, outcomes as well as their associated contexts

(negation, subject, temporality, uncertainty) to provide a richer clinical profile to infer each patient's disease state [17–21]. We will leverage the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) to promote interoperability with entities both within and outside UPHS.

Create a User-Friendly Interface to Promote Adoption

The current implementation of Carnival leverages server-side technology. There is currently a text-based command line interface that supports basic operations. Future work will include the addition of a web-services API and a browser-based Javascript user interface [22]. We will integrate Neo4j-based user interfaces to the graph (e.g., Neo4j Browser) [23]. We will also employ Javascript libraries (e.g., D3) [24] for graph visualization and user-driven data exploration.

Conclusions

We conclude that graph-based technologies can be utilized to integrate and query disparate patient health data to support complex, clinical research studies.

References

[1] B.-H. Yoon, S.-K. Kim, and S.-Y. Kim, Use of Graph Database for the Integration of Heterogeneous Biological Data, *Genomics Inform.* **15** (2017) 19–27.

[2] H. Ulrich, A.-K. Kock-Schoppenhauer, P. Duhm-Harbeck, and J. Ingenerf, Using Graph Tools on Metadata Repositories, *Stud. Health Technol. Inform.* **253** (2018) 55–59.

[3] A. Lysenko, I.A. Roznovăț, M. Saqi, A. Mazein, C.J. Rawlings, and C. Auffray, Representing and querying disease networks using graph databases, *BioData Min.* **9** (2016) 23.

[4] V. Touré, A. Mazein, D. Waltemath, I. Balaur, M. Saqi, R. Henkel, J. Pellet, and C. Auffray, STON: exploring biological pathways using the SBGN standard and graph databases, *BMC Bioinformatics*. **17** (2016) 494.

[5] W.S. Campbell, J. Pedersen, J.C. McClay, P. Rao, D. Bastola, and J.R. Campbell, An alternative database approach for management of SNOMED CT and improved patient data queries, *J. Biomed. Inform.* **57** (2015) 350–357.

[6] S. Mughal, I. Moghul, J. Yu, UKIRDC, T. Clark, D.S. Gregory, and N. Pontikos, Pheno4J: a gene to phenotype graph database, *Bioinformatics*. **33** (2017) 3317–3319.

[7] M. Brochhausen, J. Zheng, D. Birtwell, H. Williams, A.M. Masci, H.J. Ellis, and C.J. Stoeckert Jr, OBIB-a novel ontology for biobanking, *J. Biomed. Semantics*. **7** (2016) 23.

[8] Ontology for Biobanking | NCBO BioPortal, (n.d.). https://bioportal.bioontology.org/ontologies/OBIB (accessed November 25, 2018).

[9] Basic Formal Ontology (BFO) | Home, (n.d.). http://basic-formal-ontology.org (accessed November 25, 2018).

[10] C. Stoeckert, D. Birtwell, H. Freedman, M. Miller, and H. Williams, ICBO_2018_12: Transforming and Unifying Research with Biomedical Ontologies: The Penn TURBO project, in: International Conference on Biomedical Ontology (ICBO 2018), International Conference on Biological Ontology, 2018. http://icbo2018.cgrb.oregonstate.edu/.

[11] O.T. Wg, Monarch Disease Ontology. http://www.obofoundry.org/ontology/mondo.html (accessed November 25, 2018).

[12] The TURBO Ontology, *Turbo-Documentation*. https://pennturbo.github.io/Turbo-Documentation/turbo-ontology.html (accessed November 25, 2018).

[13] J.M. Hancock, OBO Foundry, in: Dictionary of Bioinformatics and Computational Biology, 2004.

[14] ChEBI (chemical entities of biological interest data based), in: Encyclopedia of Genetics, Genomics, Proteomics and Informatics, 2008: pp. 321–321.

[15] P. de Matos, R. Alcántara, A. Dekker, M. Ennis, J. Hastings, K. Haug, I. Spiteri, S. Turner, and C. Steinbeck, Chemical Entities of Biological Interest: an update, *Nucleic Acids Res.* **38** (2009) D249–D254.

[16] S. Velupillai, D. Mowery, B.R. South, M. Kvist, and H. Dalianis, Recent Advances in Clinical Natural Language Processing in Support of Semantic Analysis, *Yearb. Med. Inform.* **10** (2015) 183–193.

[17] OHDSI, OHDSI/CommonDataModel, *GitHub*. https://github.com/OHDSI/CommonDataModel (accessed November 25, 2018).

[18] D.L. Mowery, S. Velupillai, B.R. South, L. Christensen, D. Martinez, L. Kelly, L. Goeuriot, N. Elhadad, S. Pradhan, G. Savova, and W. Chapman, Task 2: ShARe/CLEF eHealth evaluation lab 2014, *CLEF 2014 Online Working Notes*. **1180** (2014) 31–42.

[19] D.L. Mowery, P. Jordan, J. Wiebe, H. Harkema, J. Dowling, and W.W. Chapman, Semantic annotation of clinical events for generating a problem list, *AMIA Annu. Symp. Proc.* **2013** (2013) 1032–1041.

[20] D.L. Mowery, B.E. Chapman, M. Conway, B.R. South, E. Madden, S. Keyhani, and W.W. Chapman, Extracting a stroke phenotype risk factor from Veteran Health Administration clinical reports: an information content analysis, *J. Biomed. Semantics*. **7** (2016) 26.

[21] D.L. Mowery, S. Velupillai, and W.W. Chapman, Medical diagnosis lost in translation: analysis of uncertainty and negation expressions in English and Swedish clinical texts, in: 2012 Workshop on Biomedical Natural Language Processing, Association for Computational Linguistics, 2012: pp. 56–64.

[22] M. Löbe, User Expectations of Metadata Repositories for Clinical Research, *Stud. Health Technol. Inform.* **253** (2018) 60–64.

[23] The Neo4j Browser: A User Interface Guide for Beginners, *Neo4j Graph Database Platform*. (n.d.). https://neo4j.com/developer/guide-neo4j-browser/ (accessed November 25, 2018).

[24] M. Bostock, D3.js - Data-Driven Documents, (n.d.). https://d3js.org/ (accessed November 25, 2018).

Address for Correspondence

David Birtwell: birtwell@pennmedicine.upenn.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190179

Text Classification to Inform Suicide Risk Assessment in Electronic Health Records

André Bittar[a], Sumithra Velupillai[a,b], Angus Roberts[a], Rina Dutta[a,c]

[a] *Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK,*
[b] *School of Electrical Engineering and Computer Science, KTH, Stockholm, Sweden*
[c] *South London and Maudsley NHS Foundation Trust, London, UK*

Abstract

Assessing a patient's risk of an impending suicide attempt has been hampered by limited information about dynamic factors that change rapidly in the days leading up to an attempt. The storage of patient data in electronic health records (EHRs) has facilitated population-level risk assessment studies using machine learning techniques. Until recently, most such work has used only structured EHR data and excluded the unstructured text of clinical notes. In this article, we describe our experiments on suicide risk assessment, modelling the problem as a classification task. Given the wealth of text data in mental health EHRs, we aimed to assess the impact of using this data in distinguishing periods prior to a suicide attempt from those not preceding such an attempt. We compare three different feature sets, one structured and two text-based, and show that inclusion of text features significantly improves classification accuracy in suicide risk assessment.

Keywords:

Suicide, Risk Assessment; Natural Language Processing

Introduction

Suicide is a serious public health problem, with almost one million people ending their lives worldwide each year [1]. In the United Kingdom, although suicide rates have dropped slightly since the early 1980s, in 2017 the Office for National Statistics nevertheless registered 5,821 suicides [2]. More than a quarter are in receipt of mental health services at the time of death [3], yet suicide risk remains immensely difficult for clinicians to assess, given the wide range of contributory factors, with the majority (88%) judged to be at 'low or no immediate risk' of suicide by clinicians at their final service contact. Current clinical methods for assessing when someone is at risk of a suicide attempt have been reported to be little better than chance [4]. New approaches to individualised risk assessment that integrate data from different sources are needed. With the availability of population-level patient data in the form of electronic health records (EHRs), novel methods for suicide risk assessment based on data mining and machine learning have been explored in recent years [5]. Indeed, machine learning models can capture complex associations between variables, making them particularly well-suited to the task of predictive analysis.

Different approaches to modelling the problem of suicide risk assessment have been explored. Several recent studies have focused on developing models that detect the suicide risk of individual patients in large historical population samples. For instance, Barak-Corren et al. (2017) trained a Naïve Bayes classifier on a cohort of more than 1.7 million patients (16,588 of whom had recorded suicidal behaviour) using a rich set of structured EHR features, including demographic, diagnostic, procedure and medication data [6]. Simon et al. (2018) combined structured EHR data and standard health questionnaire responses in a logistic regression model to predict suicide attempt and death in a cohort of nearly 3 million patients [7]. Similar work has also been carried out using state-of-the-art neural networks. For example, Bhat and Goldman-Mellor (2017) trained neural network classifiers on historical structured EHR data to detect the presence of a suicide attempt by individual patients in a given year [8].

A common factor in the aforementioned work and, more generally in the majority of research using machine learning for suicide risk assessment, is that it has relied exclusively on features derived from the structured fields of EHRs. However, much valuable data about patients is stored in EHRs as unstructured text [9]. To date, relatively little research on suicide risk has been done on mining this rich data source for features, although momentum is building. For example, Metzger et al. (2017) tested a series of machine learning classifiers to determine the prevalence of suicide-related emergency department admissions in a French hospital [10]. They used structured EHR data, as well as features extracted from clinical notes. Ben-Ari and Hammond (2015) performed a text search to identify Gulf War veterans who have made a suicide attempt [11]. They then used textual and structured features with a Random Forest classifier to predict first suicide attempts by patients in this cohort within a given year. McCoy et al. (2016) also used text-based features along with structured EHR data to gauge suicide risk after discharge [12]. They used an "off-the-shelf" Natural Language Processing (NLP) tool to extract the polarity of valence-conveying words (positive or negative) within the records and used this in regression models, finding that positive valence words were correlated with reduced suicide risk. Downs et al. (2017) used NLP to identify suicide-related mentions in a cohort of adolescents with Autism Spectrum Disorder, including a previously-developed negation detection module [13, 14]. Finally, Fernandes et al. (2018) used NLP to identify and classify mentions of suicidal ideation and suicide attempts in mental health records [15].

Methods

Classification for Suicide Risk Assessment

Most epidemiological case-control studies have used data spanning much wider time periods, typically years, and the risk factors have either been static (e.g., male gender, family history of psychiatric disorder) or lifetime ever variables (e.g., previous attempted suicide, any misuse of alcohol or drugs) [16]. In this study, we take a rather different approach, exploring whether it is possible to predict suicide attempts by building classification models based on structured and textual data from the 30-day period leading up to the event, when

there is an opportunity to intervene as it can be a time of crisis. To the best of our knowledge, our approach is novel in using this critical time period. We tested a machine learning classifier to distinguish the 30-day periods prior to a hospital admission linked to a suicide attempt (hereafter referred to as 'suicidal window') and similar periods not preceding a suicide attempt from age- and sex-matched control patient records ('non-suicidal windows') extracted from a large database of EHRs. We used a linear-kernel Support Vector Machine (SVM) [17] classifier given this algorithm's well established performance in dealing with the high-dimensionality of text data [18]. We modelled the task using supervised binary classification. A further novel aspect of our work lies in our assessment of the impact of three different sets of features, namely, structured fields from the EHRs, a rich set of binary features derived from the clinical notes, and a bag-of-words representation of the full text of these same documents.

CRIS Clinical Cohort

We studied the de-identified EHRs of over 250,000 patients from the South London and Maudsley (SLaM) NHS Foundation Trust using the Clinical Record Interactive Search (CRIS) computer system comprising both structured data and over 3.5 million text documents [19]. Data from CRIS has been linked with the UK Hospital Episode Statistics (HES) data for Admitted Patient Care within a secure 'safe haven', and it is through this linkage that admission information was extracted. The documents in CRIS have been substantially enhanced, in particular, through the application of NLP (e.g. to identify symptoms) [20].

Our dataset was derived from the EHRs of 17,640 patients. It consisted of 21,175 suicide-related (case) and non-suicide-related (control) admissions, sampled according to a 1:4 case-control ratio. Cases were defined as any admission (acute physical or specialist mental health) where there was a suicide attempt (indicated by the presence of any of the following ICD codes: X6*, X7*, X80-4*, Y1*, Y2*, Y30-4*, Y87*) with the admission lasting at least 24 hours (starting and ending on different dates). Only admissions with a start date after the 1st of April 2006 and an end date before or including 31st March 2017 were considered. Of these case admissions only those which had at least 1 document in the 30 days prior and including the date of the hospitalised suicide attempt were retained. We also removed admissions with empty documents[1]. This left a total of 4,235 suicide-related admissions in the final dataset. Each control was matched by sex, had to be alive at the admission start date of the case, and were grouped into the same age group as cases (5 year age bands < 16, 16-19, 20-24 to 80-84, 85+ years). Each control also had at least one document in the 30 days prior to and including the date of their matched case's hospitalised suicide attempt. The total number of controls was 16,940. The controls were chosen to be as representative as possible of the population from which the cases were drawn and the ratio was based on the epidemiological principle that little statistical power is gained by further increasing the number of controls beyond approximately 4 per case [21]. Key descriptive characteristics of the dataset are shown in Table 1.

Features for Classification

Features examined included standard sociodemographic and clinical descriptors selected by a clinical academic psychiatrist (RD) *a priori* (e.g. ethnic group, marital status, employment status), as well as those shown in prior work to be associated with suicide attempts (e.g. past and current substance abuse) [22]. We used 14 (categorical) features from structured fields of the EHRs, including total document counts for the 30 days prior to but excluding the day of admission, based on the hypothesis that the volume of documentation would increase prior to a risk event such as a suicide attempt due to greater service use. We also included features derived from NLP applications routinely run on CRIS. Most of these applications are built using GATE, an open source NLP toolkit [20, 23]. All of these applications combine both rule-based pattern matching algorithms and supervised machine learning models, and some detect contextual information, such as negation or family history. We derived 68 binary features from these, one per application, each feature indicating the presence or absence of at least one match by the application on at least one document in the window, excluding the admission date. Applications included the detection of positive mentions of suicide attempt by the patient, mentions of the patient having disturbed sleep, and feelings of hopelessness and paranoia, to cite a few. Finally, we also included as features the concatenated text from all documents in each window, excluding the admission date, represented as a TFIDF (term frequency-inverse document frequency) vector or "bag-of-words". TFIDF is a statistical weighting that reflects how important a particular term is in a given document in a collection, adjusted for the fact that some words occur more frequently than others [24]. When calculating word vectors, we applied the L2-Norm [25] to scale the vectors to unit length. This compensates for discrepancies in document length, such as those in our dataset. The total number of unique words in our document set was 201,538, making it a high-dimensional feature space.

	Cases	**Controls**
Patients	2,913 (16.5%)	14,727 (83.5%)
Female	1,730 (59.4%)	8,971 (60.9%)
Male	1,183 (41.6%)	5,756 (39.1%)
Admissions	4,235 (20%)	16,940 (80%)
Female	2,598 (61.3%)	10,392 (61.3%)
Male	1,637 (38.7%)	6,548 (38.7%)
Mean age (SD) years	34.4 (15.3)	34.4 (15.4)
EHR features for 30-day pre-admission windows		
Mean tokens (SD)	3455.5 (5732.4)	1344.5 (3179.9)
Total tokens	14,634,223	22,775,227
Mean docs (SD)	16.9 (31.4)	7.5 (18.1)
Total docs	71,404	127,047

Table 1: Characteristics of the dataset for suicide attempt related cases and non-suicide-related controls. Note that each window for EHR features is a 30-day period prior to the case admission date (inclusive).

Henceforth, we refer to the set of structured data as STRUCT, the text-based features derived from NLP applications routinely run on CRIS as GATE, and the bag-of-words features as TFIDF. For STRUCT and GATE, we encoded all features either as integers (e.g. age, marital status, employment status) or binary (0 or 1) values (e.g. sex, presence of the keyword *depression*). We list details of all 82 features in an online annex[2].

[1] Text from scanned documents is not always available.

[2] https://github.com/KCL-Health-NLP/medinfo2019-sa-risk/

Results

Experimental Setup

We randomly split the data into a training set (80%) and a test set (20%), ensuring the distribution of each class was the same across both sets (i.e., 1 case to 4 controls). The test data was held out until the final run in order to reduce the risk of overfitting and to provide a realistic estimation of performance on unseen data. We scaled all features to have zero-mean and unit-variance to ensure a balanced contribution of all features. We implemented the classifier and prepared data using the Scikit-learn (version 0.20.0) machine learning library for Python [26]. Our first step was to estimate the optimal parameters for the classifier (model tuning). We did this using grid search and ten-fold cross-validation on the training data, with F1-score as the evaluation metric[3]. F1-score is the harmonic mean of precision (positive predictive value) and recall (sensitivity), a metric often used in information retrieval and NLP [27]. For each tuning instance, we varied the feature set so as to tune the models to each representation of the data. This resulted in 7 different feature combinations: STRUCT, GATE, TFIDF, STRUCT+GATE, STRUCT+TFIDF, GATE+TFIDF, STRUCT+GATE+TFIDF. A flowchart of the general architecture is provided in the online annex. To gauge whether the differences in pairwise comparisons of feature sets were statistically significant, we used McNemar's test [28] (α=0.05) for classification disagreements on the training dataset.

Finally, we examined the text features that were most significant in distinguishing the two classes. For the final run on the held-out test data, we calculated the F1-score for the suicidal and non-suicidal windows and the mean of both.

Intensity of Documentation

As shown in Table 1, the mean number of documents for cases during the 30-day window is 19.9 (SD=34.0) and a lower 8.3 (SD=19.4) for controls. This divergence in mean document counts for each class supports the decision to include the number of documents in a window as a feature for classification. See Feature Importance for further comments.

Classification

In this section, we present results obtained on the held-out test data for each of the different feature sets. We report performance in terms of precision (P), recall (R) and F1-score (F). Where we make a comparison of results obtained on two feature sets, we also report the *p*-value of McNemar's pairwise test between them. Although we calculated figures for each of the two classes (suicidal windows and non-suicidal windows) separately, we are only interested in assessing the identification of suicidal windows. The correct classification of non-suicidal windows was relatively much simpler given the prevalence of this class in the dataset (mean F1-score for this class was 0.86, SD=0.08). Therefore, the figures we report are for suicidal windows only.

The classifier's performance using only structured features (STRUCT feature set) was relatively low (P=0.26, R=0.59, F=0.36). This increased with the addition of features extracted by GATE (STRUCT+GATE feature set) (P=0.49, R=0.58, F=0.53, p<0.001). However, best performance was obtained with the addition of the bag-of-words features (STRUCT+GATE+TFIDF feature set) (P=0.61, R=0.63, F=0.62, p<0.001), showing a good balance between precision and recall. Interestingly, bag-of-word features alone (TFIDF) provided the next best results (P=0.59, R=0.61, F=0.60, p=0.0042), only slightly lower than the combination of all feature sets, suggesting the importance of the bag-of-words representation. The use of GATE features on their own performed less well (P=0.50, R=0.57, F=0.54, p<0.001). This indicates that the TFIDF features captured a signal in the data

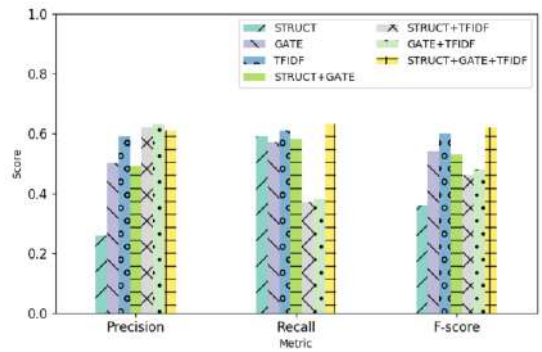


Figure 1 – Performance on the positive class (suicide attempt) in terms of precision (P), recall (R) and F1-score (F) on the test set for all feature sets.

that the targeted NLP applications did not. Using the combination of GATE features and bag-of-words (GATE+TFIDF), recall was heavily penalised, resulting in a significant reduction to the model's F1-score, despite increased precision (P=0.63, R=0.38, F=0.48, p<0.001). The combination of structured and bag-of-words features (STRUCT+TFIDF) provided a significant improvement over structured features alone (P=0.62, R=0.37, F=0.46, p<0.001), but also with a sharp increase in precision to the detriment of recall. Figure 1 provides a visual comparison of these results while Figure 2 shows a summary of all pairwise McNemar's tests across all feature sets.

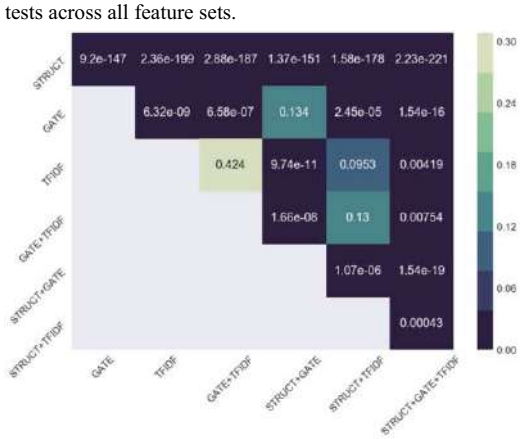


Figure 2 – Full results of pairwise McNemar's tests across feature sets on the training data. Cells show p-values (α=0.05), dark cells are more statistically significant, while lighter cells indicate the converse.

Feature Importance

During training, the (linear) SVM calculates a maximum-margin hyperplane to separate (as much as possible) the two classes in the data. The feature weights representing the coordinates of the vector orthogonal to the hyperplane are stored and their direction indicates the predicted class. We

[3] All tuning parameter ranges and the final tuned configurations are provided in the online annex.

compared the values of these weights for each of the feature sets. The number of documents within the window is among the top-ranking features in the STRUCT feature set, indicating the importance of this feature in discriminating between classes. The most discriminating feature in the GATE feature set was the presence of a positive mention of a suicide attempt by the patient. These were also the top features in the STRUCT+GATE feature set. The top two features for the TFIDF feature set were the words *overdose* and *self-harm*, while other words, such as the adjective *suicidal*, and the medication names *paracetamol* and *zopiclone*, ranked among the top 15 discriminating features. More diverse keywords, as well as other features, ranked highly when TFIDF was combined with either STRUCT or GATE, or both. Given these results, we quantified and visualised the relative frequency of these terms within the texts of the suicidal and non-suicidal windows. To offset the imbalance in the dataset (differing mean number of tokens between the two groups) we calculated per-token frequencies for each day and divided term counts for control by 4 (to adjust for the 1:4 case-control ratio). Figure 3 shows a comparison for the term *overdose*. Mentions of this term are approximately four to six times more frequent for cases than for controls across the 30-day window. Plots for the other aforementioned terms showed the same trend. These preliminary findings on the TFIDF feature set suggest that significant mentions of suicide-related behaviour (overdose, self-harm, etc.) have been recorded in the suicidal windows, but not in the non-suicidal windows. Furthermore, despite the relatively low per-token frequency of the significant terms, the TFIDF weighting allowed for these to be picked up as important features by the classifier. A more in-depth and systematic examination of the occurrences of these terms would establish whether they represent independent suicide attempts that are not recorded using ICD codes in the HES data linked to the EHRs. Nevertheless, these preliminary results support the use of such word features in classification to inform suicide attempt risk assessment.

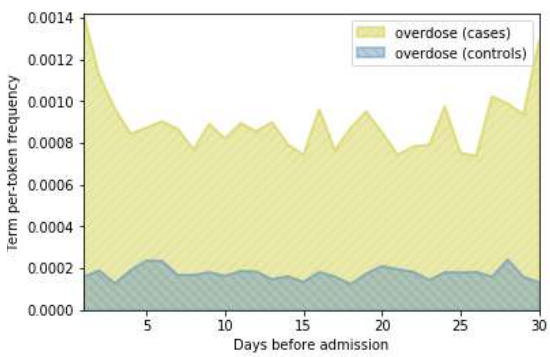


Figure 3 – Normalised relative per-token frequency of the term "overdose" for suicidal (case) and non-suicidal (control) windows.

Discussion

Using a case-control study design, we carried out an exploratory experiment on binary text classification to assess how this might help to inform suicide risk assessment. We evaluated the relative contributions of three different types of features, namely structured data, binary features derived from NLP applications routinely run on CRIS, and a TFIDF bag-of-words representation. Our results show that the use of text features significantly improves classification results, and the combination of structured and text-based features provided the best performance. An examination of the top textual features used in classification revealed the importance of certain terms for discriminating suicidal and non-suicidal windows.

The variety of ways in which the suicide risk assessment task has been modelled previously, including differences in data sets, algorithms and features, makes meaningful comparison of results between studies difficult and such was not the aim of this work.

Despite interesting results, clearly this work does have limitations. Firstly, the 14 structured features were selected to represent only a small sample of all available structured data. This means that although these features are clinically relevant for assessing suicide risk, certain features with less complete data were not tested. A broader selection of structured data may have led to better results with the STRUCT feature set. Another drawback lies in the two representations we used for the textual features. The binary (GATE) feature set is unable to account for the relative frequencies of identified terms. Thus, a single match by an application has the same "weight" as multiple matches in the document set. Although the TFIDF bag-of-words representation addresses this weakness, it does not capture the order and combination of words (e.g. multiword expressions such as *suicidal ideation*), or phenomena such as negation (e.g. *no suicidal ideation*). Furthermore, our bag-of-words approach did not enable us to distinguish terms relating to the patient from those concerning other people (e.g. *family history of suicide, father took an overdose*). Whilst this was accounted for to some extent in the GATE features, more nuanced and targeted NLP could improve performance.

Conclusions

We have shown that the inclusion of text features in classification to inform suicide risk assessment using EHR data provides a statistically significant increase in performance over a dataset containing only structured data. Including the text allows access to word features that appear to be potential markers of impending suicide risk (*overdose, self-harm, suicidal*) that are also clinically plausible. Strikingly, the intensity of documentation within the 30-day period prior to an event may also be a significant factor in determining times of increased risk.

Acknowledgements

RD is funded by a Clinician Scientist Fellowship (research project e-HOST-IT) from the Health Foundation in partnership with the Academy of Medical Sciences which also funds AB. SV is supported by the Swedish Research Council (2015-00359), Marie Skłodowska Curie Actions, Cofund,

Project INCA 600398. AR is funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London.

References

[1] World Health Organization, Mental health: suicide prevention, (2018). http://www.who.int/mental_health/suicide-prevention/en/ (accessed November 2, 2018).

[2] Office for National Statistics, Suicides in the UK: 2017 registrations, (2018). https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/suicidesinth

eunitedkingdom/2017registrations (accessed September 25, 2018).

[3] The assessment of clinical risk in mental health services. National Confidential Inquiry into Suicide and Safety in Mental Health (NCISH), The University of Manchester, 2018.

[4] J.C. Franklin, J.D. Ribeiro, K.R. Fox, K.H. Bentley, E.M. Kleiman, X. Huang, K.M. Musacchio, A.C. Jaroszewski, B.P. Chang, and M.K. Nock, Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research., *Psychol. Bull.* **143** (2017) 187–232.

[5] E. Baca-García, I. Basurte-Villamor, J.M. Leiva-Murillo, M. de Prado-Cumplido, R. Santiago-Mozos, A. Artés-Rodríguez, and J. de Leon, Using Data Mining to Explore Complex Clinical Decisions: A Study of Hospitalization After a Suicide Attempt, *J Clin Psychiatry*. (2006) 9.

[6] Y. Barak-Corren, V.M. Castro, S. Javitt, A.G. Hoffnagle, Y. Dai, R.H. Perlis, M.K. Nock, J.W. Smoller, and B.Y. Reis, Predicting Suicidal Behavior From Longitudinal Electronic Health Records, *Am. J. Psychiatry*. **174** (2017) 154–162.

[7] G.E. Simon, E. Johnson, J.M. Lawrence, R.C. Rossom, B. Ahmedani, F.L. Lynch, A. Beck, B. Waltzfelder, R. Ziebell, R.B. Penfold, and S.M. Shortreed, Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records, *Am. J. Psychiatry*. **175** (2018) 951–960.

[8] H.S. Bhat, and S.J. Goldman-Mellor, Predicting Adolescent Suicide Attempts with Neural Networks, *ArXiv171110057 Cs Stat*. (2017). http://arxiv.org/abs/1711.10057 (accessed October 25, 2018).

[9] R. Stewart, and K. Davis, 'Big data' in mental health research: current status and emerging possibilities, *Soc. Psychiatry Psychiatr. Epidemiol.* **51** (2016) 1055–1072.

[10] M.-H. Metzger, N. Tvardik, Q. Gicquel, C. Bouvry, E. Poulet, and V. Potinet-Pagliaroli, Use of emergency department electronic medical records for automated epidemiological surveillance of suicide attempts: a French pilot study: text-mining and epidemiology of suicide attempts, *Int. J. Methods Psychiatr. Res.* **26** (2017).

[11] A. Ben-Ari, and K. Hammond, Text Mining the EMR for Modeling and Predicting Suicidal Behavior among US Veterans of the 1991 Persian Gulf War, in: 2015 48th Hawaii Int. Conf. Syst. Sci., IEEE, HI, USA, 2015: pp. 3168–3175.

[12] T.H. McCoy, V.M. Castro, A.M. Roberson, L.A. Snapper, and R.H. Perlis, Improving Prediction of Suicide and Accidental Death After Discharge From General Hospitals With Natural Language Processing, *JAMA Psychiatry*. **73** (2016) 1064.

[13] J. Downs, S. Velupillai, G. Gkotsis, R. Holden, M. Kikoler, H. Dean, A. Fernandes, and R. Dutta, Detection of Suicidality in Adolescents with Autism Spectrum Disorders: Developing a Natural Language Processing Approach for Use in Electronic Health Records, *Proc. AMIA Annu. Symp.* (2017) 641–649.

[14] G. Gkotsis, S. Velupillai, A. Oellrich, H. Dean, M. Liakata, and R. Dutta, Don't Let Notes Be Misunderstood: A Negation Detection Method for Assessing Risk of Suicide in Mental Health Records, in: Proc. Third Workshop Comput. Lingusitics Clin. Psychol., Association for Computational Linguistics, San Diego, CA, USA, 2016: pp. 95–105.

[15] A.C. Fernandes, R. Dutta, S. Velupillai, J. Sanyal, R. Stewart, and D. Chandran, Identifying Suicide Ideation and Suicidal Attempts in a Psychiatric Clinical Research Database using Natural Language Processing, *Sci. Rep.* **8** (2018).

[16] K. Hawton, C. Casañas i Comabella, C. Haw, and K. Saunders, Risk factors for suicide in individuals with depression: A systematic review, *J. Affect. Disord.* **147** (2013) 17–28.

[17] C. Cortes, and V. Vapnik, Support-vector networks, *Mach. Learn.* **20** (1995) 273–297.

[18] T. Joachims, Text categorization with Support Vector Machines: Learning with many relevant features, in: C. Nédellec, and C. Rouveirol (Eds.), Mach. Learn. ECML-98, Springer Berlin Heidelberg, Berlin, Heidelberg, 1998: pp. 137–142.

[19] G. Perera, M. Broadbent, F. Callard, C.-K. Chang, J. Downs, R. Dutta, A. Fernandes, R.D. Hayes, M. Henderson, R. Jackson, A. Jewell, G. Kadra, R. Little, M. Pritchard, H. Shetty, A. Tulloch, and R. Stewart, Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) Case Register: current status and recent enhancement of an Electronic Mental Health Record-derived data resource, *BMJ Open*. **6** (2016).

[20] R.G. Jackson, R. Patel, N. Jayatilleke, A. Kolliakou, M. Ball, G. Gorrell, A. Roberts, R.J. Dobson, and R. Stewart, Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project, *BMJ Open*. **7** (2017).

[21] D.A. Grimes, and K.F. Schulz, Compared to what? Finding controls for case-control studies, *The Lancet*. **365** (2005) 1429–1433.

[22] M.A. Ilgen, M.L. Burnette, K.R. Conner, E. Czyz, R. Murray, and S. Chermack, The association between violence and lifetime suicidal thoughts and behaviors in individuals treated for substance use disorders, *Addict. Behav.* **35** (2010) 111–115.

[23] R.G. Jackson, M. Ball, R.P. BMBCh, R.D. Hayes, R.J.B. Dobson, and R. Stewart, TextHunter – A User Friendly Tool for Extracting Generic Concepts from Free Text in Clinical Research, (2014) 10.

[24] A. Rajaraman, and J.D. Ullman, eds., Data Mining, in: Min. Massive Datasets, Cambridge University Press, Cambridge, 2011: pp. 1–17.

[25] Roger A. Horn, and Charles R. Johnson, Norms for Vectors and Matrices, in: Matrix Anal., Cambridge University Press, Cambridge, England, 1990.

[26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, and D. Cournapeau, Scikit-learn: Machine Learning in Python, *Mach. Learn. PYTHON*. (n.d.) 6.

[27] N. Chinchor, MUC-4 Evaluation Metrics, in: Proc. 4th Conf. Message Underst. MUC4 92, Stroudsburg, PA, USA, 1992: p. 8.

[28] Q. McNemar, Note on the sampling error of the difference between correlated proportions or percentages, *Psychometrika*. **12** (1947) 153–157.

Address for correspondence

André Bittar, andre.bittar@kcl.ac.uk

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190180

An Automated Detection System of Drug-Drug Interactions from Electronic Patient Records Using Big Data Analytics

Guillaume Bouzillé[a], Camille Morival[b], Richard Westerlynck[c], Pierre Lemordant[a], Emmanuel Chazard[d], Pascal Lecorre[b], Yann Busnel[c], Marc Cuggia[a]

[a] Univ Rennes, CHU Rennes, Inserm, LTSI – UMR 1099, F-35000 Rennes, France,
[b] Laboratoire de Pharmacie Galénique, Biopharmacie et Pharmacie Clinique, IRSET U1085, Faculté de Pharmacie, Université de Rennes 1, F-35043, Rennes Cedex, France,
[c] IMT Atlantique, F-35576, Cesson-Sévigné, France,
[d] Univ Lille, CHU Lille, CERIM EA2694, F-59000 Lille, France

Abstract

The aim of the study was to build a proof-of-concept demonstratrating that big data technology could improve drug safety monitoring in a hospital and could help pharmacovigilance professionals to make data-driven targeted hypotheses on adverse drug events (ADEs) due to drug-drug interactions (DDI). We developed a DDI automatic detection system based on treatment data and laboratory tests from the electronic health records stored in the clinical data warehouse of Rennes academic hospital. We also used OrientDb, a graph database to store informations from five drug knowledge databases and Spark to perform analysis of potential interactions betweens drugs taken by hospitalized patients. Then, we developed a machine learning model to identify the patients in whom an ADE might have occurred because of a DDI. The DDI detection system worked efficiently and computation time was manageable. The system could be routinely employed for monitoring.

Keywords:

Computing Methodologies, Drug Interaction, Machine Learning.

Introduction

Drug-drug interactions (DDIs) are a critical issue in patient care because they can lead to adverse events and ultimately increase care costs and patient mortality. Therefore, these events must be identified and prevented as early as possible [1]. However, many new drugs are released each year, and therefore, it is very difficult for healthcare professionals to be informed and to consider all DDIs. Moreover, the alarm functionalities of drug computerized physician order entry (CPOE) systems are frequently not used because they do not focus on clinically relevant DDIs and lead users to alarm fatigue. Although focused on specific interactions, studies on DDI prevalence show the existence of risks for polymedicated patients and highlight the importance of pharmacovigilance programmes [2,3].

With the unprecedented development of digital health and hospital clinical data warehouses (CDW), data produced during the healthcare process are now easily reusable [4]. Electronic health records (EHR) contain real-time information on drug prescription/regimens during hospitalization as well as all clinical information. Such data could be analysed to estimate DDI prevalence, to facilitate health professionals' practice assessment and to detect the occurrence of DDI-linked adverse drug events (ADE). In France, pharmacovigilance currently relies mainly on the spontaneous reporting by physicians or/and detection of diagnoses that could be related to ADE from the hospital billing system (diagnosis related group, DRG, database). New data sources, such as national claim databases, are also leveraged to improve DDI and ADE detection [5,6]. EHR data-mining also could help pharmacovigilance professionals to improve drug safety assessment.

All these health-related databases fit perfectly with the big data paradigm because they contain voluminous, highly complex and heterogeneous information that is produced in real time [7]. In the last few years, many big data technologies have been developed. However, their implementation in a hospital information system for processing healthcare big data in real-world condition of use is still largely uncharted.

Here, we describe a method, which propose to use big data technology to improve drug safety monitoring in a hospital and could help pharmacovigilance professionals to make data-driven targeted hypothesis on ADEs.

Methods

Figure 1 presents the overall approach of the study and the big data technologies used in each step.

Patient Data

We used the Rennes academic hospital EHRs that are stored in a CDW called eHOP (entrepot HOPital). This CDW includes both structured data (e.g., laboratory results, drug prescriptions and regimens) and unstructured data (e.g., operative reports, discharge summaries), and is dedicated to data reuse for clinical research [8]. The eHOP's star schema architecture and graphic user interface allows researchers, even without any database language knowledge, to quickly access and efficiently search information within millions of patient records.

For this study, we used information about drug administrations (used drug(s) and regimens) and laboratory results (date, nature of the test and results: normal, abnormally high, or abnormally low).

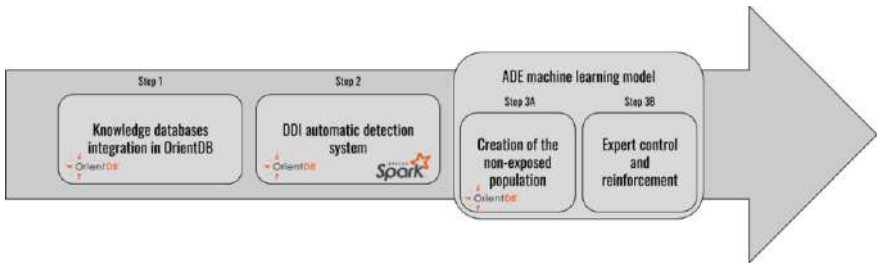


Figure 1 - Overall approach of the study

Knowledge Databases Integration (step 1)

To identify and collect information on potential DDIs, and also to compare information from different sources, we selected five drug knowledge databases: Thesaurus, Vidal, Theriaque, Micromedex and Drugs.com [9–13]. These databases are commonly used by health professionals, but are not specifically targeted to DDI detection. They are available via a web application programming interface (API) that requires a specific procedure because each database stores data with its own structure. To avoid this, we extracted the relevant information from these databases and stored it in OrientDB, a graph-oriented model database [14] that fits well with our objective because a DDI can be modelled as an edge between two drugs. Thus, once the information is stored in a single OrientDB database, no more computation is required to access such information.

DDI Automatic Detection System from Patient Records (Step 2)

For DDI identification, we collected drug data from the patient EHRs stored in eHOP and computed the active interval (i.e., the period during which a drug was effective) for all drugs taken by a patient during the hospital stay. If two active intervals overlapped (fully or partially), then analysis of the data collected in the OrientDB database allowed determining whether the two drugs interacted. In this case, the potential DDI event was stored in eHOP. As these are independent processes (each drug pair is checked independently), the Spark cluster-computing framework was used to perform distributed computing [15,16]. As all the potential DDI events can be stored in eHOP, then we could compute the prevalence of a DDI for any specific drug, molecule, or population.

Creation of a Machine Learning Model (Step 3a)

The data stored in the CDW eHOP do not allow direct confirmation of whether a patient reported a DDI-linked ADE or not. Indeed, this needs to be validated by the pharmacovigilance experts who do not have the proper means to check all the patient records. Therefore, we wanted to create a system to report to drug safety professionals only the most interesting cases among all DDIs detected by the DDI automatic detection system (i.e., patients in whom an ADE might have occurred because of a DDI).

We assume that laboratory results will change if an ADE occurs. So, we can train a machine learning model with two populations: those who experienced an ADE and those who did not. Unfortunately, we cannot identify manually who experienced an ADE. For this reason, we performed one of the research design presented by Hennessy et al. [17]: we choose to compare the population exposed to a DDI with another population non-exposed to this DDI and who did not experience an ADE, by design. There are likely many patients who do not experience an ADE in the exposed population, but the model will present only the most suspected cases and this problem will be solved with the gradual feedback of drug safety professionals: the system will adjust weights of patients in the model, giving a greater weight to the well-predicted patients.

We developed an artificial neural network system that allows us to predict an output. This system has a single hidden layer and the number of perceptrons was decided during cross-validation. Our machine learning model works in two phases. First, it uses all data available for patients who experienced a specific DDI and those who did not (exposed and non-exposed populations) to classify them as having reported an ADE or not. Then, the model is reinforced with information coming from drug safety professionals who infirm or confirm the previous classification (Fig. 2).

We then had to form the non-exposed population. Within a DDI, we called "Object" the drug under study, and "Precipitant" the other drug. Moreover, we called "Control-precipitant" any drug that has the same therapeutic use as the Precipitant, but that does not interact with the Object. For a given Object, we compared the exposed population, found with the DDI automatic detection system, to the non-exposed population. The non-exposed population included all patients, who were not in the exposed population and who had an overlap (fully or partial) between the action interval of the Object and of the Control-precipitant. We created this non- exposed population using the same process as for the exposed population.

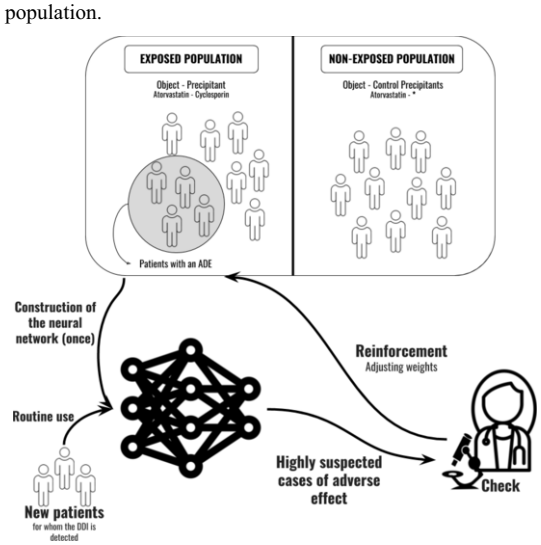


Figure 2 - Creation and use of the artifical neural network

Data processing was performed with Java 8, Spark 2.10 and OrientDB 2.2.4 on Intel(R) Xeon(R) CPU E5-2609 1,90GHz computer with 32,0 Go of RAM.

Big Data Technologies: Convenient Tools for Complex Data Processing

Here, we proposed a complete automated data treatment system, from the collection of heterogeneous data to their enhancement in a machine learning model. This system can monitor DDI prevalence and try to identify patients with a possible DDI-linked ADE, without the intervention of drug safety professionals. To achieve this, we used several convenient tools:

OrientDB is an easy-to-use tool to store pre-computed data. The OrientDB database model includes two main classes: vertices and edges that connect two vertices. In our study, the "vertex" interface represented the class "Drug" and included drug name, ID-code and half-life. The "edge" interface represented the class "Interactions" and included DDI severity level. We also specified from which drug database the information on the DDI came. Thus, via OrientDB, each drug knowledge database can be interrogated separately. The "edge" interface is also used to represent the class "Control-precipitant".

Figure 3 presents the database model through an example: Drug1 has an interaction with Drug3 according two different databases (two edges of class "Interaction"). Let consider the Object-Precipitant couple Drug1-Drug3, then Drug4 is a control-precipitant of Drug3 (one oriented edge of class "Control-precipitant"). An example of query would be: "give all the drugs that have an interaction with Drug1 according Micromedex and where the severity level is 1".

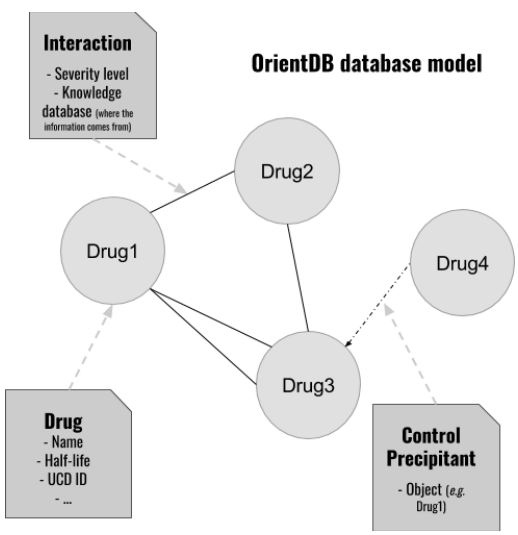


Figure 3 - OrientDb database model

The query language was very close to the structured query language (SQL) and allowed searching a vertex that walks along edges to another vertex, according to the chosen conditions. Data uploading is fast and based on a convenient Java Graph API. We manipulated a graph Java object that is automatically committed at the end of the process. Moreover, if access to a part of the graph is required (e.g., all the drugs that interact with pravastatin according to a severity level of 2), we used this object as a temporary store before processing.

Switching between different knowledge databases, stored in the same OrientDB database, involves only a variable on an edge. Ultimately, the little amount of time spent for the pre-calculation facilitates the storage and the access to multiple data sources. Only one kind of request is needed for all five databases. We could easily add information from other data sources (for example, composition of a drug and half-life of the active substances), or more precise information about DDI-linked ADEs (such as the relevant laboratory tests). Nevertheless, this task demands a manual work for each group of drugs [18].

Spark allows parallel processing easily. As many processes are independent from each other, their parallel treatment with Spark leads to a big time saving[19].

Evaluation (Step 3B)

To evaluate our DDI detection system (step 2 in Fig 1), we focused on a class of drugs called statins that are prescribed (long-term treatment) to patients with cardiovascular diseases, and particularly to elderly patients who are usually polymedicated and consequently prone to DDIs. We selected the study population (i.e., all patients taking statins) from all patients included in eHOP from January, 1 2015 to July, 8 2016. It included 10,506 hospitalized patients with a median hospitalization of 7 days, and a median age of 72 years (range: 19 to 98 years).

We defined statins as the "Object" and all the drugs that interact with them were considered as candidate "Precipitants". We selected as Control-precipitants (symbolized by * in fig 3) all the drugs that are in the same fifth level (i.e., chemical substance) as the Precipitant in the Anatomical Therapeutic Chemical (ATC) classification [20], but do not interact with the Object. Thus, Control-precipitants have the same (or a similar) therapeutic usage as the Precipitant. We stored all these data in OrientDB because each DDI is a link (i.e., edge) between drugs (i.e., two vertices).

Concerning the action intervals, we chose a period of seven half-lives for each statin molecule and arbitrarily selected one day for the Precipitant, because this information could not always be extracted automatically from the five drug knowledge databases.

To determine how well the machine learning model can identify patients who may have a DDI- linked ADE (step 3B in Fig 1), we evaluated the model prediction error using the out-of-bag (OOB) error method: several models are built with a bootstrapped dataset, the OOB error is the mean of the errors computed with non-used data in each model.

The neural network gives the probability to belong to a class. We used cross-validatin resampling to optimize the threshold separating the two class. We chose to study a specific DDI in which atorvastatin was the Object and cyclosporine the Precipitant (i.e., exposed population). The non-exposed population consisted of patients who took atorvastatin and a Control- precipitant (Fig 3). The used variables were: demographic data, pathologies (ICD-10 codes) and laboratory test results. We used all the laboratory test results included between the beginning of the event and 3 days later. If a laboratory test appeared more than once, we took the mode of the results.

The reinforcement phase was not evaluated because it is currently under construction in collaboration with drug safety specialists.

Results

DDI identification with the automatic detection system was very fast due to the use of a graph- oriented model. For instance, for the simple query "is there an interaction between these two drugs?", or the more complex query "select all drugs that interact with this specific drug", the OrientDB database was always faster (less than 20ms) than the Theriaque SQL database (several seconds). Moreover, switching to another drug knowledge database was very easy with OrientDB because it only needed to change a condition in the query (which database = 'Theriaque').

The time required to create these graph databases was reasonable: for instance, the information coming from the Theriaque database, which is equivalent to 18,800 vertices and 23 million edges, was integrated in one hour. Afterwards, data access was immediate.

Once the OrientDB database was ready, from the eHOP CDW, we checked the DDI occurrence for all drug couples in the study population. To this aim, we computed all the fully or partially overlapping action intervals for all drug couples involving a statin. For each patient, we visualized all the detected DDIs: between 22.5% and 52.2% (depending on the drug knowledge database) of the 10,506 patients who were taking statins presented at least one DDI involving a statin.

Computation time was reduced with the use of the Spark framework: the processing time of 800,000 rows of patient records decreased from 60 minutes initially to only 12 minutes with Spark.

To test the ADE prediction performance of the machine learning model, we then focused only on one specific DDI (atorvastatin-cyclosporine) to create the training sample. We could identify 102 patients with atorvastatin-cyclosporine DDIs (i.e., the exposed population) and 150 patients without this DDI (i.e., the non-exposed population) (Table 1).

Table 1- Demographic data of the exposed and non-exposed population samples

	Exposed population (n=102)	**Non-exposed population (n=150)**
Age (mean ± Sd)	72.1 ± 11.6	72.9 ± 10.9
Sex (% of men)	79.8	83.5
Cardiac pathology (%)	38.2	37.8

For the optimal threshold, the neural network out-of-bag error was 17.06%, sensitivity and specificity were 90.20% and 78% respectively, and the AUC was 0.757. The processing time was short (less than 30 seconds) and could be easily performed again during the reinforcement phase.

Discussion

DDI Automatic Detection System: A New Source of Refined Data for Drug Safety Professionals

With this DDI detection system and the CDW, we can compute the overall DDI prevalence for any drug pairs, and also according to a chosen interaction severity level, or for a specific population subset. These data are useful for drug safety monitoring/research and have been already used in a study on the use of statins [21-22]. Moreover, currently, pharmacovigilance studies use different case report databases [23]. We find DDIs directly in the patient EHRs. Therefore, after DDI detection, we can link this information to other data included in the EHR (e.g., demographic data, laboratory test, etc.) to contextualize the case.

However, our DDI detection system cannot identify all DDIs. This could be due to several reasons. First, the choice of the drug knowledge database is important, and we actually observed heterogeneity between these databases that might lead to variability in DDI detection [22]. Moreover, with more information concerning the changes in the blood concentration (and half-life) of the involved drugs, we could compute more precise action intervals, thus improving the identification of overlapping treatment periods. However, this would require extensive manual search of literature data. Finally, our system cannot detect a DDI caused by a drug prescribed/administered outside the hospital. For instance, the regular treatment is usually stopped when a patient is hospitalized in the emergency service and is recorded in the emergency report. Accessing this information requires a specific treatment of unstructured text. Another option could be to link data on the drugs prescribed in primary care settings (i.e., the national health insurance database) to the hospital data (e.g., eHOP). Despite the linkage problems and the issues due to the national health insurance database features (data only on refundable drugs and only on the drug purchase but not the regimen), the analysis of the entire patient path could bring useful information on treatment ruptures, which could suggest DDIs.

A machine learning model for search reinforcement

The automatic way used to create the non-exposed population works and selects a population similar to the exposed group in terms of demographics and pathology. If the sample is big enough, we can ask the system to select the most similar patients.

Although the study of the temporal correlations between laboratory test changes and drug administration is relevant for ADE detection [24,25], we chose a robust prediction-oriented machine learning model that can work without requiring too many adjustments. Indeed, we expect that clinical variables in the exposed population will change in the presence of a DDI. However, we do not know whether the detection of a DDI implies automatically an ADE, and accessing the information to confirm the ADE involves a considerable work for drug safety professionals that we want to avoid. Therefore, to automate the monitoring of DDI-linked ADEs, we took the data immediately available from eHOP.

As they have very similar demographic characteristics, comparing exposed and non-exposed populations seemed to be an effective way to initialize the system. An improvement would be to take into account also the information included, for example, in ADE report databases. However, this system can be easily improved even without more data. Indeed, the model predicts candidate ADE cases that are likely to have been caused by DDIs and proposes them to drug safety professionals. If these cases are confirmed by drug safety professionals, they are included in the training sample to automatically enhance the model.

On the other hand, and like for any automatic detection model, our neural network model does not allow understanding which anomaly led to the prediction of an ADE and for this the analysis of the patient record is required. A machine learning model requires a lot of work, especially the choice of the model and the features engineering. In particular, a larger sample could ollow other resampling strategies to be used, that do not require the out of bag error, which is prone to overestimation of

the true prediction error [26]. These questions need a suitable study including a better evaluation with drug safety specialists.

Conclusions

This study shows how to employ healthcare data for automated DDI monitoring and ADE prediction. It involves the complete data processing chain: data collection, processing and enrichment as well as the creation of a machine learning model. The developed statistical model is the first step for a simple and convenient use of data, and could be enriched with additional information from other databases that must be integrated (more specific drug knowledge databases, ADE report databases …).

Although no drug safety professional is required during the monitoring, their expertise is essential to properly understand the data and put them into context. Their recommendations were also important to build the monitoring system and to improve the model.

Acknowledgements

We would like to thank the French National Research Agency (ANR), for funding this work inside the INSHARE (INtegrating and Sharing Health dAta for Research) project (grant no. ANR-15-CE19-0024).

References

[1] F. Meier, R. Maas, A. Sonst, A. Patapovas, F. Müller, B. Plank-Kiegele, B. Pfistermeister, O. Schöffski, T. Bürkle, H. Dormann, Adverse drug events in patients admitted to an emergency department: an analysis of direct costs, Pharmacoepidemiol. Drug Saf. 24 (2015) 176–186.

[2] C. Bonnet, P. Boudou-Rouquette, E. Azoulay-Rutman, O. Huillard, J.-L. Golmard, E. Carton, G. Noé, M. Vidal, G. Orvoen, et al., Potential drug-drug interactions with abiraterone in metastatic castration-resistant prostate cancer patients: a prevalence study in France, Cancer Chemother. Pharmacol. 79 (2017) 1051–1055.

[3] E. Ramirez, A.J. Carcas, A.M. Borobia, S.H. Lei, E. Piñana, S. Fudio, J. Frias, A Pharmacovigilance Program From Laboratory Signals for the Detection and Reporting of Serious Adverse Drug Reactions in Hospitalized Patients, Clin. Pharmacol. Ther. 87 (2010) 74–86.

[4] J. Price, What Can Big Data Offer the Pharmacovigilance of Orphan Drugs?, Clin. Ther. 38 (2016) 2533–2545.

[5] K. Martin-Latry, B. Bégaud, Pharmacoepidemiological research using French reimbursement databases: yes we can!, Pharmacoepidemiol. Drug Saf. 19 (2010) 256–265.

[6] M.-L. Yeh, Y.-J. Chang, S.-J. Yeh, L.-J. Huang, Y.-T. Yen, P.-Y. Wang, Y.-C. Li, C.-Y. Hsu, Potential drug–drug interactions in pediatric outpatient prescriptions for newborns and infants, Comput. Methods Programs Biomed. 113 (2014) 15–22.

[7] G. Bouzillé, R. Westerlynck, G. Defossez, D. Bouslimi, S. Bayat, C. Riou, Y. Busnel, C. Le Guillou, J.-M. Cauvin, C. Jacquelinet, P. Pladys, E. Oger, E. Stindel, P. Ingrand, G. Coatrieux, M. Cuggia, Sharing health big data for research - A design by use cases: the INSHARE platform approach, in: 16th World Congr. Med. Health Inform. MedInfo2017, Hangzhou, China, 2017.

[8] G. Bouzillé, E. Sylvestre, B. Campillo-Gimenez, E. Renault, T. Ledieu, D. Delamarre, M. Cuggia, An Integrated Workflow For Secondary Use of Patient Data for Clinical Research., Stud Health Technol Inf. 216 (2015) 913.

[9] ANSM: Agence nationale de sécurité du médicament et des produits de santé. http://ansm.sante.fr (accessed August 1, 2017).

[10] VIDAL - La base de données en ligne des prescripteurs libéraux. https://www.vidal.fr (accessed August 1, 2017).

[11] Thériaque. http://www.theriaque.org (accessed August 1, 2017).

[12] Micromedex. https://www.micromedexsolutions.com (accessed August 1, 2017).

[13] Drugs.com | Prescription Drug Information, Interactions & Side Effects, Drugs.Com. https://www.drugs.com.

[14] OrientDB - Distributed Graph/Document Multi-Model Database. httpss//orientdb.com (accessed August 1, 2017).

[15] M. Zaharia, M. Chowdhury, M.J. Franklin, S. Shenker, I. Stoica, Spark: Cluster computing with working sets., HotCloud. 10 (2010) 95.

[16] M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. McCauley, M.J. Franklin, S. Shenker, I. Stoica, Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing, in: USENIX Association, 2012: pp. 2–2.

[17] S. Hennessy, C. Leonard, J. Gagne, J. Flory, X. Han, C. Brensinger, W. Bilker, Pharmacoepidemiologic Methods for Studying the Health Effects of Drug-Drug Interactions (DDIs), Clin. Pharmacol. Ther. 99 (2016) 92–100.

[18] A. Neubert, H. Dormann, H.-U. Prokosch, T. Bürkle, W. Rascher, R. Sojer, K. Brune, M. Criegee-Rieck, E-pharmacovigilance: development and implementation of a computable knowledge base to identify adverse drug reactions, Br. J. Clin. Pharmacol. 76 (2013) 69–77.

[19] A.G. Shoro, T.R. Soomro, Big data analysis: Apache spark perspective, Glob. J. Comput. Sci. Technol. 15 (2015).

[20] WHOCC - Structure and principles. https://www.whocc.no/atc/structure_and_principles/ (accessed August 1, 2017).

[21] J. Stausberg, International prevalence of adverse drug events in hospitals: an analysis of routine data from England, Germany, and the USA, BMC Health Serv. Res. 14 (2014) 125.

[22] C. Morival, R. Westerlynck, G. Bouzillé, M. Cuggia, P.L. Corre, Prevalence and nature of statin drug-drug interactions in a university hospital by electronic health record mining, Eur. J. Clin. Pharmacol. (2017) 1–10.

[23] F. Kaguelidou, F. Beau-Salinas, A.P. Jonville-Bera, E. Jacqz-Aigrain, Neonatal adverse drug reactions: an analysis of reports to the French pharmacovigilance database, Br. J. Clin. Pharmacol. 82 (2016) 1058–1068.

[24] A. Newe, S. Wimmer, A. Neubert, L. Becker, H.-U. Prokosch, M.W. Beckmann, R. Fietkau, C. Forster, M.F. Neurath, G. Schett, T. Ganslandt, Towards a Computable Data Corpus of Temporal Correlations between Drug Administration and Lab Value Changes, PLOS ONE. 10 (2015) e0136131.

[25] A. Newe, Dramatyping: a generic algorithm for detecting reasonable temporal correlations between drug administration and lab value alterations, PeerJ. 4 (2016) e1851.

[26] S. Janitza, and R. Hornung, On the overestimation of random forest's out-of-bag error, *PLoS One*. **13** (2018).

Address for correspondence

Bouzillé Guillaume: guillaume.bouzille@univ-rennes1.fr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190181

Unveiling Online Suicide Behavior: What Can We Learn About Mental Health from Suicide Survivors of Reddit?

Ashwin Karthik Ambalavan[a], Bilel Moulahi[b], Jérome Azé[c], Sandra Bringay[c,d]

[a] *Arizona State University, Tempe, AZ, USA*
[b] *Acelys Informatique, Montpellier, France*
[c] *LIRMM, University of Montpellier, CNRS, Montpellier, France*
[d] *University Paul Valéry Montpellier, Montpellier, France*

Abstract

Suicide is a growing public health concern in online communities. In this paper, we analyze online communications on the topic of suicide in the social networking platform, Reddit. We combine lexical text characteristics with semantic information to identify comments with features of suicide attempts and methods. Then, we develop a set of machine learning methods to automatically extract suicide methods and classify the user comments. Our classification methods performance varied between suicide experiences, with F1-scores up to 0.92 for "drugs" and greater than 0.82 for "hanging" and "other methods". Our exploratory analysis reveals that the most frequent reported suicide methods are drug overdose, hanging, and wrist-cutting.

Keywords:

Suicide, attempted; Natural Language Processing; Social Media

Introduction

Social media platforms, such as Twitter, Facebook, and Reddit, are bringing important challenges for a variety of social- and health-related phenomena. The widespread use of these modern means of communication creates a potential to identify and characterize users behaviors by analyzing content shared online. Recent studies have shown that people are more likely to seek support from informal online resources, rather than seeking formal treatment from professionals [1] and reported a positive correlation between suicide rates and the volume of social media posts related to suicidal ideation [2]. This research suggests that online sources may contain valuable information about specific cohorts, such as vulnerable populations, and their suicidal behavior.

In this paper, we focus on the content analysis of forums dedicated to a specific population -- suicide survivors, or which is defined here as those who have had suicide attempts. Our work aims to contribute to the literature by understanding and analyzing communications on the topic of suicide in the social networking platform, Reddit.[1] As of February 2018, Reddit is a highly trafficked website[2] with more than 542 million monthly visitors, 725.85 million comments, and 6.89 billion up-votes from its users.[3] Reddit provides users anonymity while sharing and discussing information about almost everything, without any bias on the expressed opinions and feelings. In contrast to other social media, it provides domain-specific discussion forums that are moderated by a number of users and that carry health-related knowledge expressed by various cohorts.

In our work, we analyze a serious Reddit thread[4] detailing extremely personal and traumatic experiences of suicide survivors. This thread is moderated by 41 users and includes only serious comments about the topic of suicide attempts. We argue that the study of this dataset is an important step towards the development of automated methods to identify suicide risk factors that could help mental health professionals and psychiatrists to understand and prevent suicide.

Our main objective is to identify methods used to commit suicide such as drug overdose, hanging and to quantify their evocation in social media.

Our research contributions include: (1) extraction and analysis of linguistic characteristic and key aspects of the language used by suicidal individuals to describe suicide methods; and (2) development of a set of machine learning algorithms to extract methods used by suicide attempters and classify the Reddit comments.

State of the Art

Suicide is one of the most common health issues impacting the world's population. According to the Centers for Disease Control and Prevention [3], more than 40,000 suicides were reported in the United States in 2012, which positioned Suicide as the 10th leading cause of death in the country.

A rich collection of work has been done on social media as an effort to identify and understand communications about mental health problems, including depression, mental disorder and suicide. While most of the research from literature is done on suicide notes [4], more recent work demonstrated that evaluating suicidal risk factors in social networks can be used to prevent suicide and detect suicidal ideation in its early stages.

For example, the Durkheim[5] project studied the online activities and shared content of group of US war veterans on Twitter, Facebook, and LinkedIn in order to identify markers of harmful behaviour. The group developed linguistics-driven prediction models to estimate the risk of suicide using text from the clinical notes [5]. Results showed that people who committed suicide frequently recorded behaviours indicative of fear, agitation and delusion, with around 65% accuracy.

[1] www.reddit.com
[2] https://redditblog.com/2015/12/31/reddit-in-2015/
[3] https://en.wikipedia.org/wiki/Reddit
[4] https://redd.it/4e8oip/
[5] http://www.durkheimproject.org/research/

In the same line of research, Sueki [6] examine a panel of young (early 20s) Twitter users to evaluate the association between suicide-related tweets and suicidal behaviour. The authors investigated the linguistic features of suicidal ideation and identified the most important markers of future suicide. For example, phrases such as *"want to commit suicide*" were found to be strongly associated with lifetime suicide attempts, while phrases that suggest suicidal intent, such as *"want to die*," were found to be less strongly associated. For a complete summary of literature on suicidal thoughts and behaviours, the reader may refer to the recent pioneering work of [7], who conducted a meta-analysis of 365 studies from the past 50 years.

Most closely related to the current paper, Ghitsis [8] investigate the linguistic characteristics of Reddit posts that need urgent attention. The authors focussed on subreddits in which users post comments about *Addiction, Anxiety, Dementia, Depression, self-harm* and *suicide ideation.* However, the work does not focus specifically on serious suicide posts, and, most importantly, does not investigate the classification of methods of suicides, which in turn could allow the understanding of several potential warning signs.

With the present work, we consider a broad range of linguistic, lexical, and semantic features to facilitate the task of methods classification. We specifically address the following research questions:

1. How do suicidal individuals communicate about the method used during suicide attempt experiences?
2. What are the most popular methods used by suicidal individuals to end their life?
3. How to recognize a comment dealing about suicide method?

Methods

Step 1: Data acquisition

Our dataset consists of all the comments of the Reddit thread, "*[Serious] Suicide survivors of Reddit, what was your first conscious thought after you realized that you hadn't succeeded?*"[6] We have scraped all the comments using the Python Reddit API Wrapper. The topic has been tagged with the keyword "[Serious]", which means that it is monitored by a group of moderators who remove any off-topic and irrelevant comments in the thread, keeping only serious comments about suicidal ideation. In Reddit, all the comments are organized as a big tree of comments with the parent comment as the top level node and the comments about the parent comment as their child nodes. Table 1 reports statistics about the dataset.

The lack of informed consent given by social media users for data usage leads to ethical questions. In particular, confidentiality with respect to the publication of research results is an issue. We adhere to the guidelines of [9]. Results are presented with a degree of detail that does not permit drawing conclusions on individual users.

Step 2: Data Preprocessing

We initially preprocessed the data to remove HTML tags, white spaces, pipe symbols and carriage returns. We expand English word contractions such as won't, can't, to will not, cannot, respectively using regular expressions. These steps are essential to tokenize words properly and for negation identification. We lemmatize the words using both the Python's NLTK Wordnet lemmatizer.

Table 1– Dataset statistics

Total number of comments	6,229
Number of top-level comments	1,833
Number of sentences	12,782
Average number of sentences per comment	7
Number of unique users	3,58

Step 3: Feature Extraction

In the following, we define the features used in the classification process: Trigrams, NLTK POS Tags and Customised POS Tags.

Trigrams: Lemmas on their own have very little meaning and cannot be used as an independent criteria for annotating a sentence according to a method. Hence, we used trigrams to explore sentences and find recurring patterns. Some examples include: have to kill, going to bleed, tried to slit.

NLTK POS Tags: We tokenized and tagged all the words using NLTK POS tags. We obtained a total of 5,229 verbs, 10,901 adjectives and 24,815 nouns. We used NTLK POS-tagger.

Customised POS Tags: We have also manually defined Customized POS Tags described in Table 2. Due to space restrictions, we only reported the most frequent words associated with these tags. All the Customized POS Tags are self-explanatory, except for 'Ability' categories which represent the words that usually tend to be used by people who are expressing support need such as 'I need support'.

Table 2– Customized POS Tags

Customized POS TAG	Meaning	Common Tokens
FPRP	First Personal Pronouns	I, my, myself
SPRP	Second Personal Pronouns	You, him, he, it
WPRP	Group First Personal Pronouns	We, us, our
NEG	Negations	Not, note, never
SWR	Swear words	Wtf, ass, poop
INTSFR	Intensifiers	Really, so, very
NEGINTSFR	Negative intensifiers	Awfully, horrifyingly
POS	Positive words	Adore, affirmation
NEG	Negative words	Depression, sacrifice
ABLT	Ability words	Handle, support
METHOD	Suicide methods	Hang, drug, shotgun
POSEMO	Positive Emojis	Lol, LMAO, ROFL

extracted from an internal API using 8 online synonym dictionaries[7] and we keep a synonym only if it is find at least 3 in 3 dictionaries. After manually removing irrelevant tokens, we obtain a list of 218 tokens mapped to 15 suicide methods as mentioned in Table 3.

[6] https://redd.it/4e8oip/

[7] Reverso www.reverso.net, Bab.la fr.bab.la/dictionnaire, Atlas dico.isc.cnrs.fr, Thesaurus www.thesaurus.org, Orto-lang www.cnrtl.fr/synonymie/, SensAgent diction-naire.sensagent.com/synonyme/en-fr/, The FreeDictionary www.thefreedictionary.com and the Synonym www.synonym.com, all retrieved on July 13, 2018

Table 3– Most frequent tokens associated with each suicide method label and its distribution in the dataset

Methods	Most frequent tokens	Distributions
Alcohol	Drunk, Liquor, boost	9,9
Bleed	Bleeding, blood	6
Carbon monoxide	Monoxide, gas-poisoning	0,2
Cut	Slit, cleave, tear	10,5
Starve	Starvation, famine	0,2
Disease	Illness, Sickness	4,7
Drown	Drowning, underwater	1,6
Drug	Overdose, pills, dope	36,1
Electrocute	Electrocution, electrocute	0,1
Gun	Shotgun, shoot, revolver	6
Hang	Hanging, noose, strangle	17,8
Hypothermia	Hypothermia	0,1
Jump	Leap, jump-over, jumping	4,2
Vehicle	Collision, speed, car	2,5
Suffocate	suffocation	0,1

FPRP, SPRP, WPRP and ABLT are self defined. POS and NEG list have been compiled from 4 lexicons [1,10-12].

The tag 'NEGINTSFR' is used when a word occurs in both the Negative words dictionary and the Intensifier dictionary such as 'awfully', 'horrifyingly'.

To build the METHOD list, first, we defined a list of 34 suicide methods obtained from dedicated mental health websites[8]. Then we enriched this list by using synonyms

The POSEMO list contains regular expression rules to detect positive social networking emotions including emojis and common slang such as *lol, lmao*... We also use regular expressions in order to include repeated letters as *looooooool, lol.*

Step 4: Classification

We applied supervised machine learning algorithms in order to automatically predict and extract suicide methods at the comment level. Because a comment could be annotated with one or more of the labels listed in Table 3, the task is a multi-label classification task. Every comment containing at least one annotation was extracted. We obtained a training dataset of 874 annotated comments.

We used a One VS Rest Classifier with four different models: Support Vector Machine, Logistic Regression, SGD and Perceptron.

We perform automatic cross validation with 611 comments used as training set and 263 comments used as test set (30% of the dataset).

We tuned the hyper-parameters of the model for the four different models. Global system performance is measured with micro-averaged F1-score because of the use of a skewed dataset.

[8] https://en.wikipedia.org/w/index.php?title=Suicide_methods&oldid=788573333 http://regretfulmorning.com/2011/08/the-7-most-common-drugs-people-overdose-on/ both retrieved on July 13, 2018 Reverso www.reverso.net, Bab.la fr.bab.la/dictionnaire, Atlas dico.isc.cnrs.fr, Thesaurus www.thesaurus.org, Orto-lang www.cnrtl.fr/synonymie/, SensAgent diction-naire.sensagent.com/synonyme/en-fr/, The FreeDictionary www.thefreedictionary.com and the Synonym www.synonym.com, all retrieved on July 13, 2018

Results

POS Tagged Sentence Sequence:

We tagged all the sentences with the Customised POS Tags mentioned above and look for regularities. In the following, we list two prominent examples:

- 'FPRP', 'VBD', 'TO', 'METHOD' (First Personal Pronoun followed by a verb, a preposition and the method) : "I decided to hang", "I tried to - slit, strangle, jump, or drown".
- 'FPRP', 'VBD', 'INTSFR', 'NEGATIVE' (First Personal pronoun followed by a verb, an Intensifier and a Negative Word) gives us the emotional state the person: "I felt more alone", "I became more depressed", "I was completely helpless"...

Customized POS Tagged Sequences are very useful to identify and extract patterns to characterize the way individuals communicate about the methods they used to commit a suicide attempt (RQ1).

Method Labelling

The distribution of the labels in the training set is also presented in Table 3. It is apparent that some methods to commit suicide are much more frequent compared to others. Some of the most frequent labels are, in order, *drug, hang, cut* and *alcohol*. It is interesting that most Reddit users who have commented on this submission preferred to commit suicide through *drug overdose* (RQ2).

Comment classification

Table 4 gives the comparison of the 4 used estimators along with their classification features (RQ3).

Table 4 – Classifier performances

Estimator	Accuracy	Micro-precision	Micro recall	F1-Score
Logistic Regression	0.452	0.726	0.534	0.615
Perceptron	0.631	0.863	0.709	0.778
SGD Classifier	0.642	0.856	0.722	0.783
Linear SVC	0.684	0.912	0.722	0.806

As seen in many different NLP tasks, Linear SVC has the best performance in comparison to Logistic Regression, Perceptron and SGD Classifier. Table 5 gives the classification report of Linear SVC for each individual suicide method label.

Label frequency determines classifier performance considerably. For all comments with support above 20, we can see that we have an F1-score of at least 50%. Unsurprisingly, classifiers for rare examples such as *Carbon Monoxide, Electrocute, Drown* perform worst because of the low number of samples. Classifier performance would likely improve for the low-frequency labels if more training data were obtained, without the need for new features.

Some labels (*gun, jump* and *disease*) perform better than would be expected from frequency alone, indicating that they are easier to learn than others, possibly because they are lexicalized more often, or more consistently.

Table 5 – Classification performance per methods

Methods	Precision	Recall	F1-score	Support
Alcohol	0.97	0.57	0.71	53
Bleed	1	0.52	0,69	42
Carbon monoxide	0.00	0	0	0
Cut	0.94	0.59	0,72	51
Disease	1	0.39	0,56	23
Drown	0	0	0	7
Drug	0.89	0.95	0,92	154
Electrocute	0	0	0	0
Gun	1	0.6	0,75	15
Hang	0.86	0,79	0,82	71
Hypothermia	0.94	0,81	0,87	21
Jump	1	0,5	0,67	2
Vehicle	1	1	1	1
Suffocate	1	0,29	0,44	7

Discussion

Based on the feature extraction process, we find that there are many prominent methods used by leading individuals to attempt suicide. Our analysis reveals that most of these causes appear in sentences either as a negative word or after an intensifier word. In table 6 we report some examples of comments with the number of votes as scored by Reddit. The user names of the authors are removed in order to preserve privacy.

Now, it would be interesting to combine this information about suicide methods with other dimensions such as sentiment analysis or causes such as "bullying" (User 1), "ptsd," "depression," and "divorce" (User 3). These others dimensions could be extracted from the text with a similar approach to the one used for the methods. Such signs could be used as an important signal to prevent suicide.

Table 6– Examples of comments

Author	Comment	Score
User 1	Well I **had attempted to hang myself** when **I was 15**. I have a **birth defect** (deformed legs) and **just could not take the bullying anymore**. So...	7,455
User 2	**I'am a diabetic.** I can remember deciding **to use insulin to go**. Figured that passing out and dying of a seizure due to hypoglycemia would be a **quick and easy way to go.**	1,460
User 3	**I deal with ptsd and depression** at the time I was dealing with my issues that I did after coming home from over seas **my divorce** and **I was getting close to losing my apartment I got drunk as helle got in a fight...**	709

On the other hand, the distribution of suicide method labels in the dataset (Table 3) shows that drug overdose is the most common method used to commit suicide followed by hanging, cutting, and alcohol poisoning, respectively. Since drug overdose is a growing cause of suicide, suitable precautionary methods should be taken to avoid it as much as possible. Some suggestions would be to identify the most common drugs used to commit suicide and to prescribe them to individuals only after throughly assessing their mental health.

The Reddit comment classification reveal that classifier performance depends on frequency of labels in the dataset. In order to limit the annotation effort, methods such as active learning [13] could be used to select the example to annotate in order to improve the model more rapidly.

Conclusions

In this paper we developed a set of methods based on natural language processing and machine learning in order to study the suicidal behavior of individuals who attempted suicide. We built a set of linguistic, lexical, and semantic features that help in capturing relevant language clues. These new features have been successfully used to improve the classification of suicidal thoughts, experiences, and, most importantly, suicide methods. We found that POS tagging features are important to identify and understand users' communications in forums especially when used as features within a machine learning method.

We also created and annotated a dataset that can be used by researchers to investigate others questions about suicide attempts. For future research, we plan to undertake large-scale experimental evaluation to assess the language of suicide across different types of data, such as microblogs and forums. It may be also instructive to determine whether users' classification could reveals similar profiles, suggesting similar suicide methods and comparable risk factors. The classification of the language used could also be informative.

References

[1] D. Rickwood, F.P. Deane, C.J. Wilson, and J. Ciarrochi. Young people's help-seeking for mental health problems. *Australian e-Journal for the Advancement of Mental Health (AeJAMH)*, **4** (2005) 218-251.

[2] P. Burnap, W. Colombo, and J. Scoureld. Machine classication and analysis of suicide-related communication on twitter. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media,* HT '15, pages 75-84, New York, NY, USA, 2015. ACM.

[3] S.L. Murphy, K.D. Kochanek, J. Xu, and E. Arias. Mortality in the United States. *NCHS Data Brief.*, **229** (2014) 1-8.

[4] J.P. Pestian, P. Matykiewicz, and J. Grupp-Phelan. Using natural language processing to classify suicide notes. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, BioNLP '08, pages 96-97, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics.

[5] C. Poulin, B. Shiner, P. Thompson, L. Vepstas, Y. Young-Xu, B. Goertzel, B. Watts, L. Flashman, and T. McAllister. Predicting the risk of suicide by analyzing the text of clinical notes. *PLoS ONE*, **9** (2014).

[6] H. Sueki. The association of suicide-related twitter use with suicidal behaviour: A cross-sectional study of young internet users in japan. *Journal of Affective Disorders*, **170** (2015) 155-160.

[7] J.C. Franklin, J.D. Ribeiro, K.R. Fox, K.H. Bentley, E.M. Kleiman, X. Huang, K.M. Musacchio, A.C. Jaroszewski, B. P. Chang, and M.K. Nock. Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin,* **143** (2017) 187-232.

[8] G. Gkotsis, A. Oellrich, T.J.P. Hubbard, R.J.B. Dobson, M. Liakata, S. Velupillai, and R. Dutta. The language of mental health problems in social media. In *Proceedings of*

the 3rd Workshop on Computational Content Analysis of Texts by Suicide Survivors of Reddit 15 Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, CLPsychNAACL-HLT, pages 63-73, 2016.

[9] R.T Kraut, J. Olson, M. Banaji, A. Bruckman, J. Cohen, and M. Couper. Psychological research online. Report of board of scientic affairs' advisory group on the conduct of research on the Internet. *American Psychologist*, **59** (2004) 108-117.

[10] B. Liu, M. Hu, and J. Cheng. Opinion observer: Analyzing and comparing opinions on the web. In *Proceedings of the 14th International Conference on World Wide Web,* WWW'05, pages 342-351, New York, NY, USA, 2005. ACM.

[11] S.M. Mohammad and P.D. Turney. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence,* **29** (2013) 436-465.

[12] M. Ueda, K. Mori, T. Matsubayashi, and Y. Sawada. Tweeting celebrity suicides: Users' reaction to prominent suicide deaths on twitter and subsequent increases in actual suicides. *Social Science & Medicine*, **189** (2017) 158-166.

Address for correspondence

Sandra Bringay, E-mail: bringay@lirmm.fr
Phone: +336 83 24 7933

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190182

Electrocardiogram Beat-Classification Based on a ResNet Network

Cláudia Brito, Ana Machado, António Sousa

HASLab, INESC TEC & University of Minho, Braga, Portugal

Abstract

When dealing with electrocardiography (ECG) the main focus relies on the classification of the heart's electric activity and deep learning has been proving its value over the years classifying the heartbeats, exhibiting great performance when doing so. Following these assumptions, we propose a deep learning model based on a ResNet architecture with convolutional 1D layers to classify the beats into one of the 4 classes: normal, atrial premature contraction, premature ventricular contraction and others. Experimental results with MIT-BIH Arrhythmia Database confirmed that the model is able to perform well, obtaining an accuracy of 96% when using stochastic gradient descent (SGD) and 83% when using adaptive moment estimation (Adam), SGD also obtained F1-scores over 90% for the four classes proposed. A larger dataset was created and tested as unforeseen data for the trained model, proving that new tests should be done to improve the accuracy of it.

Keywords:

Electrocardiogram, Deep Learning, Arrhythmia

Introduction

Real-time monitoring has become one of the most important and clinically relevant tasks in medical settings, yet one of the most repetitive and tiresome tasks is the analysis of 24-hour ECG records. One of the ways to automate this long task is to convert this process into a real-time process with the automatic classification of the heart rate and with this, the classification of arrhythmias.

Arrhythmias are the most common diagnoses in this medical area and are composed of electrical changes that cause the normal heart rhythm to change. These changes can cause the heart to beat faster (tachycardia), slower (bradycardia), or at an irregular beat. Arrhythmias are widely classified, and their classification may depend on the factors described above and, where they occur, ventricles or atria. These changes can even lead to sudden death from stroke or cause other types of damage because of the inability of the heart to pump enough blood into the body and consequently cause damage to the brain, heart, or other organs [1,2].

To this extent, the importance of monitoring systems increases with the extra goal of improving patient care as well as the speed with which such care is provided.

Over the years, several approaches to this topic have been built. Going from the detection of the R-peak with high precision [3] to the creation of frameworks for this subject [4,5], the heart rhythm has been deeply studied. More recently, researchers have jumped from the classification of the heart beats with traditional methods to machine learning methods and even further, deep learning methods [6,7].

It is palpable the need to not only disclose the heart rhythm as tachycardia, bradycardia or irregular rhythms but to classify each of the beats into a defined category. Although being a deafening task, several works have presented great results when performing this task. Many researchers have spent their time around this subject using several methodologies as Roopa, C. and Harish B. [7] and Salem, A. et al [8] made notice in their survey. Support vector machines [9], genetic algorithms [10], rough set theory, and hidden markov models [11] and more lately neural networks [12–19], several other works mixing different methodologies have also been proposed [20].

In the second semester of 2017, Rajpurkar et al. [21] proposed a ResNet architecture of 34 layers to classify ECG batches of 30 seconds. This work exceeded the performance of high qualified cardiologists in a dataset 500 times larger than the overall datasets and set the classification task a step further to the automated analysis.

On the other side, researchers have also been focusing their efforts in patient-specific methodologies. The neural networks are trained individually for each patient allowing to classify future holters from the same patient [22,23]. Both works show promising outcomes and present a basis for future studies.

We intend to propose a new deep learning model to classify three distinct types of heart beats (four different classes) while analysing differents perspectives from the related works hereby addressed. This paper presents an overview of the dataset used as well as the architecture of the deep learning model built, providing insights on how the model was trained, and the results obtained with prospects of future work to be done. The dataset created, from records obtained from a local hospital, provided new insights about the model. The results obtained for the MIT-BIH Arrhythmia Database were promising with the arrhythmia classification yielding 96% accuracy in the classification of each beat in each one of the four classes used. The same network was used in the larger dataset, being able to classify the arrhythmias with an accuracy of 81%.

Materials and Methods

Data Selection

The MIT-BIH Arrhythmia Database presents 48 records, where the last 23 records became online only in 2005. The 48 records were chosen from a set of over 4000 long-term Holters recorded in the laboratories of the late Boston's Beth Israel Hospital, now known as Beth Israel Medical Center, between 1975 and 1979. The first 23 records were chosen randomly from inpatients and the 25 other records were chosen from the same set yet to contain a diversity of uncommon but clinically important

conditions. The two groups have different purposes since the first one is to serve as a representative sample of waveforms and artifacts which most of the arrhythmia detectors might encounter normally. While the second set of records presents more complex arrhythmias and other conduction abnormalities.

The signals were sampled at 360Hz and recorded with a two-channel recorder. The annotations were made based on a simple slope-sensitive QRS detector and by two cardiologists, who added additional beat labels missed by the QRS detector and changed all the labels of abnormal beats. Nevertheless, during the following years many records had their beats relabeled by users who reported errors in the annotations [24,25]. These records are composed by three files, an .hea format file, .dat format file and a .atr format file.

The new dataset proposed comprises 113 records from 24-hour holters from a local hospital. These records were sampled at 125Hz. This dataset presents 2172 hours of data, while the first dataset, comprises only 48 records with 30 minutes each making a total of 24 hours of data. While the first dataset, after some preprocessing and discarded beats ended up with 97737 beats on all 4 classes, each record from the local hospital presented 100000±20000 beats (Figure 1). The created dataset was first analysed by the system's software and corrected by a technician. Then, each signal's classification was validated by a cardiologist.

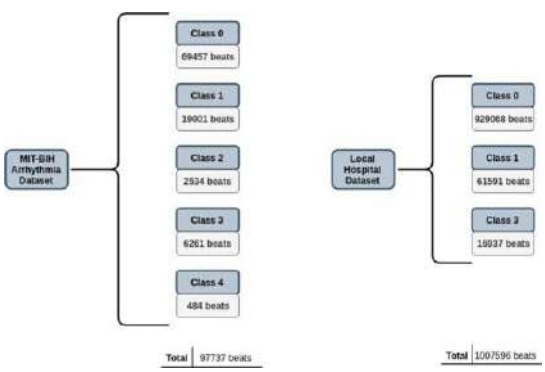


Figure 1 – Summary of Number of Beats of Each Class for Both Datasets

Later, it was disclosed the need to perform data augmentation in the classes proposed. The second and third classes were augmented since the number of beats belonging to these classes were low.

Data Treatment

The first step to analyze an ECG record begins with the load of the record and posterior filtering. These two processes were performed with the help of the CardIO library, an open-source Python library which was built to create "end-to-end machine learning models for deep research of electrocardiograms" [4].

The CardIO library relies in the WFDB package [26] to read and load files in the MIT-BIH format, this library can be executed via command line and also as a python library (the code is publicly available on Github). Using the capabilities of CardIO library, the signals were resampled to a frequency, of 125Hz. Then the signal was filtered, the filtering used in this preprocessing was based on several works that proved to be efficient [27-30]. Thus, it uses a band-pass filter, a finite impulse response filter (FIR filter [31]) which uses a frequency of 0.5Hz and 60Hz, in accordance with the theoretical foundations, yeat the new dataset was filtered with a frequency of 50 Hz and not 60Hz due to the specifications of the country's baseline wandering.

Nevertheless, in order to build the final datasets for training and testing, after these transformations, the ECG is sliced by beats and labeled with the annotations available creating the datasets for training and testing. The annotations were carefully reviewed and converted accordingly to the types of arrhythmias that were intended to classify. Therefore, the annotations were converted into four classes: normal beats as 0, other rhythms as 1, atrial premature contractions as 2 and premature ventricular contractions as 3, as seen in Table 1.

Table 1 – Representation of Each Type of Beat According to its Class

Type of Beat	Class
Normal beats	0
Other Rhythms	1
Atrial Premature Contractions	2
Premature Ventricular Contractions	3

The records were then split, based on the correct detection of the R-Peak and since it is sampled at 125Hz, it was decided to use a window with a size of 120 data points.

Model

We built a ResNet architecture (Figure 2 presents the high-level architecture of the network) based on Convolutional 1D layers for the classification task since this was the model that

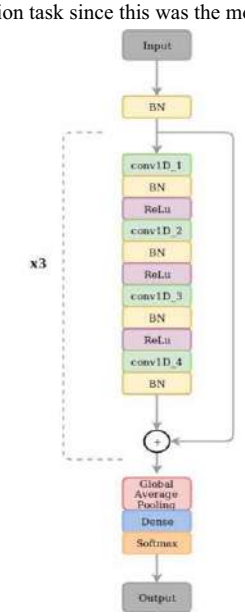


Figure 2 – Network Structure of the Model for the Classification Task. Overall the network contains 12 layers of convolution followed by a fully-connected layer and a Softmax.

surpassed the first tests performed. Initially, it was created a multilayer perceptron model and a version with 2D convolutional layers of the current model. It takes as input a time-series of 120 data points that represents the beat and outputs its label prediction. The model is composed by an initial input in a BatchNormalization layer, followed by four blocks of

Convolution1D layers (conv1D_1, conv1D_2, conv1D_3 and conv1D_4) with BatchNormalization, ReLU as activation function and a stride of 1. The conv1D_1, conv1D_2, conv1D_3 and conv1D_4 layers have a filter size of 8, 5, 3 and 1, respectively. The model could also use a Dropout schema, but considering the model complexity, we believe that we could obtain good results with BatchNormalization.

Evaluation Metrics

The effectiveness of a deep learning model for classification is usually measured by several parameters. Four of the most common are i) accuracy, ii) precision, and iii) recall, as well as the iv) F1-score [32–34]. These parameters are based on the results of true positives, true negatives, false positives and false negatives.

- True Positive (TP): correctly classified positive instances
- True Negative (TN): correctly classified negative instances
- False Positive (FP): incorrectly classified as positive
- False Negative (FN): correctly classified as negative

Accuracy

The Accuracy measures the rate of True Negatives and True Positives on all classified instances, being generically represented for binary classification by Equation 1. This result may induce in error or can hide important details since it does not distinguish between the number of correct labels of different classes. It can be understood as the number of correct predictions on top of the total number of prediction. The problems of misclassification may arise even though the accuracy results are high. In the medical field, the misdiagnose of an unhealthy subject may cost its life. When dealing with multi-classification, this formula reverts to the number of well predicted samples in relation to the total number of samples.

$$Accuracy = \frac{TN + TP}{TP + TN + FN + FP} \quad (1)$$

Precision

Precision (Equation 2) equalizes to the rate of true positive instances in the correctly classified instances.

$$\boldsymbol{Precision = \frac{TP}{TP + FP}} \quad (2)$$

Recall

Recall or sensitivity (Equation 3) gives higher scores when a high number of true positives is achieved while avoiding false negatives, this defines the true positive rate.

$$\boldsymbol{Recall = \frac{TP}{TP + FN}} \quad (3)$$

F1-Score

The F1-score is another general metric that is broadly used for evaluating these systems, the Equation 4 combines the precision and recall into a single number and is seen as one of the most reliable metrics for evaluating machine learning results.

$$F1_{score} = \frac{2 * (Precision * Recall)}{Precision + Recall} \quad (4)$$

Experimental Results

The tests performed for training the model were optimized with SGD and Adam, both with a learning rate of 0.1. Other optimisers could have been used, however, we relied on SGD and Adam based on the related work and their broad use[35]. To understand the evolution of the models with the number of epochs, both systems were tested with 10, 25, 50, 100 and 150 epochs, the maximum number of epochs was defined after the first tests, where any test over 150 epochs demonstrated that the model stopped learning. As seen in Figure 3, the first dataset was randomly sliced in 80% for training and 20% for testing, with the training set being further separated into train and validation (65/35%).

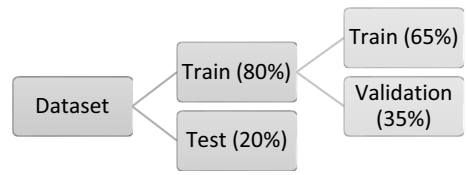


Figure 3 – Modeling the First Dataset for Training and Testing the Model

Several tests were performed for tuning the hyperparameters and adapting the learning rates to the behavior the model was exhibiting. Since this is a multi-classification task, the metrics for each class are calculated.

Figure 4 presents the overall accuracy of the models with a different number of epochs, a batch size of 2000 beats and as we are dealing with multi-classification, it used the categorical cross-entropy as loss function. The accuracy of the model reached 96% for the SGD optimizer with 150 epochs

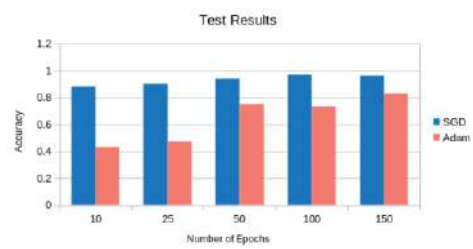


Figure 4 – Testing Results Regarding the Accuracy of the Model using Stochastic Gradient Descent and Adaptive Moment Estimation as Optimizers.

On the other hand, Figures 5 and 6 rely on the F1-score regarding the test results with the 20% of the dataset created.

The model behaves accordingly to what was expected, presenting good results for classifying all the instances initially defined. However, it can be noticed the higher results of the model optimized with stochastic gradient descent. These results are an effect of how both these optimizers work.

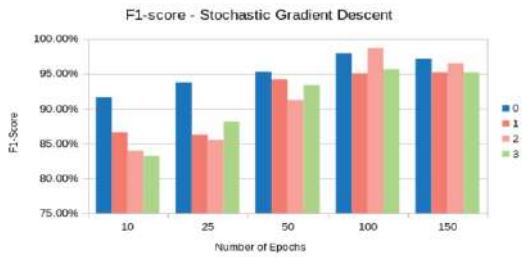


Figure 5 – Testing Results Regarding the F1-Score of the Model using Stochastic Gradient Descent as Optimizer

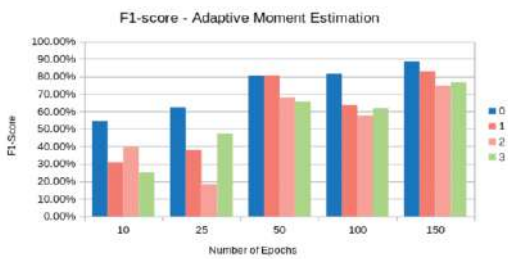


Figure 6 – Testing Results Regarding the F1-Score of the Model Using Adaptive Moment Estimation as Optimizer

The results obtained did not present overfitting, the error was minimal, and the test set obtained good results both in accuracy and F1-score. The model was able to classify all four classes and these results increased with the number of epochs.

When scrutinizing the two types of arrhythmia classified it is possible to disclose the reasons why atrial premature contractions (class 2) is the least accurate arrhythmia when using the Adam optimizer. This results can be due to the computation of gradients using this optimiser, as so, the gradient may reach a local minima but not the lowest point. In this case, it may be because of the small number of beats of this class on the test dataset. On the other side, SGD is able to find better gradients for the classification of all the classes. However, it can be seen that when the model was trained with 100 epochs it was able to reach better results when classifying the atrial premature contractions (class 2).

After the best model was trained and evaluated, the new dataset, with unforeseen data for the model, was tested. These tests were performed record by record, where 21 records achieved more than 90% of accuracy. The only problem detected was when the model tried to predict class 2 where there was no class 2 in any of those records. Tacking this into account, a new model was trained, comprising the 11 least accurate records for training and the remaining 5 records with an accuracy below 30% for testing, we decide not to use all the records due to the lack of computational power. Table 2 presents the precision, recall, and F1-score values of the last evaluation of the new model.

This model was able to accurately classify the arrhythmias with an accuracy of 81% while the F1-scores for the normal class (class 0) reached 89.9%, the other two classes performed poorly, reminding the need to perform data augmentation to balance the data to train the model.

On the other hand, several limitations raised during the development of this project, such as, the lack of computational power to diminish the training time, which took over 15 hours for the subset of only 11 records from the second dataset as well as the imbalanced dataset. These issues present a hindrance in the success of these approaches, nevertheless it is important to emphasise it presents an advance in relation to typical systems were the annotations have to be carefully reviewed 100% of the time.

Table 2 – Results Obtained for the New Dataset Based on 11 Records from the Hospital for Training and the Five Worst Records from the Hospital for Testing

Labels	Precision	Recall	F1 Score
Class 0	98.9%	82.3%	89.8%
Class 1	8.8%	47%	14.9%
Class 3	4.1%	35.3%	7.4%

In comparison with the related work, namely the work performed by [16], where the authors performed a beat-to-beat classification using the MIT-BIH arrhythmia database and obtained an accuracy of 83.4% in classifying the signal as normal or abnormal (arrhythmic), we were able to outperform their approach in 10% while classifying into four different classes.

Conclusions

In this paper, we developed a deep learning model that was able to accurately classify the heartbeat into four different classes. Focusing on two types of arrhythmia, the results obtained for this classification task were promising showing that this path for beat classification should be further investigated and providing a basis for future studies. Researchers have been dealing with beat classification using deep learning, however, the results were slowly reaching 90% of accuracy and many of these results were based on two-dimensional layers. This paper presents a ResNet with one-dimensional convolutional layers and was able to reach over 90% of accuracy and F1-scores for the four classes proposed, without falling on the rabbit-hole of overfitting.

A new dataset was created, having records 48 times bigger than the records from the MIT-BIH Arrhythmia Database. This allowed us to further explore these results, increasing our training data to create a new and better model, able to classify these 24-hour record accurately and also to try to balance the data available. Since deep learning is a methodology data-driven, the larger the dataset, the better the results. With this in min, this dataset will allow us to increase the spectrum of classified types of arrhythmia in order to create a fully automatic system without neglecting the precision and the importance of the outcome.

Acknowledgements

This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project: UID/EEA/50014/2019.

References

[1] A.C. Guyton, and J.E. Hall, TextBook in Medical Physiology, Elsevier Saunders, 2006.

[2] G.J. Tortora, and B. Derrickson, Principles of Anatomy & Physiology 14th Edition, 2014.

[3] J. Pan, and W.J. Tompkins, A Real-Time QRS Detection Algorithm, *IEEE Trans. Biomed. Eng.* **BME-32** (1985) 230–236.

[4] K. R., I. E., K. A., and P. D., CardIO library for deep research of heart signals, (2017).

[5] Welcome to BioSPPy — BioSPPy 0.4.0 documentation, (n.d.). https://biosppy.readthedocs.io/en/stable/ (accessed February 1, 2018).

[6] S.H. Jambukia, V.K. Dabhi, and H.B. Prajapati, Classification of ECG signals using machine learning techniques: A survey, *2015 Int. Conf. Adv. Comput. Eng. Appl.* (2015) 714–721.

[7] C.K. Roopa, and B.S. Harish, A Survey on various Machine Learning Approaches for ECG Analysis, **163** (2017) 25–33.

[8] A.-B.M. Salem, K. Revett, and E.-S. El-Dahshan, Machine learning techniques in electrocardiogram diagnosis, *Int. Multiconference Comput. Sci. Inf. Technol.* (2009) 429–433.

[9] A. Batra, and V. Jawa, Classification of Arrhythmia Using Conjunction of Machine Learning Algorithms and ECG Diagnostic Criteria, *Int. J. Biol. Biomed.* **1** (2016) 1–7.

[10] V. Priyadharshini, and S. Saravana, An Enhanced Approach on ECG Data Analysis using Improvised Genetic Algorithm, (2015) 1248–1256.

[11] T. Barman, R. Ghongade, and A. Ratnaparkhi, Rough Set based Segmentation and Classification Model for ECG, (2016) 18–23.

[12] M. Zihlmann, D. Perekrestenko, and M. Tschannen, Convolutional Recurrent Neural Networks for Electrocardiogram Classification, (2017) 1–4.

[13] B. Pyakillya, N. Kazachenko, and N. Mikhailovsky, Deep Learning for ECG Classification, *J. Phys. Conf. Ser.* **913** (2017) 012004.

[14] P.A. Warrick, and M. Nabhan Homsi, Ensembling convolutional and long short-term memory networks for electrocardiogram arrhythmia detection, *Physiol. Meas.* (2018) 0–20.

[15] M. Limam, F. Precioso, and U. Côte, Atrial Fibrillation Detection and ECG Classification based on Convolutional Recurrent Neural Network, *Comput. Cardiol. (2010).* **44** (2017) 1–4.

[16] S. G, S. K P, and V. R, Automated detection of cardiac arrhythmia using deep learning techniques, *Procedia Comput. Sci.* **132** (2018) 1192–1201.

[17] S. Savalia, and V. Emamian, Cardiac Arrhythmia Classification by Multi-Layer Perceptron and Convolution Neural Networks, (2018).

[18] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling, (2014) 1–9.

[19] G. Sannino, and G. De Pietro, A deep learning approach for ECG-based heartbeat classification for arrhythmia detection, *Futur. Gener. Comput. Syst.* **86** (2018) 446–455.

[20] M. Ayar, and S. Sabamoniri, An ECG-based feature selection and heartbeat classification model using a hybrid heuristic algorithm, *Informatics Med. Unlocked*. (2018) 1–9.

[21] P. Rajpurkar, A.Y. Hannun, M. Haghpanahi, C. Bourn, and A.Y. Ng, Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks, (2017).

[22] S. Kiranyaz, T. Ince, and M. Gabbouj, Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks, *IEEE Trans. Biomed. Eng.* **63** (2016) 664–675.

[23] K. Luo, J. Li, Z. Wang, and A. Cuschieri, Patient-Specific Deep Architectural Model for ECG Classification, *J. Healthc. Eng.* **2017** (2017).

[24] G.B. Moody, and R.G. Mark, The impact of the MIT-BIH arrhythmia database., *IEEE Eng. Med. Biol. Mag.* **20** (2001) 45–50.

[25] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, and H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals, *Circulation*. **101** (2000) e215–e220.

[26] Physionet, The WFDB Software Package, (n.d.). https://physionet.org/physiotools/wfdb.shtml (accessed April 2, 2018).

[27] R. Rodrıguez, A. Mexicano, J. Bila, S. Cervantes, and R. Ponce. Feature extraction of electrocardiogram signals by applying adaptive threshold and principal component analysis. *Journal of Applied Research and Technology*, 13(2):261–269, 2015.

[28] Xinqi Louis Wang and J Mikael Eklund Smieee. A real-time ECG feature detection algorithm. *Ieee*, pages 5–7, 2014.

[29] Abhinav Vishwa, Mohit K. Lal, Sharad Dixit, and Pritish Vardwaj. Clasification of Arrhythmic ECG Data Using Machine Learning Techniques. *International Journal of Interactive Multimedia and Artificial Intelligence*, 1(4):67, 2011.

[30] Ali Isin and Selen Ozdalili. Cardiac arrhythmia detection using deep learning. *Procedia Computer Science Procedia Computer Science Procedia Computer Science*, 120(00):268–275, 2017.

[31] A.V. Oppenheim, A.S. Willsky, and S.H. Nawab, Signals and Systems, Prentice Hall, 1997.

[32] S.H. Jambukia, V.K. Dabhi, and H.B. Prajapati, Classification of ECG signals using machine learning techniques: A survey, *2015 Int. Conf. Adv. Comput. Eng. Appl.* (2015) 714–721.

[33] T.J. Jun, H.J. Park, N.H. Minh, D. Kim, and Y.H. Kim, Premature ventricular contraction beat detection with deep neural networks, *Proc. - 2016 15th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2016*. (2017) 859–864.

[34] A. Vishwa, M.K. Lal, S. Dixit, and P. Vardwaj, Clasification Of Arrhythmic ECG Data Using Machine Learning Techniques, *Int. J. Interact. Multimed. Artif. Intell.* **1** (2011) 67.

[35] A. Wilson, R. Roelofs, M. Stern, N. Srebro, and B. Recht, The marginal value of adaptive gradient methods in machine learning, *Advances in Neural Information Processing System.* (2017) 4148-4158.

Address for correspondence

Cláudia Brito: claudia.v.brito@inesctec.pt

António Sousa: als@di.uminho.pt

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190183

Initial Experiments for Pharmacovigilance Analysis in Social Media Using Summaries of Product Characteristics

Leonardo Campillos-Llanos[a], Cyril Grouin[a], Agnès Lillo-Le Louët[b], Pierre Zweigenbaum[a]

[a] *LIMSI, CNRS, Université Paris-Saclay, F-91405 Orsay, France*
[b] *CRPV Paris-HEGP, Hôpital Européen George Pompidou, 20 rue Leblanc, 75015 Paris, France*

Abstract

We report initial experiments for analyzing social media through an NLP annotation tool on web posts about medications of current interests (baclofen, levothyroxine and vaccines) and summaries of product characteristics (SPCs). We conducted supervised experiments on a subset of messages annotated by experts according to positive or negative misuse; results ranged from 0.62 to 0.91 of F-score. We also annotated both SPCs and another set of posts to compare MedDRA annotations in each source. A pharmacovigilance expert checked the output and confirmed that entities not found in SCPs might express drug misuse or unknown ADRs.

Keywords:
Natural Language Processing; Pharmacovigilance; Social media

Introduction

According to the World Health Organization (WHO), pharmacovigilance is "the science and activities relating to the detection, assessment, understanding and preventing of adverse effects or any other drug-related problem".[1] The drug development process has several steps, from discovery in a laboratory, to preclinical research and clinical development involving patients. Nonetheless, after approval, complete and definitive information about drug safety is not available. Moreover, the drug use may change, the benefit/risk may evolve and health authorities need any information available.

Self-reports of adverse drug reactions (ADRs) are scarce: the French National Agency for Drug Safety (ANSM) evaluated that only 5% are reported by patients. Because ADRs are known late, except in case of highly publicized events (e.g., H1N1 flu), social media is used to improve pharmacovigilance efficiency [1,2]. However, web-based data generally contain colloquial jargon that is hard to process with common Natural Language Processing (NLP) tools, calling for dedicated approaches [3]. Pharmacovigilance also addresses abuse and drugs misuse, which involves "situations where a medicinal product is intentionally and inappropriately used not in accordance with the terms of the marketing authorization" [4].

We present an ongoing work on pharmacovigilance analyses in health fora written in French, with a special focus on drug misuse. We present the experiments we made based on a comparison of the Medical Dictionary for Regulatory Activities (MedDRA) codes annotated in web messages and the codes identified in Summaries of Product Characteristics (SPCs). As far as we know, this source has not been commonly used for NLP in health social media; this is our main contribution. We annotated pathological conditions in social media and SPCs using an NLP pipeline designed to identify such entities [5]. This tool was improved for normalizing annotations based on codes from MedDRA [6] and the Anatomical Therapeutic Classification (ATC) [7]. These methods identify potential drug misuse and unknown ADRs.

Background

Social media are useful to identify both adverse effects and drug misuse [1,2,8]. Different types of drug misuse may occur; Bigeard and colleagues report a comprehensive typology [9]. Misuse may be related to whether the drug is prescribed by a practitioner, or taken as self-medication. In case of a prescription drug, misuse may involve interaction with other drugs, an incorrect frequency, duration or dose of medication intake. Other situations involve not respecting the drug intake (e.g., if levothyroxine is not taken on an empty stomach), or problems of conservation (e.g., to keep an eye drop solution open for more than 2 weeks). Indication misuse involves the intake of a medication for an unrelated pathology (e.g., using baclofen for alcohol addiction). Another situation of interest is drug abuse, which occurs when users take drugs without any pathology but searching for a specific effect (e.g., a psychotropic effect, as is the case with the *purple drank* cocktail based on codeine cough syrup with soda).

Detection of drug misuse in social media is hard to identify automatically, since knowledge on a given medical drug indications or posology are needed to infer *unexpected* user's intake behaviors. The few works undertaken to detect drug misuse have applied supervised methods and annotated data by pharmacovigilance experts [10].

In the absence of enough available annotated data, we resorted to methods for generating training data over unlabeled samples. Distant supervision approaches [11] are close to that paradigm and have been applied in the medical domain [12]. Our method, however, relies on the annotation of knowledge

[1] https://bit.ly/2FtIYX5 [Accessed March 2019]

sources (SPCs in our context of pharmacovigilance) and then compares the output to annotations in web fora.

Methods

We conducted two types of experiments. First, we conducted a *supervised experiment* with a small set of messages annotated by pharmacovigilance experts regarding positive or negative misuse. Second, we tested to which extend comparing MedDRA codes found in messages from web fora and in SPCs provide cues of drug misuse and unknown side effects.

Corpus for Supervised Experiments

Corpus 1 contains posts written in French by consumers, extracted from several public fora from the health domain: AlloDocteurs,[2] Atoute,[3] Baclofène,[4] Doctisimo Médicaments,[5] eSanté,[6] Journal des Femmes,[7] and Vulgaris.[8] In these posts, different drugs and pharmacological substances are discussed:

Agomelatine (Valdoxan®): an antidepressant, which may be misused for insomnia, panic and anxiety attacks.

Baclofen: a muscle relaxant, used to treat addictions (especially alcoholism) out of any marketing authorization.

Duloxetine (Cymbalta®): an antidepressant drug.

Exenatide (Bietta®): an antidiabetic injection; a dose misuse (i.e., users took more injections than what is prescribed) has been detected by pharmacovigilance experts.

Myolastan (Tetrazepam®): a muscle relaxant; both misuse (indication and posology) and drug abuse has been reported.

Messages were collected with a framework for extracting data from web fora [13] and then processed using an NLP pipeline to annotate pathological and medical drug entities. Pharmacovigilance experts—pharmacists, or physicians qualified in Pharmacology—revised a selection of 1178 posts and manually annotated them as expressing misuse or not.

Corpus for Comparing Internet Posts and SPCs

Corpus 2 gathers posts from similar kind of consumers' web. We chose three specific and actual drug topics (regarding the media coverage and patients' interest), each in different fora:

1. Baclofen: data come from the Atoute website and were used in semi-automatic methods to detect drug misuse.
2. Levothyroxine: this drug replaces or provides more thyroid hormone, and a new formulation was marketed in August 2017. Messages come from the Vivre sans thyroïde forum.9
3. Vaccines: in France, only 3 vaccines (diphtheria, tetanus, poliomyelitis) were mandatory until January 1st, 2018. Since that date, 11 vaccines are compulsory: tetanus, diphtheria, poliomyelitis, and 8 additional ones: pertussis, polio, measles, mumps, rubella, hepatitis B, haemophilus influenza bacteria, pneumococcus, and meningococcus C. We used posts from the Doctissimo website.10

For each topic, we extracted 100 messages according to two inclusion criteria: presence of drug name and pathology names, and a limit of words (we avoided long generic discussions). In case of lack of messages with drug names and pathology entities, we extracted new messages to gather up to 100 messages per topic. We also used SPCs available from the French authorities[11] for baclofen, levothyroxine, and the twenty marketed products to perform the newborns immunization for the eleven vaccines: ACT-HIB®, Boostrix Tetra®, Engerix B®, Fendrix®, HBVaxPro®, Hexyon®, Imovax Polio®, Infanrix Tetra®, Infanrix Quinta®, Infanrix Hexa®, M-M-RVaxPro®, Menjugate®, Neisvac®, Pentavac®, Pneumovax®, Prevenar®, Priorix®, Repevax®, Revaxis®, and Tetravac®. We focused on sections describing indications, counter-indications or adverse drug reactions.

Annotation of Messages with an NLP Pipeline

We applied an NLP pipeline on both corpus: 1178 posts annotated by pharmacovigilance experts, and 300 concerning levothyroxine, baclofen and vaccines. As previously explained [5], the pipeline has modules for normalization, tokenization, Part-of-Speech tagging and concept annotation based on machine learning, namely Conditional Random Fields (CRF) [14]. Because the tool was improved since the first evaluation made on the first set of web fora [5], we evaluated the annotations of the 300 posts. We manually checked those annotations to build a gold standard, using BRAT (Figure 1) [15]. We evaluated the annotations through pre-cision, recall and F1-score (F1) metrics using BRATeval [16].

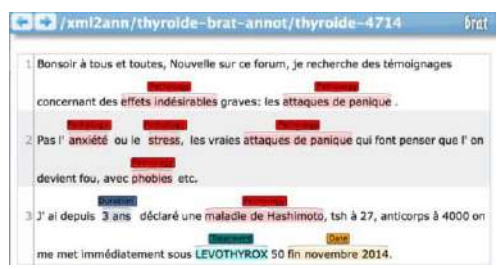


Figure 1 - Sample of forum message annotated with BRAT

We also applied the annotation tool on the summary of product characteristics (SPCs) detailed before. Thus, pathological entities in indications covered by those selected drugs are annotated and labeled with MedDRA codes. We also revised the annotations of sections Indications and Side effects in SPCs, using the same methodology to revise annotated posts.

Experiments in a Supervised Context

We followed the procedures applied by Bigeard and colleagues [10]. We tested Naïve Bayes (NB) and

[2] www.allodocteurs.fr [Accessed March 2019]
[3] www.atoute.org [Accessed March 2019]
[4] www.baclofene.com [Acccessed March 2019]
[5] http://medicament.doctissimo.fr [Accessed March 2019]
[6] https://www.e-sante.fr [Accessed March 2019]
[7] www.journaldesfemmes.fr [Accessed March 2019]
[8] www.vulgaris-medical.com [Accessed March 2019]
[9] www.forum-thyroide.net [Accessed March 2019]
[10] www.doctissimo.com [Accessed March 2019]
[11] https://bit.ly/2UNLuN6 [Accessed March 2019]

Multinomial Naïve Bayes (MBN) algorithms on the 1178 posts annotated as expressing misuse or not by pharmacovigilance experts. We tested different features:

Word tokens in message

Word roots in message

Anatomical Therapeutic Classification (ATC) codes of medical drugs in each message

MedDRA codes of pathologic entities in message

3-grams

3-character-grams

We used ScikitLearn Software for this set of experiments [17]. In all contexts, we applied 10-fold cross-validation and used an 80% and 20% ratio of training and test sets, respectively.

We also tested different subsets of messages with regard to the number of messages annotated as misuse. In real-life contexts, most posts will not contain any misuse behavior or ADR. The datasets will suffer from *class imbalance* (i.e., most samples will not express misuse nor ADRs), a common problem with supervised machine-learning algorithms [18]. In a context where most posts bear a negative class, a random classifier or a classifier labeling all samples as negative will certainly have good accuracy, even though it does not make use of any linguistic or knowledge-based feature. We thus tested different ratios of messages annotated as positive or negative misuse:

The full corpus of messages (1178 posts), 111 messages annotated as positive misuse (~10:1 ratio)

A subset of 336 messages: 111 classified as positive misuse and 225 as negative misuse (~2:1 ratio)

A subset of 246 messages: 111 classified as positive misuse and 135 as negative misuse (~1:1 ratio)

Comparing Messages in Social Media and SPCs

Once we applied the NLP tool to annotate the pathological entities in the SPCs, we assumed these coded pathologies set up the list of correct uses to be found in messages over the Internet. Conversely, all pathologies related to one of those drugs found in a message (missing in that list of *expected* pathologies) may be a drug misuse or unexpected ADR. Because we used the same annotation schema and tool, we could compare MedDRA codes in both sets and extracted a list of candidate terms. Finally, a pharmacovigilance expert and coauthor of this work (ALL)—a physician, qualified in Pharmacology, with 20 years of expertise—checked the selected entities to confirm misuse behavior or unknown ADRs.

Normalization

We used MedDRA for coding terms of pathological entities. Following Bousquet *et al*. [19], we coded Lower-level terms (LLT) for expressions of pathologies (*verbatim terms*), but also mapped these LLTs to preferred terms (PT) for pharmacovigilance analyses. We used the UMLS® [20] Concept Unique Identifiers (CUIs) to map term variants referring to the same concept. Normalization rules were applied considering inflection (singular/plural, diacritics, syntactic variants of multiwords); and Levenshtein distances were used to get the candidate term with closer string distance from a term variant.

Results

Evaluation of the NLP Pipeline

We annotated a total of 2249 pathologies and medications in forums, and 6772 in the SPCs (Table 1; we report the count of both annotations and types, i.e., different annotated items). Table 2 shows the evaluation results of each subset of posts.

Table 1 - Number of annotated entities (total items and types)

	# words	# annotations (types)	
		Pathologies	**Medical drugs**
Levothyroxine	18274	511 (230)	314 (67)
Baclofen	12941	259 (133)	386 (52)
Vaccines	10336	390 (185)	389 (83)
Total posts	41551	1160 (480)	1089 (192)
[avg per post]	[138.5]	[3.87 (1.6)]	**[**3.63 (0.6)**]**
Total in SPCs	88919	3687 (815)	3085 (320)
[avg per SPC]	[4041.8]	[167.6 (37.1)]	[140.2 (14.5)]

Table 2 - Evaluation of the NLP annotation pipeline (P: Precision; R: Recall ; F1: F-score; Avg: average)

	Pathologies			**Medical drugs**		
	P	**R**	**F1**	**P**	**R**	**F1**
Levothyroxine	0.80	0.68	0.74	0.98	0.85	0.91
Baclofen	0.84	0.70	0.76	0.89	0.80	0.84
Vaccines	0.84	0.72	0.77	0.82	0.78	0.80
Avg (posts)	0.83	0.70	0.76	0.93	0.83	0.88
SPCs	0.92	0.90	0.91	0.76	0.90	0.82

In web posts, higher F1 scores were obtained when annotating medications rather than pathological entities. This is mainly due to the higher number of different pathological entities (480) and also to the difficulty in annotating expressions of pathological conditions in *patient language* (e.g., *crevé,* 'worn out' stands for *fatigue*). Annotation results of messages concerning the baclofen and levothyroxine show higher F1 scores; this might be due to the fact that messages discussing newborns vaccination contain more different drug names (83). Results of pathological entities in posts regarding the levothyroxine might also be due to a higher number of different pathological entities in this forum (230). Because SPCs feature a lower degree of *patient language,* annotation of pathologies achieved a higher F1 score than in web fora.

Results of Supervised Experiments

As expected, the best results were obtained on the full corpus, either using Naïve Bayes or Multinomial Naïve Bayes (Table 3; we only report the results of the best features on the test set). The experiments on the other corpora configurations helped us better understand the features that *really* helped the classifier to learn and distinguish positive and negative

misuse. The best results were mainly obtained with these features: ATC codes, MedDRA codes and word roots. We had similar results as those reported by Bigeard *et al*. [10].

Table 3 - Results of classifiers in supervised context: Naïve Bayes (NB, above) and Multinomial Naïve Bayes (MNB, below). The label ratio is the proportion of posts annotated as positive or negative misuse. 2:1 stands for 2 negative posts per 1 positive; P: Precision; R: Recall; F1: F-score

	Label ratio	Features	P	R	F1
NB	~ 10:1 (all data)	ATC codes + MedDRA codes + word roots + 3-grams	0.94	0.93	**0.91**
	~ 2:1	ATC codes + MedDRA codes + word roots	0.63	0.67	0.62
	~ 1:1	Tokens + 3-grams + ATC codes; or Tokens + 3-grams	0.88	0.88	0.88
MNB	~ 10:1 (all data)	ATC codes + MedDRA codes + word roots	0.93	0.92	**0.89**
	~ 2:1	ATC codes + MedDRA codes + word roots	0.71	0.70	0.62
	~ 1:1	Tokens + 3-grams + ATC codes ; or Tokens + 3-grams	0.82	0.82	0.82

Results of comparing messages in social media and SPCs

We extracted from posts 301 pathological entities that were not documented in SPCs. Most were related to Levothyroxine (166) and Baclofen (103). Pathologic conditions related to vaccines (32) concerned Boostrix®, Engerix®, Infanrix®, Neisvac®, Pentavac®, Prevenar®, Priorix®, Repevax® and Revaxis®. We observed that some vaccines did not appear in the selected posts (ACT-HIB®, Fendrix®, HBVaxPro®, Hexyon®, Imovax Polio®, Infanrix Tetra®, M-M-RVaxPro®, Pneumovax® and Tetravac®). The pharmacovigilance expert considered that only 3 cases might be misuse related to Baclofen (1.3% of candidate items). However, one case is ambiguous: we cannot state if, when the user mentioned the unexpected pathology, he/she meant to link it to an indication related to the intake of Baclofen. No other misuse cue was confirmed with regard to other drugs. The expert identified 68 undocumented ADRs (22.6% of selected items; 28 need more context to be confirmed). Unknown ADRs mostly involved levothyroxine (52 cases), baclofen (6), Engerix® (2), Infanrix® (2), Pentavac® (3), Priorix® (1) and Repevax® (2).

Qualitative Evaluation

We analyzed messages to detect linguistic cues expressing drug indication or misuse (Table 4). Misuse due to incorrect dose might be expressed with specific verbs. However, in the Corpus 1, we observed that experts did not always annotate as misuse some contexts with those cues. Exact validation of drug doses reported by web users might indeed be within the range of correct doses that could be a *user-perceived misuse*. We estimate a linguistic analysis needs to be complemented by knowledge-driven approaches; e.g., analyses of doses in posts may be compared with ranges of doses approved by authorities. Likewise, we noticed that possible adverse drug reactions or misuse events were not detected due to the lack of MedDRA terms, especially when users write narrative descriptions of events or use non-technical expressions.

Table 4 – Samples of linguistic cues of indication, misuse or unknown adverse drug reactions (ADRs)

Misuse	*surdosé* (‘overdosed’) *mauvais dosage* (‘bad dose’) *tu ingurgites le triple de ce qui est recommandé,* ‘you take 3 times more than what is recommended’ *prendre des doses de cheval de X* (‘take a raging / strong dose of X’)
Indication	*je prends X pour …* (‘I take X for’) *X utilisé comme / pour* (‘X used as / for’)
ADRs	*X m’empechait de dormir* (‘X keeps me awake’)

Results of the Normalization Step

Applying the normalization rules on Corpus 2, we mapped to CUIs 263 out of the 344 different types of pathologies (76.4% of types), and 199 to MedDRA codes (57.8% of types). Errors were due to spelling, syntactic variation (*treatment failure* vs. *failure of treatment*), inflection (*panic attack* vs. *panic attacks*), derivation (*depressive* vs. *depression*), abbreviations (*rgo* vs. *reflux gastroesophagique*) or errors in CUI mappings.

Discussion

Communication in Internet fora is asynchronous and asymmetric, without specific interlocutors. This impacts the way medical information is expressed: incomplete, informal and creative expressions for health conditions abound, which make it difficult concept normalization and automatic analyses through NLP. Comparing pathological entities documented in SPCs medical drugs and unexpected pathologies in social media needs quality term detection and normalization. Our work is thus preliminary and suffers from the limitation that terms in patient language remained still unannotated, or entities were not normalized to accurate terms or CUIs. Moreover, we did not check the quality and correctness of the normalization step. We would like to explore normalization techniques based on word-embeddings and deep-learning. The comparison method was weak for detecting misuse, but helped in finding new ADRs; this opens the door to future work.

Regarding our supervised experiment, we lack enough data for training our model and generalizing our predictions on drug misuse to new datasets. We want to annotate SPCs of more medical drugs to gather a database of annotations to be

used in future work, especially for distant supervision approaches.

Using social media brings up other limitations related to: 1) the fact that users may not necessarily post their misuse behavior or ADRs; and 2) privacy concerns: despite users post contents to be publicly available, careful anonymization protocols are required, as we applied in the project [13].

Conclusions

We presented a method and initial experiments on pharmacovigilance analyses on social media based on NLP annotations of web fora. Through a supervised experiment with a minimal set of data, we showed that classification models might perform adequately. However, lacking of enough data to address current medications of interest, we resorted to Summary of Product Characteristics (SPCs) to overcome the data bottleneck. This approach, as far as we know, has not commonly being used and might be a source of knowledge for contrasting data reported in SPCs and unexpected users' health or conditions. A pharmacovigilance expert confirmed possible cases of misuse and some unknown ADRs. We make available the annotated data of SPCs.[12]

We highlight that our methods do not aim at replacing human decisions concerning users' behavior that could be drug misuse. These need pharmacovigilance experts to be validated; luckily, automated methods supporting these tasks make this validation faster and easier to be conducted.

Acknowledgements

The Agence Nationale de Sécurité du Médicament funded this work (grant ANSM-2016-S-076, PHARES project). We greatly thank the LIMICS team (U. Paris 13) for the data, and the pharmacovigilance team for annotating those posts.

References

[1] A. Sarker, R. Ginn, A. Nikfarjam, *et al.*, Utilizing social media data form pharmacovigilance: A review, *J Biomed Inform* **54** (2015), 202-212.

[2] A. Nikfarjam, A. Sarker, K. O' Connor, *et al.*, Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labelling with word embedding cluster features, *J Am Med Inform Assoc* **22(3)** (2015), 671-681

[3] D. Cameron, G.A. Smith, R. Daniulaityte, *et al.*, PREDOSE: A semantic web platform for drug abuse epidemiology using social media, *J Biomed Inform* **46** (2013), 985-997.

[4] European Medicines Agency (EMA). Guideline on good pharmacovigilance practices (GVP) Module VI. Available at: https://goo.gl/NuGcdi [Accessed March 2019]

[5] F. Morlane-Hondère, C. Grouin, and P. Zweigenbaum, Identification of Drug-Related Medical Conditions in Social Media, *Proc of LREC*, Portoroz, Slovenia (2016).

[6] E.G. Brown, L. Wood, and S. Wood, The Medical Dictionary for Regulatory Activities (MedDRA), *Drug Safety* **20** (1999), 109-117.

[7] World Health Organization (WHO), Anatomical therapeutic chemical (ATC) classification index (2006), WHO Collab. Centre for Drug Statistics Methodology

[8] J. Kalyanam, T. Katsuki, G.R. Lanckriet, *et al.*, Exploring trends of nonmedical use of prescription drugs and polydrug abuse in the Twittersphere using unsupervised machine learning, *Addictive Behavior* **65** (2017), 289-295.

[9] E. Bigeard, N. Grabar, and F. Thiessard, Typology of Drug Misuse Created from Information Available in Health Fora, *Stud Health Technol Inform,* **247** (2018), 351-355.

[10] E. Bigeard, N. Grabar, and F. Thiessard, Detection and analysis of drug misuses. A study based on social media messages. *Frontiers in Pharmacology* **9** (2018), p. 791.

[11] M. Mintz, S. Bills, R. Snow and D. Jurafsky, Distant supervision for relation extraction without labeled data. *Proc. of ACL* (2009)

[12] H. Poon, K. Toutanova, C. Quirk. Distant supervision for cancer pathway extraction from text. *Pacific Symposium on Biocomputing* (2014), 120-131.

[13] B. Audeh, M. Beigbeder, A. Zimmermann, *et al*. Vigi4Med scraper: a framework for web forum structured data extraction and semantic representation. *PloS one*, **12**(1), (2017), e0169658.

[14] T. Lavergne, O. Cappé, and F. Yvon, Practical very large scale CRFs, *Proc. of ACL* (2010), 503-514.

[15] P. Stenetorp, S. Pyysalo, G. Topić, *et al*., BRAT: a web-based tool for NLP-assisted text annotation, *Proc. of the Demonstrations at EACL* (2012), 102-107.

[16] K. Verspoor, A. Jimeno Yepes, L. Cavedon, *et al*., Annotating the biomedical literature for the human variome. *Database*, (2013).

[17] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al*, Scikit-learn: Machine learning in Python, *J Machine Learning Res* **12** (2011), 2825-2830.

[18] S. Santiso, A. Casillas, and A. Pérez, The class imbalance problem detecting adverse drug reactions in electronic health records, *Health Inform J* (2018).

[19] C. Bousquet, É. Sadou, J. Souvignet, *et al.*, Formalizing MedDRA to support semantic reasoning on adverse drug reaction terms, *J Biomed Inform* **49** (2014), 282-291.

[20] O. Bodenreider, The Unified Medical Language System (UMLS): integrating biomedical terminology, *Nucleic acids research* **32**(suppl_1) (2004), D267-D270.

Addresses for correspondence

Leonardo Campillos-Llanos, leonardo.campillos@limsi.fr

Cyril Grouin, cyril.grouin@limsi.fr

[12] https://bit.ly/2OlddCa [Accessed March 2019]

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190184

Expanding Evolutionary Terminology Auditing with Historic Formal and Linguistic Intensions: A Case Study in SNOMED CT

Werner Ceusters[a] and Sarah Mullin[a]

[a] Department of Biomedical Informatics, University at Buffalo, Buffalo NY, USA

Abstract

A method is described to use SNOMED CT's history mechanism as a means to compute how the formal and linguistic intensions of its concepts change over versions. As a result of this, it is demonstrated that the intended principle of concept permanence is not always adhered to. It is shown that the evolution of formal intensions can be monitored fully automatically and that the proposed procedure includes a method to suggest missing subsumers in a concept's transitive closure set by identifying mistakes that have been made in the past. Changes in linguistic intensions were found to be much more labor-intensive to identify. It is suggested that this could be improved if the history mechanism would come with more detailed motivations for change than the current and insufficiently used annotation to the effect that a fully specified name 'fails to comply with the current editorial guidance'.

Keywords:

Systematized Nomenclature of Medicine; Algorithms; Semantics

Introduction

SNOMED CT is a biomedical terminology which is anchored in an ontology of 'concepts' which are defined and related to each other using a combination of formal logic and editorial rules as specified in SNOMED CT's concept model [10, p24]. Fundamental in SNOMED CT is that each of its concepts comes with a unique identifier (SCTID) and is intended to have a unique meaning. What this meaning is intended to be, is conveyed formally by means of description logic-based assertions and informally by means of human-readable 'terms' which are called '*fully specified names*' (FSN) and which are connected to the concepts by means of '*descriptions'* [10, p12]. Each FSN is since 2003 formed by what we henceforth call a '*term proper'* and a '*semantic tag*' which indicates to which subhierarchy of SNOMED CT the concept belongs. For example, the concept with SCTID '87612001' has the FSN 'Blood (substance)' where 'Blood' is the term proper and 'substance' is the semantic tag. Concept 87612001 is an example of a concept for which the term leaves not much room for misinterpretation of the intended meaning. Its formal definition, however, specifies the intended meaning only partially: blood is a subtype of 'Body fluid (substance)' and of 'Blood material (substance)'. Semantic tags also disambiguate concepts for which the term proper is homonymous. This is for instance the case for the term 'hematoma' which comes in two flavors: 'Hematoma (morphologic abnormality)' (SCTID: 35566002) and 'Hematoma (disorder)' (SCTID: 385494008) which is formally defined in terms of the former.

SNOMED CT is since its inception in 2002 updated twice annually, for instance to include missing relevant content [7] or to remove content that was erroneous because of editorial mistakes [9] or because of mismatches in intended meaning between formal definitions and terms [11]. Also changes in the concept model itself require updating which impacts both the formal components and the FSNs. A unique feature of SNOMED CT is that it comes with a history mechanism involving certain formal metadata components that – to some degree – describe what and when changes have been made, what the reasons for these changes were (for example inconsistency with editorial rules), and how impacted components relate to each other after the change (for example what concepts, if any, replace inactivated concepts) [5].

One central principle that is intended to be maintained over versions is code or concept permanence [6]: '*Once assigned a meaning, a code must not change its meaning. Refinements, due to changes in the state of knowledge, may lead to inactivation of codes from SNOMED CT. An inactivated code may be replaced by a new, more precisely defined code*' [9, p203]. The objectives of the work presented here are twofold. The first one is to assess the extent to which the principle of concept permanence is adhered to and whether adherence to this principle can be quantified by resorting to SNOMED CT's history mechanism. The 2nd one is to find methods using this quantification to improve on prior efforts in Evolutionary Terminology Auditing which attempts to find mistakes in the last version of an ontology on the basis of errors made in the past [2]. Our hypothesis is that the stability of a concept's position in the hierarchy over distinct versions and the formal representation of reasons for change [3] contribute positively to quantification while changes in the concept model and in the FSNs contribute negatively [1].

Methods

Since SNOMED CT is an ontology that does not explicitly adhere to a view based on Ontological Realism [12], the meaning of a SNOMED CT concept can be thought of as what is conveyed by means of three aspects: (1) a *linguistic intension* as conveyed through its *label(s)*, (2) a *formal intension*, i.e. the properties implied by it as exhibited, for instance, by means of the formal relations it holds with other concepts and (3) an *extension*, i.e. the collection of data elements in, for instance, electronic medical record systems annotated with the concept [13]. Whether two concepts have the same meaning can then be determined by applying appropriate similarity functions to each of the three aspects followed by an assessment of whether the similarities are sufficiently high. For systems like SNOMED CT that maintain explicit identity over versions, concept permanence – or the opposite: *concept drift* [13] – can then be

Table 1 – Transitive closure history for SCTID:10001005 with most recent FSN: 'Bacterial sepsis (disorder)'

	Historic Formal Intension (HFI)					Historic subsumers' SCTIDs and most recent FSNs	
Ref.	**S1**	**S2**	**S3**	**S4**	**S5**		
(1)	1111	111111111	11	111111111	1111111111	91302008	Sepsis (disorder)
(2)	1111	111111111	11	111111111	1111111111	87628006	Bacterial infectious disease (disorder)
(3)	1111	111111111	11	111111111	1111111111	64572001	Disease (disorder)
(4)	1111	111111111	11	111111111	1111111111	40733004	Infectious disease (disorder)
(5)	1111	111111111	11	111111111	!!!!!!!!!!	105592009	Septicemia (disorder)
(6)	----	111111111	11	111111111	1111111111	404684003	Clinical finding (finding)
(7)	----	---------	11	111111111	0000000000	431950004	Bloodstream finding (finding)
(8)	----	---------	11	111111111	0000000000	431193003	Infection of bloodstream (disorder)
(9)	----	---------	11	111111111	0000000000	301811001	Bacterial infection by site (disorder)
(10)	----	---------	11	111111111	0000000000	301810000	Infection by site (disorder)
(11)	----	---------	11	111111111	0000000000	123946008	Disorder by body site (disorder)
(12)	----	---------	11	111111111	0000000000	118234003	Finding by site (finding)
(13)	----	---------	--	111111111	0000000000	434156008	Infectious agent in bloodstream (finding)
(14)	----	---------	--	---------	1111111111	238149007	Systemic inflammatory response syndrome (disorder)
(15)	----	---------	--	---------	1111111111	128139000	Inflammatory disorder (disorder)

Legend: '**Ref.**': reference. **S1**, … **S5**: stable history segments.

Table 2 – Historic signatures of some example concepts and their Fully Specified Names

Concept ID **Ref.**	**Historic Signature**	**Fully Specified Name (FSN)**	**RS**	**RV**	**S**
373981005	-111111111111111111111111111111111				
(1)	-111111111111111111111111111111000	Triazolam 0.125mg tablet (product)	S1	V3	0
(2)	----------------------------100	Product containing triazolam 0.125 mg/1 each oral tablet (clinical drug)	S1	V2	0
(3)	----------------------------10	Product containing only triazolam 0.125 mg/1 each oral tablet (clinical drug)	S1	V3	0
(4)	----------------------------1	Product containing precisely triazolam 125 microgram/1 each conventional release oral tablet (clinical drug)			0
103497003	111111111111111111111111111111111				
(5)	111111000000000000000000000000000	Streptococcus penumoniae 3 (organism)			1
(6)	------100000000000000000000000000	Streptococcus penumoniae serotype 3 (organism)			1
(7)	-------111111111111111111111111110	Streptococcus pneumoniae serotype 3 (organism)		V3	0
(8)	----------------------------1	Streptococcus pneumoniae Danish serotype 3 (organism)			0
290537002	111111111111111111111111111111111				
(9)	111110001111111111111111111111111	Accidental prilocaine poisoning (disorder)			1
(10)	-----111000000000000000000000000	Accidental prilocaine poisoning (event)			1
120713006	111111111111111111111111111111111				
(11)	111111111111111000000000000000000	Parainfluenza virus II antibody (substance)			0
(12)	----------------11111111111111111	Human parainfluenza virus 2 antibody (substance)			0

Legend: '**Ref.**' = reference. '**RS**' = annotated in a SNOMED CT reference set ('S1': SCTID-900000000000490003: '*Description inactivation indicator attribute value reference set*'). '**RV**': Value associated with the description annotated in RS ('V2': SCTID-900000000000485001: '*a component that contains a technical error*'; 'V3': SCTID-723277005: '*A component that fails to comply with the current editorial guidance*'). '**S**': Semantic similarity function result ('0' = not considered semantically equivalent, '1' = 'semantically equivalent').

evaluated by determining (1) whether changes in the concept's labels lead to different interpretations, (2) whether the formal intension is kept constant, and (3) whether the same sorts of data elements are annotated irrespective of version.

In absence of data annotated by means of different versions of SNOMED CT, we focused our efforts on linguistic and formal intensions. Using the history information from the July 2018 version of SNOMED CT International, we computed for all concepts, all descriptions specifying an FSN containing any semantic tag introduced since 2003 and all relations a *historic signature* as a 34-character string, using one character for each version since January 2002. A character '1' in position *n* indicates the presence of that component in version *Vn* (*n* ranging from 1 to 34), '-' absence of the component prior to its introduction, and '0' absence of the component because of its deactivation. For relationships between two concepts, the extra character '!' was used to indicate that a relationship does not hold anymore because of deactivation of the target concept.

For the formal intension aspect of a concept's meaning, we followed a suggestion from [13] and defined the *formal intension of a concept in Vn* (FIn) as the collection of all subsumers in its transitive closure in *Vn* and *the historic formal intension of a concept* (HFI) as the union of any subsumer ever encountered in some version. We further defined the *rigid formal properties set of a concept* (RFP) as the subset of transitive closure subsumers that is present in all versions in which the concept is active. Finally, we defined a *stable history segment of a concept* (SHSC) as a segment of the concept's history during which the FIn remains constant over successive versions.

Consistent with our proposal advanced in [4], we identified two types of *suspicious events* for historic signatures of concept-subsumer pairs. Concepts C1 and C2 form a concept-subsumer pair if and only if C2 is in the historic transitive closure set of C1. A historic signature for a concept-subsumer pair is *suspiciously gapped* for any transitive closure subsumer which becomes reactivated after having been deactivated: during the gap, the subsumer was thus unjustifiably absent ('missing'). A concept-subsumer pair is *suspiciously annulled in S* whenever there is a stable history segment S for which (1) the historic signature of that concept-subsumer pair contains only '0's and (2) when there is at least one other concept-subsumer pair for that concept of which the historic signature contains only '!' in S: this might be an indication that the subsumers marked by '0' are missing from the transitive closure set because of the removal of any property marked by '!'. To clarify these definitions, Table 1 provides an example of a concept with a HFI of 15, and an RFP of 4. Its HFI has 5 stable segments of which S5 is suspicious; this is because the historic signature of the properties (7) … (13) are suspiciously annulled in S5 due to the historic signature of property (5) in S5 containing '!'. This suggests that the subsumers marked by '0' are missing, i.e. unjustifiably absent, because of the – most likely justified – removal of the duplicate concept 'septicemia' (marked by '!') which was subsumed by these concepts.

Variables used in the analyses thereafter are the sizes of HFI and RFP as well as their ratio, the lifetime of a concept, and its number of suspicious events. A random number generator was used to select two random samples of each 100 concepts that were active since the first version. The samples were manually inspected for possible missing subsumers. The 1st sample was drawn from concepts which were marked as having a suspiciously gapped subsumer but not as having a suspiciously annulled subsumer. The 2nd sample consists of concepts which have at least one suspiciously annulled subsumer but are not marked as having a suspiciously gapped subsumer. Decisions for whether a subsumer is truly missing were based on SNOMED CT's editorial guidelines [9].

For the linguistic intension of a concept's meaning, we collected the historic signatures of all its FSNs over time and indicated for each change from one FSN to the next one whether a reason for the change was specified in one or other reference set distributed as part of SNOMED CT's metadata components (Table 2). Changes were syntactically qualified as having occurred in the term proper, in the semantic tag (ST), or in both. Changes from one ST to another were semantically qualified as being different, thus suggesting a distinct linguistic intension for the concept under scrutiny. To identify whether syntactic changes in the term proper for FSNs with the same ST would qualify as constituting a semantic change as well, we implemented a simple rule-based string transformation algorithm based on 99 rules. This algorithm processes each FSN in the history of a concept by iterating over a manually constructed knowledgebase sanctioning the substitution of certain character sequences (case insensitive). If at the end of the process an identical string is obtained for some FSNs, then these FSNs are considered semantically equivalent. It takes advantage of the fact that FSNs and subsequent changes thereof follow certain patterns. The example in Table 3 works for any HLA-*X*, e.g. HLA-Cw2, HLA-DQw8. Possible non-intended changes as in 'Chlamydia' → 'Cmydia' are innocent for our purposes as they would happen in each FSN of that concept. But obviously, it renders this algorithm inappropriate for computing the semantic similarity of distinct concepts on the basis of their linguistic intensions. A random sample of 200 concepts exhibiting at least one FSN change for which the algorithm failed to conclude semantic similarity was manually inspected for verification.

Changes in semantic tags (ST) were further analyzed by computing transition probabilities from one ST to another ST, and by performing agglomerative hierarchical clustering on larger trajectories and including activation and deactivation, for example finding → event → inactive, or substance →product → medicinal product. The result was assessed using the Ward (minimization of residual variance), average (averages of distances), and complete (minimization of diameter of each new group) methods from R cluster.

Table 3 – String transformation algorithm example

Search string	**Replacement**	**Rule**
"human leukocyte antigen"	→ " "	R1
"antigen"	→ " "	R2
"hla"	→ " "	R3
"-"	→ ""	R4
" "	→ ""	R5
String transformation sequence		**Rule**
'hla-dr8 antigen'		
'hla-dr8 '		R2
'-dr8 "		R3
'dr8 '		R4
'dr8'		R5
'human leukocyte antigen hla-dr8 antigen'		
' hla-dr8 antigen'		R1
' hla-dr8 '		R2
' -dr8 '		R3
' dr8 '		R4
'dr8'		R5
'human leukocyte antigen dr8'		
' dr8'		R1
'dr8'		R5

Results

Our analysis involved 403,360 concepts that were active for at least one version, 340,639 (84.45%) of which are still active in the July 2018 version (Table 4).

The size of the historic formal intensions of concepts ranged from 2 (204 concepts) to maximally 152 (1 concept, most likely not the one most frequently found in an EHR: SCTID:35057008 - *Nonvenomous insect bite of penis with infection (disorder)*). The number of stable history segments ranged from 1 (7,961 concepts) to 25 (15 concepts). Only 61,001 concepts of all concepts (15.13%) exemplified a rigid formal property set (RFP) constituting 100% of its historic formal intension, while 51,936 concepts (15.25%) do so for all currently active concepts. 39,771 (=340,639-300,868, 11.68%) active concepts have at least one suspiciously annulled formal property. 7,583 concepts (403,360-395,777) have at least one suspiciously gapped subsumer. 2,706 concepts exhibited both.

Manual inspection of the samples for possible missing subsumers revealed that 83 of the 100 concepts with the suspicious gap criterium and 91 of those selected on the basis of suspicious annulment of a subsumer did, in our opinion, miss at least one subsumer. Some examples are provided in Table 5.

The number of concepts involved in changes in linguistic intensions, separated in semantic tag changes and term proper changes, are displayed in Table 6. Only 19% of these changes were found to be documented by means of a reference set. Of the remaining 81%, 91% could be eliminated through our term transformation algorithm, thereby still leaving over 48,000 term changes to be manually inspected.

Table 4 - Descriptive statistics for Historic Formal Intension related variables

Descriptive Statistic	*HFI*	*SHSC*	*ACTIVE*	*LIFETIME*	*RFP*	*RFP%*	*Susp. Ann.*	*Susp. gapped*
Mean	22.708	6.249	0.845	25.836	9.975	49.935	1.384	0.381
Standard Error	0.028	0.006	0.001	0.018	0.016	0.048	0.009	0.027
Median	17	5	1	34	7	43	0	0
Mode	7	2	1	34	4	100	0	0
Standard Deviation	17.784	3.861	0.362	11.417	10.026	30.703	5.627	17.164
Kurtosis	2.247	0.577	1.615	-0.609	6.597	-1.091	311.509	63,995.461
Skewness	1.428	0.952	-1.901	-0.999	2.256	0.371	13.736	202.515
Minimum (Min)	1	1	0	1	0	0	0	0
Maximum (Max)	152	25	1	34	125	100	326	6,643
Confidence Level(95.0%)	0.055	0.012	0.001	0.035	0.031	0.095	0.017	0.053
N concepts with Max	1	15	**340,639**	224,447	1	**61,001**	1	1
N concepts with Min	204	7,961	**62,721**	9,689	9,241	9,241	**300,868**	**395,777**
N Active concepts with Max	1	15	340,639	224,447	1	**51,936**	1	1
N Active concepts with Min	180	7,961	0	6,505	8,274	8,274	238,147	333,066

Table 5 – Examples of missing (i.e. once present, but deleted) subsumers in the transitive closure of SNOMED CT concepts

88425004: Congenital anomaly of nervous system (disorder)
- 299735001: Neurological lesion (finding)
- 102957003: Neurological finding (finding)

87290003: Congenital anomaly of head (disorder)
- 204223000: Ear, face and neck congenital anomalies (disorder)

83502000: Operation on tendon sheath (procedure)
- 118667007: Procedure on skeletal muscular system (procedure)

11381005: Acne (disorder)
- 95320005: Disorder of skin (disorder)
- 80659006: Disorder of skin and/or subcutaneous tissue (disorder)
- 106076001: Skin finding (finding)
- 301857004: Finding of body region (finding)
- 19660004: Disorder of soft tissue (disorder)

19838004: In-vitro immunologic test (procedure)
- 103693007: Diagnostic procedure (procedure)
- 362961001: Procedure by intent (procedure)
- 127789004: Laboratory procedure categorized by method (procedure)

230179001: Chronic viral encephalitis (disorder)
- 102957003: Neurological finding (finding)

116316008: Finding of foot region (finding)
- 250171008: Clinical history and observation findings (finding)

118835007: Procedure on ileum (procedure)
- 174035000: Lower gastrointestinal procedure (procedure)

29857009: Chest pain (finding)
- 250171008: Clinical history and observation findings (finding)
- 118222006: General finding of observation of patient (finding)
- 250171008: Clinical history and observation findings (finding)

10002003: Resection of stomach fundus (procedure)
- 38829003: Partial excision (procedure)

116175006: Proximal subtotal gastrectomy (procedure)
- 38829003: Partial excision (procedure)

287812001: Repair of stomach and/or duodenum (procedure)
- 118821005: Procedure on digestive organ (procedure)
- 118717007: Procedure on organ (procedure)

Table 6 – Changes in Fully Specified Names

		Concepts with Semantic Tag changes				
		0	**1**	**2**	**3**	**Total**
Concepts with Term Proper changes	**0**	333,712	33,697	3,833	56	371,298
	1	23,631	3,026	468	19	27,144
	2	1,748	1,704	1,186	0	4,638
	3	46	145	82	0	273
	4	1	1	5	0	7
Total		359,138	38,573	5,574	75	403,360

Table 7 – Examples of changes in linguistic intension without concept deactivation

SCTID: 374142001
1 Product containing miglitol 25 mg/1 each oral tablet (clinical drug)
2 Product containing only miglitol 25 mg/1 each oral tablet (clinical drug)
3 Product containing precisely miglitol 25 milligram/1 each conventional release oral tablet (clinical drug)
SCTID: 100191000119105
1 Acquired asymmetry of prostate (finding)
2 Asymmetry of prostate (finding)
SCTID: 102549009
1 Night cramps (finding)
2 Cramp in lower leg associated with rest (finding)
SCTID: 106109006
1 Number of previous abortions (finding)
2 Number of previous induced termination of pregnancy (finding)
SCTID: 302828001
1 Syringoma (disorder)
2 Syringoma of skin (disorder)

In our sample of 200 concepts with at least one such non-documented change, we discovered 15 concepts with an FSN change exhibiting a clear shift in meaning. Some examples are shown in Table 7. Hierarchical clustering revealed statistically significant (1) that findings typically transition to events, (2) that multiple semantic tags (context dependent category, finding, procedure, regime/therapy, and disorder) transition to situation semantic tags, and (3) that substance and product have transitioned to medicinal product form, clinical drug, or became significantly more inactive.

Discussion

Computing the historic formal intension of all concepts in SNOMED CT requires a thorough understanding of the meta-components, but is algorithmically straightforward. As to the question of what formal properties should be included in it – direct subsumers only, stated relationships separate from inferred ones, the complete transitive closure with or without all associative relationships – there is no agreed upon answer [13]. Our preference for using the full transitive closure set made it possible to identify for those concepts whose formal intension changes under that criterion, subsumers that were possibly inadvertently removed as a consequence of rightfully removing some subsumed concept. While many assessments are straightforward, a problem for the evaluation, however, is that not enough textual definitions are provided for terms and concepts. What is, for instance, the scope of 'partial' in a subsumer? A 'resection of stomach fundus' (Table 5), whether complete or partial to the fundus is for sure partial for the stomach. Also the use of 'and' and 'or' is problematic. SNOMED CT's editorial guide comes in here handy, but it seems that the application thereof by SNOMED CT's authors is not rigorously followed.

The same holds for evaluating FSN changes that are suspicious for changes in the linguistic intension (Table 7). We can't imagine that clinicians who used in an earlier version a concept of the form 'Product containing X', would consider that equivalent to 'Product containing *only* X' and 'Product containing *precisely* X'. Over 300 change-sequences of this sort have been made in 2018 despite deactivation of the concepts involved seems to have been the more logical choice in light of concept permanence. Finding such meaning changes turned out not to be straightforward precisely because the reason for change mechanism is insufficiently used. Our algorithm for comparing linguistic intensions can for sure be improved, but more practical would it be if SNOMED CT would include a much more detailed list of reasons for change, and why not, a formal representation of all those conditions which make a component follow – or not – the 'current editorial guidelines'. This includes changes related to the concept model itself. Our findings related to the transitions involving semantic tags are consistent with those obtained via another methodology in [1]. It is in the first place an incomplete anchoring of the semantic tags into the formal hierarchy that poses a problem.

A limitation of the work presented here is that more manual analysis of discrepancies found is required in order to produce clear cut precision and recall values for our proposed algorithm. Also more experimenting with alternatives for historic formal intension computation is needed. Finally, it is worth exploring which missing subsumption relations detected through our effort are found as well through other methods [7; 8; 14].

Conclusions

Our results demonstrate that SNOMED CT's intended adherence to the criterion of concept permanence can be quantified but that, unfortunately, this criterion is not sufficiently applied. That changes in the concept model as expressed through changes in semantic tags have a negative effect on the automatic interpretation is re-confirmed: it cannot be formally computed whether for any given concept a semantic tag is changed because of a *local* mistake in the interpretation of that concept or a *global* change at the level of the concept model. Changes in linguistic intensions quantified on the basis of changes in the term proper of FSNs are detectable as well, but currently only with low estimated recall and precision.

Without doubt, our work demonstrates that SNOMED CT's history mechanism is a formidable resource from which valuable knowledge can be extracted to prevent mistakes in the future. It is our opinion that a mechanism like this should be standardly available in any ontology worth the name.

Acknowledgements

This work was supported in part by Clinical and Translational Science Award NIH 1 UL1 TR001412-01 from the National Institutes of Health and the Department of Veterans Affairs award #80307 'Identifying clinical and logical shortcomings in SNOMED CT', Projects #1144333 and #18114806.

References

[1] J.P. Bona and W. Ceusters, Mismatches between major subhierarchies and semantic tags in SNOMED CT, *J Biomed Inform* **81** (2018), 1-15.

[2] W. Ceusters, Applying Evolutionary Terminology Auditing to SNOMED CT, *AMIA Annu Symp Proc* **2010** (2010), 96-100.

[3] W. Ceusters, SNOMED CT's RF2: Is the future bright?, *Stud Health Technol Inform* **169** (2011), 829-833.

[4] W. Ceusters, SNOMED CT revisions and coded data repositories: when to upgrade?, *AMIA Annu Symp Proc* **2011** (2011), 197-206.

[5] W.M. Ceusters, K.A. Spackman, and B. Smith, Would SNOMED CT benefit from realism-based ontology evolution?, *AMIA Annu Symp Proc* (2007), 105-109.

[6] J.J. Cimino, Desiderata for controlled medical vocabularies in the twenty-first century., *Methods of Information in Medicine* **37** (1998), 394-403.

[7] L. Cui, O. Bodenreider, J. Shi, and G.Q. Zhang, Auditing SNOMED CT hierarchical relations based on lexical features of concepts in non-lattice subgraphs, *J Biomed Inform* **78** (2018), 177-184.

[8] L. Cui, W. Zhu, S. Tao, J.T. Case, O. Bodenreider, and G.Q. Zhang, Mining non-lattice subgraphs for detecting missing hierarchical relations and concepts in SNOMED CT, *J Am Med Inform Assoc* **24** (2017), 788-798.

[9] International Health Terminology Standards Development Organisation, SNOMED CT Editorial Guide, in, 2018.

[10] International Health Terminology Standards Development Organisation, SNOMED CT Starter Guide, in, 2018.

[11] J.M. Rodrigues, S. Schulz, B. Mizen, A. Rector, and S. Serir, Is the Application of SNOMED CT Concept Model sufficiently Quality Assured?, *AMIA Annu Symp Proc* **2017** (2017), 1488-1497.

[12] B. Smith and W. Ceusters, Ontological realism: A methodology for coordinated evolution of scientific ontologies, *Applied Ontology* **5** (2010), 139-188.

[13] S. Wang, S. Schlobach, and M. Klein, Concept drift and how to identify it, *Journal of Web Semantics* **9** (2011), 247-265.

[14] D. Wei, H. Helen Gu, Y. Perl, M. Halper, C. Ochs, G. Elhanan, and Y. Chen, Structural measures to track the evolution of SNOMED CT hierarchies, *J Biomed Inform* **57** (2015), 278-287.

Address for correspondence

Werner Ceusters, Department of Biomedical Informatics, 77 Goodell street, Bufallo NY – 14203, USA. Email: ceusters@buffalo.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190185

Generation of Surrogates for De-Identification of Electronic Health Records

Aipeng Chen[a], Jitendra Jonnagaddala[b,c], Chandini Nekkantti[d], Siaw-Teng Liaw[b,c]

[a] *Prince of Wales Clinical School, UNSW Sydney, Australia*
[b] *School of Public Health and Community Medicine, UNSW Sydney, Australia.*
[c] *WHO Collaborating Centre for eHealth, UNSW Sydney, Australia*
[d] *CGD Health Pty Ltd, Australia*

Abstract

Unstructured electronic health records are valuable resources for research. Before they are shared with researchers, protected health information needs to be removed from these unstructured documents to protect patient privacy. The main steps involved in removing protected health information are accurately identifying sensitive information in the documents and removing the identified information. To keep the documents as realistic as possible, the step of omitting sensitive information is often followed by replacement of identified sensitive information with surrogates. In this study, we present an algorithm to generate surrogates for unstructured electronic health records. We used this algorithm to generate realistic surrogates on a Health Science Alliance corpus, which is constructed specifically for the use of development of automated de-identification systems.

Keywords:

Algorithms, data anonymization, electronic health records

Introduction

Unstructured electronic health records (EHR) such as discharge summaries, encounter notes, pathology reports and radiology are valuable source of information for basic, clinical and translational researchers [1-4]. These documents are often shared with researchers for clinical and biomedical research purposes. However, these documents may contain sensitive protected health information (PHI) such as patient name and unique patient identifiers. When private information is removed, patients are willing to share their medical records for research use [5]. To protect the privacy of patients, researchers need to share or access these documents in de-identified manner. Traditionally, the PHI was manually identified and removed before they are shared. However, with large number of documents, the manual process is tedious and not feasible to scale. In a study by Dorr et al [6], it was reported that average time required to manually de-identify one document (7.9 types of PHI per document) is 87.3 seconds. Various studies have shown that it is possible to automatically de-identify unstructured EHRs with acceptable accuracy [7-10].

Automated de-identification can be employed to replace traditional manual process. The process often consists of two main stages: identifying PHI and replacing identified PHI with surrogate information. Surrogates are realistic replacements of PHI that are close to real PHI but do not contain any private information. There are a few challenges in surrogate generation.

First, surrogates are supposed to be as realistic as possible. Since the documents are mostly intended to be shared for research, readability of the text need to be maintained. Researchers need to understand the text and ineffective surrogates can be distracting. Additionally, natural language processing systems also need the documents to be realistic so that the documents can be used to train the models to detect PHI in unseen documents. Second, the context of the document needs to be preserved. Some words, especially names and locations, can appear in different forms in one piece of text. For example, a name 'John Smith' can appear in the text as 'Smith, John', 'JS', 'Mr. Smith' or just 'Smith', they refer to one person and failing to maintain these complexities can lead to biased training during the development of de-identification systems.

Third, the temporal information also needs to be preserved well so that the order of occurrence of the clinical events is maintained. Lastly, PHI can carry not only private information, but also information that could be useful in research. For example, a person's name could imply one's gender, which might be useful when no structured gender information is not available.

Some datasets use placeholders or other forms of obfuscation to remove PHI [11, 12]. Stubbs & Uzuner [13] discussed the challenges of generating realistic surrogates and described their algorithm to generate surrogates. Multiple strategies were applied to PHI of different categories in doing this algorithm. Alphabet and date shift were introduced to maintain the consistency of the context in the document. However, these methods were specific to the US and not directly translatable to Australian setting. As a result, the readability of the surrogated documents is low, and the performance of automated de-identification systems trained using this surrogated data might not perform well on Australian EHR. In this study, we improved previous strategies for surrogate generation to make it more relevant to Australian setting. Several matching strategies for different kinds of PHI were developed to match PHI with same meaning expressed in different formats. We have used the Health Science Alliance (HSA) biobank de-identification corpus. HSA biobank is an institutional biobank at UNSW Sydney for translational research.

Methods

HSA Biobank Corpus

We constructed a large corpus of pathology reports that was annotated specifically for the use of development of automated de-identification systems for unstructured EHRs. The corpus consisted of 2100 pathology reports from 1833 patients. There were 38414 pieces of PHI identified in the corpus. Most of the PHI was tagged as names, locations, dates and IDs. Only few of

the PHI were related to PHI categories: contact and age. There was no category of profession or other information observed in the HSA corpus. EHRs and their annotations were stored in the format of XML files as shown in the Figure 1.

Figure 1— Annotated Sample Document from the HAS Corpus in XML Format

PHI Categories

The definition of PHI categories was the same as it was in the i2b2 2014 de-identification shared task [14]. This in turn was developed based on the HIPAA (Health Insurance Portability and Accountability Act) established by the USA which defines 18 categories of PHI. The guideline expanded the original 18 categories to include more information. All patients' ages were included despite that only ages above 89 years were considered PHI in the HIPAA. These categories had been grouped into 8 main categories and 25 sub-categories. Detailed categories and sub-categories along with examples of each categories are presented in Table 1.

Table 1 – Detailed PHI Categories and Sub-Categories

PHI category	Sub-category	Example
Name	patient, doctor, username	John Doe, Dr. Max, Mr. Smith
Profession	none	lawyer, teacher
Location	room, department, hospital, organization, street, city, state, country, zip, other	peri-operative unit-pow, macquarie ward – rhw, 12 abc street
Age	none	23, 98
Date	none	24/12/1987, September 26th
Contact	phone, fax, email, url, ipaddress	+61-421123456 abc@gmail.com 194.223.1.1
IDs	social security number, medical record number, health plan number, account number, license number, vehicle id, device id, biometric id, id number	mrn: 9174338 id number: 12r1500257
Other	none	finger print, company logo

HSA Biobank Surrogate Algorithm

Names were collected from the Internet for the use of surrogate generation so that the surrogates can be closer to reality. Names are stored in separate files by categories. Names collected were names of individuals and locations. Names included first names and surnames. Location names included Australian states, cities, streets, organizations and hospital names. These files were loaded into the memory as lists and were later used to generate surrogates. The names and location information were specific to Australia. When a PHI entity was not found in the existing constructed surrogates map, PHI was considered as first-time occurrence, and a new surrogate was generated in various ways according to the PHI's category. In many cases, such as IDs, phone numbers, URLs and emails, it was easier to generate surrogates since they were merely combination of strings of digits and letters and sometimes some special characters such as commas, periods, parentheses. We simply replace them with randomly generated strings with the original format kept. There were some other PHI categories such as individual names, locations names lists were used so that the surrogates can appear more realistic. We formulated some rules to pick surrogates from these lists for some categories to improve the performance, which is discussed in later sections. The HSA biobank surrogate algorithm is available at https://github.com/TCRNBioinformatics/PHISurrogates/.

Name and Location

Surrogate generation for names turned out to be the most challenging part. Mapping names that could appear in different forms and preserving as much information as possible were the key challenges.

A name can appear in many different forms in one given document. For example, if a comma existed in the name, we considered the name included both first and last name. Accordingly, the algorithm removed the comma and replaced original information with surrogate names from the lists we constructed. After this replacement, names could appear as a full name, or combination of initials, or a combination of the initial of one name and another name as full. In the algorithm, we took the first two letters from both parts of the name as the key in the map. For example, "JOSM" in the case of "John Smith". If an abbreviation was observed in the text, a space character was used. For example, "J(space)SM" for "J Smith" and "JO (space)(space)" for "John". The reason for not using only the first initial letter was that different names might have identical initials and it could cause mistake in mapping. For example, "John Smith" and "Jack Scott" had the same initials "JS". On the other hand, "John" and "Johnny" could stand for the same person if "Johnny" was used as a nickname for "John". So, taking the whole name as a key could also cause mismatching in the map. With the two initials as key, a simple rule can be applied: if more than half of the letters in the key are same, they are considered as same name. For example, "JOSM", "J(space)S(space)", "JOS(space)", "J(SPACE)SM" and "J(space)S(space)" were all considered one name. As for the names as initials, we did not look up in the map but proceeded to the surrogate generation directly as the algorithm was able to map a full name's initials to the initials of its possible surrogate with the method describe above. For example, "JS" can be directly mapped to "LU" without looking up in the map or the name table so it is faster. The titles (i.e., "Dr.", "Mr.", "Miss.", "Mrs.") were extracted out initially. These titles were then added back to the surrogate names generated later.

In order to preserve context, two rules were applied when generating surrogates. Firstly, to preserve the gender information, we constructed and split the names dictionary into three types; male first names, female first names, and surnames. A first name was first searched in both male and female first names list to determine its gender. For those that belonged to both, or not found in both, a random gender was given. Then a surrogate was picked from the dictionary according to the name's gender. Then alphabet shift of fixed length was applied. With

this shift, the initial letter of a name was mapped to another letter in the alphabet, then a name starting with this letter was picked from the names dictionary lists as surrogate. We maintained various dictionaries for different sub-categories of locations, including countries, states, cities, streets, hospitals and organizations. Surrogates of locations were randomly picked from dictionaries accordingly.

Date and Age

A randomly generated date shift in the range of 1 to 730 days was applied. Dates and ages in records from same patient were shifted with the same length to maintain temporality. Similar to the names PHI category, dates could appear in various forms. For example, '18/12/2017' could appear as '18.12.2017' or '2017-12-18'. Similarly, a date '18/12' could be written as "December 18" or "Dec 18". Another challenge was the ambiguity of date strings caused by different date notation styles. For example, a date string "03/04/05" could possibly in the format of "DD/MM/YY" or "YY/MM/DD" or "MM/DD/YY". Since all the reports were retrieved from Australian hospital, we considered all the date strings are of either "DD/MM/YY" or "DD/MM/YYYY" format. We normalized all the non-standard date variations into ISO-8061 standard, "YYYY-MM-DD". The parsed dates were then shifted by adding a time shift in the format. As for the age information, applying a date shift was relatively easier. The date shift in days was converted into years and added.

ID and Contact

Surrogates for IDs were relatively easier to generate. For BIOID and IDNUM, surrogates started with two digits from 10-99, followed by an "N" and an "R" respectively, then seven digits from 1000000-9999999, as similar format was observed in our documents. Medical record numbers with seven digits varied between 1000000 and 9999999, followed by three alphabets. As for fax and phones, all surrogate numbers were generated in the Australian format: two digits from 01-08, followed by seven digits from 0000000 to 9999999. For emails, the format was a random string of length 32 plus "@gmail.com". Surrogates found from the list map or newly generated information was then used to replace the original PHI.

Results

We presented an algorithm for surrogate generation for de-identification of unstructured EHRs. This algorithm can be used either during the de-identification corpus construction or during the development of automated de-identification systems for unstructured EHRs. We tested our algorithm on 2100 annotated pathology reports from the HSA corpus. The algorithm took a total time of 2.94 seconds to process the 2100 documents and 0.0014 seconds per document on average. There was 38,414 pieces of annotated PHI in the corpus, 18.29 in each file on average. Most of the surrogates generated were names and locations (Table 2). Among all PHI entities in the HSA corpus, a total number of 1085 pieces of PHI were replaced with the previously generated surrogates found in the existing map before a new surrogate was generated (Table 2). Though it is not feasible to validate every surrogate generated, two authors have manually verified 5% of the documents from the HSA corpus to assess the readability and contextual information of the documents after surrogate generation. Our algorithm has generated surrogates as intended but few issues were observed, as we discuss in the next section.

Table 2— Count of Surrogates Generated by Categories and Found in the Lists Developed

Category	Count of surrogates	Count of surrogates from lists developed
Name	11789	284
Age	141	0
Contact	7	0
Location	9861	104
Date	7665	321
ID	8951	376
Professions	0	0
Other	0	0
Total No. of documents	2100	1085

Discussion

We discussed in detail strategies applied to tackle the key challenges in surrogate generation. Lists of names and locations were collected and used to make the surrogates as realistic as possible. Maps of existing surrogates were used to make sure that same surrogates were used to replace the same objects so that the context could be maintained. Formats of dates, IDs and contacts were saved according to the Australian standards since the EHRs were all retrieved from Australian hospitals. For names, a key of first two letters of both part of names was designed to deal with co-reference and ambiguity. Also, a date shift was applied for the generation of dates and ages so that the progress of date in the context can be preserved. The information of names was preserved by determining the possible gender of name by searching in the real name lists.

Alphabet shift was applied on names to maintain the context. There are advantages mapping the letters with a fixed alphabet shift. We could easily replace the initials with the shifted letters without concerning about the ambiguity of initials, such as "JS" for either "John Smith" or "Jack Scott" would both be replaced by "LU" with a shift of length 2. Additionally, mistakes in, mapping initials to names could be reduced too. A consistent shift of initials could make sure that the surrogates are distributed more evenly so that the surrogated documents can have a better performance when training automated de-identification systems. It is possible that the shift of initial letters in names can be inferred according to the frequency of letters in the context. It is possible that an initial can be used to identify a person, but since there are a limited number of names that appear in the documents, and the alphabet shift is generated randomly per document, it is almost impossible to deduct the original initials from the surrogate names. Even if the alphabet shift pattern is identified, only the initials of the names can be obtained, so the risk of re-identification is still negligible.

The findings in this study are subjected to several limitations. There was no profession observed in the corpus, the surrogate generation process on professions was not applied and tested. Patients' health status might be related to their profession and a surrogate need to preserve such information so that the potential relationship could be used for research. Also, as these documents are generated by clinicians, there is a possibility of spelling errors which could impact surrogate generation process. Our algorithm doesn't consider this and as a result it is possible that a misspelt PHI entity could be replaced with a non-contextual based surrogate. Date shift was applied to dates and ages in the algorithm. However, some date information remained in the documents such as holiday names, like Christmas and New

Year's Eve. This information can be used along together with the generated surrogates to infer the date shift and therefore get the original date and age. In future, we would like to address these limitations by improving our surrogate generation algorithm in turn reducing the risk of re-identification.

Conclusion

In summary, we presented an algorithm to generate realistic contextual surrogates for unstructured EHRs de-identification systems. The algorithm can also be used to construct a corpus for the development of de-identification systems. In the algorithm, replaced the PHI with realistic surrogates in order to maintain the quality and context of the documents. Australian names and date formats were used in this study. Our findings suggest that the algorithm presented in this study is capable of processing large number of documents within few seconds. However, the documents need to have PHI information already identified. Different strategies were applied to tackle different challenges. An existing surrogate map is maintained to make sure that same PHI, or PHI with the same meaning are replaced with the same surrogate so that the context is preserved. However, professions and PHI annotated as 'other', which can possibly be an important source for research use, and a critical risk of identification leak, are not observed in our HSA corpus. This study suffers from various limitations such as failure to handle misspellings and holiday information. Future work in this area is required, especially to reduce the risk of re-identification.

Acknowledgements

This study was conducted as part of the electronic Practice Based Research Network (ePBRN) and Translational Cancer research network (TCRN) research programs. ePBRN is funded in part by the School of Public Health & Community Medicine, Ingham Institute for Applied Medical Research, UNSW Medicine and South West Sydney Local Health District. TCRN is funded by Cancer Institute of New South Wales and Prince of Wales Clinical School, UNSW Medicine.

References

[1] Berman, J.J., Concept-match medical data scrubbing: how pathology text can be used in research. Archives of pathology & laboratory medicine, 2003. 127(6): p. 680-686.
[2] Pitson, G., et al., Developing a Manually Annotated Corpus of Clinical Letters for Breast Cancer Patients on Routine Follow-Up. Studies in health technology and informatics, 2017. 235: p. 196-200.
[3] Stubbs, A. and Ö. Uzuner, Annotating risk factors for heart disease in clinical narratives for diabetic patients. Journal of Biomedical Informatics, 2015. 58: p. S78-S91.
[4] Jha, A.K., The promise of electronic records: around the corner or down the road? Jama, 2011. 306(8): p. 880-1.
[5] King, T., L. Brankovic, and P. Gillard, Perspectives of Australian adults about protecting the privacy of their health information in statistical databases. International Journal of Medical Informatics, 2012. 81(4): p. 279-289.
[6] Dorr, D.A., et al., Assessing the difficulty and time cost of de-identification in clinical narratives. 2006. 45(03): p. 246-252.
[7] Golle, P. Revisiting the uniqueness of simple demographics in the US population. in Proceedings of the 5th ACM workshop on Privacy in electronic society. 2006. ACM.
[8] El Emam, K., et al., The re-identification risk of Canadians from longitudinal demographics. 2011. 11(1): p. 46.
[9] Institute, D.L.A.N.U.A.P.H.C.R., Contentious crop: harvesting information from electronic health records. 2005.
[10] Carrell, D., et al., Hiding in plain sight: use of realistic surrogates to reduce exposure of protected health information in clinical text. 2012. 20(2): p. 342-348.
[11] Kushida, C.A., et al., Strategies for de-identification and anonymization of electronic health record data for use in multicenter research studies. 2012: p. S82-S101.
[12] Chakaravarthy, V.T., et al. Efficient techniques for document sanitization. in Proceedings of the 17th ACM conference on Information and knowledge management. 2008. ACM.
[13] Stubbs, A., et al., Challenges in Synthesizing Surrogate PHI in Narrative EMRs, in Medical Data Privacy Handbook. 2015, Springer. p. 717-735.
[14] Stubbs, A. and Ö. Uzuner, Annotating longitudinal clinical narratives for de-identification: The 2014 i2b2/UTHealth corpus. Journal of Biomedical Informatics, 2015. 58, Supplement: p. S20-S29.

Address for correspondence

Dr. Jitendra Jonnagaddala; z3339253@unsw.edu.au

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190186

Fan-Beam Based Virtual Fluoroscopy for Navigated Catheterization in Interventional Radiology

Pierrick Guiral[a], Patrick Pittet[b], Yannick Grondin[c], Patrice Jalade[d], Jean-Marc Galvan[b], Guo-Neng Lu[b], Laurent Desbat[a], Philippe Cinquin[a]

[a] *Univ. Grenoble Alpes, CNRS, Grenoble INP, TIMC-IMAG, F-38000 Grenoble, France*
[b] *Université de Lyon, CNRS, Université Lyon 1, INL, F-69100 Villeurbanne, France*
[c] *SurgiQual Institute, F-38240 Meylan, France*
[d] *HCL, Centre Hospitalier Lyon Sud, Service de Radiophysique et Radiovigilance, F-69310 Pierre-Bénite, France*

Abstract

Personalized medicine implies reducing invasiveness of therapeutic procedures. Although interventional radiology proved a very interesting alternative to surgical procedures, it still raises concerns due to the irradiation dose received by the medical team (and by the patient). We propose a novel concept allowing to reduce very significantly the irradiation dose during the phases where tools inserted in the patient have to be tracked with respect to previously acquired images. This implies inserting a miniaturized X-ray detector in the tip of the tools, and reducing the dose by a "rotating collimator". We demonstrate that real-time processing of the signals allows accurate localization of the tip of the tools, with a dose reduction of at least ten times.

Keywords:
Image Processing, Computer-Assisted; Radiology, Interventional; Catheterization

Introduction

Interventional radiology is nowadays widely used in many medical fields such as cardiology, neurology, vascular surgery, gastroenterology, urology, gynaecology and orthopedics. In 2010, 545,000 interventional radiology (IR) procedures were reported in France for both diagnoses and treatments [1]. The benefits of interventional radiology are numerous and well established.

Among these procedures, catheterization consists in introducing a catheter in the patient's body through arteries or vessels to a certain region of the body for diagnostic or therapeutic purposes. Catheterization is a delicate procedure and it is most often performed under continuous 2D X-ray imaging, i.e. fluoroscopy guidance, to help the physician localize the catheter tip in the vascular system.

However, since fluoroscopy is performed each time the catheter moves, the staff (mostly the operating physician) and the patient can be significantly exposed to ionizing radiations, particularly in long procedures. Deterministic effects threshold due to high doses may be reached in such long procedures [2].

To follow the guidelines of the ALARA principle (As Low As Reasonably Achievable) [3], there has been a proposed approach for fluoroscopy guidance to reduce the dose received by staff and patients during IR procedures [4]. As compared to conventional fluoroscopy guidance which implements an X-ray cone beam, this approach suggests Virtual Fluoroscopy (VF) guidance based on a rapid scanning of the imaging field by an X-ray fan beam.

In this paper, we report the implementation of this approach as a proof of concept. Firstly, the operating principle of the method and its implementation are described. Then, the in-vitro exprimentation carried out in clinical conditions is presented. Finally, measured results are analyzed and discussed before conclusion.

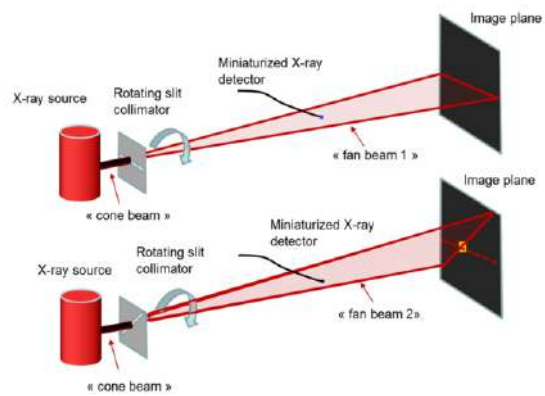


Figure 1 – Principle of the Proposed Virtual Fluoroscopy Method Based on X-Ray Fan Beam Scanning of the Imaging Field

Methods

Operating Principle

The proposed VF approach makes use of an X-Ray fan-beam to scan the image field during each fluoroscopy pulse. A rotating-slit collimator is positioned at the cone beam output of the X-ray tube. This allows the obtaining of a rotating fan beam, as illustrated in Figure 1. A miniaturized real-time X-ray detector in a guidewire is placed between the slit collimator and the image panel. The rotation speed of the collimator disk is set to be equal to the fluoroscope pulse rate. During the fan beam scan of the image field, the detector probe is irradiated two times at two specific brief moments corresponding to two projected slit positions on the image plane, $l_1(t_1)$ and $l_2(t_2)$. By detecting the detector-irradiated moments t_1 and t_2, one can determine the detector position projected in the image plane. Figure 1.a and b illustrates the two detected moments t_1 and t_2,

with corresponding slit projections l_1 and l_2 in the image plane. By calculating intersection of l_1 and l_2, the detector position projected in the collimator plane, x, is first determined and then, the corresponding position in the image plane, x', is obtained by projection along the line defined by x and the X-ray source focus. Thus, during the IR procedure, the detector position can be real-time displayed at the fluoroscopy pulse rate.

Implementation

Figure 2 implements this VF approach using several major parts presented as follows.

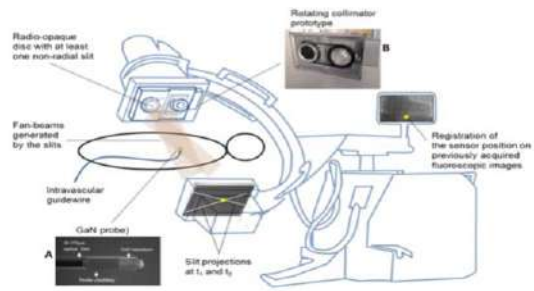


Figure 2 – Implementation of Virtual Fluoroscopy Guidance

Rotating Slit Collimator

The collimator is a 1mm-thick disk, made of Cadmium Tungstate (CdWO4), a high-density material (2'' wafer from Saint Gobain Cristal). It is noted that 1mm-thick CdWO4 absorbs more than 95% of the photon fluency at 100 keV. Moreover, CdWO4 is a scintillator with relatively high light yield, and its prompt luminescence can be detected for the synchronization of the collimator rotation on the fluoroscopy pulse rate.

The disc has a single off-axis straigth-line slit positioned at a distance *R*=2mm from its axis, as shown in Figure 3. The slit has a width of 220µm. The collimator is driven using a brushless motor (80280007, Crouzet, France) for a rotation speed up to 3000rpm (which corresponds to a fluoroscopy pulse duration down to 20 ms).

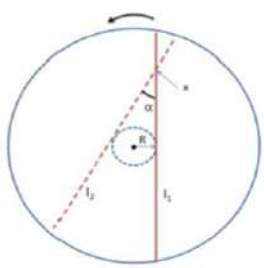


Figure 3 – Design of a Collimator with a Single Off-Axis Slit

Miniaturized Detector Probe

The detector probe is made of a small-volume Gallium Nitride (GaN) radioluminescent (RL) transducer and a coupled optical fiber to collect RL signal and to transmit it to a photodetection module. The GaN transducer has a very high RL yield with a response time in the order of ns [5]. The use of a small volume GaN transducer can produce detectable RL signal, which is suitable for designing a miniaturized detector probe.

The designed detector probe has an outer diameter smaller than 400µm, incorporating at its tip end a Si-doped GaN crystal of ~0.1mm^3 (Saint Gobain Cristals, France).

The detector probe is connected to a photodetection module for its operation. The associated photodetection module mainly consists of a photon counting head (H10682-210, Hamamatsu photonics Corp., Japan). It outputs a pulse of 10ns-width per photon and provides a pulse-pair resolution of 20ns. This signal is processed in real time by an FPGA-based processing unit.

This unit is designed to perform differents tasks: i) synchronous real-time processing of detected signal (sliding counting window of 100µs with a 10 µs resolution), ii) motor speed control and iii) instantaneous collimator position tracking.

System for In-Vitro Testing

Figure 4 illustrates a VF testing system making use of different designed parts presented above. It employs a cylindrical Plexiglas Phantom of 12cm in diameter. The detector probe is placed on the axis of this phantom. The X-ray cone beam comes from the kV On-Board-Imager (OBI) embedded on the radiotherapy system (Truebeam, Varian, Inc) installed at Centre Hospitalier Lyon Sud. This clinical equipment allows accurate positioning and alignment of the collimator and the phantom. OSL dosimeters are also employed at the top surface of the phantom for absolute dose measurements.

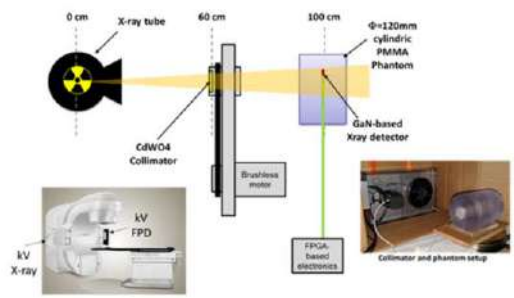


Figure 4 – System for In-Vitro Testing

Results

Dose Reduction

The proposed VF approach allows dose reduction by the ratio between the fan-beam field and the cone-beam field. By assuming an homogeneous dose in the cone beam field (i.e. $D_{CB}{\sim}constant$), the dose at a given Point of Interest (POI) in the fan beam field can be expressed in function of the cone-beam dose as:

$$\begin{cases} \boldsymbol{D_{FB}(r){\sim}D_{CB}\frac{cos^{-1}\left(\frac{R-0.5w}{r}\right)-cos^{-1}\left(\frac{R+0.5w}{r}\right)}{\pi}\ for\ r>R} \\ \boldsymbol{D_{FB}(r)=0\ for\ r{\leq}R} \end{cases} \quad (1)$$

where w is the width of the slit.

Figure 5 shows the virtual fluoroscopy dose evaluated according to equation 1.

Due to the off-axis position of the slit in the collimator, central area of the field is never exposed (blind area). For any POI positioned at an off-axis distance larger 2.5mm, the calculated dose behind the collimator is more than 20 times smaller than the one in front of the collimator.

Experimentally, we acquired images with the OBI X-ray imaging system while the collimator was in rotation at 1000rpm. We used a high dose setting of the tube to obtain a signal strong enough on the imager (i.e. 70kV, 155mA, 644ms). An X-ray image is shown in Figure 5 as well as the grey level profile along the yellow line. The central black disk of 4mm in diameter corresponds to the blind area. The bright white ring surrounding this area is due to the quasi-tangential position of the slit for these positions. The measured grey level profile is very consistent with the calculated profile shown in Figure 6.

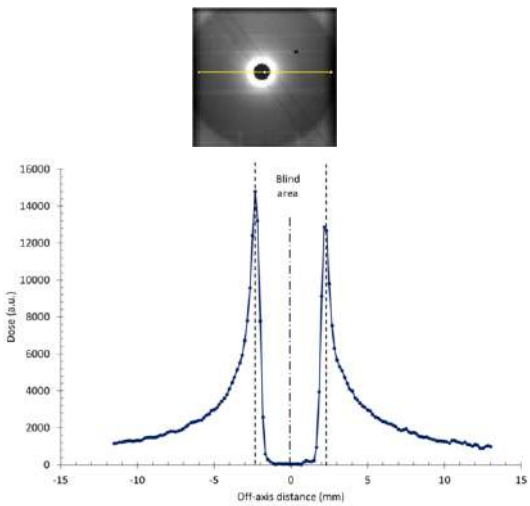


Figure 5 – Off-Axis Dose Variation Evaluated on the Rotating Collimator Image with the Gray Levels Profile Measured Along the Yellow Line

The absolute doses were measured by using the nanoDot OSL dosimeters (Landauer Inc, USA). Each consists of a 5mm diameter, 200µm thick Al2O3:C disk. The dosimeters were placed at distances D1=8mm and D2=18mm from the collimator rotation axis, on the top surface of the phantom. Each dosimeter was exposed for 10 images (70kV, 100mAs per image). The doses measured without the collimator were of 64mSv and of 62mSv for D1 and D2 positions, respectively. The dose with the collimator in operation (at 1000 rpm) decreased down to 984µSv and 681µSv, respectively. It corresponds to a dose reduction by a factor of 66 and 94 for D1 and D2, respectively.

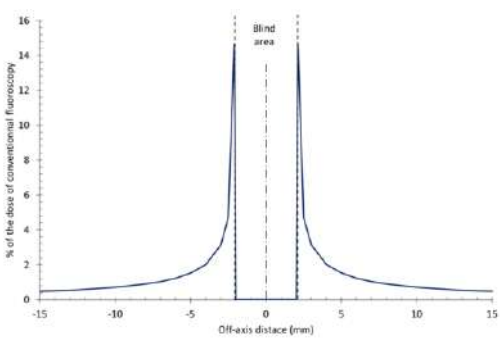


Figure 6 – Evaluated Fan-Beam Dose Versus the Off-Axis Distance of the POI (expressed as percentage of the cone beam dose)

Detection of the X-Ray Detector Probe Exposure

We placed the X-ray detector probe at a depth of 60 mm in the PMMA phantom and successively at 4mm, 7mm, and 12mm from the rotation axis (projected positions in the collimator plane). The obtained images and results are shown in Figure 7. The collimator rotation speed was set at 1000rpm and 1500rpm, respectively.

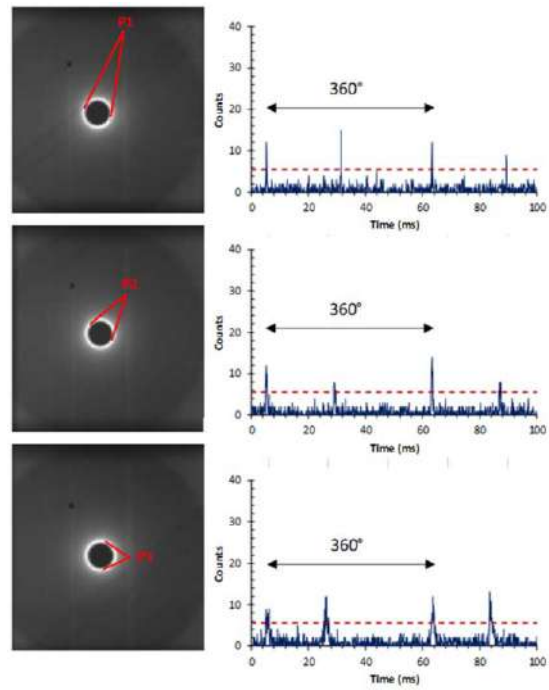


Figure 7 – X-ray Detector Probe Signal for Probes Placed at 60mm Depth in the PMMA Phantom and at Distances of a) 4mm, b) 7mm and c) 12mm from the Rotation Axis.

Figure 7 shows the signal acquired for the 3 detector probe positions. The noise floor is much higher than the dark count noise of the detector but remains sufficiently low to identify the signal peaks resulting from the detector irradiation. The slit rotation speed can be measured as the time between the i^{th} and $(i+2)^{th}$ peaks. It was measured at ~58ms/turn (i.e. 1035rpm) and ~38ms/turn (i.e. 1569rpm) for 1000rpm and 1500rpm setups, respectively.

The time between two successive peaks linearly depends on the collimator rotation angle, β required to move from slit position l_1 to the l_2 position. The angle between these two slits positions is given by $\alpha = \pi - \beta$. The radial distance of the detector from the rotation axis is given by $r = \frac{R}{\sin(\alpha)}$. We also extracted the value of α from the images shown in Figure 7. Results are shown in Table 1. For the different tests (3 probe positions and 2 collimator rotation speeds), the α values obtained by signal and image processing are consistent.

Discussion

Collimator Optimization

The testing results confirm significant dose reduction (from one to two orders of magnitudes). However, with the proposed slit

Table 1 – Results Summary

	Time /rotation (ms)	Rotation speed (rpm)	Time between consecutive probe exposures (ms)		Signal processing			Image analysis	
			Mean	Standard deviation	Rotation Angle	Alpha	Estimated radial position (mm)	Alpha	Estimated radial position (mm)
Position P1	57.9	1035	261.1	0.8	162.2°	17.8°	12.9	18.7°	12.3
	38.4	1561	171.4	4.1	160.5°	19.5°	11.8		
Position P2	57.9	1036	237.4	2.3	147.5°	32.5°	7.2	32.0°	7.3
	38.2	1571	156.8	4.8	147.8°	32.2°	7.2		
Position P3	58.1	1034	203.0	2.5	125.9°	54.1°	4.4	58.6°	4.1
	38.1	1575	132.8	7.9	125.5°	54.5°	4.4		

geometry, the dose reduction factor strongly depends on the distance of the POI from the collimator rotation axis. Thus, the dose profile versus the distance from the rotation axis (see Figures 5 and 6), show two asymptotic trends, a steep dose gradient for distances close to the slit off-axis distance and low slope for larger distances. One possible improvement may be to propose two slits as shown in Figure 8, to prevent significant dose exposure at POI close to the off_axis position of the slits. For instance, for an off-axis distance of $R = 2mm$, it could be suitable to use slits without any aperture closer than 3mm from the rotation axis. This geometry will increase the blind area (6mm in diameter instead of 4mm with the single slit geometry) but will guarantee a more homogeneous dose over the imaging field.

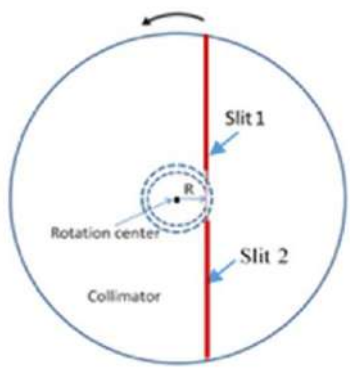


Figure 8 – Optimized Collimator Geometry with Two Off-Axis Slits

Miniaturized X-ray Detector Sensitivity

The GaN-based detector probe was operated in photon counting mode over a 100µs window. With the OSL dosimeter, we measured at the surface of the phantom a dose of 681µSv for an X-ray irradiation lasting 6440ms (10 images), i.e. a dose of ~10nSv over the counting window of 100µs. Moreover, the GaN transducer was placed at 60mm depth in the phantom where the X-ray fluence was only 19% of the fluence at the phantom surface (for the considered 70kV X-ray beam). Thus, a dose of less than 2nSv at the level of the GaN-transducer was sufficient to give more than 6 counts at the output of the detector as shown in Figure 7. This shows that the GaN-based miniaturized X-ray detector has a suitable sensitivity and a bandwidth well adapted to the VF application. Some optimizations remain to be done for contrast enhancement since the signal baseline was much higher than the dark count rate of the detector: this could be due to an insufficient shielding against radiations or parasitic scintillation light. It is worth mentioning that the corresponding dose behind the phantom, i.e. at 120mm depth was ~0.4nSv for 100µs and was not suffficient to activate the flat panel detector at this time resolution.

Conclusions

The VF method for IR procedures has been implemented and tested. The experimental results have shown that the dose can be reduced by one or two order of magnitudes by the use of a rotating slit collimator. Further studies need to be carried out on fluoroscopy clinical systems to assess dose reduction and localization resolution achieved by VF guidance as compared to conventional fluoroscopy guidance. Moreover, this study also confirms that the GaN-based miniaturized detector probe allows sufficient IN/OUT field contrast and bandwidth for this application (activated within 100µs by a dose smaller than 2nSv). These results establishes the proof of concept for the proposed VF approach. This novel method will be used for interventions such as embolization interventions, coronary angioplasty procedures, cryotherapy or radiofrequency ablations, biventricular cardioverter-defibrillator implantations. These instances represent several millions of interventions per year world-wide, so that we expect a major impact on Public Health.

Acknowledgements

This work was partially funded by ANR Labcom CQFD (ANR-13-LAB3-0002), IDEX Univ. Grenoble Alpes, Carnot Institute LSI, and SurgiQual Institute. Authors thank Gilles Simon (ILM) and Jérôme Degouttes (INL) for their helpful contributions for the design and realization of the mechanical components of the system prototype. Authors also thank Arnaud Gaudu for his significant contribution for absolute dose measurements.

References

[1] C. Etard, E. Bigand, C. Salvat, V. Vidal, J.P. Beregi, A. Hornbeck, and J. Greffier, Patient dose in interventional radiology: a multicentre study of the most frequent procedures in France, *Eur. Radiol.* **27** (2017) 4281–4290.

[2] W. Jaschke, M. Schmuth, A. Trianni, and G. Bartal, Radiation-Induced Skin Injuries to Patients: What the

Interventional Radiologist Needs to Know., *Cardiovasc. Intervent. Radiol.* **40** (2017) 1131–1140.

[3] K.J. Strauss, and S.C. Kaste, The ALARA (as low as reasonably achievable) concept in pediatric interventional and fluoroscopic imaging: striving to keep radiation doses as low as possible during fluoroscopy of pediatric patients--a white paper executive summary., *Pediatr. Radiol.* **36 Suppl 2**

[4] Y. Grondin, P. Augerat, P. Cinquin, L. Desbat, and M. Decrouez, Rotary collimator for determining the position of an element provided with sensors in an x-ray imaging system, EP3368919A1, 2016.

[5] T. Yanagida, Y. Fujimoto, and M. Koshimizu, Evaluation of Scintillation Properties of GaN, *E-Journal Surf. Sci. Nanotechnol.* **12** (2014) 396–399.

Address for correspondence

Philippe Cinquin
Philippe.Cinquin@univ-grenoble-alpes.fr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190187

Romedi: An Open Data Source About French Drugs on the Semantic Web

Sébastien Cossin[ab], Luc Lebrun[b], Grégory Lobre[a], Romain Loustau[a], Vianney Jouhet[ab], Romain Griffier[a], Fleur Mougin[b], Gayo Diallo[b], Frantz Thiessard[ab]

[a]*Bordeaux university hospital, Pôle de santé publique, Service d'information médicale, Unité Informatique et Archivistique Médicales, F-33000 Bordeaux, France*
[b]*Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, team ERIAS, UMR 1219, F-33000 Bordeaux, France*

Abstract

The W3C project, "Linking Open Drug Data" (LODD), linked several publicly available sources of drug data together. So far, French data, like marketed drugs and their summary of product characteristics, were not integrated and remained difficult to query. In this paper, we present Romedi (Référentiel Ouvert du Médicament), an open dataset that links French data on drugs to international resources. The principles and standard recommendations created by the W3C for sharing information were adopted. Romedi was connected to the Unified Medical Language System and DrugBank, two central resources of the LODD project. A SPARQL endpoint is available to query Romedi and services are provided to annotate textual content with Romedi terms. This paper describes its content, its services, its links to external resources, and expected future developments.

Keywords:

Pharmaceutical Preparations
Semantics
Vocabulary, Controlled

Introduction

Drug information is spread over multiples sites on the Internet. Summary of Product Characteristics (SPCs) are documents produced by pharmaceutical companies and approved by public health agencies. In France, ANSM (Agence Nationale de Sécurité du Médicament et des Produits de Santé), the French Medicines Agency, publishes SPCs approved by itself or the European Medicines Agency (EMA) on a website. An SPC contains key information about a marketed drug like the therapeutic indication(s), the posology, dosage adjustment, drug-drug interactions, and contraindications [1,2].

However, other additional knowledge related to marketed drugs can be found on the Internet and is not clearly connected. For example, the French thesaurus of drug-drug interaction (DDI), edited by ANSM, is the official reference document on this topic. This document, available as a PDF file, describes potential DDI (PDDI) between molecules. The links between molecules in the SPC and molecules in the reference document are not explicit and cannot be linked automatically due to semantic and syntactic interoperability issues. Another example is information about drugs' safety during pregnancy. The CRAT [3] (Centre de référence sur les agents tératogènes) is a French public organization especially involved in this public health issue. It provides free access to information about risks of drug intake during pregnancy that often disagrees with the SPC documents [2]. Still, connections and comparisons between these two sources can only be made by humans.

Furthermore, specific or general international sources, like DrugBank [4] and DBpedia [5], deliver supplementary information about drugs marketed abroad or about characteristics of molecules. DrugBank is a comprehensive, freely accessible, online database containing a large amount of information on molecules (e.g., chemical structure, half-life). DBpedia provides structured, machine-understandable knowledge extracted from Wikipedia articles. Many drugs and molecules are described by DBpedia contributors. The Unified Medical Language System [6] (UMLS) is a compendium of a large number of national and international vocabularies. In particular, UMLS contains RxNorm [7], a standard nomenclature developed by the United States National Library of Medicine (NLM) in the field of medications and the MeSH [8], a comprehensive controlled vocabulary for the purpose of indexing scientific articles.

The scattering of information is a significant issue for information retrieval which hampers the reusability of up-to-date knowledge on the web. Tim Berners-Lee, the inventor of the World Wide Web, suggested a 5-star deployment scheme for sharing information on the web (figure 1).

Figure 1– 5-star deployment scheme for Open Data [1].

The idea is to use W3C standards when publishing data in order to create a semantic graph that is capable of interlinking information of various datasets distributed over the web.

The W3C project, "Linking Open Drug Data" (LODD), focuses specifically on linking various sources of drug data together [9]. Participants of the LODD project have already made dozens of datasets relevant to pharmaceutical research and development available as linked data. In this paper we present Romedi, a new open dataset on French drugs linked to other related resources on the semantic web. In the methods section we discuss the Romedi data model and how it was linked to external resources. In the results section, examples of use cases are described.

[1] Source : https://5stardata.info/en/

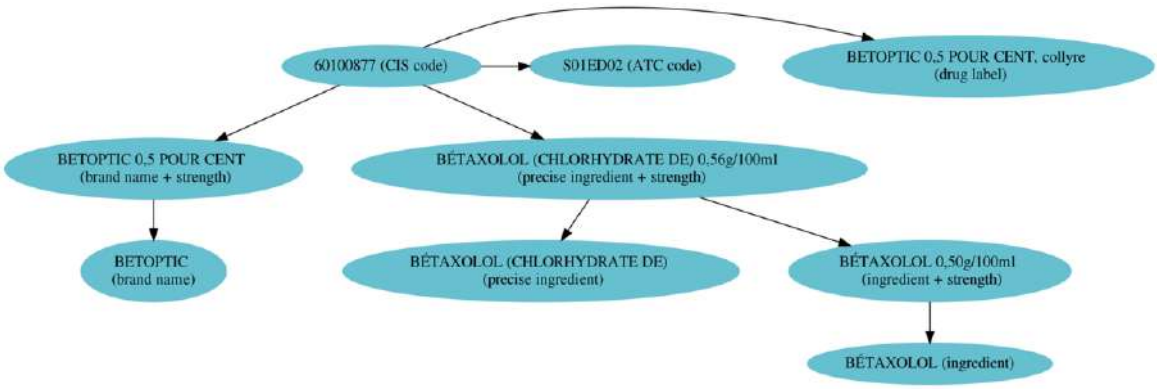


Figure 2– The Romedi data model instantiated with the drug "BETOPTIC 0.5 POUR CENT, collyre". The normalization step extracted the brand name (BETOPTIC) and the strength from the drug label. The ingredient was also normalized and linked to its corresponding UMLS concept C0005320 and DrugBank concept DB00195

Methods

Data Model

The Romedi data model is close to the RxNorm terminology [7]. It contains similar classes like the brand name (BN), ingredient (IN), and precise ingredient (PIN). The CIS code is an identifier of a marketed drug in France, and the URL (Uniform Resource Locator) to access the SPC depends on this code. For example, "BETOPTIC 0.5 POUR CENT, collyre" (Figure 2), is the label of a drug identified by the CIS code 60100877. The SPC of this drug is accessible at http://base-donnees-publique.medicaments.gouv.fr/affichageDoc.php?specid=60100877&typedoc=R. Each marketed drug has one or several ATC codes. The Anatomical Therapeutic Chemical (ATC) is a widely used system of alphanumeric codes developed by the World Health Organization (WHO)[2] for the classification of drugs.

In France, the main data source for marketed drugs is the "base de données publique des médicaments", a freely accessible database available in a text file format[3] which is updated every month by national health authorities. It contains all currently marketed drugs and drugs withdrawn from market in the last three years. The database contains details on marketed drugs like CIS codes, drug labels, and molecules. A normalization and transformation process is needed to instantiate the Romedi model as the concept of brand name is not present in this database. For example, "INEXIUM 20mg, comprimé gastro-résistant" is a drug label but the brand name "INEXIUM" is not present and must be extracted. In addition, only the precise ingredient is present for some drugs. For example, "pravastatine sodique" appears but the term "pravastatine", the ingredient, is missing. This normalization step was done by using regular expressions algorithms and an interface for manual validation by a pharmacist. This step was fully described elsewhere [10].

After model instantiation, ingredient instances were linked to UMLS [6] (Unified Medical Language System) and DrugBank [11]. The mapping between the French and international reseources was done as follows: the mapping is automatic if two terms have a perfect match, semi-automatic with a validation interface when a partial match is found, and manual when a French term cannot be found.

Ingredients were also linked to the French thesaurus of DDI. The automatic extraction and transformation of the PDF document to a CSV file was done with an R package[4]. The first use case to exploit external resources links was to compare French reference document on DDI with international ones. Ayvaz et al. [12] managed to gather information about DDI from publicly available sources, and the authors used DrugBank identifiers to describe couples of DDI interaction from different sources. The authors made a merged-PDDI database that can be downloaded online. The external links to DrugBank were used to integrate French knowledge to the merge PDDI database and to automatically compare the French national reference with other sources.

Medication Extraction Module

Automated identification of drugs in unstructured data, like social media or clinical notes, is essential for post-marketing drug safety surveillance that aims to detect signals of drug misuse or adverse event effects [13,14]. The natural language processing goal is not only to identify the terms in a corpus that correspond to drug entities, but also to map these entities to a well-established knowledge base. 'Semantic annotation' is the name given to this task by Jovanovic et al. [15].

IAMsystem, a general semantic annotation tool, was developed to facilitate the identification of Romedi terms in textual content. It was initially developed for the DoMINO project (Drug Misuses In Networks) that aims to detect drug misuse in fora [16]. The program performances were evaluated on a shared task for disease detection using death certificates [17], and the program was described in-detail at this occasion [18]. IAMsystem is open-source and available on GitHub[5]. It takes a set of terms as inputs, normalizes them, and stores them in a tree data structure for fast dictionary look-up. It handles abbreviations, a set of which were manually added for drug detection (e.g., "ac" for "acide", "vit" for "vitamine"). The typo module for drug detection is a logistic regression trained on a manually created gold standard of 3,438 potential spelling errors of brand names and molecules. Three explanatory variables are used:

[2] https://www.whocc.no/atc_ddd_index

[3] http://base-donnees-publique.medicaments.gouv.fr

[4] https://github.com/scossin/IMthesaurusANSM

[5] https://github.com/scossin/IAMsystem

1. The length of the potentially misspelled word
2. The similitude of the first letter between the potential misspelled word and the dictionary word
3. Levenshtein's distance between the phonetic transformation of two words

This last variable is computed with the French "phonetic" algorithm of the Talisman program[6]. The model is able to predict a typo of a brand name or an ingredient with a specificity of 0.93 and a sensitivity of 0.60. The performances of the annotator were evaluated in clinical texts, and the results are presented in the following section.

Romedi terms can be used to detect French drugs in textual content. Links to the CRAT website were established by detecting brand names in the alphabetic index of the web page. The graph model permits easy retrieval of information about the detected drug , such as the ATC code(s) or external resources links.

Results

Romedi Content

The first version of the Romedi terminology contains 13,661 French marketed drugs, 4,277 brand names, and 2,109 ingredients after the normalization step. Among the ingredients, 1,918 (91%) were linked to a UMLS concept and 1,434 (71%) to a DrugBank concept. 954 (95%) of the molecules in the French DDI thesaurus were mapped to a Romedi ingredient or precise ingredient. Since DrugBank contains French synonyms of ingredients and the French version of MeSH is integrated in the UMLS, most of the mappings were done automatically.

A web interface that makes the Romedi content available is accessible at https://www.romedi.fr. Like RxNav [19], the interface displays links between clinical drugs, active ingredients, brand names, ATC code(s), and external resources. Drug information from external resources is retrieved using SPARQL queries or application programming interfaces (API). For example, the DBpedia definition of a molecule is retrieved by a SPARQL query to its endpoint[7] and by using DrugBank links to DBpedia. Romedi resources are identified by their Uniform Resource Identifiers (URIs). For example, https://www.romedi.fr/romedi/CIS60100877 is the URI of the Romedi drug "BETOPTIC 0.5 POUR CENT, collyre", and is also a valid URL. Retrieving a representation of a resource identified by a URI is known as dereferencing a URI, and it can be used to obtain a representation that can be perceived by a user.

A SPARQL endpoint[8] is also available to query Romedi content. In addition, the terminology is freely downloadable as an RDF file and can be reused under an open license.

Automatic Comparison of the French DDI Reference Document with International

Mapping national ingredient concepts to international ones allows users to automatically compare drug-drug interaction information. Applied on more than 7 million drugs deliveries in France, the main discrepancy between the French thesaurus and international sources was with the couple "escitalopram – flecainide." This drug pair was considered contraindicated by an international source, and no risk of interaction is described in the French thesaurus although it contains four levels of severity. Full results of this work are available elsewhere [20]. The French thesaurus was integrated in the merged-PDDI dataset[9] in a linked open format.

Medication Extraction Module

Brand names and molecule identification performances were evaluated using the 'current medication' section in electronic health records (EHR) form from Bordeaux Hospital's emergency department. Among the 6,070 drugs detected (brand names or molecules), the specificity/sensitivity were 0.99/0.92 and 0.99/0.96 without and with the typo module respectively [10]. The performance of these models can be explained by the grocery list type of the input data and disregarding medication attributes (e.g., dosage, strength, and route). Further evaluation is required in narrative clinical notes. The programs can be installed locally[10] and the annotator is also available in an R package[11].

Discussion

Open Data has the potential to provide significant benefits to society. The Link Data movement aims to share and integrate knowledge on the Web. Still, many resources containing important information on drugs remain inaccessible to machine and hampers automatic comparison of knowledge.

Romedi is a French open dataset that connects French and international information resources about drugs.

First, the French dataset on marketed drugs (base de données publique des médicaments) was normalized and integrated in an RxNorm-like data model. Then, national and international sources were linked, including the French thesaurus of PDDI, the French reference for evaluating drugs safety during pregnancy (CRAT), DrugBank, and UMLS. An interface was developed to navigate Romedi terminology, and a SPARQL endpoint is accessible to query its content.

Developers can use Romedi API to retrieve information about French drugs for their website. Researchers can use Romedi services to extract drugs in textual content for post-marketing drug safety surveillance. Using the same terminology, data can be shared with other researchers to promote collaboration and enhance pharmacovigilance discovery.

Romedi URIs are currently used to index drugs in clinical notes in Bordeaux Hospital's i2b2 [21] data warehouse and to ease information retrieval. Searches can be performed by brand name, molecule, or ATC code. Medication extraction and indexation helps patient phenotyping and cohort identification tasks.

Future Developments

The current version of Romedi is 2.2 and the updating process is not fully automated. The biggest challenge will be maintainance of this terminology, especially correcting errors identified by users and managing updates of French-marketed drugs and external resources. Our aim is to build an engaged community of users and developers around Romedi.

So far, the Romedi data model contains no description logics. It would be feasible to infer logical equivalences between Romedi and RxNorm drugs by using a formal model. A mapping between Romedi and RxNorm beyond the ingredient level would facilitate French drug data integration in the Observational Health Data Sciences and Informatics (OHDSI) common data model that uses RxNorm as an international standard [22]. This mapping will be important to involve French researchers in international collaborations and research

[6] https://github.com/Yomguithereal/talisman/tree/master/src/phonetics
[7] https://dbpedia.org/sparql
[8] http://www.romedi.fr:8890/sparql
[9] https://www.dikb.org/Merged-PDDI/
[10] https://github.com/scossin/RomediApp
[11] https://github.com/scossin/RIAMsystem/

projects on drugs, and it may help our national agency to exchange information about post-marketing surveillance.

Conclusions

In this paper we have presented Romedi, an open dataset about French drugs linked to multiple resources on the semantic web. Web services are provided to annotate textual documents and query Romedi content to retrieve additional information about detected drugs.

Acknowledgements

This project is part of the Drugs Systematized Assessment in real-liFe Environment (DRUGS-SAFE) research platform, funded by the French Medicines Agency (Agence Nationale de Sécurité du Médicament et des Produits de Santé, ANSM). This platform aims to provide an integrated system allowing the concomitant monitoring of drug use and safety in France.

References

[1] M. Rougemont, S. Ulrich, C. Hiemke, E. Corruble, and P. Baumann, French summaries of product characteristics: Content in relation to therapeutic monitoring of psychotropic drugs, *Fundamental & Clinical Pharmacology* **24** (2010), 377–384.

[2] B. Arguello, T.M. Salgado, and F. Fernandez-Llimos, Assessing the information in the Summaries of Product Characteristics for the use of medicines in pregnancy and lactation, *Br J Clin Pharmacol* **79** (2015), 537–544.

[3] E. Elefant, C. Vauzelle, and D. Beghin, [Centre de référence sur les agents tératogènes (CRAT): A pioneer center], *Therapie* **69** (2014), 39–45.

[4] V. Law, C. Knox, Y. Djoumbou, T. Jewison, A.C. Guo, Y. Liu, A. Maciejewski, D. Arndt, M. Wilson, V. Neveu, A. Tang, G. Gabriel, C. Ly, S. Adamjee, Z.T. Dame, B. Han, Y. Zhou, and D.S. Wishart, DrugBank 4.0: Shedding new light on drug metabolism, *Nucleic Acids Res* **42** (2014), 1091–1097.

[5] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, DBpedia: A Nucleus for a Web of Open Data, in: Proceedings of the 6th International The Semantic Web and 2Nd Asian Conference on Asian Semantic Web Conference, Springer-Verlag, Berlin, Heidelberg, 2007: pp. 722–735.

[6] O. Bodenreider, The Unified Medical Language System (UMLS): Integrating biomedical terminology, *Nucleic Acids Res* **32** (2004), 267-270.

[7] S.J. Nelson, K. Zeng, J. Kilbourne, T. Powell, and R. Moore, Normalized names for clinical drugs: RxNorm at 6 years, *J Am Med Inform Assoc* **18** (2011), 441–448.

[8] S. Kim, L. Yeganova, and W.J. Wilbur, Meshable: Searching PubMed abstracts by utilizing MeSH and MeSH-derived topical terms, *Bioinformatics* **32** (2016), 3044–3046.

[9] M. Samwald, A. Jentzsch, C. Bouton, C.S. Kallesøe, E. Willighagen, J. Hajagos, M.S. Marshall, E. Prud'hommeaux, O. Hassenzadeh, E. Pichler, and S. Stephens, Linked open drug data for pharmaceutical research and development, *J Cheminform* **3** (2011), 19.

[10] S. Cossin, R. Loustau, V. Jouhet, F. Mougin, G. Evrard, C. Giljardine, G. Diallo, and F. Thiessard, ROMEDI, une terminologie médicale française pour la détection des médicaments en texte libre, Artificial Intelligence Platform, Nancy (2018).

[11] D.S. Wishart, C. Knox, A.C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, and J. Woolsey, DrugBank: A comprehensive resource for in silico drug discovery and exploration, *Nucleic Acids Res* **34** (2006), 668-672.

[12] S. Ayvaz, J. Horn, O. Hassanzadeh, Q. Zhu, J. Stan, N.P. Tatonetti, S. Vilar, M. Brochhausen, M. Samwald, M. Rastegar-Mojarad, M. Dumontier, and R.D. Boyce, Toward a complete dataset of drug-drug interaction information from publicly available sources, *J Biomed Inform* **55** (2015), 206–217.

[13] C.L. Ventola, Big Data and Pharmacovigilance: Data Mining for Adverse Drug Events and Interactions, *P T*. **43** (2018), 340–351.

[14] S. Sohn, C. Clark, S.R. Halgrim, S.P. Murphy, C.G. Chute, and H. Liu, MedXN: An open source medication extraction and normalization tool for clinical text, *J Am Med Inform Assoc* **21** (2014), 858–865.

[15] J. Jovanović, and E. Bagheri, Semantic annotation in biomedicine: the current landscape, *J Biomed Semantics* **8** (2017).

[16] E. Bigeard, Construction de lexiques pour l'extraction des mentions de maladies dans les forums de santé, *19es REncontres jeunes Chercheurs en Informatique pour le TAL (RECITAL)* (2017), 15-27

[17] A. Névéol, A. Robert, F. Grippo, C. Morgand, C. Orsi, L. Pelikan, L. Ramadier, G. Rey, and P. Zweigenbaum, CLEF eHealth 2018 Multilingual Information Extraction Task Overview: ICD10 Coding of Death Certificates in French, Hungarian and Italian, in: CLEF, 2018.

[18] S. Cossin, V. Jouhet, F. Mougin, G. Diallo, and F. Thiessard, IAM at CLEF eHealth 2018: Concept Annotation and Coding in French Death Certificates, *ArXiv:1807.03674 [Cs]* (2018).

[19] K. Zeng, O. Bodenreider, J. Kilbourne, and S.J. Nelson, RxNav: A Web Service for Standard Drug Information, *AMIA Annu Symp Proc* (2006), 1156.

[20] S. Cossin, Intéractions médicamenteuses, données liées et applications, Thèse d'exercice, Université de Bordeaux, 2016.

[21] S.N. Murphy, M.E. Mendis, D.A. Berkowitz, I. Kohane, and H.C. Chueh, Integration of clinical and genetic data in the i2b2 architecture, *AMIA Annu Symp Proc* (2006), 1040.

[22] G. Hripcsak, J.D. Duke, N.H. Shah, C.G. Reich, V. Huser, M.J. Schuemie, M.A. Suchard, R.W. Park, I.C.K. Wong, P.R. Rijnbeek, J. van der Lei, N. Pratt, G.N. Norén, Y.-C. Li, P.E. Stang, D. Madigan, and P.B. Ryan, Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers, *Stud Health Technol Inform* **216** (2015), 574–578.

Address for correspondence

sebastien.cossin@u-bordeaux.fr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190188

Clinical Text Mining on FHIR

Philipp Daumke[a], Kai U. Heitmann[b,c], Simone Heckmann[b,c], Catalina Martínez-Costa[d], Stefan Schulz[a,d]

[a] Averbis GmbH, Freiburg, Germany, [b] Geyfra GmbH, Hürth, Germany, [c] HL7 Germany, Berlin, Germany
[d] Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria

Abstract

Semantic standards and human language technologies are key enablers for semantic interoperability across heterogeneous document and data collections in clinical information systems. Data provenance is awarded increasing attention, and it is especially critical where clinical data are automatically extracted from original documents, e.g. by text mining. This paper demonstrates how the output of a commercial clinical text-mining tool can be harmonised with FHIR, the leading clinical information model standard. Character ranges that indicate the origin of an annotation and machine generates confidence values were identified as crucial elements of data provenance in order to enrich text-mining results. We have specified and requested necessary extensions to the FHIR standard and demonstrated how, as a result, important meta-data describing processes generating FHIR instances from clinical narratives can be embedded.

Keywords:

Electronic Health Records, Natural Language Processing, Semantics

Introduction

Semantic Interoperability was defined in 2000 as "…integrating resources that were developed using different vocabularies and different perspectives on the data". Semantic interoperability requires that "systems must be able to exchange data in such a way that the precise meaning of the data is readily accessible and the data itself can be translated by any system into a form that it understands" [1].

Nearly twenty years later, the lack of semantic interoperability continues being an obstacle to a more rational and effective data and information management in healthcare and biomedical research. The authors of [1] had already distinguished between vocabularies and perspectives, highlighting the division between ontology ("what there is") [2] and epistemology ("what we can know") [3]. On the level of current health informatics standards, this has driven the evolution of two genres of semantic resources:

- Terminology systems (vocabularies, thesauri, formal ontologies, classifications), which attach meaning to domain terms and elaborate on necessary and sufficient properties of (classes of) domain entities;
- Information models, which are artefacts that provide standardized structure (section, entry, grouping, etc.) and context (diagnosis, past medical history, medication order) for clinical recording scenarios.

In this paper, we will focus on the latter, particularly on HL7 FHIR [4], a standard for healthcare information exchange and sharing, characterized by its straightforward approach to implement interfaces between Electronic Health Record (EHR) data and data consuming applications.

FHIR is based on interoperable building blocks named *Resources*, small data model components defining sets of properties that describe and provide structure for domain data acquisition. Currently there are approximately 150 resources, uniquely identified with Uniform Resource Identifiers (URIs).

Examples are *MedicationRequest* (prescription), *AdverseEvent*, *Procedure* and *Condition* (problem). They constitute a graph of clinical data by explicit inter-resource references [5]. For instance, a *MedicationRequest* resource explicitly references its prescriber (a FHIR *Practitioner* resource), its patient (*Patient* resource), and the drug prescribed (*Medication* resource). A built-in extensibility mechanism can enrich existing resource definitions.

Semantic interoperability within FHIR is provided by explicit, detail-oriented, prescriptive guidance, with its interoperable meaning anchored in external terminology standards, e.g. SNOMED CT, LOINC, or ICD-10. FHIR provides a granular way to exchange data using a RESTful style approach [6]. Its focus is on providing models for frequently occurring documentation and information exchange tasks. FHIR resources can be serialized in JSON and XML.

Looking at EHR systems in use, there is still a persisting gap between structured and coded information on the one hand and a much larger amount of semi- or unstructured narratives on the other hand. It is unrealistic that this textual documentation will be largely substituted by structured documentation [7]. This is where natural language processing (NLP), in particular text mining, comes in.

This paper addresses the issue how the output of clinical text mining systems can be harmonized in FHIR. We identify gaps in the current FHIR specification related to content essential for text mining. We introduce and discuss specific FHIR extensions that allow text-mining results to be represented in FHIR without loss.

Materials and Methods

The Text Mining System *Health Discovery*

The clinical text mining technology under scrutiny is *Health Discovery* [8] by Averbis GmbH. *Health Discovery* contains over fifty different text-mining annotators for the recognition of diagnostic statements, medical procedures, lab values, drugs, anatomy, morphology, scores and others. It is available for several languages, including English and German. *Health Discovery* bundles annotators in predefined text mining pipelines to facilitate the analysis of text genres such as discharge summaries or pathology reports. It has been successfully used for various use cases, e.g., data driven patient recruitment for clinical trials [9], automated coding and billing [10], filling of tumour registries, rare disease identification [11], antibiotic resistance monitoring [12], radiology report analysis [13] and health data de-identification [14].

UIMA

Health Discovery is based on the *Unstructured Information Management Architecture* (UIMA) [15], a flexible and extensible text-mining framework for the analysis and processing of unstructured text. UIMA allows defining text-mining pipelines consisting of a set of annotators called analysis engines. The analysis engines communicate in a pipeline by adding or modifying meta-information stored in the Common Analysis System (CAS), which, in addition to the metadata, also contains the currently processed document.

In UIMA, all elements to be extracted from texts must be predefined. This is done in a so-called *Type System*, where types and their attributes (features) are defined. The most common type is *Annotation*. It comes with the attributes *begin* and *end*, which specify the text position to which an annotation relates. These attributes are inherited by all subtypes of *Annotation*.

UIMA is only a framework and does not ship a rich selection of components. There are, however, component repositories like DKPro [16], which provide analysis engines of well-known text mining components. Prominent applications in Healthcare built upon UIMA are the DeepQA system Watson [15], the clinical Text Analysis and Knowledge Extraction System (cTAKES) [17], the medication entity and attribute extraction and normalisation tool MedXN [18], and MedTime [19], a software to extract and normalize TIMEX3-based temporal expressions from clinical text.

Making a Text Mining System FHIR Compatible

Currently, no standardized UIMA-based type systems for healthcare have found worldwide acceptance. Each text-mining provider usually defines its own type system. Interoperability is created by exchanging type systems between text mining systems – in UIMA, several type systems can be used simultaneously as long as their namespaces are unique – and mappings between type systems.

To make a text mining system FHIR-compatible, two basic procedures can be chosen:

- Defining a FHIR-compatible type system [20], or
- Exposing a FHIR compatible interface.

We pursue the second approach for the following reason: Healthcare data models continue to be the subject of intensive development. In the last 2-3 decades, different versions of the FHIR standard like Version 2, Version 3, Clinical Document Architecture (CDA) and FHIR have been published. There are also other healthcare standards like OpenEHR [21]. In order to support different standards and their versions in parallel, it is not feasible to create and maintain different versions of the text mining system with a different data model for each standard. Instead, it is sufficient and more effective to provide external interfaces to these standards and versions while keeping the data model of the underlying text mining system stable.

In this paper we describe how a FHIR-compatible interface looks like and how FHIR needs to be extended to meet the requirements of Text Mining. In particular, we will answer the following questions:

1. Which are the right FHIR resources and containers to return a text mining result?
2. How can the document text and text-mining annotations such as conditions and observations be linked together?
3. How can typical text mining features, particularly (i) text span of an annotation and (ii) its confidence score be specified in FHIR?
4. How can the types of our text mining system be mapped to FHIR resources and which challenges occur?

Results

FHIR Resources and Container for Text Mining Results

Typically, a text-mining result consists of the document text and associated text-mining annotations. The document text is usually returned back by the text mining system, mainly for the following reason: Text mining annotations contain a *begin* and an *end* attribute, which specify their exact position in the text at character level. Instead of pure text files, the sender often wants to analyse text-based documents such as PDF or MS Word. The text is then extracted during the text mining analysis. The *begin* and *end* attributes always refer to the text returned by the text mining system, not to the original document. The appropriate resource for the document text is a *Binary* resource. A *Binary* can contain any content, whether text, image, pdf, zip archive, etc.

Text Mining annotations can be represented in a corresponding clinical *FHIR Resource*, such as *Condition*, *MedicationStatement*, *Observation* etc. These resources are used to record detailed information about conditions, diagnoses, consumed drugs or any other observation that has direct or indirect impact on the patient's health.

In order to return a *Binary* together with its clinical *Resources* in a single artefact, it has to be grouped into a *Bundle*. A *Bundle* is a collection of resources useful for a variety of different reasons, e.g. returning search results or sending messages. Accordingly, different types of *Bundle*s exist, e.g., *document*, *searchset* or *collection*. In our case, we need a *Bundle* of type *collection*. A *collection Bundle* is a set of resources collected into a single package for ease of distribution.

It is important to understand that a text mining result is not a *document Bundle* in the FHIR sense. FHIR *document Bundles* represent solely structured documents that are authored and assembled in FHIR.

Linking Document Text and Text Mining Annotations: FHIR Provenance

In FHIR, resources in a *Bundle* have an independent existence – they can, e.g., also be accessed using the RESTful API. If the *Binary* was not linked to the clinical resources, it would not be possible to identify the origin of a text-mining determined resource in subsequent processing steps. Therefore, we need another FHIR resource to link both resources, the *Provenance*. *Provenance* contains, among others, three attributes: *entity* gives details about the source of the data (like a PDF, text…),

agent gives details about the algorithm that extracted the data (software name, vendor, version...) and *target* points at the created resource (e.g. *Condition*, *Observation*, *MedicationStatement*...). For each clinical resource, we also specify a *Provenance* resource. Figure 1 illustrates the resulting *collection Bundle* including clinical and associated *Provenance* resources.

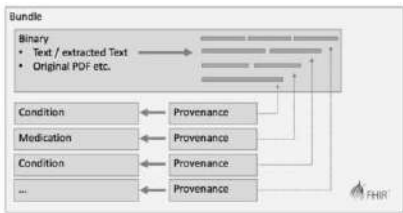


Figure 1– Simplified View of the Provenance Information Conveyed in a FHIR Bundle of Type Collection. The Result Is Conveyed in a FHIR Bundle of Type Collection. A Binary Resource Holds the Source Text (Character Stream) And/or the Extracted Text with the Original PDF Where the Text Was Extracted. Subsequently All Clinical Resources That Come out of the NLP Analysis Are listed. Each of these Resources Is Accompanied by a Provenance Resource.

Specification of Text Spans and Confidence in FHIR

As mentioned above, text mining annotations usually contain begin and end attributes that specify the exact position of an annotation in the text. Since there is no FHIR representation for these attributes yet, we specified two FHIR extensions, currently under approval by the international HL7/FHIR committee.

Extension *Character Range*

The rationale for this extension is to specify, within a source text, the place in which an annotation was found. Thus, the origin of a text mining annotation can be specified not only at document level, but also at word and character level. Traceability is an important feature of text mining systems to help users gain confidence in text mining applications and clinical research based upon text mining results.

The extension is named *character-range*. It contains an attribute named *valueRange* of type *Range*. *valueRange* itself contains the attributes *low* and *high* of type *SimpleQuantity* with the purpose to specify begin and end of annotations in text. *SimpleQuantities* always need a unit. Since characters are countable, their unit in FHIR is "1":

```
"extension": [{
  "url": "http://example.com/StructureDefinition/character-range/0.9",
  "valueRange": [{
    "low": {"value": 201,
  "unit": 1},
    "high": {"value": 219,
   "unit": 1}}]}]
```

Extension Confidence

Text mining results may be biased, incomplete or erroneous, their enrichment with a confidence score is therefore crucial. Especially text-mining systems based on machine learning, often provide a confidence value, which estimates the probability for the correctness of the annotation. Thus, the requested extension (ii) is *confidence*, specialising the type *Quantity*, stating the accuracy of annotations. The value range of a confidence is often, but not necessarily, between 0 and 1:

```
"extension": [{
  "url": "http://example.com/StructureDefinition/confidence/0.9",
  "valueQuantity": {"value": 0.1}}]
```

Mapping Text Mining Types to FHIR Resources

Now that the container, the resources and the necessary extensions have been defined, the mapping of text mining annotations to FHIR resources can be performed in a straightforward way. Both *Health Discovery* and *FHIR* know the notation of a *Concept* or *CodeableConcept*. Concepts exist to encode a medical statement with a term from a terminology. They typically consist of a code, a preferred term, a code system and the version of the code system. In *Health Discovery*, the synonym that is responsible for a match in a text (=matchedTerm) is also an attribute of *Concept*. Table 1 shows the mapping between *HealthDiscovery.Concepts* and *FHIR.CodeableConcepts*. Table 2 gives an overview of the terminology bindings used by *Health Discovery* for different *Concept* types in US English and German. In this work, we focus on the mapping of diagnoses, medications and laboratory values. Table 3 gives an overview of the mappings between *Health Discovery* and *FHIR*.

Table 1 – Mappings between HealthDiscovery.Concept and FHIR.CodeableConcept

Health Discovery	FHIR Resource
Concept	**CodeableConcept**
.conceptid	.code.coding.code
.dictCanon	.code.coding.display
.matchedTerm	.code.text
.source.split('_')[0]	.code.coding.system
.source.split('_')[1]	.code.coding.version

Table 2 – Terminology Binding in Health Discovery for US-English and German

Type	US English	German
Diagnosis	ICD-10 Clinical Modification [22]	ICD-10 German Modification [23]
Drugs	RxNorm [24]	ABDAMed [25]
Ingredients	WHO ATC [26]	German ATC [27]
Lab Values	LOINC [28]	LOINC [28]

Table 3 – HealthDiscovery Types Mapped to FHIR. Instances of (Codeable) Concept Are Marked with an Asterisk (*). The Mapping Between Medication.doseFrequency and MedicationStatement.dosage.timing (**) Is Not Further Specified, as It Is Rather Complex and Beyond this Paper

Averbis Health Discovery	FHIR Resource
Diagnosis	**Condition**
.concept*	.code*
.clinicalStatus	.clinicalStatus
.verificationStatus	.verificationStatus
.belongsTo	.subject.display
.side	.bodySite.text
Medication	**Medication**
.drug.ingredientConcept*	.ingredient*
.doseForm*	.form*
	MedicationStatement
.status	.status
.drug.strength.value	.dosage.doseQuantity.value
.drug.strength.unit	.dosage.doseQuantity.unit
.doseFrequency**	.dosage.timing**
.administrations*	.dosage.method*
LaboratoryValue	**Observation**
.parameter*	.code*
.fact.value	.valueQuantity.value
.fact.unit	.valueQuantity.unit
.interpretation	.interpretation.text

.upperLimit.normalizedValue	.referenceRange.high.value
.upperLimit.normalizedUnit	.referenceRange.high.unit
.lowerLimit.normalizedValue	.referenceRange.low.value
.lowerLimit.normalizedUnit	.referenceRange.low.unit

Assembling the JSON Response

Figure 2 shows an excerpt of a text mining response for the text "The patient had a cold". "Cold" is recognized as entity of the type *Condition* and annotated with the ICD-10 code "*J00 Acute nasopharyngitis* [common cold]".


```
{"resourceType": "Bundle",
 "id": "628320",
 "meta": {
  "versionId": "1",
  "lastUpdated": "2018-11-24T15:00:45.182+00:00"},
 "type": "collection",
 "entry": [{
   "fullUrl": "urn:uuid:f844ec9a-ef45-11e3-8bb6-
00aa004d0001",
   "resource": {
    "resourceType": "Binary",
    "contentType": "text/plain",
    "content": "UGF0aWVudCBoYWQgYSBjb2xk"}},
  {"fullUrl": "urn:uuid:f844ec9a-ef45-11e8-8bb6-
00aa004d0001",
   "resource": {
    "resourceType": "Device",
    "manufacturer": "Averbis GmbH",
    "model": "Health Discovery",
    "version": "5.6.0"}},
  {"fullUrl": "urn:uuid:f844ec9d-ef45-11e8-96f5-
00aa004d0001",
   "resource": {
    "resourceType": "Condition",
    "clinicalStatus": "active",
    "verificationStatus": "unknown",
    "code": {
     "coding": [{
      "system": "http://hl7.org/fhir/sid/icd-10-cm",
      "version": "2018",
      "code": "J00",
      "display": "Acute nasopharyngitis [common cold]"}]},
    "subject": {
     "display": "anonymous patient"}}},
  {"fullUrl": "urn:uuid:f844ec9d-ef45-11e8-96f3-
00aa004d0001",
   "resource": {
    "resourceType": "Provenance",
    "target": [{
     "reference": "urn:uuid:f844ec9d-ef45-11e8-96f5-
00aa004d0001"}],
    "recorded": "2018-11-23T18:34:03.184859+01:00",
    "agent": [{
     "whoReference": {
      "reference": "urn:uuid:f844ec9a-ef45-11e8-8bb6-
00aa004d0001"}}],
    "entity": [{
     "extension": [{
      "url":
"http://example.com/StructureDefinition/confidence/0.9",
      "valueQuantity": {
       "value": 0.1}},
     {"url": "http://example.com/StructureDefinition/character-
range/0.5",
      "valueRange": {
       "low": {
        "value": 15,
        "unit": "1"},
       "high": {
        "value": 19,
        "unit": "1"}}}],
     "role": "source",
     "whatReference": {
      "reference": "urn:uuid:f844ec9a-ef45-11e3-8bb6-
00aa004d0001"}}]}}]}
```


Figure 2– JSON Response.

Discussion

Provenance aspects of clinical data have been awarded major emphasis in recent years. The growing amount of EHRs provides unprecedented opportunity for its re-use in many tasks. However, there are various caveats to the use of such data, including inaccuracy, incompleteness and unknown provenance [29]. This is even more the case when EHR content is furthermore processed, such as by text mining systems. A dataset in FHIR created by structured data entry done by physicians requires a different interpretation compared to a seemingly identical dataset produced by text mining analysis of an unstructured narrative. For instance, the former one might meet the quality level required for triggering clinical decision support algorithms, whereas the latter one might not, although it may be perfectly suited for supporting cohort selection or outcome analyses. *Confidence* values that qualify each text mining generated data element can here be used as a valuable filter.

Quality assurance of text mining systems also requires formative assessment cycles in which human experts iteratively check extracted data elements against the source in order to assess their accuracy. If FHIR-based resources were not able to preserve this link between source and extract, this gap would have to be bridged outside the FHIR standard. The new *character-range* specification allows embedding text mining quality assurance completely within FHIR.

Our work is original in the sense that it extends FHIR to incorporate provenance and quality information. With the two extensions for *confidence* and *character-range*, FHIR now contains all specifications to represent text mining generated EHR extracts. Apart from this, two recent publications are worth mentioning, that combined FHIR with text mining. In [30], a FHIR data model automatically structures prostate pathology reports in several natural languages. In [18], the FHIR data model was also used to automatically structure free text and, in addition, to integrate structured data. In contrast to these systems, we argue that – due to different HL7 dialects and other standards – it is not necessary to base the underlying text mining system internally on FHIR, but to ensure that its external interface provides FHIR compatibility, as we have demonstrated. Thus, different text mining systems can operate in parallel or in combination on different HL7 dialects, feeding their results in a standardised information model.

Conclusions

FHIR is increasingly accepted as a universal clinical information model. It supports provenance assertions of data as a critical foundation for authenticity, trust and reproducibility. Provenance is especially relevant if datasets are created by machines, e.g. by text mining and machine learning. We have demonstrated how the output of a commercial clinical text mining tool can be harmonised with FHIR, preserving important provenance information like character span and confidence value. Necessary extensions to

the FHIR standard are currently being processed by the international HL7/FHIR committee.

Acknowledgements

The consortia SMITH and MIRACUM are funded by the German Federal Ministry of Education and Research (BMBF) within the Medical Informatics Funding Scheme (FKZ 1ZZ1606A and 01ZZ1609A).

References

[1] J. Heflin and J. Hendler, *Semantic interoperability on the web*. Maryland Univ. College Park Dept. of Computer Science, 2000.
[2] D. Kalra and D. Lloyd, EN13606 *Electronic Health Record Communication Part 1: Reference Model*. CEN TC/251, Brussels, 2007.
[3] T. Beale, Archetypes: Constraint-based domain models for future-proof information systems. OOPSLA 2002 workshop on behavioural semantics (Vol. 105).
[4] HL7 FHIR Resource list: https://www.hl7.org/fhir/resourcelist.html [accessed March 13, 2019]
[5] J.C. Mandel, D.A. Kreda, K.D. Mandl, I.S. Kohane, and R.B Ramoni, SMART on FHIR: a standards-based, interoperable apps platform for electronic health records, *JAMIA* 23 (2016), 899-908.
[6] D. Bender and K. Sartipi, HL7 FHIR: An agile and RESTful approach to healthcare information exchange. *IEEE 26th International Symposium on Computer-Based Medical Systems* (2013), 326-331.
[7] S.M. Meystre, G.K. Savova, K.C. Kipper-Schuler, and J.F. Hurdle, Extracting information from textual documents in the electronic health record: a review of recent research. *IMIA Yearbook of Medical Informatics* (2008), 128-144.
[8] https://averbis.com/health-discovery/ [accessed March 13, 2019]
[9] http://www.trinetx.com/nlp/ [accessed March 13, 2019]
[10] S. Schulz, T. Seddig, S. Hanser, A. Zaiss, and P. Daumke. Checking coding completeness by mining discharge summaries. *Stud Health Technol Inform.* (2011), 594-598.
[11] C. Catarino, A. Grandjean, S. Doss, M. Mücke, S. Tunc, K. Schmidt, J. Schmidt, P. Young, T. Bäumer, C. Kornblum, M. Endres, P. Daumke, T. Klopstock, and B. Schoser (2017), mineRARE: Semantic text-mining of electronic medical records as diagnostic decision support tool to search for rare neurologic diseases such as Pompe disease, Fabry disease and Niemann-Pick type C disease. *European Journal of Neurology* 24 (2017), 75.
[12] D. Schober, R. Choquet, K. Depraetere, F. Enders, P. Daumke, M.C. Jaulent, D. Teodoro, E. Pasche, C. Lovis, and M. Boeker, DebugIT: Ontology-mediated layered Data Integration for real-time Antibiotics Resistance Surveillance. *Semantic Web Applications and Tools for Life Sciences*, Berlin 2014.
[13] A. Gerstmair, P. Daumke, K. Simon, and E. Kotter. Intelligent image retrieval based on radiology reports. *Eur Radiol*. 22 (2012), 2750-2758
[14] K. Tomanek, P. Daumke, F. Enders, J. Huber, K. Theres, and M.L. Müller, An Interactive De-Identification-System. *SMBM – 5th International Symposium on Semantic Mining in Biomedicine* (2012), 82-86.
[15] D.A. Ferrucci, E.W. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. Kalyanpur, A. Lally, J.W. Murdock, E. Nyberg, J.M. Prager, N. Schlaefer, Nico and C.A. Welty. Building Watson: An overview of the DeepQA project." *AI magazine* 31.3 (2010), 59-79.
[16] I. Gurevych, M. Mühlhäuser, C. Müller, J. Steimle, M. Weimer, and T. Zesch, Darmstadt Knowledge Processing Repository Based on UIMA. *Proceedings of the First Workshop on Unstructured Information Management Architecture at Biannual Conference of the Society for Computational Linguistics and Language Technology.* Heidelberg: Springer (2007), 89.
[17] G.K. Savova, J.J. Masanz, P.V. Ogren, J. Zheng, S. Sohn, K.C. Kipper-Schuler, and C.G. Chute, Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, Component Evaluation and Applications. Journal of the American Medical Informatics Association: *JAMIA* 17 (2010), 507-513.
[18] S. Sohn, C. Clark, S.R. Halgrim, S.P. Murphy, C.G. Chute CG, and H. Liu. MedXN: an open source medication extraction and normalization tool for clinical text. *JAMIA* 21 (2014), 858-865.
[19] Y.K. Lin, H. Chen, and R.A. Brown, MedTime: a temporal information extraction system for clinical narratives. *J Biomed Inform* 46 (2013), Suppl:S20-8.
[20] N. Hong, A. Wen, F. Shen, S. Sohn, S. Liu, H. Liu, and G. Jiang, Integrating structured and unstructured EHR data using an FHIR-based type system: a case study with medication data. *AMIA Summits on Translational Science Proceedings* (2017), 74.
[21] https://openehr.org/ [accessed March 13, 2019]
[22] https://www.cdc.gov/nchs/icd/icd10cm.htm [accessed March 13, 2019]
[23]https://www.dimdi.de/dynamic/de/klassifikationen/icd/icd-10-gm/ [accessed March 13, 2019]
[24] https://www.nlm.nih.gov/research/umls/rxnorm/ [accessed March 13, 2019]
[25] http://abdata.de/datenangebot/abdamed/ [accessed March 13, 2019]
[26] https://www.whocc.no/atc_ddd_index/ [accessed March 13, 2019]
[27] https://www.dimdi.de/dynamic/de/arzneimittel/atc-klassifikation/ [accessed March 13, 2019]
[28] https://loinc.org [accessed March 13, 2019]
[29] W.R. Hersh, M.G. Weiner, P.J. Embi, J.R. Logan, P.R. Payne, E.V. Bernstam, H.P. Lehmann, G. Hripcsak, T.H. Hartzog, J.J. Cimino, and J.H. Saltz, Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Car*e 51 (2013), S30-37.
[30] P. Prinsen, R. van Ommering, M. Janssen, M. Loos, A. Tahmasebi, S. Thompson, S. Coons, P. Henrik Vincent, C. Lindh, O. Akre, A. Rastinehad, L.A. Mynderse, and P.J. Chang, A fully automatic solution for structuring prostate pathology reports. *Journal of Clinical Oncology* 36 (2018), e18586-e18586.

Address for Correspondence

Philipp Daumke, Averbis GmbH, Averbis GmbH, Salzstraße 15, 79098 Freiburg, Germany, info@averbis.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190189

MDRCupid: A Configurable Metadata Matching Toolbox

Noemi Deppenwiese[a], Petra Duhm-Harbeck[a], Josef Ingenerf[a,b], Hannes Ulrich[a]

[a] *IT Center for Clinical Research, Lübeck (ITCR-L), University of Lübeck, Germany*
[b] *Institute of Medical Informatics, University of Lübeck, Germany*

Abstract

Metadata matching is an important step towards integrating heterogeneous healthcare data and facilitating secondary use. MDRCupid supports this step by providing a configurable metadata matching toolbox incorporating lexical and statistical matching approaches. The matching configuration can be adapted to different purposes by manually selecting algorithms and their weights or by using the optimization module with corresponding training data. The toolbox can be accessed as a web service via programming or user interface. For every selected metadata element, the metadata elements with the highest similarity scores are presented to the user and can be manually confirmed via the user interface, while the programming interface uses a similarity threshold to select corresponding elements. An HL7 FHIR ConceptMap is used to save the matches. Manually confirmed matches may be used as new training data for the optimizer to improve the matching parameters further.

Keywords:
Metadata, Data Curation, Health Information Interoperability

Introduction

Modern healthcare produces large amounts of medical data, from routine documentation to data from clinical trials to hospital bills [1]. This data is captured using a wide variety of input systems, from hospital information systems to clinical trial management systems to laboratory systems to accounting software. The secondary use of such data for research purposes is aimed for, but complicated to establish due to the variety of different data formats [2]. However, even if the data is converted into a common format, the same bit of information may still be represented differently. For example, one system may choose to store systolic and diastolic blood pressure in two separate fields, another system may use a single field just called blood pressure. So, while the data may have the same format, it still differs semantically.

Therefore, when pooling data from different sources, e.g., for conducting meta-studies or collecting data from different healthcare providers a shared data format is needed (syntactic interoperability) first, but it alone does not guarantee that the provided data can be used together because different systems may choose different synonyms for their data (no semantic interoperability). Systems may define how they capture data by defining metadata elements which contain information about single units of data, e.g., the datatype or the unit of measure. These metadata elements can be used to identify corresponding fields in the datasets.

Then, generating a mapping including transformation rules allows converting the collected data from one system to be compatible with the data semantics of another system. In order to develop such mappings, in a first step corresponding metadata elements must be identified. This process is referred to as matching in this work.

Metadata elements need to be in a common format to be processed. The ISO 11179-3 standard [3] defines metadata elements and is supported by some metadata repositories. Metadata repositories are used to collect, store and administer metadata elements.

Multiple tools are providing an automated or semi-automated matching or mapping functionality. CUPID [4] is a popular schema matching algorithm that takes the label, datatype, permissible values, and overall structure into account but does not support all information contained in ISO 11179 data elements. Another example is the COMA++ software for schema matching which allows the usage and parametrization of multiple string-matching algorithms and accepts SQL, XML, and OWLs input files [5]. However, existing tools are not incoporating information from value sets, which is essential in the field of clinical trials.

Other systems like the MDM-Portal work by annotating metadata elements with semantic codes (in this case UMLS) and using this information to detect correspondences [6].

The objective of the developed software is to provide a highly configurable matching tool that can process ISO 11179 conformant metadata elements. The toolbox contains several string-matching algorithms as well as implementing the statistical bag-of-words method. The service can be accessed either via an application programming (API) or a graphical user interface (GUI). The toolbox includes an optimization module which processes already matched training data, to ease finding a suitable configuration.

Methods

Toolbox

Figure 1 shows the basic structure of the toolbox. To find the best match for a given data element from a set of candidate elements, a similarity score is calculated for every possible match. This score is calculated as the sum of weighted similarity scores for single data element attributes. These attributes can be regarded as metadata of the metadata and include, for example, name and datatype. Their values (e.g., *birthday* or *date of birth* for name and *string* or *date* for datatype, as seen in the left part of figure 1) are regarded as simple string values, to which processing pipes, consisting of any number of preprocessors and a lexical matching algorithm, can be applied.

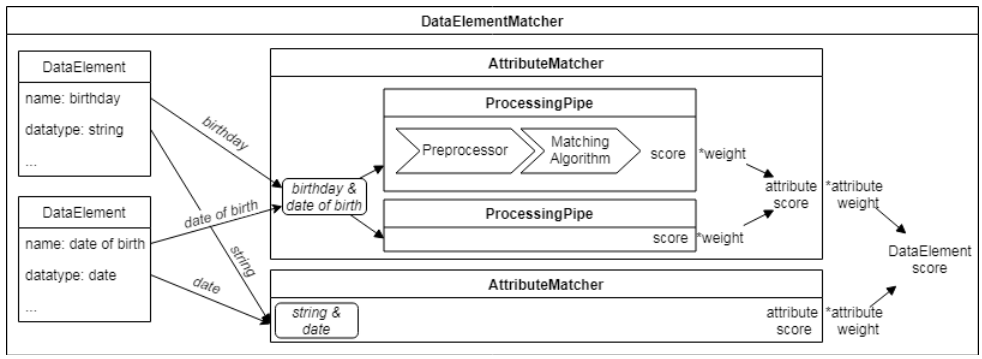


Figure 1- MDRCupid structure overview. For each metadata element attribute like name *or* datatype*, an AttributeMatcher calculates the similarity score for this attribute's values. In addition to the standard AttributeMatchers, three other calculations are performed that use information from more than one attribute (not shown).*

The toolbox provides several common lexical string matching algorithms like Levenshtein or n-gram based methods. Other string matching algorithms can be easily added provided that they fulfill two conditions:

- The algorithm needs to be normalized, and its results shall be in the range between zero and one regardless of the compared strings length.
- Higher scores shall indicate better matches. For algorithms calculating distances, this may be easily achieved by subtracting the calculated distance from one.

These conditions are necessary to ensure the optimization process works correctly. The similarity score for two string values from two data elements for the same attribute is then calculated as the sum of weighted results for all applied processing pipes. The weights for the processing pipes can be configured, as can the attribute weights mentioned before.

In addition to the comparison of simple string attributes (e.g., *name* values), similarity scores for other data element properties are calculated and contribute to the overall data element score. Since a data element may contain a list of permissible values (answer options), a similarity measure for two such lists is needed to include them in the overall score.

To calculate this measure, the similarity between every single option in data element *A* and every option in data element *B* is calculated using a configurable combination of the aforementioned lexical string matching algorithms. These scores are collected in a similarity matrix. Then, at most one number from every column and line is chosen so that the sum of these values is maximized. Every chosen value represents a match between two answer options, and the maximized combination represents the overall best matching.

Furthermore, the group elements containing a data element may hold information essential to its understanding, e.g., *inclusion* vs. *exclusion criteria*. Therefore, the similarity score between the surrounding groups influences the overall score, too.

To calculate the structural similarity between two data elements *A* and *B*, all parent group labels of *A* (a_i) are compared to all parent group labels of *B* (b_j) as well as the label of *B* itself. For every group label in *A*, the best match (highest score) is selected. These scores are then summed up and divided by the number of labels in *A* (n_a), resulting in the structure similarity score.

$$sim_{struct}(A,B) = \frac{\sum_{i=1}^{n_a} \max_{j=0}^{n_b}(sim(a_i,b_j))}{n_a}$$

The similarity between two labels is calculated analogously to the similarity between attribute values. The number of considered labels can be limited by setting the depth parameter, e.g., a search depth of one would result in only the group elements directly above the data elements being compared as formally described by (($n_{a|b} > d$) *then* ($n_{a|b} := d$)).

Finally, besides the lexical matching approaches used for the attributes, the toolbox also includes the statistical bag-of-words method [7]. Here, all values contained in one data element are treated as the text body for the data element. A dictionary-based lemmatizer is used to reduce each word to its basic form prior to generating the word count vector for every data element. The bag-of-words similarity is then calculated as the cosine distance between two data element vectors and also considered while calculating the overall data element similarity score.

A similarity threshold may be used to select valid matches. All elements with similarity scores exceeding the threshold will be considered matches, while all elements with lower scores will not.

Optimizer

Since the toolbox allows the configuration of many parameters as describes in the previous section, it also includes an optimization module to ease the process of finding the best configuration for a particular matching problem. Since the optimal configuration of attribute weights depends on the algorithms used to calculate attribute similarities, the parameters are calculated in two steps. First, the best processing pipes and their weights are determined for every attribute separately. Second, the attribute weights are calculated using the algorithm configuration from step one. A set of namespaces with at least some matching data elements needs to be provided as training data, and the correct matches need to be specified, e.g., as annotations or via an external file.

For the algorithm optimization, all possible combinations of currently implemented preprocessors and matching algorithms are built as processing pipes. To find the best combinations for a particular attribute, for every combination of two namespaces, data element similarity scores for this attribute

are calculated for each pipe separately. Then, the best three pipes are chosen per data element and rewarded three, two and one points respectively. These score points are summed up for each pipe over all training data, normalized and then used as this pipes weight.

To decide which pipes yielded the best results, one out of two different measures may be applied. Both receive a list of possible matches ranked in descending order by their similarity score to the target data element.

1. The *score distance* is calculated as the score of the first matching element (as defined by the ground truth) minus the score of the highest ranking element.
2. The *position distance* is calculated as one divided by the position of the first matching element.

Thus, larger values here imply better rankings.

To calculate how the attributes should be weighted, scores and rankings for two namespaces are calculated for each attribute separately using the pipes and weights determined in the first optimization step. Then, the attribute weights are calculated analogously to the processing pipes during the algorithm optimization. The optimization module can also be used to calculate a similarity threshold for a particular dataset. Two kinds of thresholds may be determined:

1. A *lower threshold* is calculated by summing up the similarity scores of the last correct element in the ranked result list for each training data element and dividing it by the number of training data elements. This value may be used to avoid losing correct matches.
2. Alternatively, a more *aggressive threshold* can be determined by summing up the similarity scores of the elements directly above the first incorrect match in the ranked result list. The sum is again divided by the number of training data elements and may be used to avoid false positive matches.

Interfaces

The toolbox is wrapped in a web service and can be accessed via an API or a GUI. The programming interface accepts the URLs of two Samply.MDR [8] namespaces. This open-source software is used as a source metadata repository, even though with minor adaptions other ISO 11179-3 conformant source systems should be compatible as well. Optionally, a toolbox configuration defining the pipes and weights to be used can be submitted as well. If no configuration is specified, the toolbox uses a default configuration, e.g., the result of an optimizer run. For each data element in the source namespace, the data element in the target namespace with the highest similarity score is selected as a match. Alternatively, a threshold may be used as described before. The generated matchings are returned as HL7 FHIR ConceptMaps. The ConceptMap resource provides a standardized way to represent relationships between concepts or data elements by referencing their codes, in this case, URNs assigned by Samply.MDR. ConceptMaps provide the possibility to specify the type of the relationship between the matched concepts, but since our automated matching process does not differentiate between, e.g., broader and narrower meanings, it always uses the *relatedTo* code.

Like the API, the GUI first requests the URLs of the two Samply.MDR namespaces to be matched. Then, the default mapping configuration may be edited or used as-is. In the next step, the matching is executed, and the results are displayed: For every source data element, a list of possible matches from the target namespace is shown, ranked by their similarity scores. The number of possible matches may be limited via a similarity score threshold, as described in the optimization section. The matching candidates can then be reviewed manually, and correct matches may be selected while incorrect hits may be de-selected. After confirmation, a ConceptMap containing the selected results is generated and displayed. It is also stored in an internal database to be used as further training data for the optimizer.

Results

Optimization parameters

To identify whether the score or position distance is better suited for each optimization step, the optimizer module was executed several times with different configurations on a set of training data. The metadata set originated from a biomedical research project and was split into training and test data. Elements are annotated with numeric codes, two elements with at least some common codes are considered matches. The data set has a depth of up to two nested group elements, however two thirds of the dataset have a depth of one, with an average of 11 elements per group. To evaluate the influence of the structure information, the optimization module was configured one time to omit the group label data and the other time to use all available group information. For each of this depth configurations, four configurations were calculated using all possible combinations score or position distance for the processing pipe and the attribute optimization. The obtained configurations were subsequently tested on the training and test datasets. The results are shown in table 1.

Table 1– Results of the optimized configurations using different values for the depth parameter d, on the training and test data subsets, as specified in the data column. The four rightmost columns specify whether the position (pos) or score distance was used for pipeline (first position) and attribute optimization (second position). The values indicate in which proportion of all possible matches a correct match had received the highest similarity score (Top 1) or was among the five elements with the highest scores (Top 5). The highest values per row are in bold.

d	data	error	pos-pos	pos-score	score-score	score-pos
2	train	Top 1	**0.43**	0.37	0.39	**0.43**
2	train	Top 5	0.68	0.61	0.63	**0.69**
2	test	Top 1	0.53	0.58	**0.61**	0.55
2	test	Top 5	0.78	0.82	**0.83**	0.80
0	train	Top 1	**0.44**	0.40	0.40	**0.44**
0	train	Top 5	**0.70**	0.65	0.65	**0.70**
0	test	Top 1	**0.52**	**0.52**	0.51	**0.52**
0	test	Top 5	**0.78**	0.76	0.76	**0.78**

Overall, using the score distance for pipeline and position distance for attribute optimization yields the best results and resulted in a configuration which found a correct best match for 55% of the testing data elements, in 80% of the cases suitable matches were among the five highest ranking elements. Not using the group label information yields slightly better results on the training data. However, when comparing the absolute number of correct matches, incorporating the

structure information leads to more correct matches among both the first and first five elements.

Toolbox configuration

The toolbox configuration parameters calculated by the different optimization processes are displayed in figures 2 and 3. The first figure shows the cumulated weights for pipelines over all attributes. Some pipelines with an accumulated weight lower than 0,1 for both distance measures have been omitted, among these standard edit distance measures like Levenshtein distance. Both distance measures favor similar, n-gram-based pipelines, but the position distance assigns weights more equally distributed.

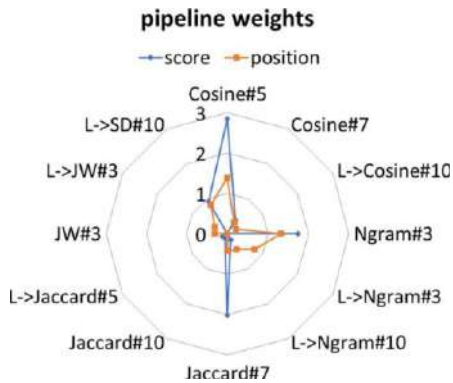


Figure 2- Pipeline weights for optimization via score and position distance, summed up over all attributes. L stands for lemmatization as preprocessing, JW for Jaro-Winkler and SD for Sorensen-Dice similarity. The numbers after the # indicate the n for n-gram-based methods.

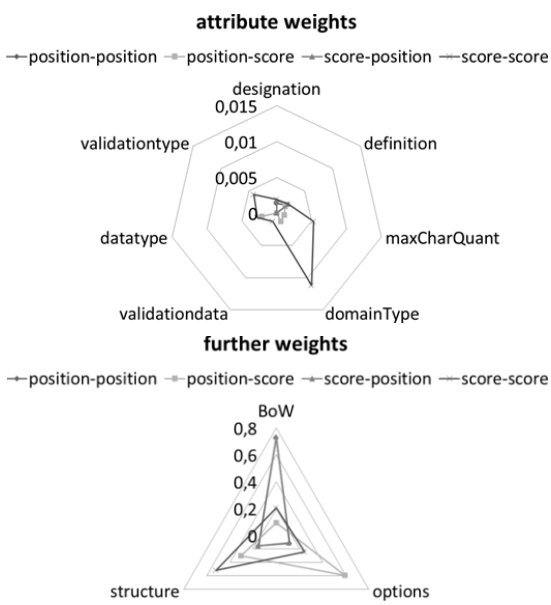


Figure 3- Calculated attribute weights by optimization method for the data element attributes (upper figure) and the computed attributes (lower figure). BoW stands for the bag-of-words score, options for the answer options similarity score and structure for the similarity score calculated from the surrounding group labels.

The weights assigned to attributes are displayed in figure 3, they are split between the data element attributes whose similarity is computed directly as string similarity (e.g., *description* or *datatype*) and the computed attributes. The computed attributes score much larger weights than the data element attributes. While both optimizers using the position distance for the attribute weight calculation step obtain similar results, the results of the optimizers using the score distance show much larger discrepancies.

Mappathon

The authors also participated in the GMDS Mappathon 2018. To comply with the competition demands, the toolbox was modified to support a Cypher-based [9] export format and external CSV files as source for ground truth. The results can be found on the corresponding website [10]. MDRCupid won the mappathon challenge in the best mapping category.

Discussion

The pipeline optimization using the score distance favors certain pipelines more clearly than the one using the position distance. This imbalance is probably due to high scores being assigned to non-matching data elements, while the three pipelines selected by the score distance tend to identify matching data elements more precise. During the attribute weight optimization, both optimizers using the position based distance for this second step calculate nearly equal weights, while the score based approach chooses very different weights. The weight that it assigns to the *domainType* attribute is remarkably high. This attribute has only two possible values . Therefore many data elements score a similarity of one and consequently a score distance of zero. The same effect occurs with option and structure similarity: here, many data elements lack an answer options list or a surrounding group and hence a similarity of zero is calculated, causing a score distance of zero if the matching element also lacks the feature in question. Thus, the score distance measure should not be used for attribute weight calculation.

Using the structure information always leads to better or equal results for the test data, but not in the training data. This effect may be caused by the lack of depth in the training and test data set. In particular, the subset of training data only contains namespaces with depths of one and zero, while the test data has depths of up to two.

While different optimization strategies like only choosing the best pipe for each attribute are feasible, trying to incorporate more matching algorithms should better avoid overfitting. In our dataset, the results of the optimized configuration were even better on test data than on training data. Still, some modifications, e.g., selecting more pipes or assigning weights directly derived from the distance measures, are conceivable and should be evaluated in further work.

One of the most common formats to represent metadata in a clinical context is the CDISC ODM widely used for clinical trials. Unlike Samply.MDR, which allows nesting as many or as little item groups as needed, CDISC ODM defines specific group layers, which additionally have an assigned meaning (e.g., *studyEvent*, *form*).

The Samply.MDR was chosen because it is open source, ISO 11179-3 conformant and allows to retrieve data elements via a RESTful API. Also, several tools exist to migrate metadata to the Samply.MDR format, e.g., from the CDISC ODM format [11].

Matching results are exported as FHIR ConceptMaps because they have multiple advantages. First, they can easily be stored, exchanged and searched by using FHIR servers. Second, multiple software libraries are supporting FHIR, easing the resource creation process. Third, ConceptMaps only store codes referring to the data elements, not the data elements themselves, avoiding problems concerning the intellectual property of the metadata elements. Fourth, they have a standardized structure and an integrated code system for metadata relationship concepts. Other formats like the Cypher-based exports required for the Mappathon challenge need to include such a code system; it is not standardized in the specification of the format.

The toolbox only supports lexical and one statistical (bag-of-words) matching methods. Another approach would include the use of semantic annotations. Although including semantic codes would allow the use of elaborate ontologies like SNOMED-CT or LOINC, annotating data elements usually requires some amount of manual work [6], while lexical and statistical methods can work with the data out of the box.

Conclusions

The secondary use of health data, e.g., using data from EHR in clinical trials, is complicated by the data models heterogeneity. Metadata matching is an essential step towards data integration because it finds correspondence in the data structure by identifying corresponding metadata elements. MDRCupid is designed to fulfill this task. It provides various lexical matching algorithms as well as the statistical bag-of-words method. The current implementation allows the user to select which combination of methods is used for each run. Preprocessing methods and string matching algorithms can be selected for each data element attribute separately. When multiple processors are applied to the same attribute, the weighting for each score can be chosen as well. Weights can be assigned to each attribute accordingly.

Besides the data element attributes per se, similarity scores are calculated for the data elements permissible values and containing item groups. Together with the attribute similarities and the bag-of-words based score, these results can be weighted and incorporated into the overall result. As an alternative to manually defining algorithms and weights, a configuration can be generated by applying the optimization module to a given set of training data. Two possible distance measures for this optimization process were evaluated on a metadata set from a biomedical research project, and the best approach resulted in a configuration which found a correct best match for 55% of the testing data elements.

While MDRCupid currently works with ISO 11179-3 conformant metadata elements, it could be adapted to support metadata with other attributes as well, e.g. the upcoming ISO 21526 standard for metadata repositories in healthcare.

The source code for MDRCupid is available on GitHub [12]. It can be deployed as a web service, including a programming and graphical user interface.

Software support for the task of metadata matching becomes more and more critical with the increase of electronically captured medical data. MDRCupid is providing this support and laying the groundwork for data integration. A set of mapped metadata elements can be used to build a comprehensive clinical repository, enabling the secondary use of healthcare data, simplifying data exchange and medical research and, ultimately, improving patient care.

References

[1] T. Benson, and G. Grieve, Principles of Health Interoperability, Springer International Publishing, 2016. doi:10.1007/978-3-319-30370-3.

[2] O.I. Ogunyemi, D. Meeker, H.-E. Kim, N. Ashish, S. Farzaneh, and A. Boxwala, Identifying appropriate reference data models for comparative effectiveness research (CER) studies based on data from clinical information systems, *Medical Care*. **51** (2013) S45-52. doi:10.1097/MLR.0b013e31829b1e0b.

[3] S.M.N. Ngouongo, M. Löbe, and J. Stausberg, The ISO/IEC 11179 norm for metadata registries: does it cover healthcare standards in empirical research?, *Journal of Biomedical Informatics*. **46** (2013) 318–327. doi:10.1016/j.jbi.2012.11.008.

[4] J. Madhavan, P.A. Bernstein, and E. Rahm, Generic Schema Matching with Cupid, in: Proceedings of the 27th International Conference on Very Large Data Bases, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2001: pp. 49–58.

[5] D. Aumueller, H.-H. Do, S. Massmann, and E. Rahm, Schema and Ontology Matching with COMA++, in: Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data, ACM, New York, NY, USA, 2005: pp. 906–908. doi:10.1145/1066157.1066283.

[6] M. Dugas, A. Meidt, P. Neuhaus, M. Storck, and J. Varghese, ODMedit: uniform semantic annotation for data integration in medicine based on a public metadata repository, *BMC Medical Research Methodology*. **16** (2016) 65. doi:10.1186/s12874-016-0164-9.

[7] F. Cady, The Data Science Handbook, John Wiley & Sons, 2017.

[8] D. Kadioglu, B. Breil, C. Knell, M. Lablans, S. Mate, D. Schlue, H. Serve, H. Storf, F. Ückert, T. Wagner, P. Weingardt, and H.-U. Prokosch, Samply.MDR - A Metadata Repository and Its Application in Various Research Networks, *Stud Health Technol Inform*. **253** (2018) 50–54.

[9] N. Francis, A. Green, P. Guagliardo, L. Libkin, T. Lindaaker, V. Marsault, S. Plantikow, M. Rydberg, P. Selmer, and A. Taylor, Cypher: An Evolving Query Language for Property Graphs, in: Proceedings of the 2018 International Conference on Management of Data, ACM, New York, NY, USA, 2018: pp. 1433–1445. doi:10.1145/3183713.3190657.

[10] Mappathon: A Metadata Mapping Challenge, *Mappathon: A Metadata Mapping Challenge*. https://www.mappathon.de/ (accessed November 19, 2018).

[11] A.-K. Kock-Schoppenhauer, H. Ulrich, S. Wagenzink, P. Duhm-Harbeck, J. Ingenerf, P. Neuhaus, M. Dugas, and P. Bruland, Compatibility Between Metadata Standards: Import Pipeline of CDISC ODM to the Samply.MDR, *Studies in Health Technology and Informatics*. **247** (2018) 221–225.

[12] IT Center for Clinical Research-Universität zu Lübeck, MDRCupid, (2018). https://github.com/itcr-uni-luebeck/MDRCupid (accessed November 19, 2018).

Address for correspondence

Noemi Deppenwiese
IT Center for Clinical Research, Lübeck;
Ratzeburger Allee 160, 23538 Lübeck, Germany
Telephone: +49 (0) 451 - 500 52702
Email: noemi.deppenwiese@uksh.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190190

A Large-Scale Analysis of Health Journalism by Reliable and Unreliable Media

Sameer Dhoju[a], Md Main Uddin Rony[a], Muhammad Ashad Kabir[b], Naeemul Hassan[a]

a Department of Computer and Information Science, University of Mississippi, USA
b School of Computing and Mathematics, Charles Sturt University, Australia

Abstract

Media outlets play crucial roles in disseminating health information. Previous studies have examined how health journalism is practiced by reliable and unreliable media outlets. However, most of the existing works are conducted over a relatively small set of samples. In this study, we investigate a large collection (about 30 thousand) of health-related news articles which were published by 29 reliable and 20 unreliable media outlets and identify several differences in health journalism practice. Our analysis shows that there are significant structural, topical, and semantic disparities in the way reliable and unreliable media outlets conduct health journalism. We argue, in this age of 'fake news', these findings will be useful to combat online health disinformation.

Keywords:

Health Communication

Introduction

Of the 20 most-shared articles on Facebook in 2016 with the word "cancer" in the headline, more than half the reports were discredited by doctors and health authorities [1]. The spread of health-related hoaxes is not new. However, the advent of Internet, social networking sites (SNS), and click-through-rate (CTR)-based pay policies have made it possible to create hoaxes/"fake news", publish at a larger scale and reach to a broader audience with a higher speed than ever [2]. Misleading or erroneous health news can be dangerous as it can lead to a critical situation. Houston reported a measles outbreak in Europe due to lower immunization rate which experts believed was the result of anti-vaccination campaigns caused by a false news about MMR vaccine [3]. Moreover, misinformation can spoil the credibility of the health-care providers and create a lack of trust in taking medicine, food, and vaccines. Recently, researchers have started to address the fake news problem in general [4; 5]. However, health disinformation is a relatively unexplored area. According to a report from Pew Research Center [6], 72% of adult internet users search online for information about a range of health issues. So, it is important to ensure that the health information which is available online is accurate and of good quality. There are some authoritative and reliable entities such as National Institutes of Health (NIH)[1] or Health On the Net[2] which provide high-quality health information. Also, there are some fact-checking sites such as Snopes.com[3] and Quackwatch.org[4] that regularly debunk health and medical related misinformation. Nonetheless, these sites are incapable of busting the deluge of health disinformation continuously produced by unreliable health information outlets (e.g., RealFarmacy.com, Health Nut News). Moreover, the bots in social networks significantly promote unsubstantiated health-related claims [7]. Researchers have tried developing automated health hoax detection techniques but had limited success due to several reasons such as small training data size and lack of consciousness of users [8-11].

The objective of this paper is to identify discriminating features that can potentially separate a reliable health news from an unreliable health news by investigating a large-scale dataset. We examine how reliable media and unreliable media outlets conduct health journalism. First, we prepare a large dataset of health-related news articles which were produced and published by a set of reliable media outlets and unreliable media outlets. Then, using a systematic content analysis, we identify the features which separate a reliable outlet sourced health article from an unreliable sourced one. These features incorporate the structural, topical, and semantic differences in health articles from these outlets. For instance, our structural analysis finds that the unreliable media outlets use clickbaity headlines in their health-related news significantly more than what reliable outlets do. Our topical analysis finds that while the reliable outlets discuss "cancer" along with research and studies, in the unreliable outlets "cancer" is associated with autism and vaccination. The semantic analysis shows that on average a health news from reliable media contains more reference quotes than an average unreliable sourced health news. We argue that these features can be critical in understanding health misinformation and designing systems to combat such disinformation. In future, our goal is to develop a machine learning model using these features to distinguish unreliable media sourced health news from reliable articles.

Related Work

There has been extensive research on how scientific medical research outcomes should be disseminated to general people by following health journalism protocols [12-16]. For instance, Lopes et al. suggest that it is necessary to integrate journalism studies, strategic communication concepts, and health professional knowledge to successfully disseminate professional findings. Some researchers particularly focused on the spread of health misinformation in social media [17]. For example, [8] analyzes Zika related misinformation on Twitter. In particular, it shows that tracking health misinformation in social media is not trivial, and requires some expert supervision. It exploited crowdsource to annotate a collection of Tweets and used the annotated data to build a rumor classification model. One limitation of this work is that the used dataset is too small (6 rumors) to make a general conclusion. Moreover, it didn't consider the features in the actual news articles unlike us. Ghenai and Mejova [9] examines the individuals on social media that are posting questionable health-related information,

[1] https://www.nih.gov/
[2] https://www.hon.ch/en/
[3] https://www.snopes.com/
[4] http://www.quackwatch.org/

and in particular promoting cancer treatments which have been shown to be ineffective. It develops a feature based supervised classification model to automatically identify users who are comparatively more susceptible to health misinformation. There are other works which focus on automatically identifying health misinformation. For example, Kinsora et al. [18] developed a classifier to detect misinformative posts in health forums. One of the limitations of this work is that the training data is only labeled by two individuals. Researchers have also worked on building tools that can help a user to easily consume health information. Kostkova et al. [10] developed the "VAC Medi+board", an interactive visualization platform integrating Twitter data and news coverage from a reliable source called MediSys[5]. It covers public debate related to vaccines and helps users to easily browse health information on a certain vaccine-related topic.

Our study significantly differs from these already existing researches. Instead of depending on a small sample of health hoaxes like some of the existing works, we take a different approach and focus on the source outlets. This gives us the benefit of investigating with a larger dataset. We investigate the journalistic practice of reliable and unreliable outlets, an area which has not been studied to the best of our knowledge.

Data Preparation

The results are presented here. Authors may choose a combination of text, tables, figures, and graphs to convey the results of their work to the reader. There are no set limitations on the number of tables, figures, and graphs that may be used in papers, posters, and proposals. Large figures and tables may span two columns. Please number tables and figures and reference them appropriately in the text.

Media Outlet Selection

The first challenge is to identify reliable and unreliable outlets. The matter of reliability is subjective. We decided to consider the outlets which have been cross-checked as reliable or unreliable by credible sources.

Reliable Media

We identified 29 reliable media outlets from three sources– i) 11 of them are certified by the Health On the Net [19], a non-profit organization that promotes transparent and reliable health information online. It is officially related with the World Health Organization (WHO) [20]. ii) 8 from U.S. government's health-related centers and institutions (e.g., CDC, NIH, NCBI), and iii) 10 from the most circulated broadcast [21] mainstream media outlets (e.g., CNN, NBC). Note, the mainstream outlets generally have a separate section for health information (e.g., https://www.cnn.com/health). As our goal is to collect health-related news, we restricted ourselves to their health portals only.

Unreliable Media

Dr. Melissa Zimdars, a communication and media expert, prepared a list of false, misleading, clickbaity, and satirical media outlets [22; 23]. Similar lists are also maintained by Wikipedia [24] and informationisbeautiful.net [25]. We identified 6 media outlets which primarily spread health-related misinformation and are present in these lists. Another source for identifying unreliable outlets is Snopes.com, a popular hoax-debunking website that fact-checks news of different domains including health. We followed the health or medical hoaxes debunked by Snopes.com and identified 14 media outlets which sourced those hoaxes. In total, we identified 20 unreliable outlets. Table 1 lists the Facebook page ids of all the reliable and unreliable outlets that have been used in this study.

Table 1– List of Facebook page ids of the reliable and unreliable outlets. Some of them are unavailable now

Reliable	everydayhealth, WebMD, statnews, AmericanHeart, BBCLifestyleHealth, CBSHealth, FoxNewsHealth, WellNYT, latimesscience, tampabaytimeshealth, philly.comhealth, AmericanHeart, AmericanCancerSociety, HHS, CNNHealth, cancer.gov, FDA, mplus.gov, NHLBI, kidshealthparents, ahrq.gov, healthadvocateinc, HealthCentral, eMedicineHealth, C4YWH, BabyCenter, MayoClinic, MedicineNet, healthline
Unreliable	liveahealth, healthexpertgroup, healthysolo, organichealthcorner, justhealthylifestyle1, REALfarmacy.com, thetruthaboutcancer, BookforHealthyLife, viralstories.bm, justhealthyway, thereadersfile, pinoyhomeremedies, onlygenuinehealth, greatremediesgreathealth, HealthRanger, thefoodbabe, AgeofAutism, HealthNutNews, consciouslifenews, HealthImpactNews

Data Collection

The next challenge is to gather news articles published by the selected outlets. We identified the official Facebook pages of each of the 49 media outlets and collected all the link-posts[6] shared by the outlets within January 1, 2015 and April 2, 2018 using Facebook Graph API. For each post, we gathered the corresponding news article link, the status message, and the posting date.

News Article Scraping

We used a Python package named Newspaper3k[7] to gather the news article related data. Given a news article link, this package provides the headline, body, author name (if present), and publish date of the article. It also provides the visual elements (image, video) used in an article. In total, we collected data for 29, 047 articles from reliable outlets and 15, 017 from unreliable outlets.

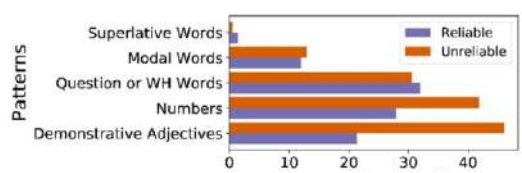


Figure 1– Distribution of clickbait patterns

Filtering non-Health News Articles

Even though we restricted ourselves to health-related outlets, we observed that the outlets also published or shared non-health (e.g., sports, entertainment, weather) news. We removed these non-health articles from our dataset and only kept *health, food & drink*, or *fitness & beauty* related articles. Specifically, for each news article, we used the document categorization service provided by Google Cloud Natural Language API to determine its topic. If an article doesn't belong to one of the three above

[5] http://medisys.newsbrief.eu

[6] Facebook allows posting status, pictures, videos, events, links, etc. We collected the link type posts only.

[7] https://newspaper.readthedocs.io/en/latest/

mentioned topics, it is filtered out. This step reduced the dataset size to 27,589; 18,436 from reliable outlets and 9,153 from unreliable outlets. We used this health-related dataset only in all the experiments of this paper. Figure 1 shows the health-related news percentage distribution for reliable outlets and unreliable outlets using box-plots. For each of the 29 reliable outlets, we measure the percentage of health news and then use these 29 percentage values to draw the box-plot for the reliable outlets; likewise, for unreliable. We observe that the reliable outlets (median 72%) publish news on health topics comparatively less than unreliable outlets (median 85%).

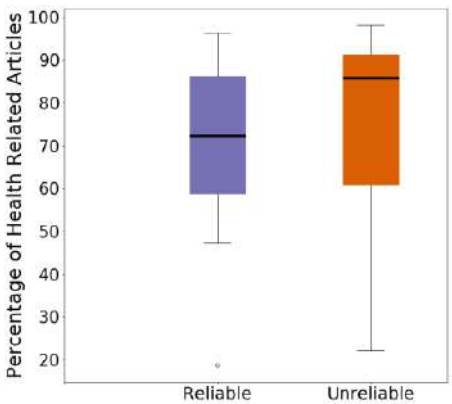


Figure 2– Comparison between reliable and unreliable outlets with respect to the presence of health-related news contents

Analysis

Using this dataset, we conduct content analysis to examine structural, topical, and semantic differences in health news from reliable and unreliable outlets.

Structural Difference

We particularly examine two structural elements- headlines and visual media of news articles.

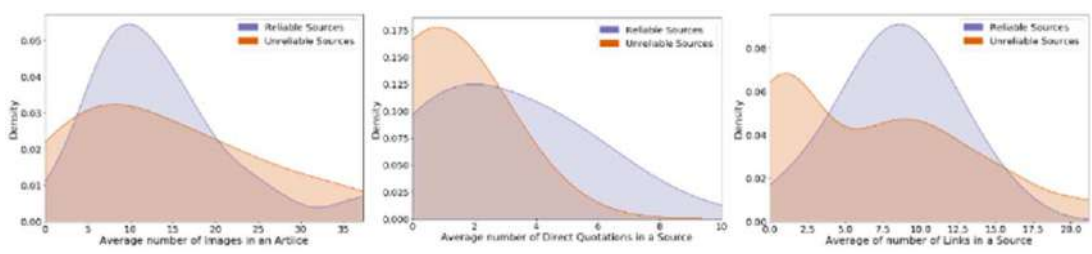
Headline

The headline is a key element of a news article. According to a study done by American Press Institute and the Associated Press [26], only 4 out of 10 Americans read beyond the headline. So, it is important to understand how reliable and short headline does. We observe that the average headline unreliable outlets construct the headlines of their health-related news. According to Breaux [27], a longer headline results in significantly higher click-through-rate (CTR) than a length of an article from reliable outlets and an article from unreliable outlets is 8.56 words and 12.13 words, respectively. So, on average, an unreliable outlet's headline has a higher chance of receiving more clicks or attention than a reliable outlet's headline. To further investigate this, we examine the *clickbaitiness* of the headlines. The term clickbait refers to a form of web content (headline, image, thumbnail, etc.) that employs writing formulas, linguistic techniques, and suspense creating visual elements to trick readers into clicking links but does not deliver on its promises [28]. Chen et al. [29] reported that clickbait usage is a common pattern in false news articles. We investigate to what extent the reliable and unreliable outlets use clickbait headlines in their health articles. For each article headline, we test whether it is a clickbait or not using two supervised clickbait detection models– a sub-word embedding based deep learning model [30] and a feature engineering based Multinomial Naive Bayes model [31]. Agreement between these models was measured as 0.44 using Cohen's κ. We mark a headline as a clickbait if both models labeled it as clickbait. We observe, 27.29% (5,031 out of 18,436) of the headlines from reliable outlets are click bait. In unreliable outlets, the percentage is significantly higher, 40.03% (3,664 out of 9,153). So, it is evident that the unreliable outlets use more clickbaits than reliable outlets.

We further investigate the linguistic patterns used in the clickbait headlines. In particular, we analyze the presence of some common patterns which are generally employed in clickbait according to [27; 32]. The patterns are- 1) Presence of demonstrative adjectives (e.g., this, these, that). 2) Presence of numbers (e.g., 10, ten). 3) Presence of modal words (e.g., must, should, could, can). 4) Presence of question or WH words (e.g., what, who, how). 5) Presence of superlative words (e.g., best, worst, never). Figure 2 shows the distribution of these patterns among the clickbait headlines of reliable and unreliable outlets. Note, one headline may contain more than one pattern. For example, this headline *"Are these the worst 9 diseases in the world?"* contains four of the above patterns. This is the reason why summation of the percentages isn't equal to one. We see that unreliable outlets use demonstrative adjective and numbers significantly more compared to the reliable outlets.

Use of visual media

We examined how often the outlets use images in the articles. Our analysis finds that on average an article from reliable outlets uses 13.83 images and an article from unreliable outlets uses 14.22 images. Figure 3a shows density plots of the average number of images per article for both outlet categories. We observe that a good portion of unreliable outlet sourced articles uses a high number of images (more than 20).



(a) Image (b) Quotation (c) Link

Figure 3–Distribution of average number of image/quotation/link per article from reliable and unreliable outlets

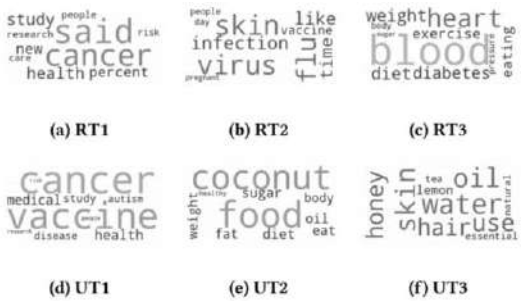


Figure 4– Topic modeling (k = 3) of articles from reliable outlets (top, denoted as RT) and from unreliable outlets (bottom, denoted as UT).

Topical Difference

The health domain is considerably broad and it covers many topics. We hypothesize that there are differences between the health topics in reliable and unreliable outlet articles. We test this hypothesis using an unsupervised & a supervised analysis.

Topic Modeling

We use *Latent Dirichlet Allocation(LDA)* algorithm to model the topics in the news articles. The number of topics, *k*, was set as 3. Figure 4 shows three topics for each of the outlet categories. Each topic is modeled by the top-10 important words in that topic. The font size of words is proportional to the importance. Figure 4a and 4d indicate that "cancer" is a common topic in reliable and unreliable outlets. Although, the words *study, said, percent, research,* and their font sizes in Figure 4a indicate that the topic "cancer" is associated with research studies, facts, and references in reliable outlets. On the contrary, unreliable outlets have the words *vaccine, autism,* and *risk* in Figure 4d which suggests the discussion regarding how vaccines put people under autism and cancer risk, an unsubstantiated claim, generally propagated by unreliable media[8,9]. Figure 4e and 4f suggest the discussions about weight loss, skin, and hair care products (e.g., essential oil, lemon). Topics in Figure 4b and 4c discuss mostly flu, virus, skin infection, exercise, diabetes and so on.

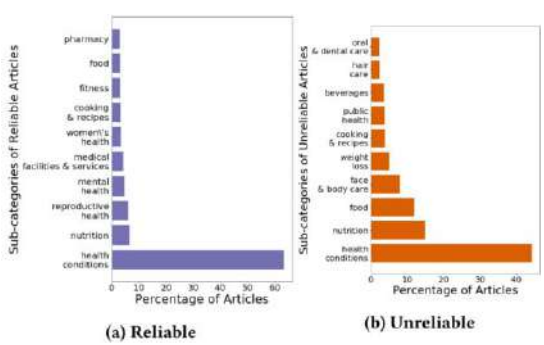


Figure 5 –Top-10 topics in reliable and unreliable outlets.

Topic Categorization

In addition to topic modeling, we categorically analyze the articles' topics using Google Cloud Natural Language API. Figure 5 shows the top-10 topics in the reliable and unreliable outlets. In the case of reliable, the distribution is significantly dominated by *health condition*. On the other hand, in the case of unreliable outlets, percentages of *nutrition* and *food* are noticeable. Only 4 of the 10 categories are common in two outlet groups. Unreliable topics have *weight loss, hair care, face & body care*. This finding supports our claim from topic modeling analysis.

Semantic Difference

We analyze what efforts the outlets make for a logical and meaningful health news. Specifically, we consider to what extent the outlets use quotations and hyperlinks. Use of quotation and hyperlinks in a news article is associated with credibility [33; 34]. Presence of quotation and hyperlinks indicates that an article is logically constructed and supported with credible factual information.

Quotation

We use the Stanford QuoteAnnotator[10] to identify the quotations from a news article. Figure 3b shows density plots of the number of quotations per article for reliable and unreliable outlets. We observe that unreliable outlets use less number of quotations compared to reliable outlets. We find that the average number of quotations per article is 1 and 3 in unreliable and reliable outlets, respectively. This suggests that the reliable outlet sources articles are more credible and unreliable outlets are less credible.

Hyperlink

We examine the use of the hyperlink in the articles. On average, a reliable outlet sourced article contains 8.4 hyperlinks and an unreliable outlet sourced article contains 6.8 hyperlinks. Figure 3c shows density plots of the number of links per article for reliable and unreliable outlets. The peaks indicate that most of the articles from reliable outlets have close to 8 (median) hyperlinks. On the other hand, most of the unreliable outlet articles have less than 2 hyperlinks. This analysis again suggests that the reliable sourced articles are more credible than unreliable outlet articles.

Conclusion and Future Work

In this paper, we closely looked at structural, topical, and semantic differences between articles from reliable and unreliable outlets. Our findings reconfirm some of the existing claims such as unreliable outlets use clickbaity headlines to catch the attention of users. In addition, this study finds new patterns that can potentially help separate health disinformation. For example, we find that less quotation and hyperlinks are more associated with unreliable outlets. However, there are some limitations to this study. For instance, we didn't consider the videos, cited experts, comments of the users, and other information. In the future, we want to overcome these limitations and leverage the findings of this study to combat health disinformation.

References

[1] Katie Forster. (accessed October 30, 2018). Revealed: How dangerous fake health news conquered Facebook. https://www.independent.co.uk/life-style/healthand-families/health-news/fake-news-health-facebook-cruel-damaging-socialmedia-mike-adams-natural-health-ranger-conspiracy-a7498201.html

[2] Mathew Ingram. (accessed October 30, 2018). The internet didn't invent viral content or clickbait journalism - there's just more of it now, and it happens faster. https://gigaom.com/2014/04/01/the-internet-didnt-invent-

[8]www.webmd.com/brain/autism/do-vaccines-cause-autism

[9]https://www.skepticalraptor.com/skepticalraptorblog.php/polio-vaccine-causescancer-myth/

[10] https://stanfordnlp.github.io/CoreNLP/quote.html

viral-contentor-clickbait-journalism-theres-just-more-of-it-now-and-it-happens-faster/
[3] Muiris Houston. (accessed October 31, 2018). Measles back with a vengeance due to fake health news. https://www.irishtimes.com/opinion/measles-back-with-avengeance-due-to-fake-health-news-1.3401960
[4] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. Science 359, 6380 (2018), 1094–1096.
[5] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter 19, 1 (2017), 22–36.
[6] Susannah Fox. (accessed October 30, 2018). The social life of health information. http://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-ofhealth-information/
[7] Gaby Galvin. (accessed October 30, 2018). How Bots Could Hack Your Health. https://www.usnews.com/news/healthiest-communities/articles/2018-07-24/how-social-media-bots-could-compromise-public-health
[8] Amira Ghenai and Yelena Mejova. 2017. Catching Zika Fever: Application of Crowdsourcing and Machine Learning for Tracking Health Misinformation on Twitter. In Healthcare Informatics (ICHI), 2017 IEEE International Conference on. IEEE, 518–518
[9] Amira Ghenai and Yelena Mejova. 2018. Fake Cures: User-centric Modeling of Health Misinformation in Social Media. In 2018 ACM Conference on ComputerSupported Cooperative Work and Social Computing (CSCW). ACM.
[10] Patty Kostkova, Vino Mano, Heidi J Larson, and William S Schulz. 2016. Vac medi+ board: Analysing vaccine rumours in news and social media. In Proceedings of the 6th International Conference on Digital Health Conference. ACM.
[11] Emily K Vraga and Leticia Bode. 2017. Using Expert Sources to Correct Health Misinformation in Social Media. Science Communication 39, 5 (2017), 621–645.
[12] Nicole K Dalmer. 2017. Questioning reliability assessments of health information on social media. Journal of the Medical Library Association: JMLA 105,1.
[13] Irja Marije de Jong, Frank Kupper, Marlous Arentshorst, and Jacqueline Broerse. 2016. Responsible reporting: neuroimaging news in the age of responsible research and innovation. Science and engineering ethics 22, 4 (2016).
[14] Marjorie Kagawa-Singer and Shaheen Kassim-Lakha. 2003. A strategy to reduce cross-cultural miscommunication and increase the likelihood of improving health outcomes. Academic Medicine 78, 6 (2003), 577–587.
[15] Gary Schwitzer. 2008. How do US journalists cover treatments, tests, products, and procedures? An evaluation of 500 stories. PLoS medicine 5, 5 (2008), e95.
[16] Miriam Shuchman and Michael S Wilkes. 1997. Medical scientists and health news reporting: a case of miscommunication. Annals of Internal Medicine 126, 12 (1997), 976–982.
[17] Felisbela Lopes, Teresa Ruão, Zara Pinto Coelho, and Sandra Marinho. 2009. Journalists and health care professionals: what can we do about it?. In 2009 Annual Conference of the International Association for Media and Communication Research (IAMCR), “Human Rights and Communication”. 1–15.
[18] Alexander Kinsora, Kate Barron, Qiaozhu Mei, and VG Vinod Vydiswaran. 2017. Creating a Labeled Dataset for Medical Misinformation in Health Forums. In Healthcare Informatics (ICHI), 2017 IEEE International Conference on. IEEE, 456–461.
[19] HEALTH ON THE NET. (accessed September 24, 2018). https://www.hon.ch/en/
[20] World Health Organization (WHO). (accessed September 24, 2018). http://www.who.int/
[21] Michael Schneider. (accessed September 24, 2018). Most-Watched Television Networks: Ranking 2016’s Winners and Losers. https://www.indiewire.com/2016/12/cnn-fox-news-msnbc-nbratings-2016-winners-losers-1201762864/
[22] Wikipedia. (accessed September 24, 2018). Wikipedia:Zimdars’ fake news list. https://bit.ly/2ziHafj
[23] Melissa Zimdars. 2016. My ‘fake news list’ went viral. But made-up stories are only part of the problem. https://www.washingtonpost.com/posteverything/wp/2016/11/18/my-fake-news-list-went-viral-but-made-up-stories-are-onlypart-of-the-problem. (2016).
[24] Wikipedia. (accessed September 24, 2018). List of fake news websites. https://bit.ly/2moBDvA
[25] informationisbeautiful.net. 2016. Unreliable/Fake News Sites & Sources. https://docs.google.com/spreadsheets/d/1xDDmbr54qzzG8wUrRdxQlC1dixJSIYqQUaXVZBqsJs
[26] American Press Institute and the Associated Press-NORC Center for Public Affairs Research. (accessed September 28, 2018). The Personal News Cycle: How Americans choose to get their news. https://www.americanpressinstitute.org/publications/reports/survey-research/how-americans-get-news/
[27] Chris Breaux. (accessed September 28, 2018). "You’ll Never Guess How Chartbeat’s Data Scientists Came Up With the Single Greatest Headline". http://blog.chartbeat.com/2015/11/20/youll-never-guess-how-chartbeatsdata-scientists-came-up-with-the-single-greatest-headline
[28] Bryan Gardiner. (accessed September 28, 2018). "You’ll Be Outraged at How Easy It Was to Get You to Click on This Headline".
[29] Yimin Chen, Niall J Conroy, and Victoria L Rubin. 2015. Misleading online content: Recognizing clickbait as false news. In Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection. ACM, 15–19.
[30] Md Main Uddin Rony, Naeemul Hassan, and Mohammad Yousuf. 2017. Diving Deep into Clickbaits: Who Use Them to What Extents in Which Topics with What Effects? In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017. ACM, 232–239
[31] Saurabh Mathur. (accessed September 24, 2018). Clickbait Detector. https://github.com/saurabhmathur96/clickbait-detector
[32] Matthew Opatrny. (accessed September 28, 2018). "9 Headline Tips to Help You Connect with Your Target Audience". https://www.outbrain.com/blog/9-headlinetips-to-help-marketers-and-publishers-connect-with-their-target-audiences/
[33] Juliette De Maeyer. 2012. The journalistic hyperlink: Prescriptive discourses about linking in online news. Journalism Practice 6, 5-6 (2012), 692–701. https://www.wired.com/2015/12/psychology-of-clickbait/
[34] S Shyam Sundar. 1998. Effect of source attribution on perception of online news stories. Journalism & Mass Communication Quarterly 75, 1 (1998), 55–68.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190191

An Interactive Timeline Visualization for Patient Cohorts in the Oncological Routine: A Use Case on Multiple Myeloma

Julia Dieter[a], Janko Ahlbrandt[a], Alexander Knurr[a], Janine Al-Hmad[a] and Frank Ückert[a]

[a] Division of Medical Informatics for Translational Oncology, German Cancer Research Center (DKFZ), Heidelberg, Baden-Württemberg, Germany

Abstract

With the growing interdisciplinarity of cancer treatment and increasing amounts of data and patients, it is getting increasingly difficult for physicians to capture a patient's medical history as a basis for adequate treatment and to compare different medical histories of similar patients to each other. Furthermore, in order to tackle the etiological mechanisms of cancer, it is crucial to identify patients exhibiting a different disease course than their corresponding cohort. Several timeline visualizations have already been proposed. However, the functions and design of such visualizations are always use case dependent. We constructed a cohort timeline prototype mock-up for a specific oncological use case involving multiple myeloma, where the chronological monitoring of various parameters is crucial for patient diagnosis and treatment. Our proposed cohort timeline is a synthesis between elements described in the literature and our own approaches regarding function and design.

Keywords:
Clinical Decision-Making, Cohort Studies, Data Mining

Introduction

For the treatment of cancer, physicians need to consider comprehensive information about the patients, including the medical history [1] as well as current events such as examination results. The information mainly comprises basic variables such as age, gender, education, and clinical variables such as details on diagnoses, therapies, staging results, laboratory analyses, and further examinations. However, those variables are usually stored in different systems within the hospital, e.g., the laboratory information system or the hospital information system. The data then have to be chronologically realigned in a manual fashion by physicians, often under time pressure. Thus, there is a great need for tools creating a chronological overview of patient data, helping to save time, and preventing errors.

To support physicians in the compaction and comparison of salient patient information as well as in the chronological realignment of data from different sources, several approaches have been developed for integration and visualization of medical patient histories. There are tools for either single patient (e.g., LifeLines, KNAVE-II; [2–4]) or for patient cohorts (e.g., LifeLines2, VISITORS; [4–7]), and they aim to display various data types, such as numerical data (e.g., MIVA), categorical data or both (e.g., WBIVS; [4,8]).

A fundamental work for the display of a single patient medical history is the LifeLines approach of Plaisant et al. [2]. It is not restricted to oncology and aims at displaying the key information of a single patient's medical history by aligning zoomable lanes on top of each other. Depending on the task at hand, they can be opened up or closed [2,4]. It focuses on categorical data. The lanes represent different areas of the patient record such as "Problems", "Diagnoses" or "Tests", while single events are encoded within the lanes. For instance, the lane entitled "Tests" harbors examination events such as "Blood" or "ECG" etc. The separate markings within the lane are positioned according to their corresponding chronological position on the x-axis, which is located at the bottom of all lanes. Design aspects such as line thickness reflect the severity of medical problems. Detailed information on an event is given in a separate window, which can be summoned by clicking on the respective event in the main view. Further interactive features include a mouseover function (i.e., displaying information when hitting a label with the cursor), zooming, alerts and a search option for keywords such as symptoms (e.g., "migraine"). KNAVE-II is another visualization architecture for time-oriented clinical patient data and has been implemented and tested in the oncological domain [3]. Similar to Lifelines, it comprises horizontally stacked views, but in contrast to the former, KNAVE-II offers domain-specific groupings as well as temporal abstractions to enable interactive data manipulation and exploration [3,9]. However, drawbacks of Lifelines and KNAVE-II might be crowding problems, i.e., the display might become too complex due to increasing timespans and number of events [2].

Approaches to the chronological display of patient cohorts have also been developed. Bernard et al. [10] developed a network of static dashboards to visualize the medical histories of a post-operative prostate cancer cohort. Each separate dashboard represents a temporal segment of the cohort, i.e., one part of time-ordered temporal data. Each dashboard is composed of different visualizations such as bar charts, pie charts or box plots. The chronological connection, i.e., the medical history of the cohort, is represented by the network that is constructed based on the connections between the single dashboards. Unfortunately, the tool has not been validated as previous visualization approaches. Thus, its efficiency in the real world setting still remains uncertain. The tool does not support the traditional timeline display, so that physicians might need extended training. In addition, the data processing steps underlying the temporal segmentation of data might be hard to follow and only a limited number of EHR elements might be shown within the dashboard due to space issues.

The LifeLines2 approach uses a classical timeline visualization, just like its predecessor LifeLines. It enables the

display of temporal categorical data for multiple patient records. The lanes are stacked vertically and represent each single cases, with event markers color-coded on the corresponding positions at the horizontal time axis [11]. All records can be aligned by a specific event in chronological display [4]. The alignment can also be achieved according to a specific events, e.g., the second diagnosis. Furthermore, users are able to filter and search for specific event sequences; and the temporal distribution of selected event types can be displayed by histogram [4,11]. In addition, approaches such as Similian, which extended LifeLines2 with a similary measure, enable the search for similar medical patient histories [4,12]. However, LifeLines2 displays one patient record per horizontal lane, so that not all records can be viewed simultaneously in larger patient cohorts due to space issues [13]. In addition, numerical data have to be preprocessed as the approach focuses on categorical data only.

The VISITORS system by Klimov et al. [7], an extension of KNAVE-II, is a visualization and exploration tool for time-oriented, categorical as well as numerical data of multiple patients. VISITORS is also constructed of horizontal lanes that are associated with a horizontal, chronological x-axis and enables the display of raw data as well as abstracted concepts for overviews. However, the tool does not enable the display of multiple numerical parameters within one lane. When several lanes for different numerical parameters need to be applied, not all lanes might be viewed within one window.

Taken together, there exist various cohort timeline visualizations for different application scenarios and data types, also exhibiting multiple interactive features. Yet, as the literature about the cohort timelines shows, their configuration and functionalities mostly depend on a specific use case for which they are constructed or adapted. Thus, there does not exist one universal solution for all use cases.

The visualizations in clinical routine is still in high demand, as no approaches are widely deplayed yet [14]. As such, we developed a timeline for cohort in clinical oncological routine, adapted to the specific use case of multiple myeloma (MM; ICD10: C90.0) within the first-line therapeutic setting. It could become apparent for the need to monitor several parameters over time and to compare the course of one patient to a corresponding cohort by applying such tool (see below).

MM can arise from two pre-stages, i.e., the Monoclonal Gammopathy of Undetermined Significance (MGUS) or the smoldering myeloma [15]. Both conditions are lacking in treatment indication and can be differentiated from the MM and monitored by several parameters such as clonal plasma cells in the bone marrow, monoclonal protein in the serum and urine as well as end organ damage [15]. The transition from the pre-stages to an MM is marked by the fulfillment of one or more of the SLiM-CRAB criteria (an acronym standing for ">60% clonal plasma cell content in the bone marrow", "free light chain ratio in the serum >100", ">1 focal lesion >1cm in MRI imaging", "Calcium", "renal insufficiency", "anemia" and "bone lesions"; [16,17]). Those criteria comprise several laboratory parameters (e.g., calcium, creatinine and hemoglobin), bone lesions as determined by imaging, the percentage of clonal plasma cells in the bone marrow, the free light chain ratio in the serum and focal lesions as determined by MRI [15].

In addition, although the steps for treatment are given in the MM treatment guideline, it might still be necessary to decide between different treatment options. Those decisions have to be made individually, for instance, the decision in favor of or against an autologous stem cell transplantation. Especially in the cases of older patients, it becomes increasingly important to weight the burden of therapeutic side effects against the potential treatment benefit. To support decision making, it might be of help to compare one patient's medical information with treatment profiles from previous patients and their treatment outcomes.

We therefore developed the MM use case-specific prototype mock-up of an interactive cohort timeline.

Methods

The construction of our cohort timeline was based on a combination of timeline functionalities and design aspects collected from a literature review and our own approaches to these issues.

The prototype mock-up of the proposed cohort timeline was developed in iterative cycles with clinical, medical informatics as well as biological inputs. We adapted the concept of horizontal lanes from previous tools, such as LifeLines, LifeLines2 and VISITORS, for visualizing the medical history of patients [2–4,7,9]. Functionalities for interactive data visualization were designed and incorporated to meet physician's needs and suggestions of previous studies [4–7,11].

The parameters in our prototype mock-up were selected based on the MM treatment guideline for diagnosis, therapy and the monitoring of the remission status, including the laboratory values, the designations of therapy steps as well as the criteria for evaluating the treatment response. In addition, the International Myeloma Working Group (IMWG) and the revised-Myeloma Comorbidity Index (R-MCI) were also included, which have been suggested to be useful in identifying geriatric risk profiles and prognostic values for functional decline among older multiple myeloma patients [18]. They are calculated by a combination of weights among various parameters including age, comorbidities and the ability to perform activities of daily living (e.g., self-care and household) [19]. Therefore, those scores are referenced as relevant decision criteria to differentiate between treatment options documented in the MM treatment guideline, especially for older patients.

We documented our general design and planned interactive funcions in a detailed prototype mock-up by a user experience specialist.

Results

The proposed interactive cohort timeline is built up of vertically stacked lanes, each comprising one key clinical component such as information on the general patient condition (IMWG score, R-MCI), laboratory data, treatment options and information on the treatment response, which are all significant for diagnosing, treating or monitoring MM. This sequence of events during patient treatment is reflected in the vertical order of the lanes from top to bottom (see Figure 1).

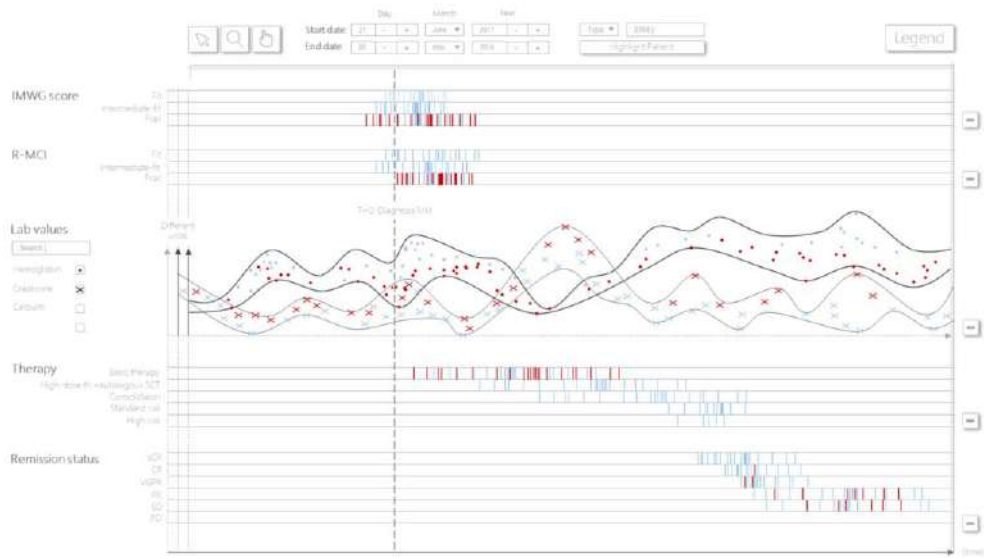


Figure 1: Prototype mock-up of the use case-specific cohort timeline

Each lane depicts all corresponding data points of the selected cohort. In case of visualizing multiple cohorts, the latter are color-coded in order to be distinguished within the same lane. The associations of laboratory values to their parameters are represented by data points, e.g., dots and crosses (see Figure 2).

The chronological reference of the visualization is given by a time-harboring x-axis at the bottom of all lanes. The vertically stacked lanes can be chronologically compared and referenced to the horizontal time axis by means of a thin vertical marker that spans all lanes (similar to the Data Point Scrubber of Faiola et al. [20]) and follows the cursor.

The visualization type for each lane is content-specific. Laboratory values are presented in an x-y-diagram-lane. Different laboratory parameters are visualized within the same lane. Upper and lower flanking lines that connect the lowest and highest values, respectively, indicating the value range. A filtering option next to the lane allows the selection of single laboratory parameters or functional parameter sets.

The occurrence of events such as treatments or evaluation of a remission status is marked in the corresponding lane by a small vertical line per patient at the corresponding time point (i.e., position on the time axis).

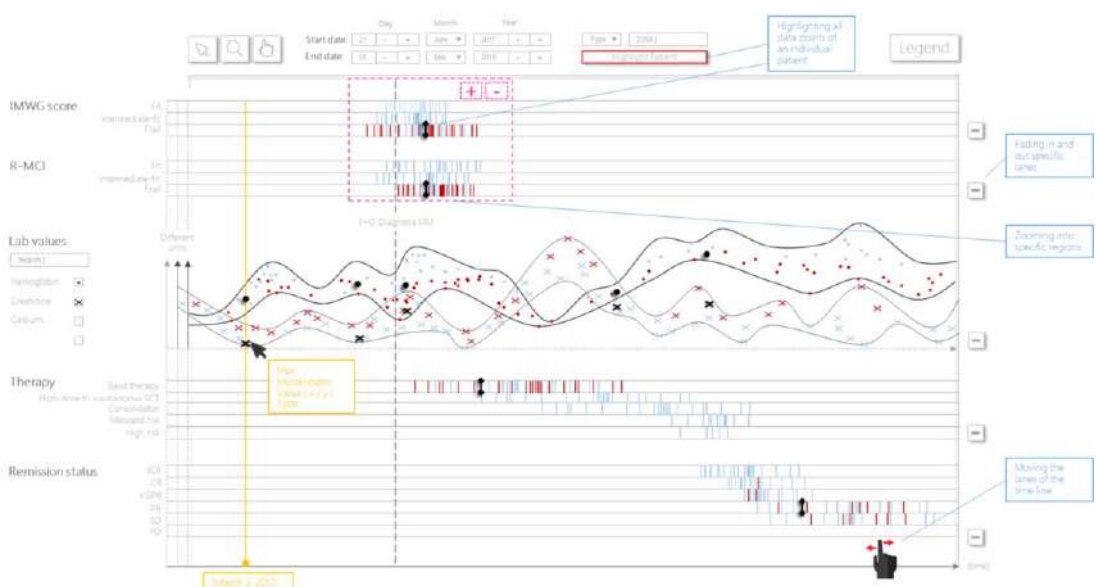


Figure 2: Proposed interactive features for the cohort timeline

Given a specific use case, a user might want to read the values in an absolute timeline or a relative one to a defined reference time point. For instance, the time of diagnosis of a specific disease might serve as reference, relative to all other events that are calculated and displayed chronologically for each individual patient. Therefore, similar to VISITORS [7], an option to choose between absolute and relative chronological display will be integrated. The options to switch between different parameters of reference [4–6] as well as the selection of a time window of interest will also be given [4,7].

Additionally, lanes can be faded in and out depending on the use case at hand [2]. Further interactive features include a hover over function (similar to Wang et al. [11]), displaying the patient-ID as well as the corresponding value when hitting a data point with the cursor. In addition, all data points from a specific patient shall be highlighted throughout the cohort timeline in order to enable the comparison to a corresponding cohort. To differentiate compact or clustered data, a zoom function will be included (e.g., [2]). In the case where a lane is too cluttered with data points or if subcohorts are to be selected, a filtering function is also added. This is enabled by checking or unchecking selection boxes on a target lane or simply by marking specific data points. In case of visualizing long time spans, a sideward scroll is included as well as an option to constrain the time window for which data are to be displayed. A retractable legend is included. Furthermore, certain users such as physicians in different specialities might have different preference on the ways to display the cohort information in timeline. And customized configurations can be saved for future use to save time. An export function for data and images will also be provided. An example of the general cohort timeline design is shown in Figure 1. The interactive functionalities are presented in Figure 2.

Discussion

The literature shows a multitude of approaches for temporal visualizations for medical histories of patient cohorts, including a variety of interactive features. However, it is also obvious that each timeline has to be adapted to a specific use case in order to see if the approach can address the specific questions in different user scenarios.

We developed a prototype mock-up for the oncological use case of diagnosing, treating and monitoring MM, as a synthesis between designs and functionalities of previous cohort timelines and our own ideas.

As described above, a physician might want to weight the burden of therapeutic side effects against the potential treatment benefit for an old MM patient. In this case, the physician plots two cohorts of previous MM patients onto the timeline: One having received stem cell transplantation as well as high-dose therapy (blue cohort in Figure 1 and Figure 2, appearing brighter in the black and white view) and another one having received basic therapy only (red cohort, appearing darker in the black and white view). Patients categorized as “frail” in the IMWG score and the R-MCI have higher creatinine and lower hemoglobin levels after therapy than fitter patients and a less optimal treatment response (e.g., remission status=stable disease).

The data points of a current patient can now be plotted, highligted and compared to the two cohorts to estimate potential treatment outcomes with and without SCT and high-dose therapy. The example patient might rather resemble the cohort not suited for SCT and high-dose therapy in our prototype mock-ups.

In a future step, the prototype mock-up presented here shall serve as a basis for an evaluation by end users. Although the prototype mock-up was constructed based on realistic data and procedures as much as possible, it has to be taken into account that the plot has not been programmed and tested with real data yet. Thus, the behavior of the plot in real live scenarios presently cannot be anticipated.

Conclusions

We present a compact, interactive cohort timeline visualization approach to display multidimensional data points of oncological cohorts in a chronological reference frame. Our prototype mock-up shall serve as a basis for a comprehensive evaluation by physicians as potential end users. Although our visualization has not been realized and tested with real data yet, it might become a promising tool for physicians to support decisions in treatment of oncological patients, hypothesis generation, and the exploration of the etiological mechanisms underlying cancer.

Acknowledgements

This work was conducted within the frame of the DataThereHouse (DTH) project in cooperation with the National Center for Tumor Diseases (NCT) Heidelberg, supported by the Federal Ministry of Education and Research (BMBF). The project is concerned with the integration of data from different sources including the Heidelberg University Hospital as well as the German Cancer Research Center (DKFZ) Heidelberg for all oncological entities.

References

[1] Drohan B, Grinstein G, Hughes K (2010) Oncology Lifeline - A Timeline Tool for the Interdisciplinary Management of Breast Cancer Patients in a Surgical Clinic. In *2010 14th International Conference Information Visualisation,* Staff I, ed. I E E E, [Place of publication not identified], pp. 233–238.

[2] Plaisant C, Mushlin R, Snyder A, Li J, Heller D, Shneiderman B (1998) LifeLines: Using visualization to enhance navigation and analysis of patient records. Proc AMIA Symp, 76–80.

[3] Shahar Y, Goren-Bar D, Boaz D, Tahan G (2006) Distributed, intelligent, interactive visualization and exploration of time-oriented clinical data and their abstractions. Artif Intell Med 38, 115–135.

[4] Rind A, Wang TD, Aigner W, Miksch S, Plaisant C, Wongsuphasawat K, Shneiderman B (2013) *Interactive Information Visualization to Explore and Query Electronic Health Records,* Now Publishers.

[5] Wang TD, Plaisant C, Quinn AJ, Stanchak R, Murphy S, Shneiderman B (2008) Aligning temporal data by sentinel events. In *The 26th Annual CHI Conference on Human Factors in Computing Systems, CHI 2008: Conference proceedings ; April 5-10, 2008 in Florence, Italy,* Burnett M, ed. ACM, New York, NY, p. 457.

[6] Wang TD, Plaisant C, Shneiderman B, Spring N, Roseman D, Marchand G, Mukherjee V, Smith M (2009) Temporal summaries: supporting temporal

categorical searching, aggregation and comparison. IEEE Trans Vis Comput Graph 15, 1049–1056.

[7] Klimov D, Shahar Y, Taieb-Maimon M (2010) Intelligent visualization and exploration of time-oriented data of multiple patients. Artif Intell Med 49, 11–31.

[8] Stevens SS (1946) On the Theory of Scales of Measurement. Science 103, 677–680.

[9] Ordóñez P, desJardins M, Lombardi M, Lehmann CU, Fackler J (2011) An animated multivariate visualization for physiological and clinical data in the ICU. In *Ihi 10 Proceedings of the 2010 Acm International Health Informatics,* Committee I'1C, ed. Association for Computing Machinery, New York, p. 771.

[10] Bernard J, Sessler D, Kohlhammer J, Ruddle RA (2018) Using Dashboard Networks to Visualize Multiple Patient Histories: A Design Study on Post-operative Prostate Cancer. IEEE Trans Vis Comput Graph.

[11] Wang TD, Wongsuphasawat K, Plaisant C, Shneiderman B (2011) Extracting Insights from Electronic Health Records: Case Studies, a Visual Analytics Process Model, and Design Recommendations. J Med Syst 35, 1135–1152.

[12] Wongsuphasawat K, Shneiderman B (2009) Finding comparable temporal categorical records: A similarity measure with an interactive visualization. In *IEEE Symposium on Visual Analytics Science and Technology, 2009, VAST 2009: 12-13 Oct. 2009, Atlantic City, New Jersey, USA ; proceedings,* Stasko J, ed. IEEE, Piscataway, NJ, pp. 27–34.

[13] Guerra Gómez, John Alexis & Wongsuphasawat, Krist & David Wang, Taowei & Pack, Michael & Plaisant, Catherine & Wang, K. (2011) Analyzing Incident Management Event Sequences with Interactive Visualization.

[14] Wanderer JP, Nelson SE, Ehrenfeld JM, Monahan S, Park S (2016) Clinical Data Visualization: The Current State and Future Needs. J Med Syst 40, 275.

[15] Wörmann, B., Driessen, C., Einsele, H., Goldschmidt, H., Gunsilius, E., Kortüm, M., Kröger, N., Ludwig, H., Mügge, L.-O., Naumann, R., Pritzkuleit, R., Röllig, C., Scheid, C., Taverna, C., Weisel, K., Weißinger, F. (2018) *Multiples Myelom: Leitlinie*.

[16] Boquoi A, Strapatsas J, Fenk R, Haas R (2018) Multiples Myelom. Klinikarzt 47, 418–425.

[17] Willenbacher E, Balog A, Willenbacher W (2018) Short overview on the current standard of treatment in newly diagnosed multiple myeloma. memo - Magazine of European Medical Oncology 11, 59–64.

[18] Engelhardt M, Dold SM, Ihorst G, Zober A, Möller M, Reinhardt H, Hieke S, Schumacher M, Wäsch R (2016) Geriatric assessment in multiple myeloma patients: Validation of the International Myeloma Working Group (IMWG) score and comparison with other common comorbidity scores. Haematologica, haematol.2016.148189.

[19] Roehrig B, Hoeffken K, Pientka L, Wedding U (2007) How many and which items of activities of daily living (ADL) and instrumental activities of daily living (IADL) are necessary for screening. Critical Reviews in Oncology/Hematology 62, 164–171.

[20] Faiola A, Boston-Clay C, Jones J, Downey M, Newlon CM (2012) Managing Patient Health Across Diverse Spaces: Using Activity Theory to Model Pervasive Decision Support, https://scholarworks.iupui.edu/bitstream/1805/6153/1/Faiola_2012_managing.pdf.

Address for correspondence

Dr. Julia Dieter
Research Associate
Division of Medical Informatics for Translational Oncology

German Cancer Research Center (DKFZ)
Foundation under Puplic Law
Im Neuenheimer Feld 280
69120 Heidelberg
Germany
phone: +49 6221 42-5115
fax: +49 6221 42-5109

j.dieter@dkfz-heidelberg.de
https://www.dkfz.de/en/mitro/

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190192

Evaluating the Impact of Text Duplications on a Corpus of More than 600,000 Clinical Narratives in a French Hospital

William Digan[ab], Maxime Wack[ab], Vincent Looten[ab], Antoine Neuraz[ac], Anita Burgun[abc], Bastien Rance[ab]

[a] *INSERM, Centre de Recherche des Cordeliers, UMRS 1138, Université Paris-Descartes, Université Sorbonne Paris Cité, France,*
[b] *Department of Medical Informatics, Hôpital Européen Georges Pompidou, AP-HP, Paris, France*
[c] *Department of Medical Informatics, Necker Children Hospital, AP-HP, Paris, France*

Abstract

A significant part of medical knowledge is stored as unstructured free text. However, clinical narratives are known to contain duplicated sections due to clinicians' copy/paste parts of a former report into a new one. In this study, we aim at evaluating the duplications found within patient records in more than 650,000 French clinical narratives. We adapted a method to identify efficiently duplicated zones in a reasonable time. We evaluated the potential impact of duplications in two use cases: the presence of (i) treatments and/or (ii) relative dates. We identified an average rate of duplication of 33%. We found that 20% of the document contained drugs mentioned only in duplicated zones and that 1.45% of the document contained mentions of relative dates in duplicated zone, that could potentially lead to erroneous interpretation. We suggest the systematic identification and annotation of duplicated zones in clinical narratives for information extraction and temporal-oriented tasks.

Keywords:

Electronic Health Records, Natural Language Processing, Algorithms

Introduction

The use of electronic health records (EHRs) and clinical data warehouses [1] (CDWs) lead to a better collection and preservation of patient information. CDWs store all kinds of data, including laboratory results, diagnostic codes, and clinical narratives (free text medical reports). In fact, CDWs are major tools for translational research. While a large portion of the information are stored in structured ways, and virtually directly reusable, a significant part of medical knowledge is stored as unstructured free text. Some studies have even shown that free text contains up to 80% [2] of overall information. With the availability of information, new ways of exploring data have emerged. For example, high-throughput phenotyping, machine learning or statistical models (including through the use of deep learning). However, free text can be subject to different types of issues (quality, typos…), that could profoundly bias results of analysis and models. One potential problem could come from duplicated sections in clinical reports [3] (created when clinicians copy/paste parts of a former report into a new one).

Duplications are common during care; information can be replicated from one document to another because of the static nature of the family history, previous treatments, and so on. However, in the case of secondary use of clinical data and more specifically, in the context of data extractions, duplications can have a strong impact on the chronology of the information. Old information can be found duplicated in a recent document. In this study, we aim at assessing if duplications have an impact on different types of models.

This study takes place in the context of big data, and where simple naïve approaches are not compatible with the volume of data considered (i.e. several months of calculation would be needed for simple tasks). A large body of work has been developed around the detection of plagiarism and duplication in clinical narratives. The volume of duplications has been evaluated as high as 80% in Northern American Hospital [4–6]. However, the exploration of duplications in French narratives remains limited, and the potential impact of such duplications is not easy to evaluate.

State of the Art

The identification of duplicated zones or plagiarism has generated a large body of work over the years. However, no open source solutions are available and able to handle the volume of text compatible with our purpose. In medicine, several studies focus on the characterization of copy and paste redundancy.

In their publication of 2013 [6], Cohen *et al*. studied the impact of 'copy and paste' redundancy in a large corpus of text. For that purpose, they developed a character based fingerprint method. This technique is inspired by BLAST [7] a bioinformatics algorithm which aims to find similar sequences. The authors considered 22,654 notes from 1604 patients. They found that clinical text had a redundancy level of 29%.

In a preliminary French study [3], D'Hondt *et al.* extend the Cohen methods and studied duplications in French clinical notes. The algorithm allowed the use of overlapping fingerprints. They also oriented documents in time. They choose fingerprint length of 30 and overlap of 10 char as their parameters. Furthermore, they identified that in clinical notes, most of the redundancy located on the footer and header section were in administrative sections. They worked on documents from three records and 361 documents. They found a redundancy level of 33% in clinical notes. Their algorithm allows finding near-duplicated and exact redundancy (at a price of a higher complexity of the algorithm).

A recent study [8] by Gabriel *et al.* was able to scale up to 1.5 million notes in 36.3 hours, regardless of the patient vector. They developed a new method base on windows of three phases. The first step is mini-hashing generation from files. Instead of looking at the character levels they look at the word level and defined a signature. This approach will not be followed in our study because we want to find exact duplication.

Goals

In this study, we aim at evaluating the duplications found within patient records at the European Hospital Georges Pompidou, a French hospital located in Paris. We adopted a strategy to enable the treatment of large quantities of text in a reasonable time. Finally, we evaluate the potential impact of duplications in two use cases: (i) the identification of treatments present in clinical narratives and (ii) the presence of relative dates.

Materials

In this section, we introduce the European Hospital Georges Pompidou and the corpus of text used for the study.

European Hospital Georges Pompidou

The European Hospital Georges Pompidou (HEGP in French) is a 700 bed hospital located in Paris. The HEGP is specialized in oncology, cardiovascular diseases and emergency medicine. The hospital has a clinical data warehouse (HEGP CDW) based on i2b2 [9] integrating virtually all the data generated by the hospital information system, which was deployed in 2008. Among the data collected, clinical narratives (comprised of clinical reports, letters, imaging reports, and so forth) represent more than 10 million items.

Corpus of Clinical Narratives

Our dataset is a subset of the corpus of the text of the HEGP CDW. We identified all the patients who received chemotherapy since the opening of the hospital 2000 (10,393 patients). We limited the selection of patients to those who had a follow-up of at least a year (i.e. patients with at least two visits distant by 365 days). Because we are interested in duplications within the record of a patient, we selected only patients with at least two distinct documents. Starting from 666,956 documents, we conserved a total of 649,651 documents after a preprocessing step (detail in the method section).

Methods

Definition of Duplicated Zones

In this manuscript, we define a duplication as an identical zone of text found conserved in at least two different documents. We focus on intra-record duplication (i.e. we search for duplication with the record of a patient, and not between patients). The document pairs are oriented in time.

Preprocessing

All documents generated in the hospital comprised administrative information (with the phone number of the service, the names of the staff, and so forth), the clinical notes themselves and footer information regarding the possible secondary use of data. We preprocessed the documents to remove the administrative zones and the footer information section. We also normalized the documents by converting the entire text to lower cases, and transforming multiple spaces into single ones.

Efficient Detection of Duplications

We aim at developing a method able to manage a substantial number of documents. We leverage the approaches developed by Cohen et al. [6] and D' Hondt *et al.* [3] to develop a mixed approach. In a nutshell, we rely on fingerprints build from the text to identify identical zones. A fingerprint is a segment of *N* consecutive letters. Fingerprints are not overlapping, if the first fingerprint is constructed from character 1 to N, the second fingerprint starts at position N+1. Similarly to D'Hondt et al., we also leveraged the notion of overlap: we add series of fingerprints with an offset of value OFFSET (i.e. starting at the OFFSETth character). OFFSETs are very similar to Open Reading Frames in DNA. Figure 1 shows a graphical summary of the approach. We detect duplicated ones by comparing fingerprints between pairs of documents. Contiguous or overlapping pairs of fingerprints (in the source and target documents) are merged together. We evaluated different sizes of fingerprints, and values of offsets to find a good compromise between the number of duplicated zones detected and the computed time needed to perform the calculation.

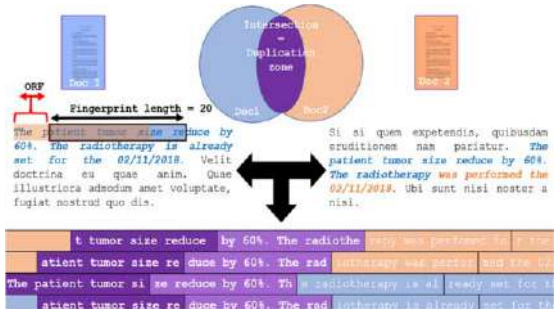


Figure 1: detection of duplicated region between two texts. The method relies on fingerprints.

Evaluating the optimal parameters: We tested combination of parameters for the values of N (size of the fingerprints), and OFFSET (value of the offset). We respectively tested the values of 30, 40 and 50 characters from the fingerprints, and 1, 5, 7, 10, 15 and 20 for offsets. In each case, we considered the offset of 1 as our baseline (fingerprints are calculated at each character position). We computed the number of common fingerprints detected, the time needed for the computation.

Computing Duplicated Zones on the Corpus of 650,000+ Documents

Using the optimal values obtained from the previous section (fingerprint length of 30, and offset of 15 characters), we computed duplicated zones on the entire corpus. Duplicated zones are detected among the documents of a single patient. We search for duplicated zones between documents oriented in time: the source document was always older than the target document. Once the pairwise duplication step has been performed, we focus on document levels, and merge all the duplicated zone detected. Figure 1 illustrates the approach.

***Filtering duplicated zones*:** After the merging steps, we identified duplicated zones with a wide variety of length, starting from 30 (the length of a fingerprint). We chose to filter out zones too small, because they likely did not correspond to copy/paste. To select a relevant threshold, we considered the number of duplications found for a given duplicated zone length.

Evaluation the volume of duplicated zones

Finally, we designed three scores to evaluate the volume of duplicated zones in the text;

Global volume of duplication in the corpus: The global duplication score

$$DupScoreGlobal = \frac{\sum_{j=0}^{Totaldocs} DuplicationLength_j}{\sum_{j=0}^{totaldocs} LengthDoc_j}$$

Average duplication score by document: The duplication score per document defined as

$$DupScoreAverageDoc = \frac{\sum_{i=0}^{Totaldocs} \frac{DuplicationLength_i}{LengthDoc_i}}{Totaldocs}$$

Average duplication score per patient: The duplication score per patient is defined as

$$DupScoreAveragePat = \frac{\sum_{k=0}^{TotalPatients} \frac{\sum_{ki=0}^{Totaldocs} DuplicationLength_{ki}}{\sum_{Ki=0}^{totaldocs} LengthDoc\ k_i}}{TotalPatients}$$

In a nutshell, the three score describe different means to measure the amount of duplications. The global duplication score ($DupScoreGlobal$) evaluate the overall amount of duplication in the corpus (in term of number of characters in duplicated zones). The average duplication score per document evaluates the impact of duplication normalized by document. Finally, the average duplication per patient measures the overall impact of duplication per patients.

Potential Impact of Duplicated Zones on Two Use-Cases

We identified two use cases that could potentially be impacted by the presence of duplicated zones:

***Detecting drugs in duplicated zones*:** We searched for occurrences of medical drugs in our corpus. We used an exact match strategy, based on a list of ingredients and brand names from the Romedi [10] resource. Romedi is a semantic web version of a French public resource of drugs made available by the French National Health Insurance. Molecules such as simple sugars (e.g., glucose), water, inorganic elements (e.g., calcium), and so forth are listed as ingredients in Romedi. However, when mentioned in the clinical narratives, these molecules rarely refer to clinical drugs. Therefore, we eliminate them from the list of drugs identified by Romedi (more precisely, we eliminated the French terms *para, olivier, alcool, sodium, potassium, calcium, glucose, magnesium, eau*). All drugs were normalized to their corresponding CUI.

We identified the number of drugs present *only* in duplicated zones, and not in the rest of the document. While the presence can be useful for the medical history and for the care of the patient, the presence in portions of text duplicated from former documents could impact machine learning models, or information retrieval processes.

***Relative dates in duplicated zones*:** Our second use case focused on temporality. One major issue when working with text is the identification of the temporality associated with the concepts identified in the text. It is always important to distinguish between events or phenotypes that occurred during or prior the encounter. We searched the duplicated zones for temporality markers using relative dates (i.e. using expressions such as yesterday, two months ago, tomorrow, today, etc.). In such cases, the reference date is assumed to be the date of the creation of the document, but because the expression is located in a duplicated zone, its actual reference date should be identified in the past. We searched the corpus for a series of 8 terms corresponding to relative dates and determined if the terms were located within a duplicated zone.

Implementation of the Pipeline of Detection

We leveraged NextFlow [11] and Docker [12]. Each portion of our pipeline uses a Docker container and Nextflow ensure the parallelization of our processes. The pipeline ran on an Ubuntu 14.04 server, with 15 cores, 64 GB of RAM, and was developed in Python 3.10. Code is accessible on our github repository: https://github.com/equipe22/duplicatedZoneInClinicalText [13].

Results

Preprocessing

A mean values of 1670 characters were eliminated in general during the preprocessing. Overall, the number of character decreased by 36%. The average length of a text before preprocessing was 4145 and 2474 characters after.

Efficient Detection of Duplications

We compared the execution time and performance with respect to the overlap for different sets of parameters of the detection duplication algorithm (see Table 1).

Table 1 – Result of parameters evaluation for 50 patients which have 30 documents in average

fingerprint length	orf size	execution time (second)	% median overlap with the baseline
20	3	653	83
	5	196	77
	7	85	69
	10	34	72
	15	**18**	**67**
	20	10	66
30	3	665	84
	5	274	82
	7	125	80
	10	46	79
	15	**22**	**78**
	20	14	75
40	3	1043	83
	5	395	81
	7	166	80
	10	63	81
	15	**28**	**78**
	20	16	72

Computing Duplicated Zones on the Corpus of 650,000+ documents.

Table 2 –Duplication detection and annotations execution time

	fingerprint generation	merge	drug annotation	time annotation
execution time	3h19	20h19	80 s	23 s

Figure 2– Distribution of the patient duplication score

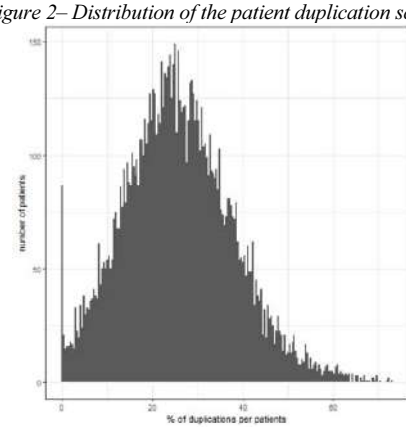


Table 3– Summary of duplication score

score	mean	Standard deviation
Global	0.33	0.33
Avg per document	0.25	0.12
Avg per patient	0.28	0.14

Potential Impact of Duplicated Zones on Two Use-Cases

***Detecting drugs in duplicated zones*:** We extracted 2,689,998 brand name and 761,611 ingredients from the corpus. 330,272 documents contain at least one drug mention. Overall, 161,067 documents had a drug detected within a duplicated region. 130,233 documents had at least one drug detected only within the duplicated zone (19.64% of our corpus).

Relative dates in duplicated zones: 45,557 documents contained at least one mention of a relative date. 9,632 documents contained a mention of a relative date within a duplicated region (21% of relative dates, 1.45% of the corpus).

Discussion

Detection of Duplication

We found that fingerprints length did not have an impact on the algorithm speed neither for the generations of fingerprints, nor the identification of duplications. The offset size did have a strong effect on both the execution time and the quality of detection. Compared to the baseline (offset of 1), the lower the offset size is, the better is the quality. However, in the spirit of a scalable approach, the processing time is incompatible with high volume of documents. We selected an offset of 15 for a fingerprint size of 30 for the reminder of our process to preserve a good quality while benefiting from a 200-fold speed improvement of the algorithm.

Filtering: We observed a large number of small-sized duplicated zones of 30 characters (more than 200 million detected duplications). 30 characters are highly unlikely to correspond to a full sentence in French. We decided to use a threshold of 1.5 fingerprints (i.e. 45 characters) to reduce the impact of artefacts that are unlikely to have been generated by a copy/paste process. Using this threshold, we found 29 million detected duplication.

Duplicated zones: Overall, the ratio of duplicated rate of duplications is 33%, in par with findings from the literature [3]. 20% of document had drugs mentioned only in duplicated zones. 1.45% of the document contained a relative data present in a duplicated zone. While the number is relatively low, the global number of documents is high: several thousands of documents for CDW with 10 million documents. The risk of misinterpretation of relative dates is high; tools such as HeidelTime [14] often used to identify mentions of temporality could provide erroneous normalization of the date since the tool would use the date of the document as a reference (instead of the data of the document source of the duplication).

Technical Significance

The performance of our heuristic allows treating a large amount of text. In this study, we managed a corpus of more than 650,000 documents within less than a day. Our CDW hosts a total of 10 million clinical narratives, some of which are the seldom report in the patient record.

The heuristic approach probably underestimates the volume of duplications. Additional fined grained approaches [3] could be applied to refine our results. We applied our approach to French, but the algorithm could be used for other languages as well.

Significance for secondary use of clinical data

The overall rate of duplicated zone (33%) is reasonable. However, we identified both drugs and relative dates were present in duplicated zones and could have a strong impact on information extractions from the text.

Duplications can have various meanings. The physician can use copy/paste to summarize the past, or to carry medical history from one document to another. Our method does not allow to identify the meaning associated with the copy/paste. However, for any application in which temporality is of importance, relative dates in duplicated zone might present an issue.

Limitation

The HEGP is specialized in oncology and cardiovascular diseases. Our selection of patients did not reflect the variety of the case present in the hospital. However, we did not filter the documents to specific sets of providers. In chronic diseases, with longer follow-ups, it would be possible for the ratio of duplication to be higher.

We used a rule based approach to clean-out the administrative sections of the document. This approach is not transposable, but proved efficient. The structure of the document is highly linked to the EHR used, the adoption of standard, etc.

We did not consider inter-patient duplications. Whereas our method could be used similarly to detect duplications among documents from different patients, it was not the purpose of our study. The detection of such zones could be interesting for quality control, or to reduce the work when annotating large corpora of texts for example.

Perspectives

Evolution of the volume of duplication over time: Because of the large variety of profiles, it is too complex to provide a good indicator of the evolution of the duplication rate over time. In our corpus, the documents were generated by many providers (medical services). We explored visually this question by representing the duplication rate over time. For comparison purpose, we normalized the time. Figure 3 provides a visualization of the duplication rate over the documents (the 0 in abscissa corresponding to the first document, and 100 to the last). A single point represents the rate of duplication of a single document for a single patient. We can see that there is visually a small trend toward an increase in the rate of duplications in the early part of the distribution, followed up by a plateau.

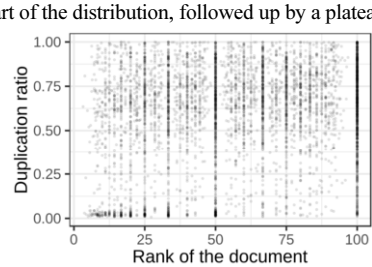


Figure 3 – Duplication representation over time per provider

We explored a visualization of the duplications within a patient records, and their organization over time in Figure 4. We leveraged the circlize [15] visualization to build a graphical summary of the duplicated zone, and their origin for a given patient. The outside circle represents the document of patients. The inner circle represents the provider of the document. Each edge represents a duplicated zone between documents. The date difference between two documents is rendered by the color; darker arcs correspond to the larger number of days than lighter arcs. In our example, Patient 1 and 2 both have 30 documents. The two patients have two distinct pathologies, and therefore two sets of distinct providers. The arcs reflect different hospitalization trajectories.

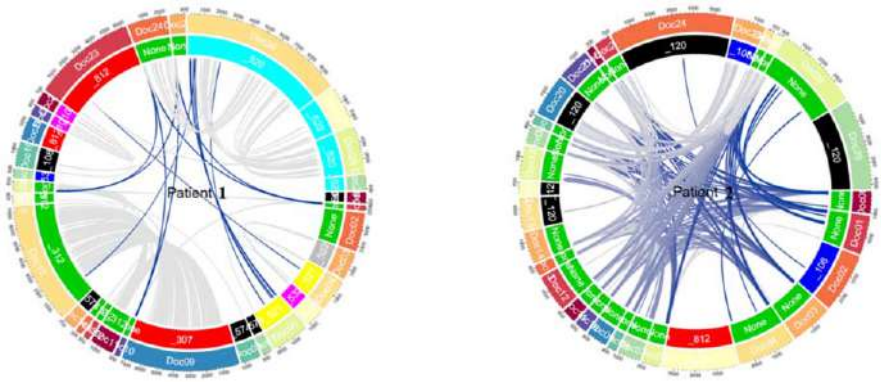


Figure 4. Duplication representation over time for two distinct patients. The length of a document reflects its number of characters, an arc between two regions translate the duplication of a portion of text.

For Patient 1, the providers reflect an oncology trajectory: digestive surgery (502, 532), imaging (312) and chemotherapy (574). For Patient 2, providers are coherent with urgent care: Internal medicine (812) and emergency medical (108) and reanimation. The systematic identification and annotation of duplicated zones are important for many aspects of data reuse. While we limited our exploration to drugs and relative dates, other semantic areas would be relevant to explore. For example, procedures and phenotypes. The annotation of duplicated zones could help identify procedures that are not relevant to the current visit

Conclusions

We developed a method to identify efficiently duplicated zones in clinical narratives. We explored a corpus of more than 650,000 documents belonging to 10376 patients. We identified an average rate of duplication of 33%, in par with value found in other studies. We evaluated the potential impact of duplications in two use-cases, the identification of drugs and the identification of relative dates. We found that 20% of the document contained drugs mentioned only in duplicated zones and that 1.45% of the document contained mentions of relative dates in duplicated zone, that could potentially lead to erroneous interpretation. We suggest the systematic identification and annotation of duplicated zones in clinical narratives for information extraction and temporal-oriented tasks.

Acknowledgements

WD is funded by the ANR PractikPharma. BR is funded in part by the SIRIC CARPEM research program.

Reference

[1] S.N. Murphy, G. Weber, M. Mendis, V. Gainer, H.C. Chueh, S. Churchill, and I. Kohane, Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2), *Journal of the American Medical Informatics Association*. **17** (2010) 124–130. doi:10.1136/jamia.2009.000893.

[2] J.-B. Escudié, B. Rance, G. Malamut, S. Khater, A. Burgun, C. Cellier, and A.-S. Jannot, A novel data-driven workflow combining literature and electronic health records to estimate comorbidities burden for a specific disease: a case study on autoimmune comorbidities in patients with celiac disease, *BMC Med Inform Decis Mak*. **17** (2017). doi:10.1186/s12911-017-0537-y.

[3] E. D'hondt, X. Tannier, and A. Névéol, Redundancy in French Electronic Health Records: A preliminary study, in: Association for Computational Linguistics, 2015: pp. 21–30. doi:10.18653/v1/W15-2603.

[4] J.O. Wrenn, D.M. Stein, S. Bakken, and P.D. Stetson, Quantifying clinical narrative redundancy in an electronic health record, *Journal of the American Medical Informatics Association*. **17** (2010) 49–53. doi:10.1197/jamia.M3390.

[5] R. Zhang, S. Pakhomov, B.T. McInnes, and G.B. Melton, Evaluating measures of redundancy in clinical texts, *AMIA Annu Symp Proc*. **2011** (2011) 1612–1620.

[6] R. Cohen, M. Elhadad, and N. Elhadad, Redundancy in electronic health record corpora: analysis, impact on text mining performance and mitigation strategies, *BMC Bioinformatics*. **14** (2013) 10. doi:10.1186/1471-2105-14-10.

[7] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res*. **25** (1997) 3389–3402. doi:10.1093/nar/25.17.3389.

[8] R.A. Gabriel, T.-T. Kuo, J. McAuley, and C.-N. Hsu, Identifying and characterizing highly similar notes in big clinical note datasets, *Journal of Biomedical Informatics*. **82** (2018) 63–69. doi:10.1016/j.jbi.2018.04.009.

[9] E. Zapletal, N. Rodon, N. Grabar, and P. Degoulet, Methodology of integration of a clinical data warehouse with a clinical information system: the HEGP case, *Stud Health Technol Inform*. **160** (2010) 193–197.

[10] ROMEDI - Référentiel Ouvert du Médicament, (n.d.). http://csfbtp.fr/telechargement.html (accessed September 10, 2018).

[11] Paolo Di Tommaso, Maria Chatzou, Pablo Prieto, Emilio Palumbo, and Cedric Notredame, Nextflow; A tool for deploying reproducible computational pipelines , (2015). doi:10.7490/f1000research.1110183.1.

[12] Docker, Inc, Docker - Build, Ship, and Run Any App, Anywhere, (n.d.). https://www.docker.com/.

[13] equipe22/duplicatedZoneInClinicalText, *GitHub*. (n.d.). https://github.com/equipe22/duplicatedZoneInClinicalText (accessed November 25, 2018).

[14] J. Strötgen, and M. Gertz, Multilingual and cross-domain temporal tagging, *Language Resources and Evaluation*. **47** (2013) 269–298. doi:10.1007/s10579-012-9179-y.

[15] Z. Gu, circlize: Circular Visualization, 2018. https://CRAN.R-project.org/package=circlize (accessed November 4, 2018).

Address for correspondence

Bastien Rance, Email: bastien.rance@aphp.fr

European Hospital Georges Pompidou, 20 rue Leblanc, 75015 Paris France. Tel: +331.56.09.59.85

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190193

Aggregation and Visualization of Laboratory Data by Using Ontological Tools Based on LOINC and SNOMED CT

Cora Drenkhahn[a], Petra Duhm-Harbeck[b], Josef Ingenerf[a,b]

[a] Institute of Medical Informatics (IMI), University of Lübeck, Lübeck, Germany,
[b] IT Center for Clinical Research, Lübeck (ITCR-L) University of Lübeck, Lübeck, Germany

Abstract

With the proliferation of digital communication in healthcare, the reuse of laboratory test data entails valuable insights into clinical and scientific issues, basically enabled by semantic standardization using the LOINC coding system. In order to extend the currently limited potential for analysis, which is mainly caused by structural peculiarities of LOINC, an algorithmic transformation of relevant content into an OWL ontology was performed, which includes LOINC Terms, Parts and Hierarchies. For extending analysis capabilities, the comprehensive SNOMED CT ontology is added by transferring its contents and the recently published LOINC-related mapping data into OWL ontologies.

These formalizations offer rich, computer-processable content and allow to infer additional structures and relationships, especially when used together. Consequently, various reutilizations are facilitated; an application demonstrating the dynamic visualization of fractional hierarchy structures for user-supplied laboratory data was already implemented. By providing element-wise aggregation via superclasses, an adaptable, graph representation is obtained for studying categorizations.

Keywords:

Biomedical Ontologies, LOINC, SNOMED CT

Introduction

As a prerequisite for any reuse of clinical data, it is essential to communicate in a semantic interoperable way so that the content's meaning is preserved independently from locally used terms and external influences like the applied software system. This is achieved by coding information explicit with a standardized terminology. The most widespread coding scheme in the laboratory domain is the *Logical Observation Identifier Names and Codes* (LOINC) terminology [1]. Each laboratory test is assigned a unique code representing its characteristics through the values of at least five axes. By using mapping methods, prepared and integrated laboratory data can provide valuable insights into patient care [2; 3].

However, activities to extend identifying LOINC codes by hierarchical structures or formal specifications have only recently begun, so that so far their use is limited to the unambiguous identification of laboratory tests with language-independent codes. Repeatedly it was demanded that LOINC codes should be enriched with a hierarchical structure., e.g. for aggregating reportable diseases in public health [4] or for improving semantic interoperability of LOINC-coded data [5].

So, the project aimed to transform the given LOINC terminology contents into a different, computer-processable representation subsequently enabling a larger scale of aggregation, analysis, and visualization of coded laboratory data.

Therefore, relevant LOINC components are converted into an ontology based on the well-known Web Ontology Language (OWL), i.e. a knowledge representation often defined as an "explicit specification of a conceptualization" [6], which uses formal logic to infer reasoner-based conclusions.

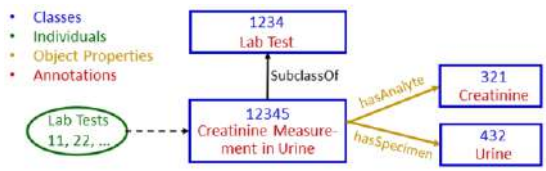


Figure 1 - Formalization of lab test classes and individuals

In order to enrich the given LOINC contents with additional information and hierarchical structures, the extensive *Systematized Nomenclature of Medicine – Clinical Terms* (SNOMED CT) shall be incorporated into the resulting ontology, as well. By using formal definitions throughout its scope, SNOMED CT is already specified as an ontology and can thus be integrated easily.

But to establish the connections between equivalent or otherwise related concepts in LOINC and SNOMED CT another source of data is needed. This is provided by the recently released official mapping between both terminologies, resulting from the publishers' cooperation agreement in 2013 [7]. Generating an OWL ontology for these mapping statements completes the first step, shown in figure 2, as a starting point for their integration and use.

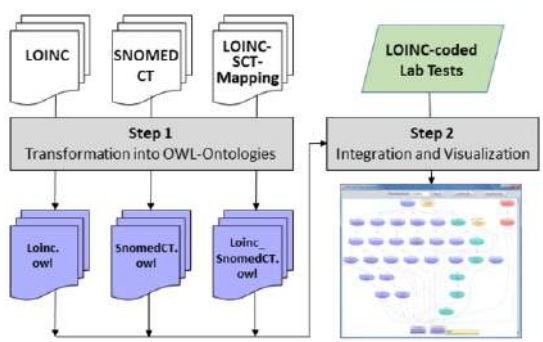


Figure 2 - LOINC aggregation and visualization tool: outline

The authors are aware that there are ontological challenges in the representation of laboratory tests, especially in connection with result values [8]. In this paper, purpose-specific LOINC

Hierarchies based on hierarchized LOINC *Part* terms (e.g. *Multi-Axial Hierarchy*), as well as logically formalized hierarchies, shall be provided via LOINC/SNOMED CT mapping in an integrated way to classify lab tests as LOINC-coded individuals per reasoner. We are convinced that LOINC- and SNOMED CT-based hierarchies serve different purposes, which complement each other but cannot replace each other [9]. This is supported by Vreeman speaking from "clinically-relevant aggregations" within LOINC by using multi-axial hierarchies and the newly introduced LOINC *Groups* [10].

Coming back to figure 2, this work concentrates on technical aspects for making use of OWL reasoners in order to integrate the three mentioned formalizations in step 2 and visualize LOINC-coded lab tests as individuals with respect to the integrated hierarchies. While there have already been attempts to build an ontology for LOINC [11; 12], this approach includes a larger scale of contents. Here, the main focus is laid onto arranging LOINC tests in hierarchical structures based on both explicit subclass relations as well as on their implicit features inferred by the LOINC/SNOMED CT-mapping.

Methods

LOINC resources

As the starting point for the project, the latest LOINC version 2.64 was utilized, published by the Regenstrief Institute in June 2018. This release contains a total of 87,863 LOINC *Terms* representing distinct laboratory tests or other clinical observations. Each *Term* is defined by a number of atomic components within. These LOINC *Parts* can be identified by their unique LP-Code and correspond to LOINC's main axes or their subcomponents describing the test's properties in detail. An example of usage is shown in figure 3. All in all, there are currently 52,000 *Parts* available which are used in about 900,000 relationships to define LOINC *Terms*.

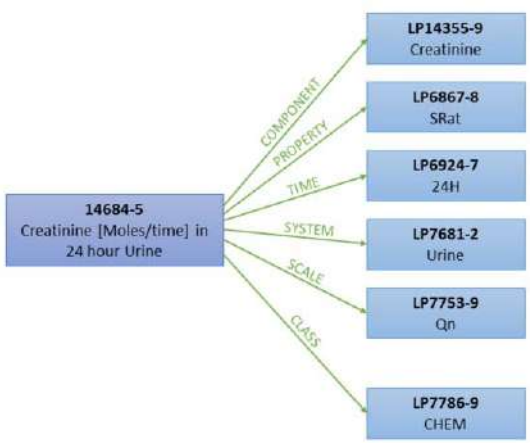


Figure 3 - Interrelation between a LOINC Term (left) and its definition by a unique combination of LOINC Parts (right).

To include even more contents into the transformation process we accessed the database containing background information within RELMA 6.23, a software offered by the Regenstrief Institute to enable searching in and mapping to LOINC. In these database tables, the information mentioned before can be found as well as more hierarchical structures. After further investigation a number of 10 so-called *Part Hierarchies* could be identified, each referencing a different aspect of laboratory tests. Even though the *Hierarchies* are built using LOINC *Parts* they are made to arrange complete LOINC *Terms* in a hierarchical structure based on their properties. As a special case the *Multi-Axial Hierarchy* is formed and made public, which uses composite LOINC *Parts* to structure *Terms* by their properties of two or more axes [13].

Converting LOINC to OWL ontology

In order to create an OWL ontology for the LOINC coding system in an automated and reusable way, an algorithmic approach was implemented using *Java*. For parsing all source files, present or previously transformed in CSV format, the Java CSV library is incorporated. Subsequent construction of OWL aspects is conducted by means of the OWL API [14].

```
Class: <http://imi.uni-luebeck.de/loinc#14684-5>

  Annotations:
    rdfs:comment "Creatinine;SRat;24H;Urine;Qn",
    rdfs:label "Creatinine [Moles/time] in 24 hour Urine"

  EquivalentTo:
    <http://imi.uni-luebeck.de/loinc#L_0>
     and (<.../loinc#hasClass> some <.../loinc#LP7786-9>)
     and (<.../loinc #hasComponent> some <.../loinc#LP14355-9>)
     and (<.../loinc #hasProperty> some <.../loinc#LP6867-8>)
     and (<.../loinc #hasScale> some <.../loinc#LP7753-9>)
     and (<.../loinc #hasSystem> some <.../loinc#LP7681-2>)
     and (<.../loinc #hasTime> some <.../loinc#LP6924-7>)
```

Figure 4 - OWL definition in Manchester Syntax for the LOINC Term shown in figure 3.

LOINC *Terms* are represented as distinct OWL classes with their code as an identifier, complemented by human-readable descriptions as label and comment. To model fully-defined LOINC *Terms* on the detailed level, they receive an equivalence statement combining all related *Parts* into one axiom. Here, an additionally defined superclass L_0, referring to LOINC Tests in general, is used as a starting point. Then, each LOINC *Part* is integrated as a specifically defined OWL class and added via an *ObjectProperty* derived from the *Part's* category. An exemplary result of this OWL translation is given in figure 4. For some LOINC *Terms,* not all required *Parts* are stated so that the combined definition is represented as subclass axiom, limiting logical inferences later on.

```
Class: <http://imi.uni-luebeck.de/loinc#LH56020>

  Annotations:
    rdfs:label "Urine Test"

  EquivalentTo:
    <http://imi.uni-luebeck.de/loinc#L_0>
     and (<.../loinc#hasSystem> some <.../loinc#LP7681-2>)

  SubClassOf:
    <http://imi.uni-luebeck.de/loinc#LH647>
```

Figure 5 - OWL concept of a hierarchy element describing all tests measured in the specimen "urine". The partially identical definition to figure 4 shall be noted.

For the ontological representation of LOINC *Hierarchies* each of their elements found as tree nodes in the source files is translated into a separate OWL class. The respectively specified parent node is referenced in a subclass axiom, building the hierarchy's backbone structure. In order to add its conceptual meaning, each element is furthermore defined using the LOINC *Part* associated with the hierarchy node in question. So, an equivalence axiom is constructed from the conjunction of the lab test basic class, the appropriate *Object-Property* and the *Part's* OWL class, as shown in figure 5. By using this kind of definition for hierarchy elements, the similarity to those used for LOINC *Terms* leads to logically inferred subsumptions of test terms into hierarchical classes.

But as there are some divergent structures in the LOINC *Part Hierarchies*, these implicit definitions can't be used throughout. Regarding both the *Component* and the *Multi-Axial Hierarchy* duplicate uses of the same LOINC *Part* can be found, causing equivalent definitions of hierarchy elements by the previously described approach. To avoid false inferences, the affected *Hierarchies* are therefore limited to primitive definitions. LOINC *Terms* are thereupon subsumed explicitly by the means of subclass axioms.

SNOMED CT resources

To include SNOMED CT contents, both the source files of the terminology in itself as well as those specifying the mapping to LOINC are needed. For the latter, we could obtain the latest version of the LOINC/SNOMED CT Cooperative package, published as Production Release in July 2017. According to the version used in these mapping files, we decided to utilize the International release 20170731 of SNOMED CT.

As the largest terminologies in medicine, SNOMED CT covers a much wider range of application than LOINC. Though laboratory tests and observations are mainly described by SNOMED CT *Procedure* and *Observable Entity* concepts, many other categories can be applied to define individual aspects of lab data as well, especially by combining them into *Post-coordinated Expressions*. Because of this, the complete SNOMED CT ontology shall be taken into consideration.

The cooperative package includes two main types of relations between LOINC and SNOMED CT. Firstly, entire LOINC *Terms* are mapped to combined SNOMED CT post-coordinated expressions, using the *Observable Entity* concept as a basic class. Any other components required to describe the LOINC *Term* are added as pairs of attributes and concepts, adhering to a compositional model. Secondly, the elementary LOINC *Parts* are associated with corresponding SNOMED CT components, once again built from attribute-concept-pairs.

Convert SNOMED CT to OWL ontology

For the transformation of SNOMED CT into an OWL ontology a Perl script is provided by SNOMED International. It was applied according to its instruction, converting given RF2 files into the OWL XML/RDF format.

Regarding the ontology generation for the mapping contents, the same procedure as for the LOINC terminology could be utilized. So, another Java algorithm using the OWL API was implemented.

```
Class: <http://imi.uni-luebeck.de/loinc#14684-5>

  EquivalentTo:
    <http://snomed.info/id/363787002>
    and (<http://snomed.info/id/370132008> some <http://snomed.info/id/30766002>)
    and (<http://snomed.info/id/704318007> some <http://snomed.info/id/118562009>)
    ...
```

Figure 6 - Excerpt of the equivalence definition for the previous LOINC Term mapped to SNOMED CT components.

Relationships of LOINC *Terms* to SNOMED CT post-coordinated expressions are hereby represented in the same way already used for their definition by combined LOINC *Parts*, as shown in figure 6. Where applicable, another equivalence or subclass statement is added to the OWL class of the LOINC *Term* accordingly, composed of the conjunction of all required SNOMED CT attributes and concepts.

The ontological representation of LOINC *Part* mappings turned out to be more difficult, resulting partly from the given unbalanced definitions, which refer to single LOINC *Parts* on one side but combined SNOMED CT expressions on the other. As a portion of each *Part's* meaning is implicitly included in its category, the corresponding *ObjectProperty* is used to form a combined expression for the LOINC *Part*. Additionally, both sides are complemented by the inclusion of the respective basic class (*LOINC Test* and *Observable Entity*) to ensure conformity with any other ontological components.

As a result of these adjustments, both terms needed for the part mapping transformation consist of complex class definitions, so that *General Concept Inclusions* (GCIs) have to be used for an appropriate OWL representation (see figure 7).

```
Class: "Observable Entity"
    and "Has Specimen" some "Urine specimen"

  EquivalentTo:
    "LOINC Test"
    and "hasSystem" some "Urine"
```

Figure 7 - Simplified OWL example of a SNOMED CT expression (above) mapped to a LOINC Part (below) using labels instead of identifiers.

Another problem occurred during the conversion of non-equivalent part mappings, with LOINC *Parts* being broader or narrower in content than the corresponding SNOMED CT expression. Due to the exactly specified compositional model of SNOMED CT, some LOINC *Parts* are mapped to a rather complicated composition of elements. In the given source files these relations were found to be not differentiated sufficiently, leading to imprecise definitions and thus false inferences. Because of this, only LOINC *Parts* with equivalent SNOMED CT expressions were considered for the envisioned prototype.

Visualization of created hierarchies

In order to demonstrate the resulting OWL ontologies' possibility of usage and to gain a graphical representation of the hierarchical structures inferred by them, another Java application was implemented afterward. It is based on the OWL API as well and uses an implementation (*ElkOwlApi*) of the ELK reasoner for computing inferences in addition [15]. Furthermore, the *JGraphX library* was chosen to provide graph drawing functionality. Accordingly, the complete graphical user interface was designed using Java *Swing*.

The application's input mostly consists of the three previously created OWL ontologies, each providing LOINC, SNOMED CT or mapping knowledge, which are then inferred conjointly by the ELK reasoner. Based on LOINC-coded laboratory individuals provided by the user, all relevant classes of these terminologies are determined. A recursive algorithm is hereby used to traverse the inferred hierarchies while collecting all visited superclasses. These are added as nodes to a tree-like graph structure, including their relationships as edges.

When the root element is reached, the completed graph is visualized on the GUI. A number of stylesheets are utilized to differentiate between elements by defining distinct representations according to source terminology or class type. Furthermore, some interactive functions are implemented in order to improve usability beyond basic operations. By folding and expanding subtrees, the graph's complexity can be reduced as needed and by highlighting a path's course becomes more concise. Finally, the implemented tool allows adding user-defined hierarchies that are transferred to a light-weight ontology which is as well included in the inference and visualization, shown as red elements on the right in figure 8.

Results

For the LOINC terminology, two different OWL ontologies were generated successfully: A full version and one comprised

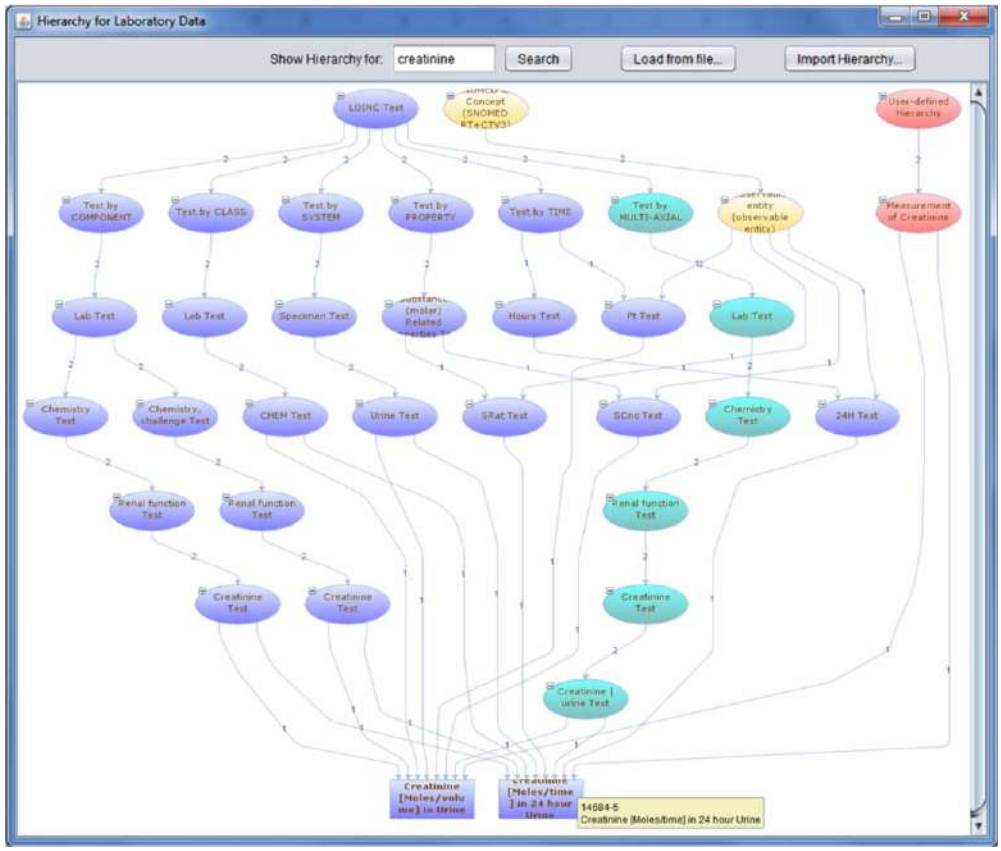


Figure 8 - Visualization of inferred hierarchies for the LOINC Terms 14684-5 and 14683-7 as leaf nodes. Each tree node represents a distinct OWL class, each edge denotes a subclass relation. The three root elements refer to the LOINC terminology (left), the SNOMED CT ontology (middle) and a user-defined hierarchy (right). The OWL ontology generated for the LOINC/SNOMED CT-Mapping does not include any named classes, so its contents is solely represented by edges.

of laboratory tests only. Both of these are divided into two main areas separating single 'LOINC Parts' from composite 'LOINC Tests'. The latter contains hereby all transformed hierarchy elements in their defined tree-structure as well as all LOINC *Terms*, which are classified into these hierarchies. A basic *Term* without subcomponents is typically included into six different hierarchies, one for each of its definitional *Parts* (except 'Scale') and the *Multi-Axial Hierarchy*.

For SNOMED CT an OWL representation could be generated easily by using the given script. The output contains the entire terminology contents keeping its extensive predefined relations and hierarchies. Another OWL ontology could be built for the cooperative package linking LOINC to SNOMED CT. Here, all of the LOINC *Term* mappings were transformed into OWL axioms using classes already defined in the other ontologies. In terms of the LOINC *Parts*, only equivalent mappings could be included as explained before, leaving one quarter of the enclosed relationships out of the result.

All of the created OWL ontologies can be applied as input in suitable software applications, either separately or combined. By importing them into the ontology editor *Protégé,* their contents and characteristics can be examined, see Table 1.

Additionally, the ELK reasoner was used to infer all ontologies, requiring about ten seconds in the case of their combined usage. Afterwards, both stated and inferred axioms were evaluated manually.

Table 1 - Extent of generated OWL ontologies

	LOINC (lab only)	**SNOMED CT**	**Mapping**
Axioms	1 004 634	1 520 034	62 155
Classes	277 159	335 225	34 272
ObjectProperties	16	97	25
AnnotationProp.	2	6	0
SubClassOf	366 292	253 406	2 355
EquivalentClasses	62 306	81 818	25 503
GCIs	0	0	5 969

No conflicts or contradictions could be found up to this point, concluding that the ontologies are consistent and valid. Regarding the inferred hierarchical order, the LOINC *Hierarchies* were found to yield a large amount of structuring information. This becomes particularly apparent through the graphs generated by the visualization application, as shown for two exemplary LOINC *Terms* in figure 8. Most of the pictured nodes refer to OWL classes generated based on LOINC contents. The same applies to the edges depicting subclass axioms, both stated and inferred.

By the inclusion of SNOMED CT knowledge and the cooperative mapping further hierarchical deductions are inferred into the arrangement of LOINC *Terms*. These are computed by the reasoner in a multi-level process, typically based on one of the *Term's Parts*, its representation in SNOMED CT and the

subclass definitions therein. Though there is more background information used in the inferred hierarchies, the graph visualization includes only two SNOMED CT elements: The root concept and the basic *Observable Entity*. All other involved SNOMED CT components are defined as post-coordinated expressions and thus not as named classes, which could be represented as tree nodes. The visualization application allows the rendering of multiple instances of LOINC *Terms*. These are displayed collectively in one hierarchy graph, taking into account their interrelations and frequencies of occurrence. As a result, an aggregated analysis in different levels of granularity is facilitated for a set of LOINC-coded laboratory tests.

Discussion

Despite the heterogeneous structure and non-formal definitions of LOINC contents, an extensive OWL ontology could be created comprised of LOINC *Terms, Parts* and *Hierarchies*. By using this formalization new insights into the terminology's hidden information are granted, whereupon the already existing hierarchies appeared to yield a particularly large potential by structuring the otherwise unsorted LOINC *Terms* according to their characteristics.

Additionally, the integration of SNOMED CT contents adds even more knowledge and relationships, enabling a wider range of application. The tree-like graphs created based on a user-supplied input of LOINC codes of interest provide a clearly arranged hierarchical structure and thus new possibilities to evaluate laboratory data. In this representation LOINC *Terms* and hierarchy elements are already displayed comprehensively, whereas contents derived from SNOMED CT or the cooperative mapping is included in a more subtle way that requires further improvement in order to present meaningful information.

For all of the generated OWL ontologies and their inferences, a profound evaluation is needed to ensure validity and to specify statistical properties. Amongst others, it is planned to use this tool within the "LOINC-300"-activities within the BMBF-funded Medical Informatics Initiative in Germany [16].

Conclusions

In this project we could develop an approach to improve the reusability of LOINC-coded laboratory data by converting relevant terminology contents into OWL ontologies, hence facilitating advanced analysis based on formally-defined representations and logical inferences.

By including the otherwise only internally used LOINC *Hierarchies* as well as SNOMED CT knowledge, extensive structures could be formed that build novel hierarchical classifications for LOINC *Terms*. For a concrete visualization, an application could be implemented that creates tree-like hierarchy graphs for LOINC-coded lab data, thereby enabling individualized and aggregated evaluations.

Acknowledgements

We thank the Regenstrief Institute for providing the structured LOINC source data and for their support.

The research was supported in part by the BMBF Medical Informatics Initiative. For using SNOMED CT source data an affiliate development license was obtained from IHTSDO.

References

[1] C.J. McDonald, et al. Case, and P. Maloney, LOINC, a universal standard for identifying laboratory observations: a 5-year update, *Clin Chem* **49** (2003), 624-633.
[2] C. Zunner, et al., Mapping local laboratory interface terms to LOINC at a German university hospital using RELMA V.5: a semi-automated approach, *J Am Med Inform Assoc* **20** (2013), 293-297.
[3] M.F. Emons, Integrated patient data for optimal patient management: the value of laboratory data in quality improvement, *Clin Chem* **47** (2001), 1516-1520.
[4] S. Steindel, J.W. Loonsk, A. Sim, T.J. Doyle, R.S. Chapman, and S.L. Groseclose, Introduction of a hierarchy to LOINC to facilitate public health reporting, *Proc AMIA Symp* (2002), 737-741.
[5] M.C. Lin, D.J. Vreeman, et al., Auditing consistency and usefulness of LOINC use among three large institutions - using version spaces for grouping LOINC codes, *J Biomed Inform* **45** (2012), 658-666.
[6] N. Guarino, D. Oberle, and S. Staab, What is an ontology?, In: *Handbook on ontologies*, 2009, pp. 1-17.
[7] D. Vreeman, New Regenstrief and IHTSDO agreement to make EMRs more effective at improving health care, *LOINC News* (2013), see https://loinc.org/news/new-regenstrief-and-ihtsdo-agreement-to-make-emrs-more-effective-at-improving-health-care/ (accessed November 5, 2018)
[8] M. Mary, et al., Ontological Representation of Laboratory Test Observables: Challenges and Perspectives in the SNOMED CT Observable Entity Model Adoption, In: *AIME 2017, LNAI 10259*, 2017, pp. 14-23.
[9] J. Ingenerf and R. Linder, Assessing applicability of ontological principles to different types of biomedical vocabularies, *Methods Inf Med.* **48** (2009), 459-467.
[10] D.J. Vreeman, Grouping data coded with LOINC, see https:// danielvreeman.com/grouping-data-loinc/ (accessed November 5, 2018).
[11] M. Mary, L.F. Soualmia, and X. Gansel, Formalisation de la terminologie LOINC et évaluation de ses avantages pur la classification des tests de laboratoire, In: *28es Journées francophones d'Ingénierie des Connaissances IC 2017*, 2017, pp. 2-13.
[12] T. Adamusiak and O. Bodenreider, Quality assurance in LOINC using Description Logic, *AMIA Annu Symp Proc* **2012** (2012), 1099-1108.
[13] Regenstrief Institute, LOINC Multiaxial Hierarchy, see https:// loinc.org/downloads/accessory-files/ (accessed November 5, 2018).
[14] M. Horridge and S. Bechhofer, The OWL API: A Java API for OWL ontologies *Semantic Web* **2** (2011), 11-21.
[15] Y. Kazakov, M. Krötzsch, and F. Simančík, The Incredible ELK: From Polynomial Procedures to Efficient Reasoning with EL Ontologies, *Journal of Automated Reasoning* **53** (2014), 1-61.
[16] S.C. Semler, F. Wissing, and R. Heyder, German Medical Informatics Initiative, *Methods Inf Med* **57** (2018), e50-e56.

Address for correspondence

Cora Drenkhahn, Institute of Medical Informatics, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany

Email: c.drenkhahn@uni-luebeck.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190194

Compatible Data Models at Design Stage of Medical Information Systems: Leveraging Related Data Elements from the MDM Portal

Martin Dugas[a], Stefan Hegselmann[a], Sarah Riepenhausen[a], Philipp Neuhaus[a], Leonard Greulich[a], Alexandra Meidt[a], Julian Varghese[a]

[a]*Institute of Medical Informatics, University of Münster, Germany*

Abstract

Compatible data models are key for data integration. Data transformation after data collection has many limitations. Therefore compatible data structures should be addressed already during the design of information systems.

The portal of Medical Data Models (MDM), which contains 20.000+ models and 495.000+ data items, was enhanced with a web service to identify data elements, which are frequently collected together in real information systems. Using Apache Solr, a fast search functionality to identify those elements with semantic annotations was implemented.

This service was integrated into the metadata registry (MDR) component of MDM to make it available to the scientific community. It can be used to build intelligent data model editors, which suggest and import frequent data element definitions according to the current medical context.

Keywords:
Health Information Interoperability, Data collection, Metadata

Introduction

Integration of medical data from different sources is a common but complicated task. There are many use cases for data integration, for instance in medical research: comparison of data from different sites, comparison of new study results with published data sources or merging of routine data with study data. In routine care, data integration is needed for quality assurance, for example to transfer a site-specific data set into a central database, or clinical decision support, for instance to apply standardized decision rules on local data sets. This list is not complete. However, there are many reports about data integration issues [1].

In many cases, collected data from different sources is similar in several aspects (e.g. regarding disease domain, patient group, assessment, therapy, outcome), but from a data analysis point of view, similar is not good enough. Instead, data shall be compatible to enable merging of data sources and joint analysis. A common approach to address this problem is data transformation after data collection. Data warehousing is a typical example for this approach: It applies extract-transform-load (ETL) to generate an integrated data set from different sources. However, the transformation step has important limitations, such as information loss during aggregation or bias due to semantic differences of data elements.

Why are there so many similar but incompatible data sources in medicine? This is caused by the semantic richness of medical terminology. SNOMED CT [2] contains more than 300.000 non-synonymous terms. Even for a small data model like a case report form with only one sheet (e.g. 40 data elements), there is an *astronomical* number of different models. As a consequence, if two medical experts design a data model for a given medical topic independently, the probability for compatible models is extremely close to zero.

Table 1 presents an example for data elements in two information systems, which are similar but incompatible: Myocardial infarction and atrial fibrillation are only a subset of heart diseases, i.e. a patient with myocardial infarction or atrial fibrillation has heart disease, but a patient with heart disease has not necessarily myocardial infarction or atrial fibrillation. The value sets of pain level are not transformable between system 1 (5 levels) and 2 (4 levels). This similarity has major drawbacks: joint data analysis of those systems is very limited and a joint algorithm for decision support is not applicable.

Table 1– Similar data are problematic for data integration: Data from systems 1 and 2 are similar, but incompatible.

System 1	System 2
heart disease yes/no	myocardial infarction yes/no atrial fibrillation yes/no
pain level 0-1-2-3-4	pain level 0-1-2-3

The similarity of medical data models is a big challenge, but also an opportunity for standardization and thereby better compatibility. Data integration would be a lot easier if data models from different sources were more compatible, i.e. had a larger proportion of compatible data elements. Overall, it would be highly desirable to foster compatible data models at the design stage, because post-hoc transformations have limitations.

The Portal of Medical Data Models (MDM-Portal) [3] addresses this problem with an open access approach to foster best practice sharing and thereby promotes re-use of existing data models. Users can search, comment, upload and download data models. An integrated metadata registry (MDR) component provides search functionality at the data element level to identify most frequent data element definitions for a medical concept such as weight or potassium.

The objective of this work is to go one step further and answer the following question: For a given data element (or a set of data elements) - what other data elements are collected most frequently together with this data element in real information systems?

This list of related data elements can contribute to the design of new information systems. It supports completeness of data models by identification of data elements, which are frequently covered in many related data sources; and, importantly, it fosters compatibility of data models by re-use of data element definitions from existing data sources wherever possible.

A concrete example: What lab values (or any other medical characteristics) are frequently collected together with "bilirubin"? And what are the precise data element definitions of these related data elements?

This type of relatedness goes beyond terminological relationship (like atrial fibrillation is a subtype of heart disease) - it is about actual data elements used in real medical data sources.

Methods

Semantic annotation of data elements

Semantic annotation is needed to identify meaningful related data elements because item names can be ambiguous [4]. Figure 1 presents an example: an attribute named "size" could refer to body height, tumor size or shoe size. By assigning a semantic code – such as a UMLS [5] or SNOMED CT code – the medical concept of an item name can be specified and a reference to a more detailed description is available.

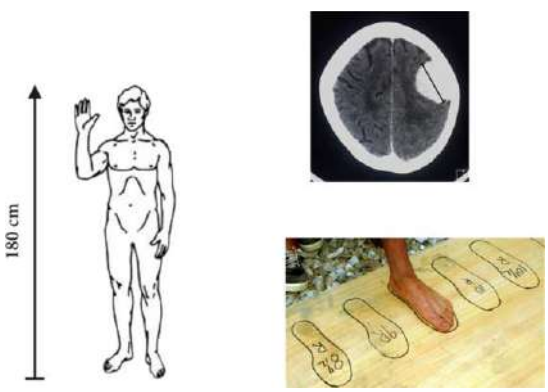


Figure 1– The need for semantic annotation: "size" is ambiguous, it can refer to body height, tumor size or shoe size.

Open access to medical data models

To identify related data elements, the MDM portal is used as a source of data models. Figure 2 presents the overall architecture of this system, which is described in more detail in [3,6]. As of March 2019, MDM contains more than 20,000 data models with approximately 495,000 semantically annotated data elements. To our knowledge, it constitutes the largest collection of medical data models in Europe. Each data element consists of name, textual description (which can be multilingual), data type, and semantic code. Optionally, a code list, value range or unit can be provided. These data elements are defined according to ISO 11179 [7], i.e. for each element, a concept domain and a value domain are specified. An MDR component provides search functionality and frequency counts based on data element name, data type, range, unit, values (code list) and semantic coding.

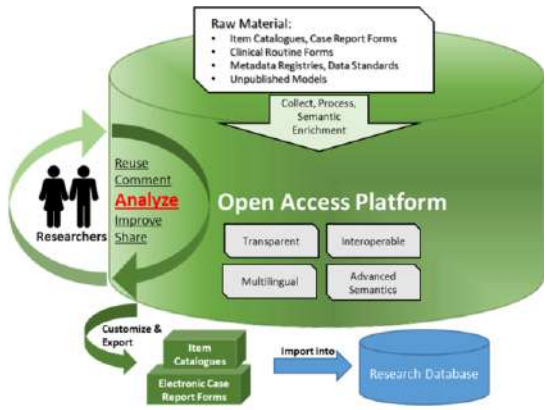


Figure 2– Overall approach of the MDM portal: an open access platform for medical data models

Relationship of data elements

Each data model consists of a list of context-specific data elements; therefore any two data elements can be defined as related, if these elements belong to the same data model. This means that these two data elements are collected together in the same documentation setting.

Results

Web service to identify related data elements

Apache Solr is used to index data elements in the MDM portal [8]. This open-source search platform is selected due to its fast full-text search and query language. A web service is implemented with the Java web framework Spring Boot. Java libraries SolrJ and JAXB are applied for the communication with Solr and to process the XML-format of ODM files within the MDM portal.

The overall architecture of the web service "related items" is depicted in figure 3. There are two variants of this service: Variant 1 identifies all data elements, which are jointly collected with a given data element (indicated by data element ID). Variant 2 identifies all data elements, which are jointly collected with a set of data elements with the same name (e.g. "age"). In both variants, results are ordered by frequency of co-occurrence.

The MDR component of MDM is enhanced with this web service using JavaScript to demonstrate the search functionality for related data elements. It is available at https://medical-data-models.org/ under actions / metadata registry (after free registration).

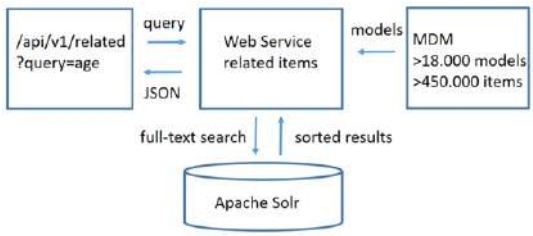


Figure 3– Architecture of the web service "related items". MDM items are indexed with Apache Solr. For a given query, jointly collected data elements are reported in JSON format, ordered by frequency of co-occurence.

Figure 4 presents the output of the MDR for search term "Bilirubin". The most frequent definition (n=30) for items named "Bilirubin" is data type float, unit mg/dl and UMLS code C0005437. By clicking on the green button "Search for related items" a list of related data elements, ordered by frequency, is displayed (figure 5). For instance, AST is on this list. This is plausible from a medical perspective, because AST and bilirubin are laboratory tests, which are frequently elevated in certain liver diseases.

Because the system architecture is web service based, this search functionality for related items can also be used by external systems.

MetaData Registry (MDR): Data Elements of MDM English

Item Name Bilirubin ↳ Search for related items

(e.g. weight, height, pulse, ASA, temperature, systolic blood pressure, age, potassium)

Showing 1 to 10 of 1,635 entries

Frequency	Data element	Data type	Range check	Unit	Values	Semantic Code	Models	
30	Bilirubin	FLOAT		mg/dL		C0005437		↳ Related
9	Bilirubin	FLOAT				C1278039		↳ Related
17	bilirubin measurement	INTEGER		mg/dl		C0344395		↳ Related
8	Bilirubin	INTEGER			Grade 0 (0) Grade 1 (1) Grade 2 (2) Grade 3 (3) Grade 4 (4) Grade 5 (5)			↳ Related

Figure 4– Search for items "Bilirubin" in the MDR component

MetaData Registry (MDR): Data Elements of MDM English

Item Name Bilirubin × Close related items

(e.g. weight, height, pulse, ASA, temperature, systolic blood pressure, age, potassium)

Showing 1 to 10 of 2,436 entries

Count	Data element	Data type	Range check	Unit	Values	Semantic Code	Models
30	AST	FLOAT		U/L		C0201899	
30	Leukocytes	INTEGER				C1271681	
30	Urea	FLOAT		mg/dl		C0523961	
30	Basophils percentage	FLOAT		%		C2237945	
30	Neutrophils	FLOAT		1000/l		C0200633	
30	Date and time sample taken	DATETIME				C0011008, C0040223, C0200345	
30	Sodium	FLOAT		mmol/L		C0337443	
30	Erythrocytes	FLOAT		Mio/l		C0014772	
30	Neutrophils percentage	FLOAT		%		C2238207	
30	Lymphocytes percentage	FLOAT		%		C2200256	

Figure 5– Related items to "Bilirubin": AST is a marker for liver disease like bilirubin

Discussion

Data from different sources should be as compatible as possible to facilitate data integration. Syntactic and semantic interoperability of Electronic Health Records (EHR) is a well-known challenge [9]. Definition of data elements is a key step in the design of medical information systems. A huge number of similar, but incompatible data models can be designed due to the semantic richness of medical terminology. For this reason, re-use of data element definitions from already existing information systems should be fostered. The MDR component of MDM supports this re-use with lists of data element definitions for a given item name.

This work goes one step beyond: It identifies data elements, which are collected frequently together with a given data element in real information systems.

The aspect of frequency is important to select most suitable elements for data collection: In a typical medical setting, data consumers want to analyze many data elements with high data quality, but data producers can provide only a limited number of data elements due to resource constraints. For example, what laboratory tests should be collected in a certain medical setting? Each additional lab value provides potentially interesting information, but is also associated with costs. The cost of data collection is a limiting factor for the number of patients, which can be documented.

With the new web service, designers of information systems can learn from already available data sources. If certain data elements were collected frequently together with a given data element, then those elements should also be considered for the new system. This facilitates joint data analysis in the future, in particular if the data element definitions from previous sources can be re-used without modifications.

Related data elements can provide a link between medical terminology and real information systems. The MDM portal provides a large collection of real data models from different sources. In contrast to terminology services, MDM contains information, which data elements are collected together. This goes beyond terminological relationships like "atrial fibrillation is a subtype of heart disease". MDM can answer the following questions: What data element definitions are available for "atrial fibrillation" and how frequent are these definitions being used? The new web service extends this functionality: What data elements are frequently collected together with the data element "atrial fibrillation"? The answer to this question provides the opportunity to harmonize and improve information systems already at the design stage.

At present, most IT frameworks for information systems provide its own data model editor. Relatively few of these editors can import external data model definitions. If it was easier to re-use data models from existing data sources, compatible data structures would be fostered and data integration would be easier. In addition, a lot of resources to maintain and use thousands of different data model editors could be saved.

A key idea of the new web service is to make data model editors more intelligent: with this web service, the editor can suggest new data elements, which are related to the current model context. For instance, suitable laboratory values for a given setting could be suggested.

There are many format options for data model definitions. In the field of clinical research, CDISC Operational Data Model (ODM) [10] is a growing standard and due to its endorsement by regulatory authorities more and more software solutions adopt it. For this reason MDM provides data models in ODM format, but also in many other formats like HL7 FHIR [11], Archetype Description Language as well as office and statistical formats.

There are many other approaches to foster compatible data models at the design stage of medical information systems. In particular, there are several initiatives to develop and consent common data elements (CDEs), such as the NIH CDE Repository [12]. Archetypes [13] are another major initiative to foster compatible data structures. Several libraries for standardized measures and instruments are being developed, such as PhenxToolkit [14] and ICHOM [15]. To our knowledge, none of these initiatives provide a service to identify related data elements from real data sources.

The current implementation of the web service for related data elements is a proof of concept and therefore has limitations, which should be addressed in future work:

1. It should be validated for a large number of data elements.
2. It relies on the contents of the MDM portal, which covers only a small subset of all medical data models.
3. It needs to be evaluated externally in the design of information systems, specifically regarding its contribution to the compatibility of data models.

Conclusions

Compatible data models should be implemented at the design stage of information systems. A novel web service to foster re-use of data element definitions is available for the scientific community. This service is based on the MDM portal and identifies data elements which are collected frequently together in real information systems.

Acknowledgements

Funding from the German Research Foundation (DFG Grant DU 352/11-2, DU 352/11-1) is acknowledged.

References

[1] R.L. Richesson, J. Krischer, Data standards in clinical research: gaps, overlaps, challenges and future directions, *J Am Med Inform Assoc*. **14** (2007), 687-96
[2] SNOMED CT, Systematized Nomenclature of Medicine Clinical Terms, available from http://www.snomed.org/ [accessed November 15, 2018].
[3] M. Dugas, P. Neuhaus, A. Meidt, J. Doods, M. Storck, P. Bruland, J. Varghese, Portal of medical data models: information infrastructure for medical research and healthcare. Database (2016), pii:bav121.
[4] M. Dugas, Missing semantic annotation in databases. The root cause for data integration and migration problems in information systems, *Methods Inf Med.* **53** (2014), 516-7.
[5] UMLS, available from http://www.nlm.nih.gov/research/umls/ [accessed November 23, 2018].
[6] J. Varghese, M. Fujarski, S. Hegselmann, P. Neuhaus, M. Dugas, CDEGenerator: an online platform to learn from existing data models to build model registries, *Clin Epidemiol*. **10** (2018), 961-970.
[7] International Organization for Standardization/International Electrotechnical Commission: ISO/IEC 11179,

Information Technology - Metadata Registries (MDR) - Part 1: Framework; 2004, page 11, available from http://standards.iso.org/ittf/PubliclyAvailableStandards/c035343_ISO_IEC_11179-1_2004(E).zip [Accessed November 23, 2018]
[8] S. Hegselmann, M. Storck, S. Geßner, P. Neuhaus, J. Varghese, M. Dugas, A Web Service to Suggest Semantic Codes Based on the MDM-Portal, Stud Health Technol Inform. **253** (2018),35-39.
[9] S. Garde, P. Knaup, E. Hovenga, S. Heard, Towards semantic interoperability for electronic health records, Methods Inf Med. **46(3)** (2007),332-43.
[10] Operational Data Model, available from https://www.cdisc.org/standards/data-exchange/odm [accessed November 23, 2018]
[11] HL7 FHIR, available from https://www.hl7.org/fhir/ [accessed November 23, 2018]
[12] NIH CDE Repository, available from https://cde.nlm.nih.gov/ [accessed November 23, 2018]
[13] openEHR, available from https://www.openehr.org/ [accessed November 23, 2018]
[14] C.M. Hamilton, L.C. Strader, J.G. Pratt, D. Maiese, T. Hendershot, R.K. Kwok, J.A. Hammond, W. Huggins, D. Jackman, H. Pan, D.S. Nettles, T.H. Beaty, L.A. Farrer, P. Kraft, M.L. Marazita, J.M. Ordovas, C.N. Pato, M.R. Spitz, D. Wagener, M. Williams, H.A. Junkins, W.R. Harlan, E.M. Ramos, J. Haines, The PhenX Toolkit: get the most from your measures, *Am J Epidemiol.* **174** (2011), 253-60.
[15] International Consortium for Health Outcome Measurement, available from https://www.ichom.org/ [accessed November 23, 2018]

Address for correspondence

Prof. Dr. Martin Dugas, dugas@uni-muenster.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190195

Word Embedding for French Natural Language in Healthcare: A Comparative Study

Emeric Dynomant[a,b,c], Romain Lelong[b,c], Badisse Dahamna[b,d], Clément Massonnaud[b], Gaëtan Kerdelhué[b,d], Julien Grosjean[b,d], Stéphane Canu[c], Stéfan Darmoni[b,d]

[a] *OmicX, 72 Rue de la République, 76140, Le Petit Quevilly, Normandie, France,*
[b] *Department of Biomedical Informatics, Cour Leschevin, CHU de Rouen, 1 Rue de Germont, 76031 Rouen, Normandie, France*
[c] *LITIS, Université de Rouen Normandie, Avenue de l'Université, 76800, Saint-Étienne-du-Rouvray, Normandie, France*
[d] *LIMICS, Campus des Cordeliers, INSERM U1142, 15 Rue de l'École de Médecine, 75006, Paris, France*

Abstract

Structuring raw medical documents with ontology mapping is now the next step for medical intelligence. Deep learning models take as input mathematically embedded information, such as encoded texts. To do so, word embedding methods can represent every word from a text as a fixed-length vector. A formal evaluation of three word embedding methods has been performed on raw medical documents. The data corresponds to more than 12M diverse documents produced in the Rouen hospital (drug prescriptions, discharge and surgery summaries, inter-services letters, etc.). Automatic and manual validation demonstrates that Word2Vec based on the skip-gram architecture had the best rate on three out of four accuracy tests. This model will now be used as the first layer of an AI-based semantic annotator.

Keywords:

Natural language processing, word processing, data mining.

Introduction

Context

The use of clinically derived data from electronic health records and other clinical information systems can greatly facilitate clinical research as well as optimizing diagnosis related groups, operational and quality initiatives. The main approach for making this data available is to incorporate the data from different sources into a joint health data warehouse that contains different kinds of natural language documents such as prescription, letters, surgery reports, etc. All documents are written in everyday language.

A Semantic Health Data Warehouse (SHDW) was developed by the Department of Biomedical Informatics of the Rouen University Hospital (RUH), Normandy, France. It is composed of three independent layers based on a NoSQL architecture: a cross-lingual terminology server, HeTOP, which contains 75 terminologies and ontologies in 32 languages [1]. Then, a semantic annotator based on Natural Language Processing (NLP) called *Multi-Terminological Concept Extractor* (ECMT) [2]. Finally, a semantic multilingual search engine [3].

To improve the ECMT, a new strategy using deep learning techniques was defined. To implement it, a new text representation had to be designed to fit the input of neural networks algorithms.

Word embedding

In NLP, a chosen representation has to keep the semantic similarities between different words from a corpus of texts. Thus, the representation of a unique token has to show its proximity to other related meaning concepts, as illustrated in the quotation "*You shall know a word by the company it keeps"* [4], now known as the *distributional hypothesis*.

In fact, a compact and precise representation of words could bring several benefits. First, computers are way better to perform operations on low-dimensional objects. Second, probabilities calculation or mathematical operations can be done on words, such as the famous "*(king – man) + woman ~ queen*". And finally, the vectors' dimensions created to represent a word can be used to fit this word in a space and thus make distance comparisons with other tokens.

Implementations

Word2Vec

The word2vec approach was the first modern embedding released in 2013 [5]. Mikolov *et al.* implemented two kinds of architectures. The Continuous Bag of Word (CBOW) architecture treats the entire context as a single observation. A hierarchical softmax was also used to reduce computational limits [6]. The input layer accepts one-hot encoding as input (a sentence is encoded as a very sparse vector composed of 0 or 1, depending on the words found in this sentence). The Skip-Gram (SG) architecture uses a sliding window to define "*context / target*" pairs (*e.g.*, "*How / you*" is the context of the word "*are"* in the sentence "*how are you?*"). The entire corpus **V** will thus be transformed into many pairs *context / target* (*i.e., input / output* of the network). To reduce the computation of such an amount of data (in a "normal" training situation, all the weights should be updated in each passing through an example), the authors brought some new tricks. Word pairs appearing always together are treated as single tokens, frequent words subsampling and negative sampling [7].

GloVe

This model was released by the Stanford University [8]. Like Word2vec, GloVe can embed words as mathematical vectors; however, it differs on the method used to capture similarity between words, GloVe being a count-based method. The idea was to construct a huge co-occurrence matrix of shape **V×C** with **V** being the vocabulary of the corpus and **C** context examples. The probability ($Vw_1 \| Vw_2$) of a word Vw_1 being close to another Vw_2 will increase during the training. This gigantic matrix is then factorized by using the log function.

FastText

FastText is a newly released model and comes from a new idea [9]. Bojanowski *et al.* consider that a word could be the result of all of the vector decomposition of this word (sub word model). Each word V_w can be decomposed into a set of *n*-grams vectors. For example, the word "*boat*" can be seen as a set of *n*-gram with $n = 3$ as [$b + bo + boa + o + oa + oat + a + at + t$]. Thus, each word is embedded in the vector space as the sum

of all vectors composing this token, incorporating morphological information into the representation. Like Word2Vec, FastText comes with the two different architectures (SG and CBOW).

Related work

For a few years, the huge interest in word embedding led to comparison studies. Scheepers, Gavves, and Kanoulas [10] compared the three word embedding methods presented here but the three models were trained on different datasets (Word2Vec on news data, while FastText and GloVe trained on more definitional data, Wikipedia and Common Crawl respectively). Bairong *et al.* [11] also performed a comparison between these three, but focused on bilingual automatic translation comparison (BLEU score [12]) and without human evaluation for all the different models. More recently, Beam *et al.* [13] produced huge publicly available word embeddings based on medical data, however they didn't compare FastText, only Word2Vec and GloVe. Finally, Wang *et al.* [14] compared word embedding training set influence on models utilization, and its impact on different NLP tasks related to medical applications.

Moreover, many different teams or companies have released pre-trained word embedding models (*e.g.,* Google, Stanford University, etc.). However, in a clinical context, the vocabulary coverage of those embeddings is quite low regarding the words used. Indeed, many misspells, acronyms of specific abbreviations are regularly found in the documents produced in the hospital. Thus, a local training on specific data is often needed, especially with languages other than English.

Contributions

The objective here is to compare these five different methods to obtain the best possible words embedding (Word2Vec SG and CBOW, GloVe, FasText SG and CBOW). This representation will then be used as the input of deep learning models constructed to improve the annotating phase actually performed by the ECMT in the SHDW. This NER phase will be the first step toward a multilingual and multi-terminologies concept extractor. The influence of the number of documents in the training set will also be assessed.

Methods

Corpus

The corpus used in this study is composed of health documents from the SHDW of the RUH, France. All these documents are in French. They are also quite heterogeneous but their type is stored in the SHDW: discharge summaries, surgery or procedure reports, drug prescriptions and letters from a general practitioner. All these documents are written by medical staff in the RUH.

Document de-identification

These documents were then de-identified to protect each identity of every patient or doctor from the RUH. Every of the first and last name stored in the RUH main databases were replaced by non-informative tokens such as <*doctor*>, <*firstname*> or <*lastname*>. Moreover, other tokens have been used such as <*email*> or <*date*>.

Pre-processing

Texts have been split into token lists, the data has been lowered (meaningless to make distinction), the punctuation was removed, and the numerical values were replaced by a meta-token <number>. We chose to not remove stop words, due to their negligible impact on the context. Indeed, their multiple apparitions in many different contexts will just create a cluster of stop words in the middle of the VSM.

Training

Models have been trained on a server powered by four XEON E7-8890 v3 and 1To of RAM located on the RUH. We based the tuning of the models' hyper-parameters on the literature and on our own experience [15]. All chosen values are listed on Table 1. The *minimum count* parameter was set higher than usual settings due to the large quantity of data in the training set.

Table 1— Hyperparameters Used to Train the Five Word Embedding Models

Parameter	Model	Value
epochs	All	100
Min. count	All	20
Window size	All	7
Learning rate	All	2.5×10^{-2}
Embedding size	All	80
Alpha rate	All	0.05
Negative sampling	Word2vec/FastText	12
Subsampling	GloVe	$1e^{-6}$

Evaluation

Cosine similarity

We compare how similar the embeddings for a pair of concepts are by computing the cosine similarity of their corresponding vectors, and then using this similarity to assess whether or not the two concepts are related.

We used two well-known validations set UMNSRS-Similarity and UMNSRS-Relatedness, containing 566 and 588 manually rated pairs of concepts respectively [16]. However, since our corpus was in French, we used the translated and aligned version of the MeSH (Medical Sub-Heading) terminology stored in the ECMT to translate these two sets. The result provides a number of 308 pairs for the UMNSRS-Sim and 317 for the UMNSRS-Rel.

Mathematical operations

Mikolov's paper presenting Word2Vec showed that mathematical operation on vectors such as additions or subtractions are possible. Mathematical operations covering a wide range of possible subjects found in the selected documents (hospital departments, human tissues, biology, drugs) were defined.

Odd one out

The odd one out similarity task tries to measure the model's accuracy by giving three different words to the model. Two of them are known as linked, not the third one. Then, the model has to output the word vector which does not cluster with the two others [17]. To create such a validation corpus, 53 pairs of concepts potentially linked in a medical text have been defined from the MeSH terminology. Then, 53 words appearing more than 1,000 times in the corpus have been randomly selected to be used as odd terms.

Human evaluation

A formal evaluation of the five methods was performed by a medical doctor (SJD). A list of 112 terms has been extracted from the Medical Sub Heading (MeSH) terminology, covering every possible branch. SJD then assessed the relevance of the top five closest word vectors returned for each of the 112 concepts by the five created models. Retrieved citations were assessed for relevance according to a three-point modality scale

used in other standard information retrieval test sets: bad (0), partial (1) or full relevance (2).

Training set influence

To go further, models are going to be trained twice. First by randomly selecting 5% of the total amount of available documents in the RUH (~600K) and all evaluation tasks were performed against the five models. Then, the entire corpus (~12M) was used as a training set, and those evaluation tasks were assigned to these newly trained models.

Results

Corpus

In total, 641,279 documents for the first phase and 11,762,100 for the second one had been extracted from the RUH. They had been de-identified and pre-processed. Regarding the vocabulary, 355,597 unique tokens are found in those 12M of documents. However, this number can be pondered with 170,433 words appearing less than 5 times in the entire corpus (mainly misspells, but also geographic locations or biological entities like genes, proteins, etc). In total, 50,066 distinct words are found more than 20 times in the corpus, thus present in the models (minimum count parameter set to 20). On average, each document contains 281.26 words (*sd* = 207.42).
These documents were decomposed using the Term-Frequency Inverse-Document-Frequency (TF-IDF) algorithm, results in a frequency matrix. Vectors have been used to clusterize those documents with a kMeans algorithm (number of classes **K = 5**). To visualize their distribution on two dimensions, the t-SNE algorithm has been used (figure 1).

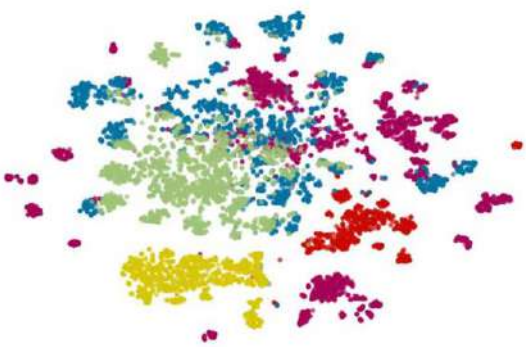


Figure 1— Two-Dimensional t-SNE Projection of 10,000 documents Randomly Selected Among Main Classes in the HDW

The five different colors correspond to the five types of documents selected (discharge summaries (green), surgery (blue) or procedure (purple) reports, drug prescriptions (yellow), letters from a general practitioner (red)).

Those main classes are well separated, the vocabulary itself contained in the documents from the HDW being sufficient to clusterize each type of text. However, discharge summaries, surgery or procedure reports are a bit more mixed because of the words used in these kind of contexts (short sentences, acronyms and abbreviations, highly technical vocabulary...). Regarding drug prescriptions and letters to a colleague or from a general practitioner, they present more specific vocabulary (drugs and chemicals, and current/sustained language respectively), involving more defined clusters for these two groups.

Training

Regarding the training time, models are very different. GloVe is the fastest algorithm to train with 18 min to process the entire corpus of 600K documents (Table 2).

Table 2— Algorithms Training Time (minutes) Regarding the Number of Documents

All time are given in minutes. GloVe is the fastest algorithm to train.

Algorithm	600K documents	12M documents
Word2Vec SG	182.0	497.9
Word2Vec CBOW	33.4	308.8
GloVe	**17.5**	**65.9**
FastText SG	1678.1	5573.8
FastText CBOW	1577.0	4974.0

GloVe performs much better in terms of computational time due to the way it handles the vocabulary. GloVe is stored as a huge co-occurrence matrix and thanks to its count-based method, which is not computationally heavy, it can be highly parallelized. It was expected that FastText would take a lot of time to train, due to the high number of word sub-vectors it creates. However, for Word2Vec, the difference between the two available sub-architectures is highly significant (33 min to 3h02 for 600K documents). This difference could come from the hierarchical soft-max and one-hot vector used by the CBOW architecture, which reduces the usage of the CPU. With SG, the minibatch parsing of all the context / target pairs highly increases the time to go through all possibilities.

Evaluation

Cosine similarity

The percentages of validated pairs from the UMNSRS datasets are presented in Table 3. FastText SG performed this task with the highest score (3.89% and 5.04% of valid pairs found for UMNSRS-Sim and UMNSRS-Rel respectively with 600K documents, 6.43% and 7.13% with 12M documents). The very low scores indicate that this kind of published dataset is useful to validate models trained on more academic texts, not those written in natural language. Some words will never be found because of the use of an acronym by health practitioners ("*HTA*" instead of "*HyperTension Artérielle*") or because of the informal form, mainly used in these kinds of documents.

Table 3— Percentage of Pairs Validated by the Five Trained Models on Two UMNSRS Evaluation Sets

	600K documents		12M documents	
	U-Sim	U-Rel	U-Sim	U-Rel
W_SG	2.92	4.10	4.73	6.92
W_CBOW	3.57	4.10	5.12	6.92
GloVe	1.29	0.94	1.35	0.94
F_SG	**3.89**	**5.04**	**6.43**	**7.13**
F_CBOW	**3.89**	3.79	**6.43**	6.65

Mathematical operations

A list of six mathematical operations has been defined with the help of a medical doctor and a university pharmacist (listed in Table 5). Each operation consists in verifying if *(term_1 – term_2) + term_3 ~ term_4* is true.

Table 4— Logical Operations on Words Having to be Retrieved with the Different Trained Models

Each is listed in English but has been performed against models in French.

1	(cardiology - heart) + lung ~ pneumology
2	(melanoma - skin) + gland ~ adenoma
3	(corpuscle - blood) + immune ~ immunoglobulin
4	(furosemide - kidney) + heart ~ fosinopril
5	(limb - lower) + upper ~ arm
6	(morphine - opioid) + antalgic ~ perfalgan

The number of validated operations performed by each model is presented on Table 4. Word2Vec SG gets the highest score on this task (5/6, regardless of the number of documents in the training set), while GloVe gets the lowest one (2/6).
Interestingly, no operation has been failed by the five models, indicating that none of them is simply not logical or just too hard to perform.

Table 5— Score for Mathematical Operation Tasks on Six Point Maximum for Each of the Five Trained Models

Algorithm	600K documents	12M documents
Word2Vec SG	**5**	**5**
Word2Vec CBOW	3	3
GloVe	2	3
FastText SG	3	4
FastText CBOW	3	3

Odd one similarity

Regarding the odd one similarity task, Word2Vec SG is the best so far with 65.38% (600K documents) and 75.5% (12M documents) of odd-one terms correctly found (Table 6).
Regarding the sub-architectures presented by both Word2Vec and FastText, the SG always performed better than the CBOW, possibly due to the negative sampling. In fact, the studied corpus is quite heterogeneous, and a word can be listed as items (*e.g.,* drugs in prescriptions) instead of being used in correct sentences. Sometimes, the complete update of vectors' dimensions generates non-sense in the models.

Table 6— Percentage of Odd One Tasks Performed by Each of the Five Trained Models

Algorithm	600K documents	12M documents
Word2Vec SG	**65.4**	**75.5**
Word2Vec CBOW	63.5	69.8
GloVe	18.5	39.6
FastText SG	44.4	45.8
FastText CBOW	40.7	41.3

Human Evaluation

The evaluation focused on 2800 terms (5 vectors × 112 MeSH concepts × 5 models), and was performed by two evaluators, CM and SJD, on models trained with 600K and 12M documents. First, the accordance between CM and SJD was assessed with a weighted kappa test (k = 0.6133). According to the literature, the agreement between the two evaluators can be considered as substantial [17]. This agreement can be retrieved in figure 2. The accord is stronger for the extreme scores (0 and 2) while the agreement about the middle score of 1 is least pronounced.
Moreover, to assess if human evaluators remained consistent regarding the cosine score computed by each model, we compared the average note given by the two evaluators with the average of the cosine distance computed for each model (table 7). Word2Vec with the SG architecture performed the highest score, regardless of the evaluator (1.469 with 600K documents and 1.420 with 12M). Interestingly, GloVe computes the shortest cosine distance in averages (0.884 on the 112 given concepts), while both evaluators gave it the lowest grade.

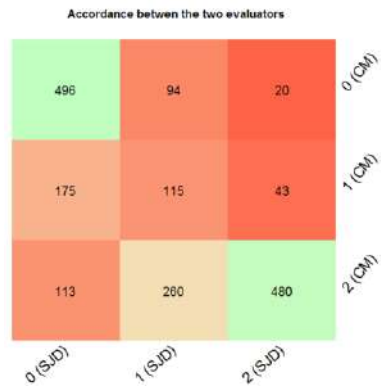


Figure 2— Global Representation of the Notation Accordance between the Two Evaluators (CM and SJD).

Notes attributed to a model output are going from 0 (bad matching) to 2 (good matching). Colors are ranging from light green (very similar) to red (completely different).

Table 7— Comparison between Cosine Distance Computed by each Model and the Human Evaluation Performed

Notes and distances are in averages on the top-5 closest vectors for 112 queries on every model.

Algorithm	600K documents		12M documents	
	Cos	Eval	Cos	Eval
Word2Vec SG	0.731	**1.469**	0.785	**1.420**
Word2Vec CBOW	0.776	1.215	0.716	1.33
GloVe	**0.884**	0.703	0.692	1.27
FastText SG	0.728	1.156	**0.963**	0.25
FastText CBOW	0.748	1.131	0.930	0.466

Discussion

In this study, the three most famous word embeddings have been compared. Word2Vec SG got the best score for three out of the four rated tasks (FastText SG is the best regarding the cosine one). These results are coherent with those obtained by Muneeb *et al.*, which compared Word2Vec and GloVe with the cosine similarity task [19]. GloVe had the worst grade, however it's the fastest to train so far. Regarding FastText, it is interest-

ing to note that the morphosyntactic similarities are kept in account in the vector space creation. Moreover, the sub-vector decomposition of words allows this kind of model to be queried by words absent from the original training corpus. We can imagine this model being used for orthographic correction or acronym disambiguation.
Interestingly, the size of the training set does not heavily influence Word2Vec (Table 7), but GloVe seems to improve its quality according to the human annotators. In fact, designed with a count-based method, this algorithm is directly affected by the amount of available data to train. FastText is highly degraded with more documents. The relevance of the returning vectors remained low because of the high proportion of morphosyntactic similarities between the sent and the returned tokens.
The corpus used as a training set comes from a real work environment. Finding a good evaluation for embedding produced in such a context is a hard task, and the performances shown by some models trained on scientific literature are often biased.
Future work will assess whether the best embedding method could help for semantic concept enrichment.

Conclusion

In our case, Word2Vec with the SG architecture got the best grade regarding three of the four rated tasks.
Any end user is now able to query the word embedding models produced on a dedicated web site as well as to download high quality dimension reduction images and test sets (URL: https://cispro.chu-rouen.fr/winter/) .

Acknowledgements

This work was partially granted by the Ph.D CIFRE number 2017/0625 from the French Ministry of Higher Education and Scientific Research and by the *OmicX* company (ED). The authors thanks Catherine Letord, pharmacist, Jean-Philippe Leroy, M.D. for their help in creating the test datasets.

References

[1] J. Grosjean, T. Merabti, B. Dahamna, I. Kergourlay, B. Thirion, L.F. Soualmia, and S.J. Darmoni, Health multi-terminology portal: a semantic added-value for patient safety, *Studies in Health Technology and Informatics*. **166** (2011) 129–138.
[2] S. Sakji, Q. Gicquel, S. Pereira, I. Kergourlay, D. Proux, S.J. Darmoni and M.H. Metzger, Evaluation of a French medical multi-terminology indexer for the manual annotation of natural language medical reports of healthcare-associated infections., *Studies in Health Technology and Informatics*. **160** (2010).
[3] R. Lelong, L. Soualmia, B. Dahamna, N. Griffon, and S.J. Darmoni, Querying EHRs with a Semantic and Entity-Oriented Query Language, *Studies in Health Technology and Informatics*. **235** (2017) 121–125.
[4] J.R. Firth, A Synopsis of Linguistic Theory, *Basil Blackwell Oxford*. **59** (1957) 168–205.
[5] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, and J. Dean, Distributed Representations of Words and Phrases and their Compositionality, *NIPS'13 Proceedings of the 26th International Conference on Neural Information Processing Systems*. **2** (2013) 3111–3119. doi:arXiv:1310.4546.
[6] A. Mnih, and G. Hinton, A Scalable Hierarchical Distributed Language Model, *Proceedings of the 21st International Conference on Neural Information Processing Systems*. **21** (2008) 1081–1088.
[7] M. Baroni, G. Dinu, and G. Kruszewski, Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors, *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. **1** (2014) 238–247. doi:10.3115/v1/P14-1023.
[8] J. Pennington, R. Socher, and C.D. Manning, GloVe: Global Vectors for Word Representation, *Conference: Conference: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. **14** (2014) 1532–1543. doi:10.3115/v1/D14-1162.
[9] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, Enriching Word Vectors with Subword Information, *Transactions of the Association of Computational Linguistics*. **5** (2017) 135–146. doi:arXiv:1607.04606.
[10] T. Scheepers, E. Gavves, and E. Kanoulas, Analyzing the compositional properties of word embeddings, *University of Amsterdam*. (2017).
[11] Z. Bairong, W. Wenbo, L. Zhiyu, Z. Chonghui, and T. Shinozaki, Comparative Analysis of Word Embedding Methods for DSTC6 End-to-End Conversation Modeling Track, *Tokyo Institute of Technology*. (2016). doi:arXiv:1706.07440.
[12] K. Papineni, S. Roukos, T. Ward, and W.J. Zhu, BLEU: a method for automatic evaluation of machine translation, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. **40** (2001) 311–318. doi:10.3115/1073083.1073135.
[13] A.L. Beam, B. Kompa, I. Fried, N.P. Palmer, X. Shi, T. Cai, and I.S. Kohane, Clinical Concept Embeddings Learned from Massive Sources of Multimodal Medical Data, *ArXiv*. (2018). doi:arXiv:1804.01486.
[14] Y. Wang, S. Liu, N. Afzal, M. Rastegar-Mojarad, L. Wang, F. Shen, and H. Liu, A Comparison of Word Embeddings for the Biomedical Natural Language Processing, *Journal of Biomedical Informatics*. **85** (2018).
[15] B. Chiu, G. Crichton, A. Korhonen, and S. Pyysalo, How to train good word embeddings for biomedical NLP, *Proceedings of the 15th Workshop on Biomedical Natural Language Processing*. **15** (2016) 166–174. doi:10.18653/v1/W16-2922.
[16] S. Pakhomov, B. McInnes, T. Adam, Y. Liu, T. Pedersen, and G.B. Melton, Semantic similarity and relatedness between clinical terms: an experimental study, *AMIA Annual Symposium Proceedings*. **2010** (2010) 572–576.
[17] J. Sinapov, and A. Stoytchev, The odd one out task: Toward an intelligence test for robots, *IEEE 9th International Conference on Development and Learning*. **9** (2010). doi:10.1109/DEVLRN.2010.5578855.
[18] M.L. McHugh, Interrater reliability: the kappa statistic, *Biochemia Medica*. **22** (2012) 276–282.
[19] M. Th, S. Sahu, and A. Anand, Evaluating distributed word representations for capturing semantics of biomedical concepts, *Proceedings of BioNLP 15*. **15** (2015) 158–163. doi:10.18653/v1/W15-3820.

Address for correspondence

Emeric Dynomant
emeric.dynomant@omictools.com
+33 6 59 90 12 49

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190196

Learning Portuguese Clinical Word Embeddings: A Multi-Specialty and Multi-Institutional Corpus of Clinical Narratives Supporting a Downstream Biomedical Task

Lucas Emanuel Silva e Oliveira[a], Yohan Bonescki Gumiel[a], Arnon Bruno Ventrilho dos Santos[a], Lilian Mie Mukai Cintho[a], Deborah Ribeiro Carvalho[a], Sadid A. Hasan[b], Claudia Maria Cabral Moro[a]

[a] Health Technology Program, Pontifical Catholic University of Paraná, Curitiba, PR, Brazil,
[b] AI Lab, Philips Research North America, Cambridge, MA, USA

Abstract

In this paper, we trained a set of Portuguese clinical word embedding models of different granularities from multi-specialty and multi-institutional clinical narrative datasets. Then, we assessed their impact on a downstream biomedical NLP task of Urinary Tract Infection disease identification. Additionally, we intrinsically evaluated our main model using an adapted version of Bio-SimLex for the Portuguese language. Our empirical results showed that the larger, coarse-grained model achieved a slightly better outcome when compared with the small, fine-grained model in the proposed task. Moreover, we obtained satisfactory results with Bio-SimLex intrinsic evaluation.

Keywords:

Natural Language Processing; Electronic Health Records

Introduction

The Electronic Health Record (EHR) was mainly designed to digitally store patients' data and improve healthcare operational efficiency. Moreover, researchers found it a rich source to support several clinical informatics applications such as medical concept extraction, disorder reasoning, and patient history summarization [1].

Natural Language Processing (NLP) and Machine Learning (ML) techniques are widely used to extract, identify, and summarize EHR data, despite their dependency on laborious manual annotation and hand-crafted features [2,3]. Recently, many studies applied Deep Learning (DL) approaches to process EHR data [4–6], achieving better performance than traditional NLP/ML methods and requiring less time-consuming feature engineering.

An important component of DL for NLP methods is the use of Word Embeddings (WE) to represent each word as a vector in a low dimensional space [7] and employ this vector as an input feature. The resulting word vectors can be used to address many other NLP-related problems like sentiment analysis [8] and paraphrase detection [9]. Several studies used WE to solve health-related tasks like drug name recognition [10], semantic similarity [11], biomedical named entity recognition (bio-NER) [12], and patient outcome prediction [6].

To the best of our knowledge, only a few studies have applied WE to the Portuguese language in the biomedical domain (e.g., [13,14]). Three main studies made a WE repository available for both European and Brazilian Portuguese (pt-br) languages using a multi-genre corpus with data from Wikipedia, GoogleNews, etc. [15–17].

Despite the success of WE in the clinical NLP domain, it is difficult to find large, representative corpora to address relevant tasks, especially based on EHR data. Wang et al. [18] provided a comprehensive set of WE training experiments from distinct resources, namely clinical notes, biomedical articles, Wikipedia, and news. They found that the WE trained from EHR had the best results in clinical information extraction tasks. The semantic similarity captured by the EHR embeddings is closer to human experts' judgments on all datasets and, together with PubMed WE, the EHR embedding model finds more relevant similar medical terms.

Roberts [19] evaluated the trade-offs between small (and representative) corpora against large (but unrepresentative) corpora for training a WE model for clinical NLP tasks. It is not easy to decide between a huge, general-purpose corpus and a small, highly-representative corpus. For instance, one can choose a medium-sized clinical notes corpus instead of a corpus of a varying set of documents, a large scientific corpus, or even a combination of both. They found that merging multiple corpora is the best option when generating embeddings.

Thus, there exists a gap in building Portuguese clinical WE models for research: we could not find a clinical WE model available for NLP tasks in Portuguese. However, to provide a consistent WE model, a set of experiments are required to prove the usefulness of the model. It is possible to evaluate the model *extrinsically* by applying it to a downstream task, or *intrinsically* by measuring the innate quality of word representations through syntactic and semantic analogies [15,20,21].

Chiu et al. [22] developed two comprehensive resources, Bio-SimLex and Bio-SimVerb, targeting the intrinsic evaluation of word representations in biomedicine. Bio-SimLex is a list of 988 word pairs (nouns) in English and their respective similarity score (defined by a group of expert annotators), which can be used to compare the similarity scores between a WE model and the ones defined in Bio-SimLex. As some studies affirm that intrinsic and extrinsic evaluation scores do not always correlate [23–25], the authors claim that their evaluation resources can serve as a predictor of performance on downstream tasks. This is especially true for the Bio-SimLex set and bio-NER task, which were highly correlated.

In this paper, we address the identified research gap in pt-br clinical WE model generation and investigate an important research question: can a clinical WE model trained with multi-specialty and multi-institutional clinical narratives achieve

good results in downstream biomedical NLP tasks? We trained a preliminary multi-institutional and multi-specialty clinical WE model and assessed its performance by (i) checking if a large, coarse-grained model applied to a Deep Learning algorithm performs as well as a small, fine-grained model for predicting a specific disease and (ii) analyzing the results on Bio-SimLex evaluation set.

Methods

In this section, we describe our training process for the pt-br clinical WE model, including the dataset, preprocessing steps, and parameter space; the deep learning algorithm used to identify Urinary Tract Infection disease (UTI); the Bio-SimLex pt-br adaptation process; and the experimental setup.

Data

To generate the WE model, we used a collection of de-identified clinical narratives obtained from a group of three hospitals in south Brazil written from January 2013 to December 2017. These narratives include 745,731 documents of different types (nursing notes, discharge summaries, and ambulatory records) and various medical specialties (Cardiology, Nephrology, Endocrinology, etc.). We will call this entire collection: GROUP-ALL. We classified GROUP-ALL as coarse-grained because of the generalized nature of its data, which does not contain data specific to one type of document, specialty, or institution.

As we intended to run an algorithm to detect UTI disease and compare the results between a fine- and coarse-grained WE model (see next sections for details), we used a subset of data from GROUP-ALL. All narratives that contained a disease name corresponding to the ICD-10 code N.39, which denotes UTI, or corresponding subclass codes were filtered for inclusion, forming the new dataset called GROUP-UTI.

Additionally, we obtained another dataset from a hospital in southeast Brazil to train the UTI disease detection algorithm (see the following sections). This dataset, named ANN-UTI, consisted of narratives annotated with corresponding ICD-10 codes related to UTI diseases. We present a few sample narratives and dataset statistics in Table 1 and Table 2.

Preprocessing

We preprocessed the narratives of GROUP-ALL dataset using the following steps sequentially: (1) sentence parsing, (2) sentence tokenization, (3) lowercasing, (4) accentuation removal, (5) numeric characters removal, and (6) stopwords removal (using NLTK stopwords).

Embedding Parameters

To train a preliminary WE model, we followed recommendations and guidelines provided by prior studies that deeply analyzed the impact of hyperparameters on word vector quality. Unlike Beam et al. [26] who strictly reproduced the parameters of Levy et al. [27], we opted to select values from clinical WE studies as the default algorithmic configuration when there were inconsistencies among the guidelines

CBOW vs. Skip-gram: Several studies affirm that, in general, Skip-gram model performs better than CBOW [26,28,29]. Hence, we used Skip-gram to train our model.

Negative sampling: Levy et al. [27] recommended using multiple negative samples [30] when using Skip-gram. We used 8 negative samples based on Boag and Kané's [31] work on training WE based on a Clinical Metathesaurus.

Minimum count: Chiu et al. [28] showed the limited effect of this parameter on overall scores. Therefore, we used the default value of 5, which reduced the GROUP-ALL vocabulary size to 56,195 unique tokens, GROUP-UTI to 5,125, and ANN-UTI to 3,203.

Sub-sampling: Chiu et al. [28] described that this parameter does not have a significant impact on extrinsic evaluation, so we set the default value of 1e-3.

Context-size: Similar to Boag and Kané [31], we set the context/window size to 8.

Vector dimension: Following Boag and Kané [31], we used 300 as the dimension size of the vectors since Chiu et al. [28] and Fanaeepour et al. [32] found little improvement using 200. Furthermore, 300 corresponds with the configuration

Table 1– Example narratives from each dataset and their granularity. Note that, narratives from ANN-UTI contains an ICD-10 code at the beginning of the note, which is the result of expert annotation i.e. labeling with the corresponding diagnosis

Dataset	Sample Narrative	Granularity
GROUP-ALL	# RETORNO PARA REAVALIAÇÃO DE GLAUCOMA # GPAA EM USO DE DUO TRAVATAN E BRIMONIDINA - EM USO IRRGEULAR! # DMRI SECA AO - PACIENTE ESTÁ USANDO TRAVATAN A NOITE E LACRIFILM 3X/DIA AV CC 20/60 - 20/40 PIO 21/16 AO ESCAVAÇÃO DE 0,9, DRUSAS EM POLO POSTERIOR ORIENTO NECESSIDADE DE USO DOS COLIRIOS - CIENTE DO PROGNOSTICO PRESCREVO NOVAMENTE DUOTRAVATAN (E FORNEÇO MAIS UMA AMOSTRA GRATIS) + BRIMONIDINA 12/12 + LACRIFIM AO SOLICITO NOVO CAMPO VISUAL. RETORNO EM 1 MÊS PARA REAVALIAR PIO	**Coarse-grained**: Various medical specialties and institutions
GROUP-UTI	PACIENTE 62A, COM QUADRO DE INCONTINÊNCIA URINÁRIA DE ESFORÇO ASSOCIADA A URGÊNCIA HÁ APROXIMADAMENTE 01 ANO, PCTE RELATA TER INFECÇÃO DO TRATO URINÁRIO NÃO TRATADA ENCAMINHO A SEU MÉDICO SOLICITO EPU	**Medium-grained**: Narratives that contain ICD-10 N39 group of diseases, multi-institutional.

requirements of the GloVe model (that is the UTI detection algorithm baseline).

DeepCoder: An Algorithm to Identify Urinary Tract Infection

UTI is defined as "an infection anywhere in the urinary tract (urethra, bladder, ureters, or kidneys)" [33] or "the clinical syndromes of acute, uncomplicated, urinary infection" [34]. ICD-10 reserves a specific class for such problems: "N39 - Other disorders of urinary system" and 7 subclasses (N39.0, N39.1, N39.2, N39.3, N39.4, N39.8 and N39.9) to provide more detailed information for clinical evaluation.

Table 2 – Dataset sizes by number of tokens and sentences

Dataset	#Sentences	#Tokens	#Unique tokens
GROUP-ALL	2,412,055	32,023,244	287,495
GROUP-UTI	26,719	319,203	17,518
ANN-	2,030	205,318	11,494

DeepCoder was developed using a GloVe-trained WE model as input, based on ANN-UTI texts following the preprocessing steps and hyperparameter configuration described in previous sections. The algorithm is composed of a neural network architecture formed with an embedding layer as input with 500 dimensions (e), four convolutional layers with kernel = 128 (k), a Rectified Linear Unit (ReLU), and window sizes of 5, 8, 10, and 12. Between each convolutional layer, there exists a 1-max pooling layer, a global max pooling layer, and a dropout of 0.2 (d) that feeds a 128-sized Dense (i) layer and softmax activation with 8 possible outputs (7 for the ICD-10 N39 subclass codes, and 1 for a generic non-UTI code).

Experimental Setup

We based our experiments on an extrinsic evaluation using the DeepCoder algorithm. In order to predict the model performance in other biomedical NLP downstream tasks, we performed an intrinsic evaluation using the Bio-SimLex set.

Extrinsic Evaluation

The Word2Vec models trained with GROUP-ALL, GROUP-UTI, ANN-UTI datasets, and the original DeepCoder GloVe model (trained with ANN-UTI) were utilized in DeepCoder to perform the extrinsic evaluation and compare relative performance. All results were calculated using a 10-fold cross-validation. Figure 1 presents the extrinsic experimental setup overview.

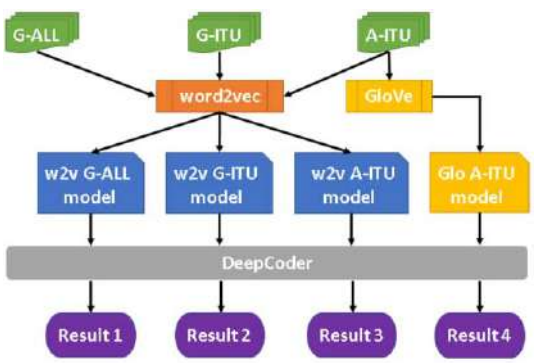


Figure 1 – Extrinsic experimental setup: each dataset was used to obtain different WE models using Word2Vec and GloVe algorithms. We used the four pre-trained WE models to create input embeddings for DeepCoder

Intrinsic Evaluation and Bio-SimLex Adaptation

To perform the intrinsic evaluation, we used the Bio-SimLex evaluation set and associated semantic similarity and relatedness scores. To evaluate a WE model, we calculated the Spearman's correlation coefficient between the similarity ratings found in the model versus the ratings defined by experts and available in Bio-SimLex.

Two researchers (one with a medical background, the other with a health informatics background) translated and adapted the terms to pt-br, observing the following instructions:

- Check if the English term exists in the Unified Medical Language System (UMLS) and has a pt-br translation. If not, discuss the best possibility between the two translators. If yes, then pick the preferable option. If multiple options, then prioritize terms:
  - Labeled as ISPREF='Y'
  - With a higher number of occurrences in GROUP-ALL dataset
- Label the word pair as General, EHR, or Biomedical – where the first stands for words from the general domain, the second represents words seen in EHR texts, and the third category contains words that are not present in EHR but are used in biomedical context (e.g., biomedical articles).

In Table 3 we present some translation examples.

Table 3 – Bio-SimLex translation examples

Term1	Term2	Score	Type
therapy/terapia	treatment/tratamento	9.32	EHR
oxide/óxido	sucrose/sacarose	0.00	BioM

It is important to highlight some issues found during the translation process, including: (i) the difficulty to translate ambiguous terms that have multiple meanings or translations, depending on the context (e.g., abstract) and (ii) word pairs without well-known synonyms in pt-br, like hindbrain and rhombencephalon, which we found only one possible word to translate both (*rombencéfalo*). In this case, we removed the word pair to avoid equal words in the evaluation set.

Due to the differences between the general domain, biomedical literature, and EHR texts, we performed three separate intrinsic evaluation runs: one using all the translated Bio-SimLex data; another with word pairs categorized as Biomedical and EHR, excluding the General category; and using only the word pairs labeled as EHR. By doing this, we tried to fairly evaluate our model which was trained exclusively with EHR text data.

Results

We summarize the results from our extrinsic evaluation in Table 4. It is possible to verify that the scores are very similar across all models. The GROUP-ALL model is the largest and least representative compared to the gold standard, and ANN-UTI is the smallest and most representative (because it contains texts from the same dataset as the gold standard). This suggests that the GROUP-ALL model takes advantage of its size to yield good performance for the UTI identification task. Likewise, the ANN-UTI models (Word2Vec and GloVe), as the more

representative models, achieve good results as expected. The GROUP-UTI embeddings yielded an inferior result, most likely due to less representative texts of less than medium-size.

The results of the intrinsic evaluation of the GROUP-ALL model, using three subsets of the Bio-SimLex are shown in Table 5. The Spearman's correlation coefficient increases when we use more specific subsets (Biomedical and EHR). Using only the word pairs labeled as EHR our model achieved a correlation score of 0.6419, which is similar to the results obtained by Chiu et al. [22] with their Skip-gram and PubMed-w2v models, varying 0.07 and 0.05 respectively.

Table 4 – Extrinsic experiments results. F1-score mean (F1 with cross-validation with 10 folds) and standard deviation (SD)

WE model	F1	SD
Word2Vec GROUP-ALL	**0.95**	**0.04**
Word2Vec GROUP-UTI	0.91	0.14

Table 5 – Intrinsic experiments results of Spearman's correlation coefficient (RHO) divided by the categories of Bio-SimLex word pairs. RHO

Categories	RHO
General+Biomedical+EHR	0.4558
Biomedical+EHR	0.5679

Discussion

The results from our empirical extrinsic evaluation confirm previous findings [19] suggesting that we have indeed a trade-off between corpus size and similarity when it comes to WE. The results imply that the answer to our research question is: yes, a large coarse-grained WE model can yield good results for a downstream biomedical NLP task.

We can highlight a few limitations of this study and consequently propose some future work. For example, the models were extrinsically evaluated for one task only; and to overcome this limitation we opted to use Bio-SimLex to emulate the results to another variety of bio-NLP tasks. But due to the difficulties found in the translation process, we think that the evaluation set lost some of its reliability, then would be better to build a specific evaluation resource for EHR pt-br data, although the obtained results are similar to Chiu et al. [22].

We would also explore hyperparameter tuning and other WE algorithms such as fastText [35], and wang2vec [36] in the future.

In this paper, we explored WE models trained with words only, although it is possible to use several approaches that focus on enhancing WE with clinical knowledge by concatenating extra information to the vector space [11,21,26,31,37,38]. It is also worth noting that despite the existence of various approaches to generate clinical embeddings, there is limited consensus among researchers on what is the state-of-the-art for each bio-NLP task.

Besides some authors [39] discuss the reliability on ICD-10 coding, our work relied on a simple annotation process containing only one disease and its specializations, which did not lead us to uncertainties and the complex ICD environment that build-up in some cases.

We also plan to enlarge our corpus by adding biomedical publications, Wikipedia, and other open source datasets in Portuguese.

Conclusions

In this study, we built WE models with different granularities and extrinsically evaluated them using a disease prediction algorithm (DeepCoder) to assess performance variation due to different word embeddings. We used an adapted version of the Bio-SimLex set to intrinsically evaluate a large, coarse-grained model in order to predict the model's performance in other biomedical downstream tasks.

We conclude that it is possible to achieve similar results using a large, coarse-grained WE model and a small, fine-grained alternative to facilitate a bio-NLP task. However, robustness of our models could be ensured by applying them to a wide range of clinical prediction tasks, as the Bio-SimLex adaptation to pt-br has some limitations and reliability issues.

Acknowledgments

Philips Research North America and Pontifical Catholic University of Paraná supported this work. "*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001*" partially financed this study also.

References

[1] B. Shickel, P.J. Tighe, A. Bihorac, and P. Rashidi, Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis, *IEEE J. Biomed. Heal. Informatics*. **22** (2017) 1589–1604.

[2] F. Liu, J. Chen, A. Jagannatha, and H. Yu, Learning for Biomedical Information Extraction: Methodological Review of Recent Advances, *Comput. Res. Repos.* (2016).

[3] C. Friedman, T.C. Rindflesch, and M. Corn, Natural Language Processing: State of the Art and Prospects for Significant Progress, a Workshop Sponsored by the National Library of Medicine, *J. Biomed. Inform.* **46** (2013) 765–773.

[4] R. Miotto, L. Li, B.A. Kidd, and J.T. Dudley, Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records, *Sci. Rep.* **6** (2016) 1–10.

[5] A.N. Jagannatha, and H. Yu, Structured Prediction Models for Rnn Based Sequence Labeling in Clinical Text, *Proc. Conf. Empir. Methods Nat. Lang. Process. Conf. Empir. Methods Nat. Lang. Process.* **2016** (2016) 856–865.

[6] Y. Feng, X. Min, N. Chen, H. Chen, X. Xie, H. Wang, and T. Chen, Patient Outcome Prediction via Convolutional Neural Networks Based on Multi-Granularity Medical Concept Embedding, in: 2017 IEEE Int. Conf. Bioinforma. Biomed., IEEE, 2017: pp. 770–777.

[7] Y. Goldberg, A Primer on Neural Network Models for Natural Language Processing, *J. Artif. Intell. Res.* **57** (2016) 345–420.

[8] A.L. Maas, R.E. Daly, P.T. Pham, D. Huang, A.Y. Ng, and C. Potts, Learning Word Vectors for Sentiment Analysis, *Proc. 49th Annu. Meet. Assoc. Comput. Linguist. Hum. Lang. Technol.* (2011) 142–150.

[9] R. Socher, E. Huang, and J. Pennington, Dynamic Pooling and Unfolding Recursive Autoencoders for Paraphrase Detection., *Adv. Neural Inf. Process. Syst.* (2011) 801–809.

[10] I. Segura-bedmar, and P. Mart, Exploring Word

Embedding for Drug Name Recognition, *Sixth Int. Work. Heal. Text Min. Inf. Anal.* (2015) 64–72.
[11] L. De Vine, G. Zuccon, B. Koopman, L. Sitbon, and P. Bruza, Medical Semantic Similarity with a Neural Language Model, *Proc. 23rd ACM Int. Conf. Inf. Knowl. Manag. - CIKM '14*. (2014) 1819–1822.
[12] Y. Wu, J. Xu, M. Jiang, Y. Zhang, and H. Xu, A Study of Neural Word Embeddings for Named Entity Recognition in Clinical Text., *AMIA Annu. Symp. Proc.* **2015** (2015) 1326–33.
[13] F. Duarte, B. Martins, C.S. Pinto, and M.J. Silva, A Deep Learning Method for ICD-10 Coding of Free-Text Death Certificates, in: Prog. Artif. Intell. EPIA 2017. Lect. Notes Comput. Sci., 2017: pp. 137–149.
[14] M. V. Treviso, C.D. Shulby, and S.M. Aluisio, Evaluating Word Embeddings for Sentence Boundary Detection in Speech Transcripts, *Proc. 11th Brazilian Symp. Inf. Hum. Lang. Technol.* (2017) 151–160.
[15] J. Rodrigues, A. Branco, S. Neale, and J. Silva, LX-DSemVectors: Distributional Semantics Models for Portuguese, in: Comput. Process. Port. Lang. PROPOR 2016. Lect. Notes Comput. Sci., 2016: pp. 259–270.
[16] N.S. Hartmann, E. Fonseca, C.D. Shulby, M. V Treviso, J.S. Rodrigues, and S.M. Aluísio, Portuguese Word Embeddings: Evaluating on Word Analogies and Natural Language Tasks, *Proc. 11th Brazilian Symp. Inf. Hum. Lang. Technol.* (2017).
[17] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, and T. Mikolov, Learning Word Vectors for 157 Languages, *Comput. Res. Repos.* (2018).
[18] Y. Wang, S. Liu, N. Afzal, M. Rastegar-Mojarad, L. Wang, F. Shen, P. Kingsbury, and H. Liu, A Comparison of Word Embeddings for the Biomedical Natural Language Processing, *J. Biomed. Inform.* **87** (2018) 12–20.
[19] K. Roberts, Assessing the Corpus Size vs. Similarity Trade-off for Word Embeddings in Clinical NLP, *Proc. Clin. Nat. Lang. Process. Work.* (2016) 54–63.
[20] T. Mikolov, K. Chen, G. Corrado, and J. Dean, Efficient Estimation of Word Representations in Vector Space, *ArXiv Prepr. ArXiv1301.3781*. (2013).
[21] Y. Choi, C.Y.I. Chiu, and D. Sontag, Learning Low-Dimensional Representations of Medical Concepts., *AMIA Jt. Summits Transl. Sci. Proceedings. AMIA Jt. Summits Transl. Sci.* (2016) 41–50.
[22] B. Chiu, S. Pyysalo, I. Vulić, and A. Korhonen, Bio-Simverb and Bio-Simlex: Wide-Coverage Evaluation Sets of Word Similarity in Biomedicine, *BMC Bioinformatics*. **19** (2018) 33.
[23] B. Chiu, A. Korhonen, and S. Pyysalo, Intrinsic Evaluation of Word Vectors Fails to Predict Extrinsic Performance, in: Proc. 1st Work. Eval. Vector-Sp. Represent. NLP, ACL, 2016: pp. 1–6.
[24] A. Gladkova, and A. Drozd, Intrinsic Evaluations of Word Embeddings: What Can We Do Better?, in: Proc. 1st Work. Eval. Vector-Sp. Represent. NLP, ACL, 2016: pp. 36–42.
[25] M. Faruqui, Y. Tsvetkov, P. Rastogi, and C. Dyer, Problems With Evaluation of Word Embeddings Using Word Similarity Tasks, in: Proc. 1st Work. Eval. Vector-Sp. Represent. NLP, ACL, 2016: pp. 30–35.
[26] A.L. Beam, B. Kompa, I. Fried, N. Palmer, X. Shi, T. Cai, and I.S. Kohane, Clinical Concept Embeddings Learned from Massive Sources of Medical Data, *Comput. Res. Repos.* (2018).
[27] O. Levy, Y. Goldberg, and I. Dagan, Improving Distributional Similarity with Lessons Learned from Word Embeddings, *Trans. Assoc. Comput. Linguist.* **3** (2015) 211–225.
[28] B. Chiu, G. Crichton, A. Korhonen, and S. Pyysalo, How to Train good Word Embeddings for Biomedical NLP, *Proc. 15th Work. Biomed. Nat. Lang. Process.* (2016) 166–174.
[29] T.H. Muneeb, S.K. Sahu, and A. Anand, Evaluating Distributed Word Representations for Capturing Semantics of Biomedical Concepts, *Proc. 2015 Work. Biomed. Nat. Lang. Process. (BioNLP 2015)*. (2015) 158–163.
[30] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, Distributed Representations of Words and Phrases and their Compositionality, in: NIPS'13 Proc. 26th Int. Conf. Neural Inf. Process. Syst., 2013: pp. 3111–3119.
[31] W. Boag, and H. Kané, AWE-CM Vectors: Augmenting Word Embeddings with a Clinical Metathesaurus, *NIPS 2017 Work. Mach. Learn. Heal.* (2017).
[32] M. Fanaeepour, A. Makarucha, and J.H. Lau, Evaluating Word Embedding Hyper-Parameters for Similarity and Analogy Tasks, *Comput. Res. Repos.* (2018).
[33] B. Foxman, Urinary Tract Infection Syndromes. Occurrence, Recurrence, Bacteriology, Risk Factors, and Disease Burden, *Infect. Dis. Clin. North Am.* **28** (2014) 1–13.
[34] L.E. Nicolle, Urinary Tract Infection, *Crit. Care Clin.* **29** (2013) 699–715.
[35] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, Enriching Word Vectors with Subword Information, *Comput. Res. Repos.*(2016).
[36] W. Ling, C. Dyer, A.W. Black, and I. Trancoso, Two/Too Simple Adaptations of Word2Vec for Syntax Problems, in: Proc. 2015 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol., ACL, 2015: pp. 1299–1304.
[37] Z. Yu, T. Cohen, E. V Bernstam, T.R. Johnson, and B.C. Wallace, Retrofitting Word Vectors of MeSH Terms to Improve Semantic Similarity Measures, *Proc. Seventh Int. Work. Heal. Text Min. Inf. Anal.* (2016) 43–51.
[38] T. Bai, A.K. Chanda, B.L. Egleston, and S. Vucetic, Joint Learning of Representations of Medical Concepts and Words from EHR Data, in: 2017 IEEE Int. Conf. Bioinforma. Biomed., IEEE, 2017: pp. 764–769.
[39] J. Stausberg, N. Lehmann, D. Kaczmarek, and M. Stein, Reliability of Diagnoses Coding with ICD-10, Int. J. Med. Inf. **77** (2008) 50–57.

Address for correspondence

Lucas Emanuel Silva e Oliveira.
tel: +55 41 992872702.
email: lucas.oliveira@pucpr.br

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190197

Query Translation Between AQL and CQL

Georg Fette[a,b], Mathias Kaspar[b], Leon Liman[a], Maximilian Ertl[b], Jonathan Krebs[a], Georg Dietrich[a], Stefan Störk[b], Frank Puppe[a]

[a] *Chair of Computer Science 6, University of Würzburg, Würzburg, Germany,*
[b] *Comprehensive Heart Failure Center, University Hospital of Würzburg, Würzburg, Germany*

Abstract

Secondary use of electronic health records using data aggregation systems (DAS) with standardized access interfaces (e.g. openEHR, i2b2, FHIR) have become an attractive approach to support clinical research. In order to increase the volume of underlying patient data, multiple DASs at different institutions can be connected to research networks. Two obstacles to connect a DAS to such a network are the syntactical differences between the involved DAS query interfaces and differences in the data models the DASs operate on. The current work presents an approach to tackle both problems by translating queries from a DAS using openEHR's query language AQL (Archetype Query Language) into queries using the query language CQL (Clinical Quality Language) and vice versa. For the subset of queries which are expressible in both query languages the presented approach is well feasible.

Keywords:

Electronic Health Records, Data Warehousing, Information Systems

Introduction

The secondary use of electronic health records (EHR) has become an important field in medical informatics [1]. Routine clinical data is reused for various scientific purposes, like prospective estimation of study cohort sizes or support of study cohort acquisition. When this routine data is scattered in various data sinks in various heterogeneous data formats, it is difficult to access. In order to improve access to the EHR, solutions have been developed to aggregate the routine data and to make it accessible via a standardized interface. Two paradigms are possible when constructing such a data aggregation system (DAS): The first method is to leave the data in its original place and to provide a standardized access by aggregating the requested data chunks on the fly using appropriate extraction transformation load (ETL) pipelines. SMART on FHIR [2] projects for example realize this paradigm by aggregating requested FHIR resources on the fly. This approach works well for patient centered access (i.e. queries operating on the data of a single patient). However, for population centered access (i.e. queries operating on data of all patients) this approach lacks indices covering the involved data sources. Without such indices, query speed performance does not scale related to the amount of queried patient data. On large volumes of EHR data this would heavily afflict the usability of FHIR related query languages like FHIR-REST-API[1] or CQL (Clinical Quality Language)[2]. The second method to provide access to aggregated routine data is to provide an additional data sink with a generic data model, into which the heterogeneous routine data is persistently transferred via an ETL pipeline. The data sink is supplied with appropriate indices, so it can be efficiently queried via a standardized query language. Examples of this architecture are i2b2 [3], openEHR [4] or OMOP [5]. HAPI[3] and VONK[4] are data sink servers that do not only use FHIR as an access interface in patient centered mode, but they also use it as the data model in which the EHR data is stored, so it can be queried in population centered mode using query languages like FHIR-REST-API or CQL.

Clinical Research Networks

In order to increase the volume of underlying patient data for larger and thus more expressive query results, DASs at different institutions can be connected to clinical research networks. Examples for such networks are PCORnet, the OHDSI research network or EHR4CR. In a DAS network a query is distributed to connected nodes, where it is independently evaluated. Each node's results are returned and combined to an aggregated result. Systems like SHRINE [6] for i2b2 or SNOW [7] for openEHR perform the query distribution and result aggregation automatically. Each network, however, only allows DASs having the same query language to be part of the network. If a DAS with a different query language is intended to be integrated into the network, the data from that DAS has to be transferred (like in [8]) into a new dedicated DAS supporting the networks query language. However, parallel support of multiple DASs containing the same redundant data at one institution creates an overhead in support and hardware.

An alternative approach is to translate queries of an incompatible DAS into the query language required by the DAS network. The current work examines the feasibility of translating queries formulated in the query language CQL into AQL (Archetype Query Language, the query language of openEHR) and vice versa. To accomplish this task, the query is first translated into an intermediate query graph model, then potentially necessary graph transformation are applied on the query graph and, finally, the graph is translated into the desired target query language.

Methods

Query Languages

Before going into detail about the translation process, the two languages shall be briefly introduced:

CQL is a functional query language. A CQL script is defined as a so called *library* in which, besides the actual queries, also

[1] https://www.hl7.org/fhir/search.html
[2] https://cql.hl7.org
[3] http://hapifhir.io
[4] https://fire.ly/vonk/vonk-fhir-server

meta information related to the queries can be defined. CQL is independent of a concrete data model, as the model to be used is explicitly defined in each CQL library. The CQL data model elements can be composed of the following data types: primitives (Booleans, Strings, numbers and timestamps), clinical codes, quantities, intervals, lists and structured types. Structured data objects, which can again contain other structured data objects, can be accessed with a path syntax (e.g. *patient.contact[0].name.family*). Besides the data model, a CQL library can define its search context, which is either *Patient* or *Population* centered. The main functional part in a CQL library are the *statements*. Each statement can be seen as an individual query. Usually a statement defines which data sources the query is operating on (e.g. *[Patient] A* or *[Observation] B*), how the elements of the queried sources are constrained (e.g. *B.valueString = 'x'*), how the various data sources have to be related to each other (e.g. *[Patient] A with [Observation] B such that B.subject = A*) and which elements of potential matches have to be returned as results (e.g. *Return Tuple {id:A.identifier, B.valueString}*). CQL provides a rich repertoire of operators (e.g. comparison, logical, arithmetic, list access, aggregation) to perform calculations on the data model.

OpenEHR's query language AQL is SQL-inspired and as well a functional language. It operates on a data model that is defined by the openEHR data modeling language ADL. OpenEHR's root data model elements are called *archetypes*. An archetype can be composed of the following data element types: primitives (Booleans, Strings, numbers and timestamps), other archetypes and generic container structures (e.g. lists). Structured data objects are accessed via a path syntax. An AQL query mainly consists of three parts: The *FROM* part defines which archetypes are queried and how these archetypes are related to each other concerning meronymity relations (e.g. *EHR A contains Observation B*). The *WHERE* part constrains the archetype elements (e.g. *B/value > 1*). The *SELECT* part defines which elements have to be returned in the results (e.g. *select A/ehr_id, B/value*). Compared to CQL, AQL provides a smaller set of operators (comparison, logical, *matches*, *exists*). For the sake of brevity archetype names in the following chapters are shortened (e.g. openEHR-EHR-OBSERVATION.LabResult.v1 → Observation[LabResult]).

Currently, the CQL translation has the following constraints: 1. The CQL queries may only contain a single statement and 2. The CQL queries are defined in the context *Population*. If a CQL query with context *Patient* has to be translated to AQL, the given patient identifier has to be integrated into the CQL statement as a constrained *Patient.identifier*.

Queries as Graphs

Queries can be seen as graphs. In order to perform the query translation, the CQL/AQL (depending on the translation direction) query to be translated is transformed into an intermediate graph model. The graph is subsequently translated into the desired target query language. The graph model is depicted in Figure 1. A graph contains a set of data model elements that are connected via relations. Although the type of these relations is not specifically defined, they could be interpreted as meronymity (i.e. *contains*) relations. Additionally, a graph contains a set of operators, which contain as parameters either other operators, data model elements or literals. A similar same approach has already been applied in [9] where AQL was translated into the query language of i2b2.

For parsing AQL queries the parser from the AQL-processor of the EtherCIS project[5] was taken and combined with a graph builder written by the authors. The parser for CQL libraries was taken from the cql-2-elm project[6] and combined with a graph builder also written by the authors. The graphs retain 1. the structure of data model elements mentioned in the query, 2. the constraints and operators on data element values and 3. which data elements have to be contained in the returned results. Whenever elements in a query are named with an alias, all references to that alias create *IsRelatedTo/ HasParameter* relations to the graph node identified with that alias.

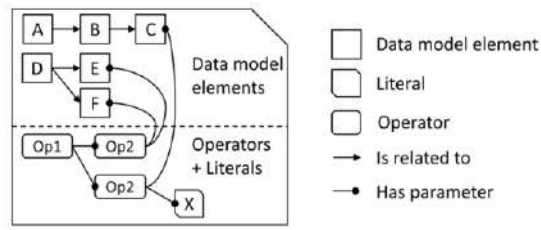


Figure 1 – Model of graphs, which are used as intermediate data model during query translation.

Graph Transformations

Because of properties of CQL/AQL that do not exist in the respective other language or non-matching properties of the incorporated data models, the graph has to be appropriately transformed, in order to fit all required properties of the target language. Currently there exist the following types of transformations: Meronymity-Equality-Transformations and their reverse, Path-Transformations, Concept-Code-Mappings, Operator-Mappings, Resolve-Quantity-Transformations, Resolve-Interval-Transformations (see Figure 2).

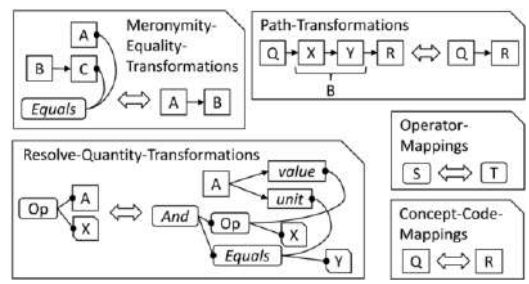


Figure 2 – Query graph transformations

Meronymity-Equality-Transformations

Meronymity-Equality-Transformations are needed because in CQL the connection between two root data model elements is represented by an equality constraint between one root element and a child element of the other root element (e.g. *[Patient] A with [Observation] B such that B.subject = A*). The definition of both involved element types (*[Observation]* and *[Patient]*) determine the type for the element *B.subject*, which would otherwise not be specified. In contrast, in AQL the main mechanism to define structural relationships between data model elements are meronymity relations, which also define the type of the contained elements (e.g. *Ehr A contains Observation[LabResult] B*). Meronymity-Equality-Transformations are parameter free. The transformation looks for any three data model elements *A*, *B* and *C*, with a relation between *B* and *C* and an *Equals* operator containing *A* and *C* as parameters. The element *C* and the *Equals* operator are deleted and the element *A* gets connected to the element *B*. The reverse Equality-Meronymity-Transformation requires a parametrization because only a defined set of related elements *A* and *B* should be transformed and the name of the newly

[5] http://ethercis.org

[6] https://github.com/cqframework/clinical_quality_language

created element *C* has to be given. For all directly connected elements *A* and *B* the connection gets removed and instead a new element *C* gets created and connected to *B*. Furthermore, a new *Equals* operator gets created receiving *A* and *C* as parameters.

Path-Transformations

Path-Transformations provide the possibility to shorten or lengthen data model element paths. They are needed because elements from the source data model, which have to be mapped to their semantically identical counterparts in the target data model, might be encapsulated in additional element wrappers. These wrappers have to be removed if they do not exist in the target data model. E.g. the intermediate element *valueQuantity* in *Observation.valueQuantity.value* (FHIR) has no counterpart in *Observation[LabResult]/value* (openEHR). The transformation can be configured to either shorten or lengthen a path. In order to shorten a path, the transformation searches for nodes *Q* and *R* connected by a list of nodes *B* and removes the intermediate nodes. The reverse transformation introduces new intermediate nodes *B* into the graph between two directly connected nodes *Q* and *R*.

Concept-Code-Mappings

Concept-Code-Mappings map element identifiers of the source data model to element identifiers of the target data model by renaming nodes.

Operator-Mappings

Operator-Mappings map operators of the source query language to operators of the target query language by renaming nodes.

Resolve-Quantity-Transformations

AQL does not contain quantities as build in types. Therefore, operators on quantity types have to be exchanged by *and*-connected clauses that check the requested value as well as the requested quantity type (e.g. *days, cm*) (e.g. *[Encounter] A where A.length > 120 days → [Encounter] A where A.length.value > 120 and A.length.unit = 'days'*). The transformation looks for data model elements *A* contained as a parameter in an operation *Op* (with potential further parameters). The node *A* get connected to two newly created data model elements *value* and *unit* and gets removed from the parameters of *Op*. Additionally, a new *equals* operator is created which gets the new *unit* node and a newly created literal *Y* as parameters. The value of *Y* has to be given as a transformation configuration and defines the unit of the quantity to be computed. The new *equals* node and the *Op* node get connected in a newly created *and.*

Resolve-Interval-Transformations

Similar to quantity types, AQL does not contain intervals as build in types. Operators on interval types have to be exchanged by *and*-connected clauses, that check the requested constraints (e.g. *[Patient] A where A.birthDate in A.contact[0].address.period → [Patient] A where A.birthDate > A. contact[0].address.period.low and A.birthDate < A. contact[0].address.period.high*).

Graphs to Queries

After all necessary graph transformations have been executed, the graph can be translated into the desired target query language via respective Graph2QueryString writers implemented by the authors.

The proposed method was tested on manually designed AQL/CQL queries and on queries contained in the AQL/CQL documentation. A query was (when translatable) translated into its respective counterpart and re-translated into its original language. The original query was compared for semantic equality to its translated counterpart as well as to its re-translation into the original query language.

Results

Translations Constraints

The proposed methodology using the currently available set of graph transformations allows the translation of all CQL/AQL queries having the following properties: A query has to contain the data type *Patient* (*EHR* in AQL) as a query source. The query may contain an arbitrary amount of additional data sources of arbitrary type. A query may return an arbitrary subset of query sources or data model elements that are reached from the data sources via paths. The paths can have arbitrary length. Data model elements can be constrained using the logical comparators listed below, parametrized with literals or with other data model elements. A query may contain the data types *Boolean*, *String*, *Integer*, *Decimal*, *Timestamp*, *Date*, *DateTime*, *Quantity* and *Interval*. Data model elements or literals can be processed with operators from the set of mutual operators listed below.

Translations are only possible for queries containing exclusively operators which exist in both query languages with the same semantics and the same interface. These operators (using their CQL-displaynames) grouped into operator categories are:

- Boolean operators *{ and, or, not }*
- comparator operators on numeric/date/timestamp types *{ =, <, <=, >, >=, != }*
- comparator operators on Strings *{ =, !=, matches }*
- operators on lists/iterators *{ in, exists, with, without }*.

All other operators either have to be substituted by appropriate graph transformations or render a query untranslatable.

CQL does not contain an equivalent to AQL's *contains* operator. This operator, which performs a type matching on a given to be contained data type, returns the matched child data element. The operator is substituted by Meronyminity-Equality-Transformations.

The mutual set of operators is reduced by the following limitations of the two languages, which could not be substituted by graph transformations:

AQL contains no arithmetic operators (e.g. *+, -, *, /*). AQL comprises a limited set of list operators (e.g. *first, last, sort, count* do not exist). It is only possible to access specific list elements via a given index or via the *matches* operator. The *contains* operator can only be used to constrain the data model structure and not to check the containment of a single element in a list (e.g. *[ProcedureRequest] A where A.notes contains 'x'*).

AQL allows no aliased valueset definitions. All valuesets have to be directly given as parameters to the operator using them, instead of having the possibility to reference a previously defined valueset with an alias. CQL, on the other hand, allows only valueset aliases based on valueset ids but no definition of valuesets by listing their contained codes. As the current implementations is restricted to CQL statements instead of CQL libraries, the usage of valuesets or therein contained codes still has to be resolved.

Table 1 shows a selection of example queries that have been automatically translated given proper graph transformation configurations.

Table 1 – Automatically translated example queries

CQL	AQL
[Patient] A where A.active = true and A.gender = 'male'	select e from EHR e where e/active = true and e/gender = 'male'
[Patient] A where exists(A.name B where B.given = 'John')	select e from EHR e contains composition a[HumanName] where a/given = 'John'
[Patient] A with [Encounter] B such that not(exists(B.discharge)) and B.subject = A	select e from EHR e contains composition a[Encounter] where not exists a/content[Discharge]
[Patient] A with [Observation] B such that B.code = 'Calcium' and B.valueQuantity > 10'mg' and B.subject = A	select e from ehr e contains observation a[LabResult] where a/code = 'Calcium' and a/valueQuantity/value > 10 and a/valueQuantity/unit = 'mg'

Figure 3 pictures an example of an CQL query being transformed into AQL. The query represents a request for patients having at least one *Calcium* laboratory measurement with a value of more than 10. The CQL query is parsed into the uppermost graph depicted in Figure 3. Successively, Meronymity-Equality-Transformations, Path-Transformations and Concept-Code-Mappings are applied to the graph (parametrized with the configurations from Table 1, CQL → AQL). The Meronymity-Equality-Transformation links the *Patient* node with the *Observation* node, which is the required representation in AQL. The Path-Transformation removes the *valueQuantity* node, because in the openEHR data model the value element is a direct child element of the *LabResult* archetype. The Concept-Code-Mappings rename the nodes *Patient* and *Observation* as *Ehr* and *Observation[LabResult]*, as those are the semantically equivalent data elements in the openEHR data model. Finally, the transformed graph is translated into an AQL query String.

Table 2 – Manually defined configurations used in the example in Figure 3. The column headings reference the nodes from Figure 2.

CQL → AQL		
Path-Shorten-Transformations		
Q	**B**	**R**
Observation[LabResult]	{valueQuantity}	value
Concept-Code-Mappings		
Q	**R**	
Ehr	Patient	
Observation[LabResult]	Observation	
AQL → CQL		
Equality-Meronymity-Transformations		
A	**B**	**C**
Patient	Observation	subject
Path-Lengthen-Transformations		
Q	**B**	**R**
Observation[LabResult]	{valueQuantity}	value
Concept-Code-Mappings		
Q	**R**	
Ehr Observation[LabResult]	Patient Observation	

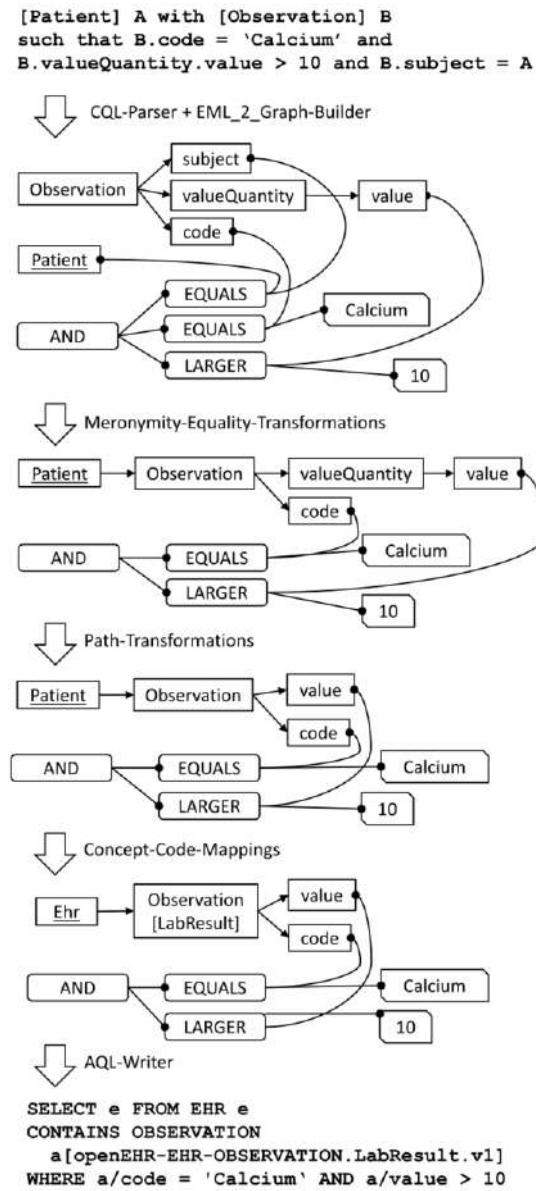


Figure 3 – Translation of an CQL query into AQL. Underlined nodes denote elements that have to be included in the result

The translation from AQL to CQL goes accordingly with the transformation configurations from Table 1, AQL → CQL. It is noted that the order of the execution of the different transformation types remains the same, so the re-translation is not according to the reverse direction in Figure 3, which has to be considered in the transformation configurations. The AQL query is first parsed into a graph. Subsequently the Equality-Meronymity-Transformation creates the child node *subject*, connects it to the *Observation[LabResult]* node, removes the relation between *Ehr* and *Observation[LabResult]* and adds a new *Equals* operator with *subject* and *Ehr* as parameters. The Path-Transformation adds the node *valueQuantity* between *Observation[LabResult]* and *value*. The Concept-Code-Mappings rename the nodes *Ehr* and *Observation[LabResult]* as *Patient* and *Observation*. Finally, the graph is translated into a CQL query String.

Discussion

The approach to translate a query into another query to improve or facilitate its evaluation is known as query rewriting [10]. The early work on query rewriting is related to view-based relational database querying and deals with the translation of queries formulated in the same language for both, source and target. In recent years query rewriting was as well used in conjunction with graph data query system, e.g. [11] resembles the presented approach, whereas there XQuery was translated into SPARQL.

The discussed graph model and the graph transformations have been specified and implemented by the authors. Both elements could be exchanged by a standard from the graph computing community like e.g. Owl to model graphs and graph transformation frameworks like GROOVE to model the transformations.

Although the used graph model is similar to the parser graphs produced by each of the systems, it was more convenient for the authors to use a separate model in order to be independent of possible system specific modeling paradigms.

The concept-mappings and path-transformations could be substituted by a mechanism using FHIR ConceptMaps, which encode the mappings from elements of one data model to equivalent data model elements in a target system.

When applying the presented approach on productive data models, the configuration of the various transformations could become cumbersome when the data models are large. It would be beneficial if some configurations could be automatically deduced by an analysis of the source and the target data model. When both data models are annotated with terminology codes, the configuration of concept-mappings could be deduced automatically by identifying equivalent data model elements annotated with the same codes.

An issue that could appear when applying query translations in combination with concrete query engines system implementations could be that the translations are semantically equivalent, but they can contain differences in their syntactical structure. E.g. the two CQL queries *[Patient] A with [Encounter] B such that B.patient = A* and *[Patient] A where exists [Encounter] B where B.patient = A* are semantically equivalent, but a query execution engine could handle the execution of a *with-such-that* expression differently than a *where-exists* with a nested sub query.

An aspect not yet covered by the presented approach is the translation of *valueset*, *concept*, *code* and *codesystem* definitions. Due to the omnipresence of valueset references in the query examples of both query languages, the presented approach could have difficulties to be applied in productive systems without this issue being solved.

A further aspect not tackled in the presented approach is how query results are returned by the respective systems. Both systems have their own method and syntactical encoding for delivering results. For that topic, proper adapters would have to be developed as well. The example from Figure 3 for example returns complete patient objects, which are serialized differently by the respective systems. In order to prevent differences in serializations, the return types of translatable queries could be restricted to primitive types (e.g. patient ids instead of complete patient objects).

As the presented approach is still work in progress, it can be extended whenever the query language specifications are changed or extended and thus set of mutual operators grows. E.g. the AQL specification already contains announcements of future language elements like arithmetic operators, an extended *matches* operator or alias definitions (i.e. *let*).

Conclusions

An approach was presented to translate Clinical Quality Language (CQL) queries into Archetype Query Language (AQL) queries and vice versa. Several examples were shown to illustrate the capabilities of the presented apprach and for one example the translation process was illustrated in detail. As CQL and AQL do not comprise the same sets of operations, the translation capabilities of the presented approach are restricted to a subgroup of possible queries, which have to be composed of a set of common operators of both query languages. Despite this limitation, queries expressible in both languages can be automatically translated, which would allow both query systems to be transparently included in a distributed research network of the other type.

Acknowledgements

An implementation of the presented approach in Java is available at https://gitlab2.informatik.uni-wuerzburg.de /gef18bg/cdw_querymapper. The authors thank B. Haarbrandt, A. Zautke and G. Vella for helpful comments. This research was funded by grant of German Federal Ministry of Education and Research (Comprehensive Heart Failure Center Würzburg, grants #01EO1004 and #01EO1504).

References

[1] Prokosch H, Ganslandt T, Perspectives for medical informatics. Reusing the electronic medical record for clinical research, *Methods Inf Med* **48(1)** (2009), 38-44.
[2] Mandel JC, Kreda DA, Mandl KD, et al, SMART on FHIR: a standards-based, interoperable apps platform for electronic health records, *J Am Med Inform Assoc* **23(5)** (2016), 899-908.
[3] Murphy S, Weber G, Mendis M et al, Serving the Enterprise and beyond with Informatics for Integrating Biology and the Bedside (i2b2), *J Am Med Inform Assoc* **17(2)** (2010), 124-30.
[4] Kalra D, Beale T, Heard S, The openEHR Foundation. *Stud Health Technol Inform* **115** (2005), 153-73.
[5] Overhage JM, Ryan PB, Reich CG et al, Validation of a common data model for active safety surveillance research, *J Am Med Inform Assoc* **19(1)** (2012), 54-60.
[6] Weber GM, Murphy SN, McMurry AJ, et al, The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories, *J Am Med Inform Assoc* **16(5)** (2009), 624-30.
[7] Hailemichael MA, Marco-Ruiz L, Bellika JG, Privacy-preserving Statistical Query and Processing on Distributed OpenEHR Data, *Stud Health Technol Inform* **210** (2015), 766-70.
[8] Haarbrandt B, Tute E, Marschollek M, Automated population of an i2b2 clinical data warehouse from an openEHR-based data repository, *J Biomed Inform*, **63** (2016), 277-94.
[9] Fette G, Kaspar M, Liman L, et al. Query Translation between openEHR and i2b2. (in press) *Stud Health Technol Inform* (2019)
[10] Calvanese D, De Giacomo G, Lenzerini M, et al. What is query rewriting?, *Proc. of KRDB* (2000), 17–27.
[11] Droop M, Flarer M, Groppe J, et al. Embedding Xpath Queries into SPARQL Queries, *ICEIS* (2008).

Address for Correspondence

Georg Fette, georg.fette@uni-wuerzburg.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190198

Interactive Machine Learning for Laboratory Data Integration

Nathanael Fillmore[a], Nhan Do[a,b], Mary Brophy[a,b], Andrew Zimolzak[a,c]

[a] *Massachusetts Veterans Epidemiology Research and Information Center, VA Boston Healthcare System, Boston, MA, USA*
[b] *Boston University School of Medicine, Boston, MA, USA*
[c] *Baylor College of Medicine, Houston, TX, USA*

Abstract

Laboratory data collected in the electronic health record as part of routine care can be used in secondary research. For example, the US Department of Veterans Affairs maintains a data warehouse covering over 20 million individuals and 6.6 billion lab tests. However, data aggregation in such a data warehouse can be difficult. In order to retrieve all or nearly all of one type of lab result with a high degree of precision, we perform clinical concept adjudication, which is the process of an expert determining which database records correspond to a target clinical concept. In this work, we develop an interactive machine learning tool to "extend the reach" of expert laboratory test adjudicators. Our tool provides access to automatic laboratory classification in a user-facing front end that covers all steps in an adjudication workflow, in order to lower barriers to collaboration, increase transparency of adjudication, and to promote efficiencies and data reuse.

Keywords:

Clinical Laboratory Information Systems; Systems Integration; Supervised Machine Learning

Introduction

Clinical laboratory data are crucial to medical research, including retrospective studies and clinical trials. Laboratory data collected in the electronic health record (EHR) during the care process can be used in research (termed "secondary use") [1]. For example, serum creatinine lab test results, collected during routine clinical care, serve as safety endpoints for a "point of care" trial comparing the efficacy of antihypertensives [2]. Research requires all serum creatinine results to be retrieved using a unique database identifier, but reality is not this simple.

The United States Department of Veterans Affairs (VA) maintains a data warehouse covering 20 million individuals and 6.6 billion lab tests (as of 2014) [3], but as others have noted "data aggregation across the VHA is highly problematic" [4]. As one example, if we search for "creatinine" in the warehouse's LabChemTest dimension table, we find >1000 matching lab test types, many of which are not truly serum creatinine. If the query requires "creatinine" followed by "serum," it retrieves a shorter (64 lab types) and much more specific list, but many true positives are missed. Therefore, in order to retrieve all or nearly all the serum creatinine results with a high degree of precision, we perform *clinical concept adjudication*, which is the process of an expert determining which database records correspond to a target clinical concept.

The clinical concept adjudication process currently used in our research center has been previously described [5] and is designed to harmonize lab test results from 144 independent VA clinical laboratories (Figure 1). In brief: (1) Clinical subject matter experts (SMEs) design a search for appropriate laboratory test names. For example, if hemoglobin is the target concept, search terms might include "HGB" or "Hemoglobin", and several Logical Observation Identifiers Names and Codes (LOINCs). (2) Database technicians retrive candidate database records. For each record, the technicians also pull associated metadata, including specimen types (e.g. whole blood, urine, cerebrospinal fluid), units, distribution of numeric results, and laboratory test names. Many SMEs do not have technical expertise to perform a database pull, or they do not have permission to execute such a pull. Database technicians deliver these results to SMEs using an Excel spreadsheet that can be sorted and filtered (Table 1). (3) Two SMEs label each record, evaluating for appropriate specimen types, units, distribution, and test names. (4) SMEs resolve disagreements, producing a consensus labelling. (5) Database IDs and consensus labels of

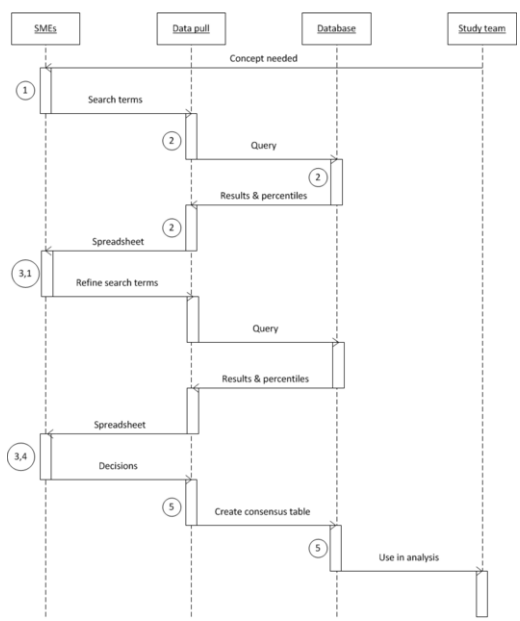


Figure 1– Sequence diagram. The workflow between subject matter experts, data technicians, the database, and the study team is illustrated.

Table 1– Example data used in adjudication. The task is to label each row, identified by the ID column, with a label, shown in the SME Label column, based on data in the remaining columns. The Lab Test Name, Topography, and LOINC columns identify the name and specimen type of the lab test. The Count column shows the number of lab test results associated with the ID, and p1, p50, and p99 (as well as others not shown) describe the first, 50th, and 99th percentiles of values of these results.

ID	SME Label	Lab Test Name	Topography	LOINC	Count	p1	p50	p99
1	Yes	SODIUM	SERUM	Missing	115053	126	140	149
2	No	RANDOM URINE SODIUM	URINE	Missing	734	6	52	194.1
3	No	SODIUM	URINE	Missing	89	5	49.5	155.9
8	Yes	SODIUM	SERUM	2947-0	126	133.2	140	144.7
9	No	SODIUM	URINE,24HR	2947-0	98	13.8	150	877.7
92	No	SODIUM	PERITONEAL	2950-4	10	124	132	138.8
461	Yes	SODIUM*IA	BLOOD	2950-4	714	125	139	170.9

"yes" or "no" for each record are stored in a new database table, and the spreadsheet is stored for future reference. Note that steps (1)-(3) are often iterated if the search terms need refinement.

Several drawbacks to this process exist. First, several steps in this process are time-consuming. Step (2) can take hours to run, because it may require computing on millions of values for each database ID. Step (3) requires SMEs ultimately to address each row (often about 1000) of a spreadsheet. The current process requires manual transmission of information and requires the SME and database technician to wait on each other. Second, the adjudicated concept can go out-of-date as new database records are added. Third, it is difficult for the final consumer (i.e. the study team performing analyses) to understand, validate, or adapt the new database table that contains the final adjudicated concept. This is partly because of the manual transmission of work products, which can result in the initial search terms (or other work products) being lost, or at least not available to final consumers.

Therefore, we sought to develop a machine learning tool to "extend the reach" of expert laboratory test adjudicators, so that they do not need to classify each data element manually. We aimed to provide access to automatic laboratory classification in a SME-facing front end for continued use by adjudicators. We intend our tool to cover all steps in the workflow (including the initial search for laboratory test names, the consensus process, and the database storage of SME decisions), in order to lower the barrier to collaboration among the group, to increase transparency of the adjudication process, and to promote efficiencies and reuse of adjudicated laboratory tests across the entire VA.

Prior work. Previous authors have faced similar lab result harmonization problems. For instance, the LOINC standard has been developed to identify clinical laboratory test results [6], and previous authors have described mapping their local data to this standard [7], but mappings of local laboratory tests to LOINC may be erroneous [8]. In another example, the Mini-Sentinel program received laboratory results from twelve data partners and encountered inconsistent units and LOINC availability, among other challenges, many of which were addressed by manual quality checking [9]. As an example from VA data, we found LOINC wrongly coded as body fluid sodium (2950-4) when it should be coded as serum/plasma (Table 1). This means that LOINC alone cannot be used to accurately retrieve all test types for a single clinical concept.

Early work has shown that lexical comparison of test names followed by subject matter expert review can integrate a more detailed set of laboratory test codes (LOINC) with a less detailed one (SNOMED procedures) [10]. To facilitate the mapping of local lab terms to LOINC, The Regenstrief LOINC Mapping Assistant (RELMA) was developed [11,12]. As a practical example of lab test harmonization, a large group of 44 European hospitals developed an internal dictionary of laboratory observations and mapped it to LOINC, but they found that some tests were not referenced in LOINC [13]. More advanced work has been been done to share results for LOINC mapping, to take advantage of "crowdsourcing" [14,15], and to perform automated mapping to LOINC using machine learning, which assigned about 70-80% of local terms to the correct LOINC [16]. Finally, work has been performed to address laboratory unit conversions [17] and standardizing laboratory test *results* across multiple institutions, as distinct from standardizing simply the type of test [18].

In our data warehouse, lab tests are already mapped to LOINC, but imperfectly. We also often need to aggregate several granular LOINCs into one more general concept. Our current lab adjudication process accomplishes this, but with many roles and manual steps (Figure 1).

Methods

We developed a machine learning system to predict whether a candidate lab test type clinically belongs within the concept of interest, based on a set of lab tests already labelled by SMEs as belonging or not. Seven lab tests were studied: alkaline phosphatase (ALP), alanine transaminase (ALT), albumin (ALB), high-density lipoprotein cholesterol (HDLC), sodium (Na), magnesium (Mg), and hemoglobin (HGB). We chose the following initial features for use by these algorithms. We used a bag-of-words encoding for longer textual fields like the test name, topography, component, and specimen. We used a categorical encoding for short textual fields, including the VA hospital site identifier, the lab result units, and the LOINC. We used as-is the numerical fields describing the number of results and their distribution. We included as a feature a Kolmogorov-Smirnov statistic that compares, for one lab test type, the distribution of its results to the distribution of all positive training examples' results [19].

Table 2– Machine learning results (accuracy) for each of 7 lab tests, using 3 different learning methods: LASSO (LR), support vector machines (SVM), and random forests (RF). Also shown are the total number of database IDs, the number of IDs labelled positive in the "ground truth" SME labelling, and the number of associated observations (lab test results).

Data Set	Number of Database IDs	Number of Positive Database IDs	Number of Observations	LR	SVM	RF
ALP	1588	747	38,297,131	0.98	0.97	0.98
ALT	716	341	84,180,150	0.98	0.94	0.92
ALB	3351	601	114,613,056	0.97	0.92	0.98
HDLC	742	358	79,767,824	0.98	0.91	0.98
Na	322	153	7,821,081	0.97	0.98	0.99
Mg	2066	502	17,768,685	0.97	0.95	0.99
HGB	1737	830	135,382,741	0.97	0.95	0.99

We evaluated this basic system using seven datasets that had been already adjudicated by VA experts (Table 2), and using three algorithms: logistic regression with an L1 penalty (also known as the least absolute shrinkage and selection operator, or LASSO), support vector machines (SVM), and random forests. Initial development and testing of learning algorithms was performed in Python using the scikit-learn package [20]. We used 10-fold cross validation to evaluate the accuracy of the system using each algorithm.

After the basic learning system was developed and validated, an interactive adjudication interface was implemented as a web application, as described in detail below. This system is available to VA SMEs in a hosted environment, as well as to the public at https://github.com/nfillmore/active-adjudication.

Results

Machine learning

Ten-fold cross-validation accuracy was high for all seven laboratory tests, and for all three methods (Table 2). For 6 out of 7 laboratory tests, random forests achieved the top cross validation accuracy, and LASSO achieved second place or better. Random forests achieved the fastest convergence per training example, and LASSO consistently achieved >90% cross validation accuracy at 100 or fewer training examples (Figure 2).

Operationalizing as web application

After the basic learning system was developed, an interface was designed as a web application, with a front-end written in JavaScript (Figure 3). We implemented the back-end (including LASSO) entirely in Java, which we found removes the need for any research group to install their own machine learning pipeline. The application covers the the laboratory adjudication workflow from initial assignment of the lab concept to a SME and generating search terms, to saving the final consensus adjudication for study use. The web application is hosted in a secure environment where VA SMEs can access it, and its source code is available for use by the public at https://github.com/nfillmore/active-adjudication.

Support for search term generation

Recall that in the current process for lab adjudication, data pull personnel must wait on SMEs to generate search terms, and SMEs then wait to receive an Excel spreadsheet. In order to avoid this inefficiency, we integrated the initial search into our interactive tool. The SME specifies search terms and can see results immediately. We accelerate this process with pre-computed value percentiles for all lab test identifiers in the VA system (this took over a month of wall clock time). After refining the search terms, the SME can begin labelling examples, although the query can be modified later without losing any labels already input.

Support for adjudication

The labeling interface allows SMEs to view both the original table of data elements and the feature matrix for the learning algorithm. These tables can be viewed separately or side-by-side, with support for sorting by any column. As the SME labels examples, statistics are recomputed in real time. The default feature matrix is as described above for the basic system, but SMEs may add or remove any features (a process termed *feature engineering*), including features corresponding to a

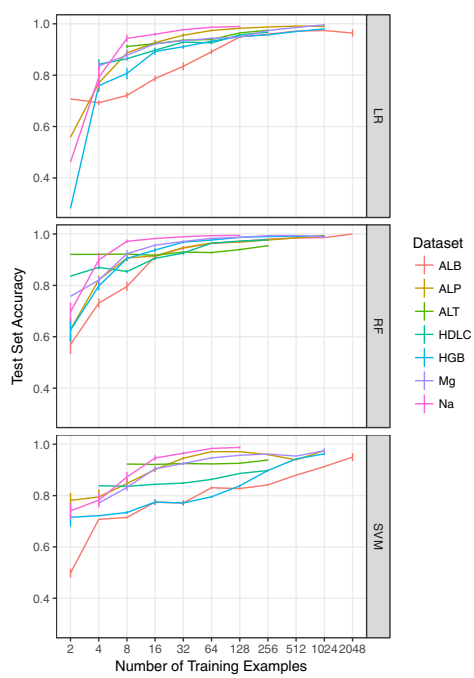


Figure 2– Learning curves for LASSO (LR), random forests (RF) and support vector machines (SVM)

single word or phrase. For example, if the SME needs to retrieve measurements of serum immunoglobulin free light chains, the phrase "free light" is more specific than either "free" or "light" by itself, and thus the SME may want to add this two-word feature, which is not available to the learning system by default.

Support for consensus labelling

The current adjudication process requires SMEs to resolve labelling conflicts. Our application facilitates this by displaying any conflicts and prompting for a consensus label. As in the main labelling interface, the application uses machine learning to predict the consensus label.

Support for finalizing and sharing results

Because lab adjudication does not involve any protected health information, adjudicated labels can be freely shared among different researchers, but there is currently no simple mechanism to do this within VA's nationwide network of researchers. Therefore, our tool offers such a mechanism. Each adjudication can be published to a searchable list of "Completed Adjudications", tagged with the target lab, the group that carried out the adjudication, the date of completion, and other information. These adjudications can be viewed and exported for use in individual research projects. In some cases, an adjudication might be close to what a research group is looking for, but not an exact match for their target concept. Thus, the tool also allows published adjudications to be "forked", i.e., copied into a new adjudication instance, where labels and search criteria can be modified as needed, exactly as if the adjudication had been created from scratch by the new researcher. The lineage of forking is tracked and displayed to users, to promote appropriate credit as well as transparency. To promote division of labor, even incomplete adjudications are viewable and forkable by other researchers, but these are clearly marked as incomplete.

Discussion

We have developed a tool that uses interactive machine learning to assist lab adjudication experts. Among the three learning methods tested, LASSO performed nearly as well as the others, with the advantage that it is easier to interpret the model's predictions. We found no large differences in classifier performance among the seven lab tests for which we simulated annotation. Because of the rapid convergence of the learning methods, only about 100 lab data elements need to be adjudicated, and the rest (about 1000, depending on the lab test) can be inferred with high accuracy, which is theoretically a 10-fold improvement in adjudication time. In practice, the expert could label all data elements with > 1000 observations and allow the system to predict the rest.

SMEs within the VA now have access to machine learning assistance for laboratory data integration, without the need to implement a learning pipeline. In addition, we have precomputed and stored percentiles of numerical laboratory values, so these are available rapidly to SMEs. Our web application covers all steps of the laboratory adjudication workflow (Figure 1, Figure 3), including capturing the initial query, and writing to the database for future use.

Our system can also be used by SMEs outside the VA, after creation of an appropriate backing database of precomputed metadata and percentiles of laboratory values. Creating such a database is straightforward and instructions are available with the software.

Figure 3– Screenshots of our tool: (A) Adjudication task creation. (B) Interactive annotation and machine learning for adjudication. (C) Sharing adjudicated results. Details on these steps are in the text.

We envision several feasible improvements to our system. First, a quantitative study of SME time saved could be undertaken. To facilitate this, we have already incorporated time-stamp logging of user actions into the software. Second, we could adapt the system to monitor the database for updates and ask for new labels as appropriate to keep concepts up-to-date. Third, we could use an approximation scheme for streaming computation of percentiles, so that the precomputed numerical distributions do not go out of date as new observations are collected [21].

In comparison to RELMA, we incorporate collaboration among multiple roles (e.g. SME, database technician), and we incorporate the aggregation of multiple potential LOINC code points. Machine learning has been applied to lab data cleaning [16], but to our knowledge *interactive* machine learning has not. Our web application covers all parts of the clinical laboratory test adjudication workflow, greatly decreasing the work required of SMEs, and it introduces further advantages by capturing work products that otherwise could be lost, increasing transparency, and encouraging sharing.

Conclusions

In this work, we developed an interactive machine learning tool for laboratory test adjudicaton. Our tool provides access to automatic laboratory classification in a SME-facing front end for continued use by adjudicators. The tool covers all steps in an adjudication workflow, and in doing so, lowers barriers to collaboration, increases transparency of adjudication, and promotes efficiencies and data reuse.

Acknowledgements

This material is based upon work supported by the United States Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, Cooperative Studies Program (CSP) and the Veterans

Affairs/National Cancer Institute Big Data-Scientist Training Enhancement Program (BD-STEP). The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.

References

[1] MIT Critical Data, Secondary Analysis of Electronic Health Records, 2016. doi:10.1007/978-3-319-43742-2.

[2] F.A. Lederle, W.C. Cushman, R.E. Ferguson, M.T. Brophy, and L.D. Fiore Md, Chlorthalidone Versus Hydrochlorothiazide: A New Kind of Veterans Affairs Cooperative Study., *Ann. Intern. Med.* **165** (2016) 663–664. doi:10.7326/M16-1208.

[3] S.D. Fihn, J. Francis, C. Clancy, C. Nielson, K. Nelson, J. Rumsfeld, T. Cullen, J. Bates, and G.L. Graham, Insights from advanced analytics at the Veterans Health Administration., *Health Aff. Proj. Hope*. **33** (2014) 1203–1211. doi:10.1377/hlthaff.2014.0054.

[4] B.P. Giroir, and G.R. Wilensky, Reforming the Veterans Health Administration — Beyond Palliation of Symptoms, *N. Engl. J. Med.* **373** (2015) 1693–1695. doi:10.1056/NEJMp1511438.

[5] S. Raju, Y.-L. Ho, A. Zimolzak, B. Katcher, K. Cho, and D. Gagnon, Validation of Laboratory Values in a Heterogeneous Healthcare System: The US Veterans Affairs Experience., *31st Int. Conf. Pharmacoepidemiol. Ther. Risk Manag. ICPE Boston 822-262015*. (n.d.).

[6] A.W. Forrey, C.J. McDonald, G. DeMoor, S.M. Huff, D. Leavelle, D. Leland, T. Fiers, L. Charles, B. Griffin, F. Stalling, A. Tullis, K. Hutchins, and J. Baenziger, Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results., *Clin. Chem.* **42** (1996) 81–90.

[7] A.N. Khan, S.P. Griffith, C. Moore, D. Russell, A.C. Rosario Jr, and J. Bertolli, Standardizing laboratory data by mapping to LOINC, *J. Am. Med. Inform. Assoc. JAMIA*. **13** (2006) 353–355. doi:10.1197/jamia.M1935.

[8] M.-C. Lin, D.J. Vreeman, C.J. McDonald, and S.M. Huff, Correctness of Voluntary LOINC Mapping for Laboratory Tests in Three Large Institutions, *AMIA Annu. Symp. Proc. AMIA Symp.* **2010** (2010) 447–451.

[9] M.A. Raebel, K. Haynes, T.S. Woodworth, G. Saylor, E. Cavagnaro, K.O. Coughlin, L.H. Curtis, M.G. Weiner, P. Archdeacon, and J.S. Brown, Electronic clinical laboratory test results data tables: lessons from Mini-Sentinel., *Pharmacoepidemiol. Drug Saf.* **23** (2014) 609–618. doi:10.1002/pds.3580.

[10] R.H. Dolin, S.M. Huff, R.A. Rocha, K.A. Spackman, and K.E. Campbell, Evaluation of a "lexically assign, logically refine" strategy for semi-automated integration of overlapping terminologies., *J. Am. Med. Inform. Assoc. JAMIA*. **5** (1998) 203–213.

[11] C.J. McDonald, S.M. Huff, J.G. Suico, G. Hill, D. Leavelle, R. Aller, A. Forrey, K. Mercer, G. DeMoor, J. Hook, W. Williams, J. Case, and P. Maloney, LOINC, a universal standard for identifying laboratory observations: a 5-year update., *Clin. Chem.* **49** (2003) 624–633.

[12] Regenstrief Institute, Inc, Common LOINC Laboratory Observation Codes, (n.d.). http://loinc.org/usage/obs.

[13] P.-Y. Vandenbussche, S. Cormont, C. Andre, C. Daniel, J. Delahousse, J. Charlet, and E. Lepage, Implementation and management of a biomedical observation dictionary in a large healthcare information system., *J. Am. Med. Inform. Assoc. JAMIA*. **20** (2013) 940–946. doi:10.1136/amiajnl-2012-001410.

[14] B.E. Dixon, J. Hook, and D.J. Vreeman, Learning From the Crowd in Terminology Mapping: The LOINC Experience., *Lab. Med.* **46** (2015) 168–174. doi:10.1309/LMWJ730SVKTUBAOJ.

[15] D.J. Vreeman, J. Hook, and B.E. Dixon, Learning from the crowd while mapping to LOINC., *J. Am. Med. Inform. Assoc. JAMIA*. **22** (2015) 1205–1211. doi:10.1093/jamia/ocv098.

[16] M. Fidahussein, and D.J. Vreeman, A corpus-based approach for automated LOINC mapping., *J. Am. Med. Inform. Assoc. JAMIA*. **21** (2014) 64–72. doi:10.1136/amiajnl-2012-001159.

[17] R.G. Hauser, D.B. Quine, and A. Ryder, LabRS: A Rosetta stone for retrospective standardization of clinical laboratory test results., *J. Am. Med. Inform. Assoc. JAMIA*. **25** (2018) 121–126. doi:10.1093/jamia/ocx046.

[18] R.G. Hauser, D.B. Quine, A. Ryder, and S. Campbell, Unit conversions between LOINC codes., *J. Am. Med. Inform. Assoc. JAMIA*. **25** (2018) 192–196. doi:10.1093/jamia/ocx056.

[19] N. Smirnov, Table for Estimating the Goodness of Fit of Empirical Distributions, *Ann Math Stat.* **19** (1948) 279–281. doi:10.1214/aoms/1177730256.

[20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, Scikit-learn: Machine Learning in Python, *J Mach Learn Res*. **12** (2011) 2825–2830.

[21] T. Dunning, and O. Ertl, Computing Extremely Accurate Quantiles using t-Digests, (2017). https://github.com/tdunning/t-digest/blob/master/docs/t-digest-paper/histo.pdf.

Address for correspondence

Nathanael Fillmore (Nathanael.Fillmore@va.gov) and Andrew Zimolzak (Andrew.Zimolzak@va.gov)

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190199

A Concept for Graph-Based Temporal Similarity of Patient Data

Matthias Ganzinger, Jens Schrodt, Petra Knaup

Institute of Medical Biometry and Informatics, Heidelberg University, Heidelberg, Germany

Abstract

Computer-based decision support systems are often used for dedicated tasks such as the detection of sepsis. However, positive predictive values for sepsis detection are reported to achieve only around 46%. In this paper we describe a novel approach to use temporal data of electronic patient records based on similarity measures. We apply the concept of case-based reasoning, which is well-established in many fields of medical informatics. Temporal patient data are organized in a time-graph structure. For the quantification of similarity between cases, we exploit graph theory based approaches. For development and evaluation of our time-graph similarity frame we use the open MIMIC III dataset. In a later phase, we envision to transfer our concept from sepsis to other diseases.

Keywords:

Medical Informatics, Computer-Assisted Decision Making.

Introduction

Computer-based clinical decision support systems (CDS) are important elements for improving the quality of patient care. A relevant clinical example is the early detection and management of sepsis. Sepsis is a systemic reaction of the human body against inflammatory infections with approximately 154,000 diagnosed cases in Germany per year leading to as many as 56,000 deaths per year [1]. Early detection and start of treatment are crucial factors for improving survival and complete recovery [2].

CDS have been a field of research in medical informatics for several decades. In recent years, the field was complemented by data science approaches analyzing vast amounts of patient-related and disease-related data in order to build computer models of diseases. Terms commonly used for such concepts are systems medicine, personalized medicine, or precision medicine. One of the ideas that are currently under discussion in systems medicine is to identify similar patients or cohorts within comprehensive patient-related databases such as electronic health record (EHR) systems [3]. However, the concept of providing clinical decision support on the basis of treatment experiences made with past patient cohorts has been established in medical informatics as case-based reasoning (CBR) already in the 1980s based on previous works in cognitive sciences [4, 5].

Currently, temporal data are rarely used in similarity-based decision support systems. To change this, it would be necessary to bring existing temporal data into a form that supports reasoning on these data. For CBR, the case of a patient would have to be described considering all relevant temporal events. Such a temporal profile or network of different event values over time could be used to find a similar patient in the case base, whose case shows a similar sequence of temporal events in addition to the similarity of attribute values. This could lead to improved clinical conclusions for diagnosis and therapy or medication. For our example of sepsis, this would mean to compare the course of disease of current patients with those who had developed a sepsis in the past in order to detect sepsis in an early stage to save time until treatment is started.

Since there are no established approaches for temporal similarity in CBR, we describe a possible model for the development of such a similarity measure in the course of this manuscript. This includes potential forms of representing temporal data for similarity and reference data sets for the development process.

Patient Similarity and CBR

Technically, historic treatment information is represented as a patient case. Each case is described by a set of attributes, often represented as structured fields in the database of an EHR system. To quantify similarity of two individual cases, a so-called local similarity measure is defined to compare the values of a certain attribute for the two cases. Global similarity between two cases is gained by computing and weighting the importance of attributes and merging local similarity measures into a single value. Great care has to be taken to choose an appropriate similarity measure that takes into account the characteristics of the respective attributes. For example, the similarity of the body weights of two patients might be calculated as the Euclidian distance between the values. For attributes with discrete value domains, such as ABO blood groups, the context for which similarity is being calculated has to be considered. For example, when looking at the risk of developing coronary heart disease (CHD), blood group values of A, B, and AB might be considered more similar to each other than compared to O. The reason is that bearers of blood group O seem to have a significantly lower risk to develop CHD than bearers of the other groups [6]. For other diseases, a different assumption of similarity might be more conclusive. In these cases, value distributions or expert knowledge can be used to establish a numeric similarity value.

Due to the development of big data analysis initiatives, more and more patient related data sources are expected to become available in a structured way for research on clinical decision support systems. Using these sources for finding similar patients is only possible if suitable similarity measures exist and are correctly chosen for the respective attributes.

Our group has conducted research on similarity measures in a systems medicine research project on multiple myeloma. For this project we developed an IT architecture for systems medicine applications that rely on the concept of patient similarity. As one result we established a novel similarity measure that considers survival data (e.g., progression free survival, or overall survival) as they are measured in clinical trials for the computation of similarity. However, in the course of our research we found additional aspects of patient similarity that have not been taken into consideration sufficiently so far

due to the lack of appropriate similarity measures. Specifically, these additional aspects are temporal relations among symptoms and events like diagnostic and therapeutic measures. In the everyday life of a physician it is often essential to know the sequence of such events to get an appropriate diagnosis [7].

Decision Support for Sepsis Diagnosis

Supporting early diagnosis and management of sepsis has been subject to previous studies. Corfield et al. describe a single, not computer-based, score as an early warning sign for sepsis [8]. In terms of computer-based CDS, one approach is to generate sepsis alerts if a number of criteria is met by entries in the electronic medical record of a patient [9]. While this system had no false negative results, its positive predictive value was only 44.7%. Amland et al. report on a similar system with a likewise positive predictive value of 46% [10]. Such results suggest the necessity for further research to improve accuracy of computer-based CDS for sepsis. Considering temporal aspects might help to fill this gap [11].

Temporal Abstraction

Temporal abstractions are qualitative representations of temporal intervals with respect to a specific context [12]. Such abstractions are used to derive higher level concepts from raw, quantitative data. For example, individual measurements of body temperature could be aggregated to intervals and labelled with qualitative descriptors like “normal temperature” or “hyperthermia”. A widely used set of temporal abstractions has been defined by Allen [13, 14]. He describes a system of thirteen basic relations such as “before”, “overlaps”, or “during”, covering all relations that are possible between two temporal intervals. Temporal abstractions can be used to reduce complexity when working with temporal dependencies.

Temporal Data Structures

Temporal relations among health related events could be an important source of information for diagnosis and prognosis of patients. Thus, such relationships in patient data are in focus of many research activities such as temporal mining. A basic question for the development of a temporal similarity measure is how data should be organized and stored for efficient processing. Traditionally, medical events are recorded in the form of tables either as paper charts or electronic database tables. To support temporal information, the 2001 edition of the SQL standard query language for relational databases supports temporal tables. However, this extension does not focus on working with temporal relations among entities, but provides means to model the time when entries are valid in the real-world context. This is similar to conserving the state of a database into a snapshot which makes it possible to reproduce the data for time points in the past. So the word ‘temporal’ has a different meaning in the SQL standard. As a consequence, temporal relations have to be implemented outside the database management system into an application with the potential requirement to transfer large amounts of data. The goal of such an application should not be to save different states of database tables over time but to save and compare temporal data like patient events in a timeline.

Time-Graphs

An intuitive representation of temporal relationships among medical data might be graphs or networks. Medical events could be presented as nodes while edges show the temporal relations among them. Interestingly, the term *temporal network* is widely used in computer science, but usually not in the sense that the network shows temporal sequences. Instead, the term describes the research on the development of networks over the time. In research on social networks, for example, it is investigated how the contacts of a person change by comparing the respective network at different points in time [15].

To avoid confusion with such temporal developments of graphs, the term *time-graph* is used for sequential graphs throughout this proposal.

A graph-based representation of temporal EHR data has been shown to work by Liu et al. in a paper on temporal phenotyping [16]. They introduce temporal phenotypes as a set of basis graphs that are combined to construct an observed temporal graph. The authors describe the application of similarity-based regularization for calculating the temporal phenotypes. However, the aim of their work is not to identify similar patients in the sense of CBR but to predict medical events such as the onset risk of heart failure.

Thus, additional research is necessary to establish time-graphs for the use of finding patients with similar temporal disease development. This could be achieved by calculating the similarity of time-graphs each representing one patient.

Graph Similarity

Graphs have been studied for a long time in graph theory. One problem related to comparing two graphs is the task of determining their structural identity or the existence of an isomorphism between them. In addition, quantifying the similarity of graphs and subgraphs has been of interest in many application areas like analyzing the world wide web, business process models (BPM), timetables, schema matching, or chemical structures [17–21]. Comparable results have not yet been researched for time-graphs. No specific time-graph similarity approach is known to the authors. Nevertheless, there are some approaches that are related to graph similarity, but it has not been investigated yet, if they can be applied for medical time-graphs.

One generic approach that has been demonstrated to work with BPM, for example, is the *graph edit distance*. It is capable of measuring similarity between two graphs by assessing the number of modifications necessary to transform one graph into the other [22]. A different method for quantifying similarity among two graphs is using an iterative framework based on the idea that two nodes or edges are similar if their neighbors are similar. Since similarity is propagated to adjacent elements during iterations, this method is also called similarity flooding [20]. Another method is *NetSimile*, which calculates different node-specific values, like the number of neighbors, the clustering coefficient of a node, and others. These attributes are finally aggregated into one value that describes the similarity of two graphs using the Canberra Distance [23]. These graph similarity approaches are candidates for being transferred to time-graphs.

Graph-Based Temporal Similarity

Taking up the concept of case-based reasoning, we develop a computer-based CDS for early detection of sepsis aiming for high accuracy. Specifically, we will develop a novel measure for calculating temporal similarities of disease progress among patients. This measure relies on graph-theory concepts. Later, we plan to extend our approach to other disease patterns.

Methods

The main research question of our project is as follows: How could temporal medical data of different patients be used to identify patients with a similar temporal profile for clinical decision support systems?

This question is divided into three sub-questions:

1. How can medical data be represented as a time-graph?

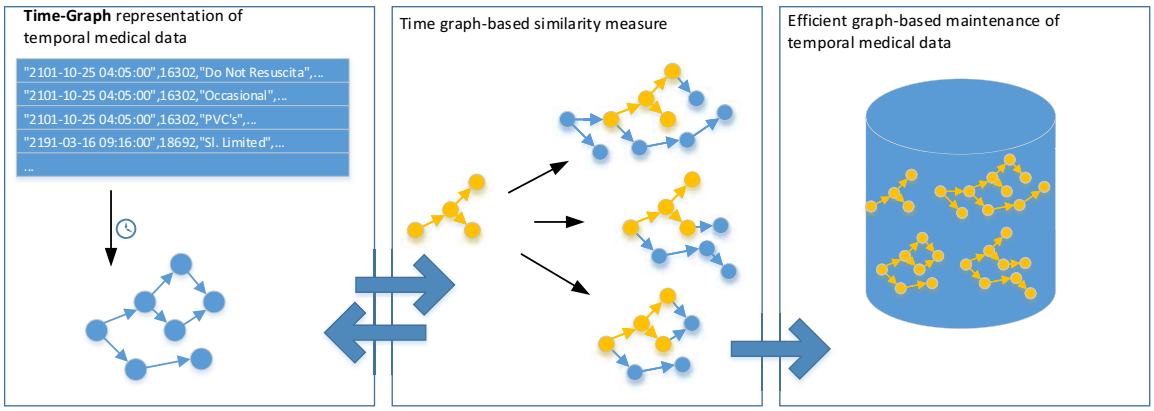


Figure 1 – Main Parts of the Time-Graph Similarity Framework.

2. What similarity measure can be used on time-graphs?
3. How can time-graphs be stored efficiently?

To answer these questions, our objective is to develop a new approach of making temporal aspects of medical data available in decision support systems, especially with respect to similarity-based approaches like case-based reasoning. We aim to establish graphs of clinical events as an innovative representation of temporal relations among these events. Here, we will consider concepts like networks, directed/undirected graphs, weights of edges, trees, and the significance of cyclic, i.e. reoccurring parts. The challenge is to find a graph-based representation of temporal medical data that provides an optimized basis for similarity quantification. To support case-based reasoning, a novel graph-based similarity measure for time-graphs in the form of a framework of methods is part of this approach. For this framework we investigate graph-theoretical approaches for the assessment of graph similarity. A set of criteria to classify these methods for applicability in the medical CBR context is established. As part of this development, graph similarity methods are tested in various disease-specific scenarios.

Finding the most similar patient with respect to his time-graph in a big case base can be very time-consuming, making efficient management of graph-based data structures essential. Thus, we investigate which kind of database management systems could form an appropriate basis for storing time-graphs. Specifically, we compare relational data base systems with temporal index structures with dedicated graph-based database systems. The exact requirements towards such database systems depend on the characteristics of the graph-based similarity measure.

Results

Our concept is comprised of three major blocks, corresponding to our research questions. First, research on the graph-based representation of temporal medical data is performed. Second, a novel similarity measure for temporal data in a graph representation is developed. In a third block, existing database technologies for efficient storage and similarity retrieval of graph-based temporal data are reviewed.

In addition, a user interface for temporal queries on the database is part of the concept. For research and testing of our approach, we use the MIMIC III database [24]. This dataset provides data of approximately 40,000 patients who were treated in an intensive care unit (ICU). All data are annotated with timestamps making them ideal for temporal research in clinical context.

Time-Graph Representation of Temporal Medical Data

As the first step, the MIMIC III database is analyzed with data mining tools for cases that can be used as blue prints for the development of a time-graph representation of the case documentation. In this phase, we consider cases suitable, if the temporal relation between events is known *a priori*. Medical treatment guidelines are helpful to identify such cases. For example, the guidelines for diagnosis and treatment of sepsis list a number of conditions that have to be met within a certain timespan. Thus, we will mine the database for patients who are diagnosed for sepsis and other diseases commonly occurring in an ICU context with temporal aspects. This will result in a set of reference cases with typical temporal disease progress is identified in the MIMIC III dataset.

Further, an appropriate graph-based representation for patient cases is established. Existing approaches will be reviewed: The most straight-forward approach is to form a timeline where all data entries are lined up in order of the respective timestamps. In this approach, each node has one edge pointing to the subsequent node. A timeline already allows to compare the temporal developments of different cases. However, in this naïve approach no difference is made between different classes of events. As a consequence, we will investigate more complex representations that also consider different classes of nodes. Nodes within each class have a linear temporal order, but are interwoven with nodes of other classes. In contrast to a timeline, such additional edges might help to better represent event-specific temporal relations. For example, known cause-effect relations such as the administration of antipyretics and the consecutive decrease of body temperature are modelled into the graphs.

Further research is made regarding the most suitable topology of the resulting graph. Possible topologies include directed graphs with or without cycles, weighted graphs, and trees.

Time-Graph-Based Similarity Measure

We investigate methods described in the literature for calculating the similarity between two graphs for their applicability on time-graphs. Aspects to be considered are the topology of the graphs and the possibility to work with sub-graphs. This is important, since the time-graph of medical events of a newly diagnosed patient is likely not as extensive as that of a patient who has finished treatment. We investigate similarity methods of different categories such as graph

isomorphism (e.g., edit distance based measures [22]), iterative methods like similarity flooding [20], and feature extraction methods like NetSimile [23].

The base methods for graph-based similarity are adapted to fulfil requirements of temporal graphs. For this development process, cases of the MIMIC III database are used as models and test cases. Criteria for the successful development of the similarity measure are defined during this task. They are based on plausibility checks and correlation analysis of the cases determined as similar by the new similarity measure. The graph-based temporal similarity measure developed in this work packages can be used by itself, but also in combination with conventional similarity measures based on attribute values. In this task, we investigate, how such different types of similarity measures can be combined to calculate a comprehensive similarity value. The approach will be based on the paradigm that global similarity is comprised as a combination of weighted local values. In this task, a suitable algorithm for combining the different similarity measures into one global value will be researched.

Efficient Graph-Based Maintenance of Temporal Medical Data

Electronic health record data are usually stored in relational databases with no specific focus on temporal data retrieval. Graph databases are potentially good candidates for high efficiency, especially in combination with a time-graph data model. In addition, in this kind of databases not only the raw data is saved like it is the case in traditional relational databases but also the connections between the different entities of the raw data are saved. This aspect makes graph databases potentially more efficient than relational databases but it still has to be investigated if this is still true for temporal data like it is used in our case.

We investigate if relational databases such as PostgreSQL can be used as an efficient storage for time-graphs. Specifically, we assess if existing SQL standard extensions support our graph-oriented similarity measure. We also evaluate proprietary temporal extensions that exist for some open-source database management systems. Further we investigate if dedicated index tables can support queries for similarity retrieval. The idea for such index tables is to pre-calculate temporal relations (respecting temporal abstractions) within cases to allow for fast retrieval without the need of calculating temporal relations at query time.

In contrast to relational databases, graph databases are designed to store graph data structures. In this task, we will investigate open source graph databases such as neo4j for their applicability in our context. Special focus lays on the efficient retrieval of similar cases. We will compare the respective query possibilities for retrieving cases based on subgraphs. One criterion for the assessment of the database management systems is the availability of features for efficient similarity retrieval. Further, the query times are compared for the different approaches.

For using the database system selected in tasks 3.1 and 3.2, data have to be prepared first. To do this, we will follow the extract, transform, and load (ETL) approach that is well-established for example in the context of data warehouses. Thus, we will develop ETL processes to make a subset of MIMIC III data available in our temporal database. Depending on the specific database management system, we will use open-source ETL software such as Talend Open Studio or KNIME to automate this process.

Discussion

In this manuscript, we present a comprehensive approach for the establishment of a graph-based similarity measure temporal patient data. The first step will be the development of a time-graph representation for temporal medical data as basis for the remaining parts. Further, efficient storage approaches and graph-based temporal mining are investigated.

In this phase, the development will be performed on the well-established open data set MIMIC III for patients suffering from sepsis. Being ICU derived data, this data set has a rather fine-grained temporal resolution compared to other fields of EHR data. After establishing this primary application, we are going to extend our research to other ICU related disease patterns as well as to other clinical areas, where documentation frequency and data types differ from ICU data. In addition, we are going to validate our findings with other clinical data sets.

References

[1] C. Engel, F.M. Brunkhorst, H.G. Bone, R. Brunkhorst, H. Gerlach, S. Grond et al., Epidemiology of sepsis in Germany: results from a national prospective multicenter study, *Intensive Care Med* **33** (2007), 606–618.

[2] A. Rhodes, L.E. Evans, W. Alhazzani, M.M. Levy, M. Antonelli, R. Ferrer et al., Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016, *Intensive Care Med* **43** (2017), 304–377.

[3] S.A. Brown, Patient Similarity: Emerging Concepts in Systems and Precision Medicine, *Front Physiol* **7** (2016), 561.

[4] J.L. Kolodner, Reconstructive memory: A computer model, *Cogn Sci* **7** (1983), 281–328.

[5] A. Aamodt, E. Plaza, Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches, *AI Communications* **7** (1994), 39–59.

[6] M. He, B. Wolpin, K. Rexrode, J.E. Manson, E. Rimm, F.B. Hu et al., ABO blood group and risk of coronary heart disease in two prospective cohort studies, *Arterioscler Thromb Vasc Biol* **32** (2012), 2314–2320.

[7] J.C. Augusto, Temporal reasoning for decision support in medicine, *Artif Intell Med* **33** (2005), 1–24

[8] A.R. Corfield, F. Lees, I. Zealley, G. Houston, S. Dickie, K. Ward et al., Utility of a single early warning score in patients with sepsis in the emergency department, *Emerg Med J* **31** (2014), 482–487.

[9] S.Q. Nguyen, E. Mwakalindile, J.S. Booth, V. Hogan, J. Morgan, C.T. Prickett et al., Automated electronic medical record sepsis detection in the emergency department, *PeerJ* **2** (2014), e343.

[10] R.C. Amland, K.E. Hahn-Cover, Clinical Decision Support for Early Recognition of Sepsis, *Am J Med Qual* **31** (2016), 103–110.

[11] J. Kim, J.M. Blum, C.D. Scott, *Temporal features and kernel methods for predicting sepsis in postoperative patients: Technical report*, University of Michigan, USA, 2010.

[12] M. Stacey, C. McGregor, Temporal abstraction in intelligent clinical data analysis: A survey, *Artif Intell Med* **39** (2007), 1–24.

[13] J.F. Allen, Towards a general theory of action and time, *Artif Intell* **23** (1984), 123–154.

[14] J.F. Allen, G. Ferguson, Actions and Events in Interval Temporal Logic, *J Logic Computation* **4** (1994), 531–579.

[15] Holme P, Saramäki J, editors, *Temporal Networks*, Berlin, Heidelberg, Springer Berlin Heidelberg, 2013.

[16] C. Liu, F. Wang, J. Hu, H. Xiong, Temporal Phenotyping from Longitudinal Electronic Health Records. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15. New York, New York, USA: ACM Press; 2015. p. 705–14.

[17] V.D. Blondel, A. Gajardo, M. Heymans, P. Senellart, P. van Dooren, A measure of similarity between graph vertices: Applications to synonym extraction and web searching, *SIAM Rev Soc Ind Appl Math* **46** (2004), 647–666.

[18] R. Dijkman, M. Dumas, L. García-Bañuelos, Graph Matching Algorithms for Business Process Model Similarity Search. In: Dayal U, Eder J, Koehler J, Reijers HA, editors. Business Process Management: 7th International Conference, BPM 2009, Ulm, Germany, September 8-10, 2009. Proceedings. Berlin, Heidelberg: Springer Berlin Heidelberg; 2009. p. 48–63 Available from: URL:http://dx.doi.org/10.1007/978-3-642-03848-8_5

[19] E.K. Burke, B. MacCarthy, S. Petrovic, R. Qu, Case-Based Reasoning in Course Timetabling: An Attribute Graph Approach. In: Goos G, Hartmanis J, van Leeuwen J, Aha DW, Watson I, editors. Case-Based Reasoning Research and Development. Berlin, Heidelberg: Springer Berlin Heidelberg; 2001. p. 90–104 (Lecture Notes in Computer Science).

[20] S. Melnik, H. Garcia-Molina, E. Rahm, Similarity flooding: A versatile graph matching algorithm and its application to schema matching. In: Proceedings 18th International Conference on Data Engineering: IEEE Comput. Soc; 2002. p. 117–28.

[21] J.W. Raymond, RASCAL: Calculation of Graph Similarity using Maximum Common Edge Subgraphs, *Comput J* **45** (2002), 631–644.

[22] H. Bunke, On a relation between graph edit distance and maximum common subgraph, *Pattern Recognit Lett* **18** (1997), 689–694.

[23] M. Berlingerio, D. Koutra, T. Eliassi-Rad, C. Faloutsos, NetSimile: A Scalable Approach to Size-Independent Network Similarity, Available from: URL:http://arxiv.org/pdf/1209.2684v1

[24] A.E.W. Johnson, T.J. Pollard, L. Shen, L.W.H. Lehman, M. Feng, M. Ghassemi et al., MIMIC-III, a freely accessible critical care database, *Sci Data* **3** (2016), 160035.

Address for Correspondence

Matthias Ganzinger
Institut für Medizinische Biometrie und Informatik
Im Neuenheimer Feld 130.3
69120 Heidelberg
Germany
Email: matthias.ganzinger@med.uni-heidelberg.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190200

Learning to Identify Severe Maternal Morbidity from Electronic Health Records

Cheng Gao[a], Sarah Osmundson[b], Xiaowei Yan[c], Digna Velez Edwards[a,b], Bradley A. Malin[a,c,d], You Chen[a]

[a]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States
[b]Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN, United States
[c]Sutter Research, Development and Dissemination, Sacramento, CA, United States
[d]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, United States

Abstract

Severe maternal morbidity (SMM) is broadly defined as significant complications in pregnancy that have an adverse effect on women's health. Identifying women who experience SMM and reviewing their obstetric care can assist healthcare organizations in recognizing risk factors and best practices for management. Various definitions of SMM have been posited, but there is no consensus. Existing definitions are further limited in that they 1) are often rooted in existing clinical knowledge (which is problematic as many risk factors remain unknown), leading to poor positive predictive performance (PPV), and 2) have limited scalability as they often require substantial chart review. Thus, in this paper, a machine learning framework was introduced to automatically identify SMM and relevant risk factors from electronic health records (EHRs). We evaluated this framework with EHR data from 45,858 deliveries at a large academic medical center. The framework outperformed a state-of-the-art model from the U.S. Centers for Disease Control and Prevention (AUC of 0.94 vs. 0.80). Specially, it improved upon PPV by 59% (CDC: 0.22 vs. our model: 0.35). In the process, we revealed several novel SMM indicators, including disorders of fluid or electrolytes, systemic inflammatory response syndrome, and acidosis.

Keywords:

Electronic health records, machine learning, severe maternal morbidity.

Introduction

Severe maternal morbidity (SMM) is a physical condition that results from, or is aggravated by, pregnancy and has an adverse effect on a woman's health. [1] SMM is an umbrella for various phenomena, including but not limited to, cardiomyopathy, hemorrhage, organ failure, pregnancy-induced hypertension and embolism, pulmonary embolism sepsis, seizure shock, severe amniotic fluid embolism, and uterine rupture. SMM can be considered a near miss for maternal mortality because, without in-time identification and treatment, these conditions can lead to maternal death [2; 3]. Thus, SMM is utilized as a new indicator of the quality of maternal health. [4] Its prevalence has steadily increased in low- and middle-income countries, as well as the developed world. [4-6] It now affects around 60,000 women in the U.S each year. [6] At the same time, the U.S. Department of Health Resource and Services Administration has increasingly focused on reducing the rate of SMM and improving its survival rate [7].

Identifying women who experience SMM and reviewing their obstetric charts can enable clinicians and healthcare organizations, more generally, to better understand the primary etiologies and contributing factors of SMM. In doing so, this can lead to a refinement of care management practices to prevent SMM or improve the quality of its management. However, there are major challenges in identifying SMM, which partially arise due to the lack of a consensus definition. For instance, the Centers for Disease Control and Prevention (CDC) has provided a definition based on 25 SMM indicators in the form of diagnosis and procedure codes [1]. The CDC definition is straightforward to specify and, thus, easy to implement in that International Classification of Disease (ICD) codes are readily available in hospital discharge data. However, it does not indicate the severity of maternal morbidity and, as a result, its application leads to many false positives and negatives. [8]

Another popular definition of SMM has been proposed by experts from the American College of Obstetricians and Gynecologists (ACOG) and the Society for Maternal-Fetal Medicine (SMFM). This definition is a guideline that allows for each birthing facility to determine their own SMM criteria [8; 9]. The guidelines are based on two primary factors: i) admission to an intensive care unit (ICU) or ii) transfusion of four or more units of blood. [7] These may be supplemented by other factors, such as intubation for at least 12 hours, organ system failure, and unanticipated surgical intervention [8]. Although SMM identification approaches built upon the guidelines can achieve high levels of performance [8], they have several deficiencies. First, the approach designed for one facility cannot be directly applied by another facility. This is because the risk factors used are often facility-specific and are unique unto a facility. [6] Second, the process of SMM identification requires substantial manual effort in the form of in-depth chart reviews, which incurs a nontrivial amount of time and effort.

Recent studies have shown that machine learning algorithms along with data in electronic health records (EHR) can be leveraged to identify or predict a variety of diseases [10-11]. For example, supervised machine learning algorithms were applied to identify different types of cancers and their progression pathways [11]. Another example [10] is the application of deep learning algorithms to predict Alzheimer's disease in its early stage. To identify SMM during delivery hospitalizations, we introduce a machine learning framework that incorporates data in EHRs. This framework has several notable benefits in that 1) it leverages standardized clinical terminologies as inputs, and thus models developed under the framework can be applied in EHRs of other facilities; 2) it can deal with over a thousand clinical concepts with reduced manual effort; and 3) it allows for the discovery of novel factors to identify SMM.

Methods

We begin this section with a description of the data workflow as shown in Figure 1. First, before feeding data into the classification model, we annotate EHRs as SMM cases and controls. The classification model is then evaluated in terms of area under the receiver operating characteristic (ROC) curve (AUC). Lastly, risk factors are ranked according to odds ratios and their contributions to the performance of the classification models.

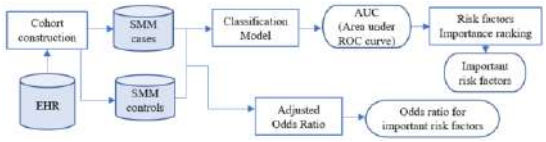


Figure 1– Data workflow

Dataset

We collected 13 years of data (2005 to 2018) from the Vanderbilt University Medical Center (VUMC) EHR system, which includes 45,858 deliveries for obstetric inpatients. This dataset includes patient demographics (e.g., age and race), admission notes, discharge summaries, and clinical concepts in the form of diagnoses (ICDs) and procedures (CPTs). A summary of the medical concepts and demographics are provided in Tables 1 and Table 2, respectively. We note that the diagnosis and procedure codes were transformed into standard terminologies defined under the Observational Medical Outcomes Partnership (OMOP) common data model, operated by the Observational Health Data Sciences and Informatics (OHDSI) consortia [12]. All these codes in this study are called medical concepts.

In Table 1, it can be seen that deliveries with SMM have more assigned diagnoses and procedures than those without SMM. This implies that deliveries with SMM are more complicated. There was a total 3,972 unique diagnosis concepts and 1,581 unique procedural concepts. In Table 2, it can be seen that there was a greater percentage of black women experiencing SMM (24%) than those without (18.5%).

Cohort construction

Table 1– Summary statistics of diagnosis and procedural concepts assigned to patients

Class	Statistiscs	# of Diagnoses Assigned	# of Procedures Assigned
SMM	Mean	31.3	22.7
	Median	27	20.5
	IQR	22	16
Non-SMM	Mean	9.7	8.1
	Median	8	8
	IQR	6	5

IQR: inter-quartile range

We use the criteria *four or more units of blood transfusion or ICU admission* developed by the ACOG to identify SMM candidates [8]. Using this criterion, 594 SMM candidates were identified. Obstetricians then reviewed the charts for these candidates and excluded 182 candidates who did not experience SMM. Our final cohort contained 412 SMM cases (0.9% of all deliveries). We assume that all of the remaining deliveries were controls. We acknowledge that the controls may contain a small number of SMM cases that cannot be captured by the ACOG criteria. A recent study [8] shows the number of missed cases is likely too small to impact the performance of identification models.

SMM identification models

Feature engineering

For each instance, we extract their assigned diagnoses and procedures represented by medical concepts using the OMOP common data model. Each medical concept is treated as a binary feature, where 1 indicates a patient was assigned the concept during the delivery and 0 otherwise.

Since the number of diagnoses (3,972) and procedures (1,581) is substantially larger than the number of positive instances, which may influence the performances of machine learning models. As such, we aim to do feature selection before applying machine learning models. To orient the system to be more generalizable, we adopt a filter feature selection algorithm without using class information (i.e., SMM vs. Non-SMM) and machine learning models. The feature selection algorithm ranks the concepts according to their frequency occurring in the investigated patient population. We incrementally add top ranking features into the machine learning models. Specifically, we train models using the top 10%, 20%, 40%, 80% and 100% of the features

Table 2– Summary statistics of patient demographics

Class	Age (years)		Race (%)	
SMM	Mean	28.1	White	259 (62.9%)
	Median	28	Black	99 (24.0%)
	Q1	23	Asian	13 (3.1%)
	Q3	33	Other	41 (10.0%)
No evidance of SMM	Mean	28.1	White	30,626 (67.4%)
	Median	28	Black	8,396 (18.5%)
	Q1	24	Asian	2,316 (5.1%)
	Q3	32	Other	4,108 (9.0%)

Q1: first quartile; Q3: third quartile

Balanced versus unbalanced cohorts

As noted earlier, the dataset is highly imbalanced - less than 1% of the women experienced SMM. As such, learning models based on the observed case to control ratio could lead to bias in the model. To mitigate this problem, we randomly sample controls (Non-SMM) from the data to investigate case:control ratios in the form of 1:1, 1:4, 1:16 and 1:64. In addition, we evaluated how the ratio impacts the performance of machine learning models to identify SMM from the natural case:control ratio.

Regularized logistic regression models

We train four types of models:

M_0: ⟨CDC-25, no learning⟩ The CDC model, which leverages 25 indicators to identify SMM. This model is considered as the baseline.

M_1: ⟨diagnoses, ridge logistic regression⟩ The features are diagnosis concepts and ridge logistic regression (L_2) is used as machine learning models. Under this category, we train regression models using top 10%, 20%, 40%, 80% and 100% of the diagnosis codes, and datasets with case:control ratios as 1:1, 1:4, 1:16 and 1:64.

M_2: ⟨procedures, ridge logistic regression⟩ The features are procedural concepts and ridge logistic regression is used as machine learning models. We use the same strategy in M_1 to train the regression models.

M_3: ⟨diagnoses and procedures, ridge logistic regression⟩ The features are diagnosis and procedural concepts and ridge logistic regression is used as machine learning models. We use the same strategy in M_1 to train regression models.

For each model, we apply 10-fold stratified cross-validation. The performance of models is measured in terms of sensitivity (SN), specificity (SP), positive predictive value (PPV), and AUC.

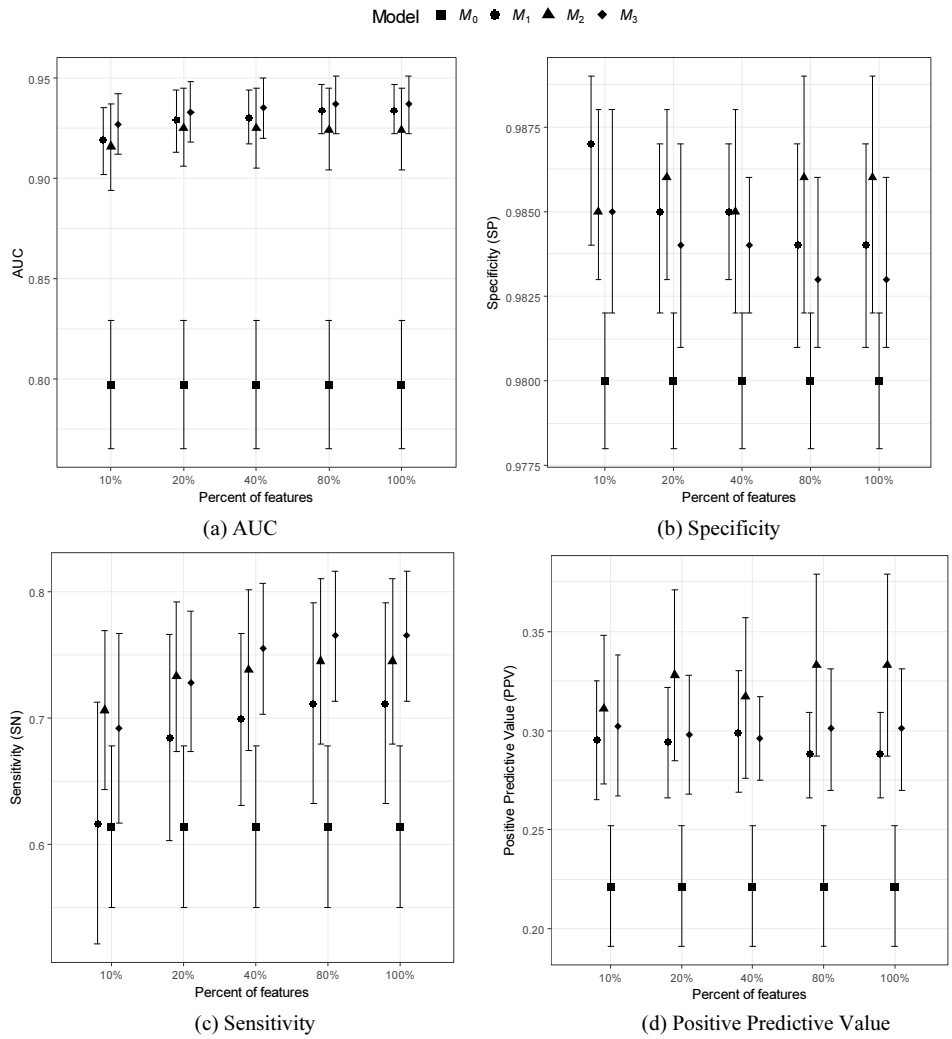
Medical concept ranking algorithm

To determine which medical concepts play more important roles in identifying SMM, we use the following equation to rank concepts:

$$F_{imp}^{j} = \sum_{i=1}^{10} \frac{AUC_i}{\sum_{i=1}^{10} AUC_i} \beta_i^j,$$

where F_{imp}^{j} is the overall importance of the j^{th} concept and β_i^j is the coefficient of the j^{th} concept for the i^{th} cross-validation model. The medical concept importance is adjusted by the model performance (AUC) and the importance of a concept within a model (coefficient). The larger the value of F_{imp}^{j} is, the more important the j^{th} concept is. For instance, imagine we train two models m_1 and m_2 on two medical concepts c_1 and c_2, such that the AUC of the models was 0.8 and 0.9, respectively. It was found that the coefficients of the two concepts in m_1 and m_2 was (0.6, 0.4) and (0.8, 0.2) respectively. According to our approach, the importance of medical concept c_1 is (0.8 / (0.8 + 0.9)) × 0.6 + (0.9 / (0.8 + 0.9)) × 0.8, or 0.71.

We learn a logistic regression model based on the medical concepts with the largest importance scores, along with age and race. Odds ratios are calculated for each feature in the SMM and non-SMM classes along with their corresponding p-values. The OR was adjusted for age and race and the p-value was subject to a Bonferroni correction.



(a) AUC (b) Specificity

(c) Sensitivity (d) Positive Predictive Value

Figure 2– SMM recognition performance as a function of the number of features in the logistic regression model

Results

Influence of the number of features

We use the natural proportion of cases to controls to train and validate models using the top 10%, 20%, 40%, 80% and 100% of concepts. The performance (in terms of AUC, SN, SP, and PPV and the corresponding 95% confidence interval) is depicted in Figure 2.

In Figure 2(a), it can be seen that M_1, M_2, and M_3 outperform the baseline M_0. Their performance improves as more features are added into these models, with an increase in AUC from 0.919 to 0.937. At each percentage of features, model M_3 exhibits the best performance of all three models. Model M_3 with 80% of features achieves the best AUC of 0.937, which is 18.6% better than M_0 (0.790).

Figures 2(b) and 2(c) show the specificity and sensitivity with respect to the percentage of features considered. It is observed that the best specificity is achieved at 10% of the features for M_1, improving by about 1% over the baseline model M_0, while the sensitivity improves by 25% from 0.614 for M_0 to 0.765 for M_3 (which occurs when 80% of the features are used). Figure 2(d) shows the PPV for the models. The PPV clearly improves through M_1, M_2, and M_3, where M_2 has the best PPV 0.347 (which occurs when 80% of the features are used), which is 57% better than that achieved by the baseline M_0.

Influence of the case:control ratio

Figure 3 shows the identification performance under different case:control ratios with all the features. It is notable that the performance does not always improve as the sampling ratio increases. M_1 and M_3 achieve the best AUC for the 1:64 sampling ratio while M_2 achieves the best AUC at 1:4.

Important concepts

We rank features using M_3 with 80% of concepts because it exhibits the best AUC under a natural ratio of cases and controls. Out of 5,553 concepts, the top 10 positive concepts are reported in Table 3. The ranks of these features are reported, along with their concept names and odds ratios.

The concepts in Table 3 can be categorized into three groups. The first group is respiratory-related, including the 4th concept. This has face value as patients with acute respiratory failure may need to be intubated or are dependent on ventilation. This leads to the second group, which includes the 1st, 2nd, and 5th concepts. These concepts indicate that most of the patients who are intubated or need ventilation assistance experience SMM. The procedural concept "*Critical care, evaluation and management of the critically ill or critically injured patient*" (3rd concept) is positively associated with SMM.

In addition, three novel risk factors (7th, 8th, and 9th concepts) were found to be associated with SMM in this study. Most of the patients that were assigned these concepts experienced SMM. To the best of our knowledge, these concepts have not been reported in the literature nor are they in the CDC definition.

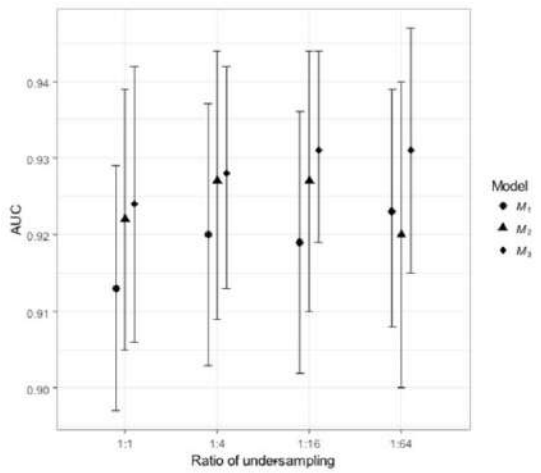


Figure 3– AUC performance as a function of SMM case:control ratio

Table 3– The most important concepts that are positively associated with SMM (p-values smaller than 10^{-7})

Rank	Concept Name	Type	OR (95% CI)	Patients with this concept	SMM Patients with this concept
1	Dependence on ventilator	Condition	1938.9 (388.3 – 35217.1)	16	15
2	Intubation, endotracheal, emergency procedure	Procedure	1461.1 (428 – 9151.5)t	25	23
3	Critical care, evaluation and management of the critically ill or critically injured patient; first 30-74 minutes	Procedure	1735.4 (1076.8 – 2969.8)	184	166
4	Acute respiratory failure	Condition	1272.3 (644 – 2892.8)	80	72
5	Ventilation assist and management, hospital inpatient/observation, initial day	Procedure	1232.5 (663.6 – 2557.3)	95	85
6	Trauma and postoperative pulmonary insufficiency	Condition	1007.1 (486.4 – 2444.6)	62	55
7	Disorders of fluid and electrolytes	Condition	685.4 (261.3 – 2351)	27	23
8	Systemic inflammatory response syndrome	Condition	561.3 (229.1 – 1682.8)	12pt	23
9	Acidosis	Condition	325.8 (201.4 – 544.4)	82	59
10	Sepsis	Condition	308.2 (162.5 – 617.8)	44	31

Discussion and Conclusions

SMM often leads to adverse outcomes, including prolonged length of stay and an increased rate of postpartum readmission. The machine learning-based framework introduced in this paper can enable the automatic identification of SMM events, based on diagnosis, procedural and demographic features, and provide intuition into risk factors associated with SMM. In particular, this work has two notable findings.

First, the framework shows that the machine learning approach outperforms the baseline model (CDC criteria) in terms of AUC, positive predictive value and sensitivity. As such, we believe that a greater number of SMM cases can be identified through our framework, with a smaller false positive rate, than that is currently achieved through current best practice. This should substantially reduce the amount of manual effort to identify SMM instances. In addition, since we rely on the OMOP common data model, the proposed framework is likely to be generalized to other institutions.

Second, our model uncover potentially novel concepts that are positively associated with SMM. 85% (23 of 27) patients diagnosed with fluid and/or electrolytes disorder experienced SMM during their delivery hospitalization. Similarly, 82% (23 of 28) patients diagnosed with systemic inflammatory response syndrome developed SMM.

Despite these findings, it should be recognized that this is a pilot study and there are several limitations that need to be addressed as this line of research moves forward.

First, we assumed that our constructed control group does not contain any SMM cases; however, this is not guaranteed. It is challenging to solve such a limitation in a pilot study. This is because it would require substantial effort to review all 45,446 control instances to identify missed SMM cases. A recent study indicates that there may exist 1 SMM cases per 10,000 samples in the control group [9]. Still, given to the small missed-case:control ratio (1:10000), this situation likely has little impact on the performance of the SMM identification models.

Second, we only applied the simplest machine learning algorithm-logistic regression in this study and many other advanced machine learning algorithms are missing. We anticipate adding neural networks, decision tree, support vector machine (SVM), and gradient boosting classifier in our next-step study to improve SMM identification performances.

Third, more advanced feature selection algorithms can be used to preselect candidate features feeding the machine learning models. In this study, we preselected candidate features feeding machine learning models based on their frequency occurring in the patient population, which does not consider correlations between features. The performance of machine learning models can be improved by removing duplicated features.

Fourth, we only considered diagnosis and procedural features and neglected many other important features including laboratory variables, medications, and vital signs, which may be related to SMM.

Fifth, the concepts assigned to patients were considered in an atemporal fashion. It can potentially boost identification performance to model temporal relation between concepts. Yet chronological information between concepts reflects how SMM develops and is handled. Such information would be useful for characterizing how at-risk patients were managed.

Acknowledgments

This research was sponsored in part by NIH grants R01LM010207 and R00LM011933.

References

[1] Centers for Disease Control and Prevention, Severe maternal morbidity in the United States at http://www.cdc.gov/reproductivehealth/MaternalInfantHealth/SevereMaternalMorbidity.html accessed on Nov. 20, 2018.

[2] E.V. Kuklina, S.F. Meikle, D.J. Jamieson, M.K. Whiteman, W.D. Barfield, S.D. Hillis, and S.F. Posner, Severe obstetric morbidity in the United States: 1998-2005, *Obstet Gynecol* **113** (2009), 293-299.

[3] K.S. Joseph, S. Liu, J. Rouleau, R.S. Kirby, M.S. Kramer, R. Sauve, W.D. Fraser, D.C. Young, and R.M. Liston, Severe maternal morbidity in Canada, 2003 to 2007: surveillance using routine hospitalization data and ICD-10CA codes, *J Obstet Gynaecol Can* **32** (2010), 837-846.

[4] W.M. Callaghan, A.A. Creanga, and E.V. Kuklina, Severe maternal morbidity among delivery and postpartum hospitalizations in the United States, *Obstet Gynecol* **120** (2012), 1029-1036.

[5] S.E. Geller, A.R. Koch, C.E. Garland, E.J. MacDonald, F. Storey, and B. Lawton, A global view of severe maternal morbidity: moving beyond maternal mortality, *Reprod Health* **15** (2018), 98.

[6] W.M. Callaghan, A.P. Mackay, and C.J. Berg, Identification of severe maternal morbidity during delivery hospitalizations, United States, 1991-2003, *Am J Obstet Gynecol* **199** (2008), 133 e131-138.

[7] W.M. Callaghan, W.A. Grobman, S.J. Kilpatrick, E.K. Main, and M. D'Alton, Facility-based identification of women with severe maternal morbidity: it is time to start, *Obstet Gynecol* **123** (2014), 978-981.

[8] E.K. Main, A. Abreo, J. McNulty, W. Gilbert, C. McNally, D. Poeltler, K. Lanner-Cusin, D. Fenton, T. Gipps, K. Melsop, N. Greene, J.B. Gould, and S. Kilpatrick, Measuring severe maternal morbidity: validation of potential measures, *Am J Obstet Gynecol* **214** (2016), 643 e641-643 e610.

[9] S. Hunt, Defining severe maternal morbidity, Nurs Womens Health 18 (2014), 169-172.

[10] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, Early diagnosis of Alzheimer's disease with deep learning, in: 2014 IEEE 11th international symposium on biomedical imaging (ISBI), IEEE, 2014, pp. 1015-1018.

[11] K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, and D.I. Fotiadis, Machine learning applications in cancer prognosis and prediction, Computational and structural biotechnology journal 13 (2015), 8-17.

[12] P.E. Stang, P.B. Ryan, J.A. Racoosin, J.M. Overhage, A.G. Hartzema, C. Reich, E. Welebob, T. Scarnecchia, and J. Woodcock, Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership, *Ann Intern Med* **153** (2010), 600-606.

Address for correspondence
cheng.gao@vumc.org

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190201

Leveraging Electronic Health Records to Learn Progression Path for Severe Maternal Morbidity

Cheng Gao[a], Sarah Osmundson[b], Xiaowei Yan[c], Digna Velez Edwards[a,b], Bradley A. Malin[a,d,e], You Chen[a]

[a]*Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States*
[b]*Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN, United States*
[c]*Sutter Research, Development and Dissemination, Sacramento, CA, United States*
[d]*Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, United States*
[e]*Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, United States*

Abstract

Severe maternal morbidity (SMM) encompasses a wide range of serious health complications that would likely result in death without in-time medical attention. It has been recognized that various demographic factors (e.g., age and race) and medical conditions (e.g., preeclampsia and organ failure) are associated with SMM. However, how medical conditions develop into SMM is seldom investigated. We hypothesize that SMM has a progression path, which is associated with a sequence of risk factors rather than a set of independent individual factors. We implemented a data-driven framework that leverages electronic health records (EHRs) in the antepartum period to learn the temporal patterns and measure their relationships with SMM during the delivery hospitalization. We evaluate the framework with two years of data from 6,184 women who had delivery hospitalizations at Vanderbilt University Medical Center. We discovered 69 temporal patterns, 12 of which were confirmed to be significantly associated with SMM.

Keywords:

Electronic health records, Pregnancy Risk Factors

Introduction

Severe maternal morbidity (SMM) such as severe hemorrhage, uterine rupture, organ failure, embolism, pulmonary embolism, and seizure is a physical condition that either result from or is aggravated by pregnancy and has an adverse effect on a woman's health.[1] It has been recognized that women with SMM often have a longer length of stay, higher medical cost at the time of delivery, and potentially worse neonatal and postpartum outcomes than those without SMM [2]. SMM has been utilized as an indicator of the quality of maternal health and the cost of healthcare.[3; 4]

Although the definitions of SMM are different from one study to another, its prevalence has steadily increased worldwide. [4; 5] Around 60,000 women are now affected in the US every year. [6; 7] Learning the risk factors of SMM, and predicting it early can potentially improve the quality of women's health and reduce healthcare costs.

Given the complexity of SMM and its worldwide prevalence, health practitioners and researchers at the national and state level have made considerable progress to learn the risk factors for SMM [9; 10]. It has been recognized that various factors including demographics (e.g., age and race), lifestyle (e.g., tobacco use and alcohol consumption), and medical conditions (e.g., preeclampsia and eclampsia, gestational diabetes mellitus, and prior cesarean section) are related to SMM. Yet almost all studies to date relied on hypothesis-driven approaches to learn risk factors, which have several drawbacks. First, they only investigated a small number of predetermined risk factors and neglected many other factors which may contribute to SMM. This is due to the fact that these studies relied on experts' experiences to determine candidate risk factors which can cost a huge amount of manual effort. Second, most studies did not consider the temporal relationships between risk factors. This is important because women with SMM, rarely transition directly from a healthy state to SMM or death. Rather they tend to progress along a continuum from one health status to another and, eventually, to severe morbidity or death.

To overcome these drawbacks, we introduced a data-driven framework to learn the progression path for SMM in the antepartum period (i.e., pregnancy before delivery). We focused on this period because our aim was to discover temporal patterns of risk as early as possible so that appropriate interventions could be applied to prevent SMM or reduce harms caused by SMM. In addition, this framework is built using on-the-shelf data and, thus,considered a large number of candidate risk factors with minimal manual effort. In particular, we leveraged data in electronic health records (EHRs) to learn temporal patterns between maternal conditions, in terms of International Classification of Disease (ICD) codes and measure the relationships of the temporal patterns with SMM. This work is notable because it shifts the investigations of SMM risk factors from atemporal to temporal manner.

Methods

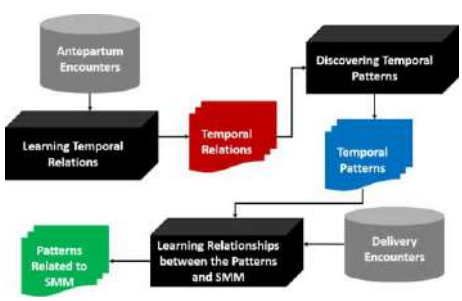


Figure 1. An overview of the data-driven framework to learn temporal patterns related to SMM.

Our data-driven framework consists of three components: i) a data mining algorithm to learn temporal relations between diagnoses in antepartum period, ii) a statistical model to

identify temporal patterns from temporal relations, and iii) a regression model to identify SMM related temporal patterns. An overview of the framework is depicted in Figure 1.

Dataset

Table 1. Statistics of age, race, and length of EHR for patients with and without SMM

Factor	SMM $N = 69$	Non-SMM $N = 6,115$
Age (mean ± SD)	29.1 ± 5.3	28.9 ± 5.5
Race (%)		
White	43 (62.3%)	4,129 (67.5%)
Black	16 (23.2%)	1,020 (16.7%)
Asian	7 (10.2%)	393 (6.4%)
Other	3 (4.3%)	573 (9.4%)
Length of EHR (in weeks)	78.2 ± 73.1	73.9 ± 68.8

The data is based on 6,184 obstetric patients who had a delivery encounter at Vanderbilt University Medical Center (VUMC) between 2015 to 2017. For each patient, the data documents: 1) demographics (age and race), and 2) clinical concepts during antepartum and delivery encounters. For each encounter, we have admission and discharge dates and the set of ICD-9 codes. The data in the antepartum encounters were leveraged to learn temporal relations between ICD-9 codes and the data in the delivery encounters were used to determine if a patient experienced SMM or not. We used 25 SMM indicators in the form of ICD-9 codes, as defined by the Centers for Disease Control and Prevention (CDC) to identify SMM cases [1]. Summary statistics of patients with and without SMM are shown in Table 1.

Encounter: An encounter corresponds to a hospital visit, which has the admission and discharge date. Encounters in this paper are denoted as $E_1, E_2, \cdots, E_n$.

Code: A diagnosis is specified as an ICD-9 code or procedure code. ICD-9 codes are used for billing services and are assigned to EHRs by physicians in patient care or by medical coders after the discharge of an encounter. The time when an ICD code is assigned to EHRs of a patient within an encounter does not necessarily represent the exact time when the patient received the corresponding diagnosis. However, ICD-9 codes across encounters should have temporal relations. For instance, if code 648.01 was assigned in one encounter ranging from February 3 to February 4, and code 655.83 was assigned in another encounter ranging from March 3 to March 5, then 655.83 was diagnosed after 648.01, and the temporal relationship exists between them. We denoted a code as c_i. If an encounter E_k has m different codes, then we will represented all codes in the E_k as $\{c_1, c_2, \cdots, c_m\}$.

Code sequence: A code sequence is an ordered series of codes coming from disparate encounters. For example, as shown in Figure 2, this patient has three antepartum encounters with 4 code sequences: $\{c_1 \rightarrow c_2 \rightarrow c_4,\ c_1 \rightarrow c_2 \rightarrow c_5,\ c_1 \rightarrow c_3 \rightarrow c_4,\ c_1 \rightarrow c_3 \rightarrow c_5\}$.

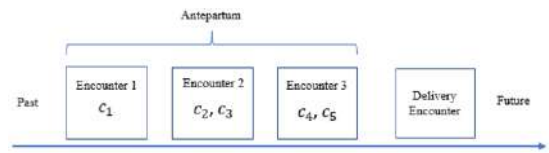


Figure 2. An example of four temporal perinatal encounters of a patient. Codes were assigned in each of the three antepartum encounters.

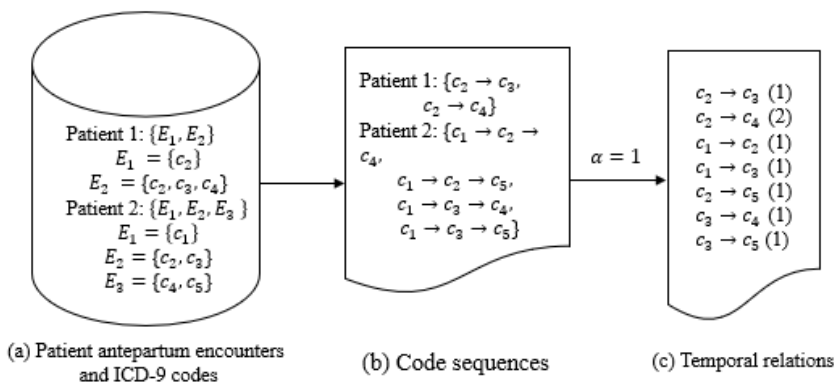


Figure 3. An example of learning temporal relations from six code sequences extracted from five encounters

Building Temporal Relations

We designed a sliding window-based algorithm to build temporal relations between diagnosis codes. The ordered relation between a pair of codes within a code sequence as:

$$Code_{relation}(c_i, c_j) = \begin{cases} \frac{1}{(p(c_i)-p(c_j))^2}, & (0 < p(c_j) - p(c_i) \leq \alpha \\ 0 \quad , \quad otherwise \end{cases} \quad (1)$$

where $p(c_i)$ is the position of a code c_i in a code sequence and α is the window size. The position of the first code of a sequence is set as 1 and the position of the last code is the length of a sequence.

Figure 3 shows an example to illustrate the algorithm. There are two patients P1 and P2, with 2 and 3 encounters, respectively. Each encounter has several diagnosis codes assigned. For the first patient, the first encounter contains one code c_2 and the second encounter contains three codes c_2, c_3, c_4. The code sequences for the first patient are $\{c_2 \rightarrow c_3,\ c_2 \rightarrow c_4\}$. It is notable that $c_2 \rightarrow c_2$ is not included in code sequences because c_2 appears in both encounters and $c_2 \rightarrow c_2$ does not convey any temporal information between encounters. The code sequences for the second patient are $\{c_1 \rightarrow c_2 \rightarrow c_4,\ c_1 \rightarrow c_2 \rightarrow c_5,\ c_1 \rightarrow c_3 \rightarrow c_4,\ c_1 \rightarrow c_3 \rightarrow c_5\}$. If we set window size α is 1, we can learn the strength of temporal relation for each pair of codes within a code sequence via Eq. (1). The total strength of the temporal relation for a pair of codes is the summary of temporal strength of that pair of codes across all sequences. Temporal

relations and their strengths are depicted in Figure 3c . For instance $c_2 \rightarrow c_4$ appears in two code sequences and the its temporal strength in each of the code sequence is 1 accoridng to Eq. (1), and thus its total strength is accumulated across the two sequences.

Discover temporal patterns

We applied statistical matching methods [12] along with statistical models to identify temporal patterns from the learned temporal relations. For an investigated temporal relation $c_i \rightarrow c_j$, we defined patients containing c_i as cases, and patients without containing c_i as controls. We selected most similar controls and cases in terms of age, race and EHR length. Based on the selected cases and controls, we used a Chi-square test to test the significance of $c_i \rightarrow c_j$ after Bonferroni correction. A code relationship was considered a temporal pattern when the p value was smaller than 0.05.

Learn relationships of temporal pattern with SMM

We used logistic regression to model relationships between antepartum temporal patterns and SMM during delivery hospitalization. Age, race, and length of EHR are included in the model as confounders. For each temporal pattern, logistic regression is applied to evaluate its adjusted relative risk (aRR) on SMM. The significant temporal patterns are reported with the 95% confidence interval of aRR.

Table 2. Temporal patterns related to the fetal abnormality

Subcategories	Temporal Patterns	Summary of the Pattern
Diabetes mellitus	648.01 → 655.83	Antepartum diabetes may lead to suspected fetal abnormality
Cervical shortening	649.73 → 644.03	Cervical shortening increases the risk of threatened premature labor
Fetal growth retardation	764.90 → 656.53	Fetal growth retardation may affect the management of mother
Uterine size date discrepancy	649.63 → 656.53	Uterine size date discrepancy is associated with poor fetal growth
Fetal abnormality	656 → 655.81	Suspected fetal abnormality and placental problem
Advanced maternal age	659.61 → 660	Women with advanced maternal age tend to have obstructed labor
Antepartum complication	644.03 → 655.73, 648.93 → 644.03	The antepartum complication may lead to threatened premature labor that is associated with decreased fetal movements

Result

Temporal relation patterns

Sixty-nine temporal patterns were found in the study. These patterns were categorized into two groups: i) fetus abnormality (Table 2) complicating pregnancy (Table 3). An example of a mother's conditions leading to fetus abnormality is 648.01 (Diabetes mellitus) → 655.83 (suspected fetal abnormality), which implies that a mother with antepartum diabetes tends to have an abnormal fetus. An example of a mother's health conditions leading to complicating pregnancy is 649.73 → 644.03, which indicates cervical shortening increases the risk of threatened premature labor.

Temporal patterns were further categorized into a set of subtypes including obesity, type 2 diabetes, hypertension, opioid dependence, tobacco use, abdominal pain, advanced maternal age, urinary tract infection, hypothyroidism and historical delivery condition such as previous cesarean section. Drug addictions lead to complicating pregnancy. In addition, mothers with drug addictions usually suffer from posttraumatic stress disorder (309.81), viral diseases (647.63) and mental disorders (648.43).

Obesity, diabetes mellitus, and hypertension are correlated with each other (642.03 (Benign essential hypertension) → 648.03 (diabetes mellitus) and 642.23 (transient hypertension) → 649.13 (obesity complicating pregnancy)), and they together complicate pregnancy. Abdominal pain (789.09), advanced maternal age (659.63), urinary tract infection (599.0), hypothyroidism (244.9) and historical cesarean section (654.23) also complicate pregnancy.

Temporal patterns related to SMM

Among the 69 temporal patterns, 12 (or 17.4%) were significantly associated with SMM. The degrees of the associations between the 12 patterns and SMM are depicted in Figure 4. For example, women that had antepartum drug dependence and mental disorder had greater than 8-fold risk of SMM in their delivery. Similarly, antepartum conditions hypertension and diabetes (642.03 → 648.03, 648.03 → 642.03 and 648.03 → 642.23) increase the risk of SMM. Other temporal patterns that may lead to SMM include obesity, advanced maternal age, viral disease, and tobacco use disorder.

Discussion

Severe maternal morbidities (SMM) during delivery often results in adverse outcomes, including a prolonged length of stay and an increase of postpartum readmissions. In this study, we introduced a data-driven framework to infer temporal relational patterns between diagnoses during the antepartum period and analyzed their associations with SMM. This work has two notable findings.

First, we demonstrated that there are temporal patterns that suggest progression paths for SMM. For instance, pre-existing health conditions including hypertension, diabetes mellitus, and

obesity appeared to develop into comorbidities in the antepartum period, complicating pregnancy. Additionally, lifestyle factors, such as drug/opioid-dependence and tobacco use disorder, may also complicate pregnancy. It was also found that the well-being of a fetus is associated with adverse antepartum conditions. For instance, mothers with diabetes mellitus tend to have abnormal fetus.

Table 3. Temporal patterns related to complicating pregnancy

Subcategories	Temporal patterns	Summary of the Pattern
Obesity	278.01 → 649.11, 278.01 → 649.13, 648.01 → 649.11, 278.00 → 649.13, 649.12 → 648.83, 642.93 → 649.13, 659.63 → 649.13, 642.23 → 649.13, 642.03 → 649.13, 649.13 → 642.03	Obesity, especially morbid obesity, complicating pregnancy
Hypertension	796.2 → 642.23, 796.2 → 649.13, 401.1 → 642.03, 401.9 → 642.03, 648.03 → 642.03, 642.03 → 649.13, 642.23 → 649.13, 250.00 → 642.23, 648.03 → 642.23, 796.2 → 642.33, 796.2 → 649.13	Pre-existing hypertension, elevated blood pressure reading, antepartum hypertension complicating pregnancy
Diabetes Mellitus	250.01 → 648.01, 250.00 → 648.01, 250.00 → 648.83, 250.00 → 642.23, 250.00 → 648.03, 648.01 → 649.11, 790.29 → 648.83, 642.03 → 648.03, 648.03 → 642.23, 648.01 → 649.11	Type I and II diabetes with abnormal glucose tolerance complicating pregnancy; antepartum diabetes may cause some suspected fetal abnormality
Drug/opioid Dependence	304 → 648.31,304 → 648.30, 304.00 → 648.30, 304.00 → 648.33, 304.00 → 648.43, 304.00 → 649.03, 304.01 → 648.31, 304.01 → 648.30, 304.01 → 648.33, 648.31 → 649.01, 648.33 → 647.63	Drug dependence and opioid dependence are related, together complicating pregnancy; drug/opioid dependence is associated with mental disorder
Tobacco	648.31 → 649.01, 305.1 → 648.30, 649.03 → 648.30, 304 → 649.03, 648.33 → 649.03	Tobacco use disorder is associated with drug/opioid dependence; both could complicate pregnancy
Other	309.81 → 648.30	Other conditions such as posttraumatic stress disorder, thyroid dysfunction, urine infection complicating pregnancy, pain
	244.9 → 648.13, 244.9 → 648.91	
	599.0 → 646.63	

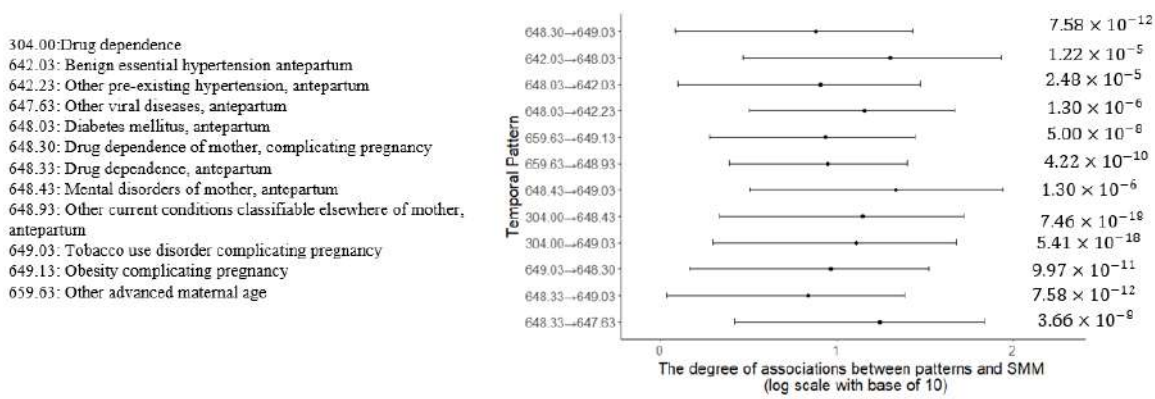


Figure 4. The adjusted relative risk of relationships between temporal patterns and SMM.

Second, we found that antepartum drug dependence and mental disorder are strongly related to SMM. Early identification of these high-risk mothers and adopt appropriate intervention strategies may improve care quality of SMM management and reduce harms caused by SMM.

At the same time, we note that this is a pilot study, and there are several limitations that need to be addressed.

First, we investigated a small number of SMM instances, which may limit our findings in this study. For instance, many potential temporal patterns may be missed in our investigated patient population. In addition, we only considered diagnoses

within antepartum periods, which may miss risk factors of SMM ahead of the antepartum period.

Second, we only studied temporal patterns between diagnosis codes and neglected other rich health information including medications, vital signs, and labs which could provide insights into SMM.

Third, this study was conducted in a single institute and should be expanded in many other institutes to conclude more general recommendations about managing SMM.

Fourth, we used a simple and natural temporal relation learning algorithm to learn temporal patterns. More advanced temporal pattern mining algorithms including Allen's algebra and graph neural networks can be applied in future studies.

Conclusion

This research broadens our current knowledge of the continuum of maternal health in the United States by inferring the association between antepartum comorbidities and SMM. Our work suggests that mothers with drug dependence, hypertension, diabetes mellitus, viral disease, and tobacco use disorder have an increased risk of SMM. While further investigation is needed, we believe that healthcare organizations should focus their attention on childbearing women with these conditions in the antepartum period.

Acknowledgments

This research was supported, in part, by the following grants from the National Library of Medicine at the National Institutes of Health: R00LM011933 and R01LM010207. The content in this work is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

[1] Centers for Disease Control and Prevention, Severe maternal morbidity in the United States at http://www.cdc.gov/reproductivehealth/MaternalInfantHealth/SevereMaternalMorbidity.html accessed on Nov. 20, 2018.

[2] E.M. Harvey, S. Ahmed, S.E. Manning, H. Diop, C. Argani, and D.M. Strobino, Severe Maternal Morbidity at Delivery and Risk of Hospital Encounters Within 6 Weeks and 1 Year Postpartum, *J Womens Health (Larchmt)* **27** (2018), 140-147.

[3] K. Joseph, S. Liu, J. Rouleau, R.S. Kirby, M.S. Kramer, R. Sauve, W.D. Fraser, D.C. Young, and R.M. Liston, Severe maternal morbidity in Canada, 2003 to 2007: surveillance using routine hospitalization data and ICD-10CA codes, *Journal of Obstetrics and Gynaecology Canada* **32** (2010), 837-846.

[4] E.V. Kuklina, S.F. Meikle, D.J. Jamieson, M.K. Whiteman, W.D. Barfield, S.D. Hillis, and S.F. Posner, Severe obstetric morbidity in the United States: 1998-2005, *Obstet Gynecol* **113** (2009), 293-299.

[5] S.E. Geller, A.R. Koch, C.E. Garland, E.J. MacDonald, F. Storey, and B. Lawton, A global view of severe maternal morbidity: moving beyond maternal mortality, *Reprod Health* **15** (2018), 98.

[6] E.K. Main, A. Abreo, J. McNulty, W. Gilbert, C. McNally, D. Poeltler, K. Lanner-Cusin, D. Fenton, T. Gipps, K. Melsop, N. Greene, J.B. Gould, and S. Kilpatrick, Measuring severe maternal morbidity: validation of potential measures, *Am J Obstet Gynecol* **214** (2016), 643 e641-643 e610.

[7] W.M. Callaghan, A.A. Creanga, and E.V. Kuklina, Severe maternal morbidity among delivery and postpartum hospitalizations in the United States, *Obstet Gynecol* **120** (2012), 1029-1036.

[8] W.M. Callaghan, A.P. Mackay, and C.J. Berg, Identification of severe maternal morbidity during delivery hospitalizations, United States, 1991-2003, *Am J Obstet Gynecol* **199** (2008), 133 e131-138.

[9] K.E. Gray, E.R. Wallace, K.R. Nelson, S.D. Reed, and M.A. Schiff, Population-based study of risk factors for severe maternal morbidity, *Paediatr Perinat Epidemiol* **26** (2012), 506-514.

[10] S. Liu, R.M. Liston, K. Joseph, M. Heaman, R. Sauve, and M.S. Kramer, Maternal mortality and severe morbidity associated with low-risk planned cesarean delivery versus planned vaginal delivery at term, *Canadian medical association journal* **176** (2007), 455-460.

[11] C.L. Roberts, J.B. Ford, C.S. Algert, J.C. Bell, J.M. Simpson, and J.M. Morris, Trends in adverse maternal outcomes during childbirth: a population-based study of severe maternal morbidity, *BMC Pregnancy Childbirth* **9** (2009), 7.

[12] J.S. Sekhon, Multivariate and propensity score matching software with automated balance optimization: the matching package for R, (2011).

[13] D. Boaz, M. Balaban, and Y. Shahar, A temporal-abstraction rule language for medical databases, in: Proceeding of the workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP), 2003.

Address for correspondence
cheng.gao@vumc.org

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190202

Building an Experimental German User Interface Terminology Linked to SNOMED CT

David Hashemian Nik[a], Zdenko Kasáč[a], Zsófia Goda[a], Anita Semlitsch[a], Stefan Schulz[a,b]

[a]Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria
[b]Averbis GmbH, Freiburg, Germany

Abstract

We describe the process of creating a User Interface Terminology (UIT) with the goal to generate a maximum of German language interface terms that are mapped to the reference terminology SNOMED CT. The purpose is to offer a high coverage of medical jargon in order to optimise semantic annotations of clinical documents by text mining systems. The first step consisted in the creation of an n-gram table to which words and short phrases from the English SNOMED CT description table were automatically extracted and entered. The second step was to fill up the n-gram table with human and machine translations, manually enriched by POS tags. Top-down and bottom-up methods for manual terminology population were used. Grammar rules were formulated and embedded into a term generator, which then created one-to-many German variants per SNOMED CT description. Currently, the German user interface terminology contains 4,425,948 entries, created out of 111,605 German n-grams, assigned to 95,298 English n-grams. With 341,105 active concepts and 542,462 (non FSN) descriptions, it corresponds to an average of 13 interface terms per concept and 8.2 per description. An analysis of the current quality of this resource by blinded human assessment terminology states equivalence regarding term understandability compared to a fully automated Web-based translator, which, however does not yield any synonyms, so that there are good reasons to further develop this semi-automated terminology engineering method and recommend it for other language pairs.

Keywords:

Natural Language Processing, Systematized Nomenclature of Medicine, Translations

Introduction

In medical documentation, user interface terminologies (UITs) bridge between the language in use by clinicians and the standardised language of reference terminologies. Interface terms vary between specialties, professional groups and dialects and undergo constant evolution due to medical progress and dynamics of language [1]. We present a cost-effective, manual, incremental approach to acquire and create interface terms with the goal to map them onto a reference terminology (SNOMED CT), both using bottom-up and top-down approaches.

Health care terminologies are crucial resources for semantic interoperability. Despite their differences in architecture and scope (e.g. ICD-10 as a classification system for diseases vs. SNOMED CT as an ontology for all aspects of electronic health record (EHR) content), there is a principal distinction between:

- Labels (fully specified names, reference terms), which aim at providing self-explaining descriptions of the meaning of concepts, often paralleled by formal definitions in ontologies or text definitions in thesauri.
- Interface terms (non-preferred synonyms), which represent the jargon used by clinical practitioners in their daily documentation and communication.

Take the SNOMED CT preferred term "*Primary malignant neoplasm of lung*" as an example. This term is precise but artificial. Screening the complete PubMed corpus yields no single occurrence of this term, and it would be very unlikely be found in clinical documents either. Its popular synonym "*lung cancer*", retrieves 120,682 documents from PubMed and is common in clinical documents and problem lists. Another example is "EKG", an acronym-term we retrieved 12,208 times in a corpus of cardiology discharge summaries, while the full term “Elektrokardiogramm” does not occur a single time. Reference terms are characterised by precision; interface terms by brevity. This highlights the need for interface terms, wherever content is automatically extracted from texts [2]. This has recently been emphasised by the European project ASSESS CT (Assessing SNOMED CT for Large Scale eHealth Deployments), which recommended broad efforts to be invested into UITs [3] linked to reference terminologies like SNOMED CT, rather than into translations proper. For humans, the possible fuzziness and ambiguity of interface terms (in this case, it may not be quite clear whether lung metastases are in the scope of "lung cancer") is a minor problem due to context and implicit understanding within a user group, whereas word sense disambiguation is still a major problem for machines.

A major use case for UITs is the provision of dictionary entries for natural language processing (NLP) systems. Other use cases are related to structured data entry using data acquisition forms, whenever the terms should be close to the user's language preferences. Limitations of UITs lie in the conceptual content of the underlying reference terminology (in this case SNOMED CT). There is still a substantial amount of interface terms that cannot be precisely mapped to terms of a corresponding domain terminology.

Interface terms can be single words like "pancreas" or "dermatology", compound words like "lymphangiosarcoma", acronyms like "ARDS" or multi word terms like "hereditary factor VIII deficiency disease". Possible sources of interface terms include:

- Automatic and manual term translations
- External, generally accessible corpora of the target language (e.g. books, publications, articles etc.);
- Institution-specific value sets and term collections;
- Clinical corpora constituted by EHR narratives (privacy protection must be taken into account).

Lexical ambiguity (e.g. "delivery" in "drug delivery" vs. "delivery of a baby") is characteristic for interface terms, and

it rarely runs parallel between languages (e.g. in German, words for the delivery of goods, substances or babies are completely distinct). Especially short acronyms are prone to ambiguities (e.g. "MI" means "myocardial infarction" in cardiology, but "macular ischemia" in ophthalmology; "DM" may be expanded to "diabetes mellitus", "diameter", "disease management", or "dermatomyositis" according to its context. Only long acronyms like "NSTEMI" can be expected to have a unique meaning. Ambiguous acronyms are mostly unproblematic once they occur as constituents of a longer term, in which they are univocal (example: "type 2 DM").

In the following we describe the ongoing construction of a German UIT for SNOMED CT, which uses the English SNOMED CT description table for the semi-automated construction of German-language interface terms.

Material and Methods

SNOMED CT

SNOMED CT [4] is the most comprehensive healthcare terminology on the globe, with use in over 50 countries. The current international SNOMED CT version (July 2018) consists of 340,659 active representational units, known as *SNOMED CT concepts*. They have an exclusive meaning and a unique machine-readable identifier. *SNOMED CT Descriptions* include a *Fully Specified Name* (FSN) for each concept, together with one or more synonyms. FSNs represents the concept with an official, ideally self-explaining name. Synonyms share meaning with FSNs but find more common use in displaying or selecting a desired FSN. The meaning of SNOMED CT concepts is, in addition, delineated by text definitions and formal axioms in Description Logics. E.g., *Gastritis* is defined as logically equivalent to a disorder with inflammatory morphology that is located at some stomach structure.

SNOMED CT's increasing adoption in German-speaking countries (with Switzerland and Austria already being members of SNOMED International, whereas Germany is still negotiating) prioritises harmonisation with local language and terminologies. This has motivated the authors to develop a semi-automatic, resource-aware and pragmatic method for creating an experimental German UIT linked to the current international version of SNOMED CT. The methodological framework developed might be applied to other languages [5]. This UIT feeds the currently largest SNOMED CT use case in the German-speaking country, *viz.* semantic annotations of clinical texts within the hospital information system of the Austrian healthcare provider KAGes, using natural language processing [6].

Procedure

The starting point of the interfacing process was the set of over 700,000 English SNOMED CT terms ("Descriptions") from the international version. This list contains, apart from the official labels (Fully Specified Names), one preferred term (dependent on the release) and zero-to-many synonyms. SNOMED CT terms are characterised by many repetitive substrings, exemplified by the term "magnetic resonance imaging of hip". That there are 1,620 occurrences of "magnetic resonance imaging" and 1,239 occurrences of "of hip" in SNOMED CT shows the potential of a modular approach to term translation.

N-gram Table as Core Translation Resource

In order to harvest such repetitive passages, a language-specific rule set was created and implemented in Python 3 to chunk decomposed terms down into shorter units, constituting word *n*-grams with *n* ranging from *1* to *6*. Here "word" encompasses, to a minor extent, also sub-word entries needed for single-noun composition, which is particularly common in German. SNOMED CT terms therefore can be constituted by one to many n-grams, e.g. "Escherichia coli" is a 2-gram, which can equally stand alone, or occur in a longer term like "Escherichia coli antibody". In addition to the n-grams with $n > 1$, also all single words were added. In summary, n-grams include complete and partial constituents of the SNOMED CT labels and descriptions. The result was an .xlsx table with currently enclosing close to 550,000 English-language n-grams, ranked by their decreasing frequency in the source, i.e. the SNOMED CT's description table. It is structuredin the following collumns:

- *"ID"* – n-gram identifier code;
- "N-gram English": from words to n-grams, ($n < 7$);
- *"Length"*: number of tokens in the n-gram (n);
- *"Count"*: n-gram frequency in the source SNOMED CT description file;
- *"N-gram German 1"*, *"N-gram German 2"*... German translations of English n-grams.

We started populating this table with German terms, started four years ago with limited resources (one part-time terminologist and several medical students). Along time it has been subject to constant optimization and quality improvement heuristics. The description of some of these heuristics (mostly developed as a sequence of trial-and-error cycles) will occupy the remainder of this section. The main goal of this description is to outline a general methodology for modular and incremental CIT developments for new language with limited resources.

Translation Heuristics

The first step was to tackle large amounts of easily machine translatable content. Method of choice was Google Translate. For surprisingly many more common medical terms, for example *"arm"*, *"status"*, or *"hormone"*, but also *"cholecystectomy"* or *"tendon sheath"*, useful translated content could be harvested. Several limitations were observed: (i) each translation only yielded one target n-gram (no synonyms are provided when inserting source terms in a batch), (ii) single English words were often undefined in terms of POS (part of speech, e.g. noun, adjective, verb), like *"set"*, *"back"*, *"general"* and therefore ambiguous, (iii) the translation of numerous words did not correspond to the medical meaning (e.g. *"delivery"* translated into *"Lieferung"* and not into the translations relevant to medical context like *"Entbindung"* or "Gabe"), (iv) less common medical or chemical terms were not translated, and the untranslated term was returned, instead (which then was preserved, waiting for manual correction).

Given SNOMED CT's richness in concepts for biological organisms, for which Latin terms are common and identical across languages, we harvested a large number of terms that occurred identically in the English and Spanish versions, assigned to the same concepts (e.g. *"Homo sapiens"*, *"Ascaris lumbricoides"*, *Angina pectoris"*, *"Anorexia nervosa"*). Thus, they could be safely added to the German version.

For the left-over of untranslatable content, manual input was required. It required mainly addition of new German n-grams and modification of the existing human or machine generated ones, addition of synonyms (e.g. *"Leber- " to "hepatisches"*), as well as appending grammar tags to both manual and machine created words (e.g. *"hepatisches|JJ"* to mark it as an

adjective and "-" to mark *"Leber-"* as a prefix-like morpheme). Humans also revised the machine generated terms by evaluating randomly sampled sets of the generated terms, often revealing major systematic errors in the machine translation. Mistranslations are deleted and revised, untranslated terms grouped by certain patterns (e.g. ending with "acid" and changed to the German ending "säure" by repetitive search / replace actions). Manual work was prioritized according to the guiding principles: (i) n-gram frequency; (ii) clinical relevance; and (iii) single words.

To ensure all SNOMED CT concepts get covered in one primary form, we set the goal of at least one translation for every single word n-gram (n=1), Such atomic units represent individual interface terms or parts of larger composite interface terms. Semantic composition is not robust multiword n-grams. Therefore, we considered multi word n-grams down to a frequency of eight. Less frequent n-grams were left out, because of the enormous amount of multiword n-grams. The decision in favour of manual translating multiword terms was positive whenever:

- The formation of a multiword term in the source language was not paralleled by a translation in the target language. E.g. "*malignant neoplasm*" translates not only into "maligne Neoplasie" but also into "*Malignom*"; and "oral solution" does not translate into "*mündliche Lösung*", but into "*Lösung zur oralen Einnahme*".
- One component of a multiword term was ambiguous and could be disambiguated in the composition. E.g. "*back pain*" in "*Rückenschmerzen*" ("back" = "Rücken"), and "*back door*" into "*Hintertür*" ("*back*" = "*Hinter-*").

The application of these rules often results in a labour-intensive walkthrough of the whole n-gram table.

Addressing German Language Features Like Inflection and Composition

German word inflections heavily depend on gender, tense, person, number, declination type and case. All entries, be they manual or automatically made, need therefore to be reviewed and enriched with grammatical information. This task relies mainly on human editing, using however machine support for pattern-based search-replace actions. Automated term creation with correct inflection suffixes requires a distinction between nouns, adjectives, verbs, prepositions and determiners, yielding the following tagging suffixes:

- Singular gender or plural tags for nouns (|NN|N, |NN|F, |NN|M, NN|P)
- Cases for verbs (|VV|N, |VV|G, |VV|D, |VV|A)
- Markers for adjectives (|JJ)
- Preposition tags depending on case (|PP|N, |PP|G, |PP|D, |PP|A)
- Definite and indefinite articles (|DET|D, |DET|I)
- Noun tags of Latin words or other nouns for which single-noun composition is not allowed (|NL|N, |NL|F, |NL|M, |NL|P)

Tags can be omitted in multi word entries where, e.g. due to a preposition, the inflection case is already determined.

In addition, all numbers were substituted by placeholders (ð, ðð, ððð etc. depending on number of digits and decimals) in the n-gram table. (While the decimal denominator in English is point in German comma is used.)

Single word composition is an important feature of the German language, e.g. with "[Fracture] of the big toe" translated into "*Großzehen[fraktur]*". Such composition patterns are characterised by inversion (with the head morpheme trailing) and by inconstant addition of infixes to the modifying morphemes. This requires the addition of translations to English prepositional phrases starting with "of", which not only translates to a structurally similar German genitive construction ("der Großzehe"), but also to morpheme-like prefixes like "Großzehen...", tagged by bracketing underscores ("_Großzehen_"). In certain instances, we use the additional tag "%VOID%". "%VOID%" is used as an empty word and can be suffixed with tags and thus gaining an inflection attribute. It also prohibits noun compositions (which are otherwise standard between neighbouring nouns). The German n-gram "zur Benutzung%VOID%|PP|G" (to use) illustrates how %VOID% can avoid unwanted word composition, while enforcing the next phrase set into genitive case. If "%VOID%" were absent, the German word "Benutzung" would potentially be attached to a noun right to it, resulting in a disallowed nominal compound. Other tags are %SWAP% and %RIGHT%, with the former swapping the right and the left portion of a phrase (due to different word order in German), and the latter putting a word to the end of a phrase, typically a separable verb prefix, e.g.: "stops" □ "hört mit|PP| %RIGHT% auf".

Top-down and Bottom Up Term Harvesting

Both top-down and bottom-up term harvesting has been used. Top-down means an intellectual effort of enriching English n-grams with additional translations. E.g., when encountering the English term "diabetes mellitus" (together with the identical German term) the editor might remember that there is a synonym "Zuckerkrankheit" to be added. One problem of this method is that it requires extensive manual editing, and is only performed when a manual reviewer goes through the list checking for missing synonyms (starting at very frequent and relevant terms), requiring an excellent command of medical terminology.

Bottom-up describes lexicon population starting with content from representative corpora (clinical texts, biomedical literature, existing terminologies in the target language), from which terms are extracted and mapped to corresponding n-grams in the source language. The advantage of this method is the adjustment of the interface term vocabulary to preselected document genres, covering specific jargon. N-gram hit lists (ordered by decreasing frequency) extracted from these corpora are excellent sources of commonly used terms in the context that is to be covered. One source prepared for our project was a hit list with 4,000 de-identified dermatology summaries, which resulted in 24,000 n-grams ($n \geq 3$). After filtering out entries already included in the master n-gram table (and their inflectional variants), terminologists walk through the hit list and try to identify a corresponding English term for each entry in the master n-gram table.

In cases of doubt, translations are marked for further analysis and discussion. Translation resources such as and offline dictionaries, online translation tools (web search for usage frequency and contexts, various online dictionaries, Linguee, Oxford German Dictionary, DUDEN Wissennetz), term clusters retrieved from the UMLS metathesaurus as well as German titles of Medline-indexed citations (marked as [tt] in the Medline records) have proven useful.

Tooling

Appropriate tooling for distributed, cooperative terminology work is still a desideratum. N-gram editing and maintenance is currently done via a Microsoft Excel table, shared via a repository, which assures that only one instance of this table

can be edited at a time. This Excel n-gram was enriched by macros written in the VBA script language. E.g. the "exploration macro" and "modify and add macro". The former retrieves already existing word and sub-term translations for a given n-gram. The addition of synonyms is supported by a "*Modify and Add Macro*". It allows batch additions of synonyms. When entering an n-gram (e.g. "*Neoplasie|NN|F*") with a synonym (e.g. "*Neubildung|NN|F*") it iterates through the complete n-gram list and suggests additions (e.g. "*Basalzellen-Neubildung*" to "*Basalzellen-Neoplasie*"). Other macros are used for plausibility checking, e.g. identifying disallowed tags.

Term Generation

Term generation is done with a series of language specific Python 3 scripts. In a nutshell, they take the chunked English description table, translate each chunk by n-gram table lookups, thus collecting n-gram translations and assembling them into complete SNOMED CT description translations. In this process, the tagged words are interpreted by a variant generator, which generates inflectional variants, following an inbuilt noun phrase grammar. It also produces single word compounds, using specific tags and rules as explained above. The problem of this generative approach is its combinatorial explosion: if a SNOMED CT term consists of four n-gram chunks, each of which with 3 translations, $3^4 = 81$ terms are generated. To mitigate the growth of generated terms, as well as to improve quality we pursued two approaches. To increase translation of longer n-grams, in order to avoid uncommon term associations, and complete translations of very long terms, which are unlikely to occur in clinical texts, such as FSNs resembling textual definitions[1].

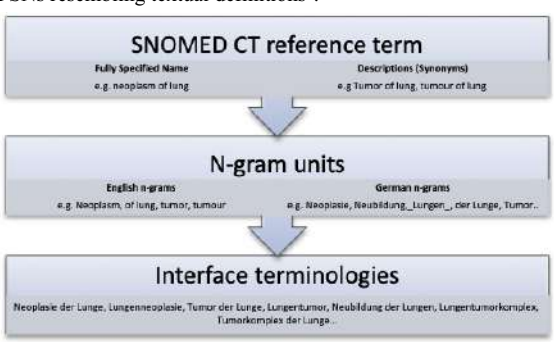


Figure 1 –Interface Term Generation

Quality Checks and Updates

Due to the large size of the UIT, only random samples are regularly quality-checked, assuming that frequent systematic errors also surface in these samples. Errors are traced back, either to the n-gram table or to the inflection and composition routines in the Python scripts. A challenge is also the biannual update of the n-gram table, together with the resulting interface term list, when a new international SNOMED CT version is released. Changes include addition, alteration and removal of concepts. During each version release, we have observed a growth of the n-gram resource. This opens the opportunity to fill in terms, e.g. more detailed pharmacological ingredients that previously were missing.

[1] Such as "pT2: Tumor invading two subsites in a single region or extending to involve an adjacent region within the nasoethmoidal complex, with or without bony invasion (nasal cavity and ethmoid sinus) (finding)"

Validation Study

The last 2018 version of the UIT was used for a blinded validation study. 200 concepts were randomly selected, and for each of them, one of the German UIT entries – together with the English description in original from – was randomly selected. For each English description, an alternative translation was created using DeepL [7], a machine translation website based on neural networks. For each of three terminologists, a custom translation list with 200 term translations, 100 from DeepL, and 100 from the German UIT was generated. The selection was done by chance, and the terminologists were not informed about the source. For each translation four pieces of information were required: (i) assessment of *Content comprehensibility* (regardless of style and grammar), (ii) *Grammar* (regardless content errors or bad word choices), (iii) *Style & Spelling*, for each of which using a five-point Likert scale (1 - very bad, 5 very good). Finally, an "ideal" manual translation for the term was added by each terminologist.

Results

The German User Interface Terminology currently encompasses 4,425,948 entries, automatically generated from a core vocabulary of 111,605 German n-grams, assigned to 95,298 English n-grams. With 341,105 active concepts and 542,462 (non FSN) descriptions, this corresponds to an average of 13 interface terms per concept and 8.2 per description.

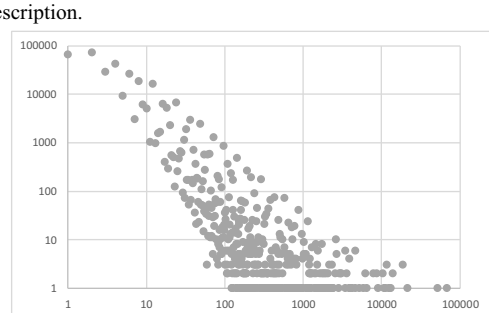


Figure 2 –Distribution of Synonym Sets in UIT. Logarithmic axes:x-Axis: size of synonym set, y-axis: frequency of sets of this size

Fig. 2 shows the distribution: most sets have one to 10 members, but there were outliers with more than 10,000 members, due to term compositionality. Table 1 provides the rating results.

Table 1 –Likert Scale Rating of Translations (Arithmetic Means of 300 Human Ratings per Category)

SNOMEC CT User Interface terminology (UIT)			**DeepL machine Translator**		
Content	Grammar	Spelling / Style	Content	Grammar	Spelling / Style
4.6	4.7	4.4	4.6	4.9	4.6

These results need to be interpreted in the light of the fact that DeepL provided only one translation per concept, whereas the UIT produced much more (mean: 12.1, median 3 translations per term). Whereas the DeepL result can be assumed to correspond to the most popular translations, our sampling

algorithm had picked out the UIT completely randomly. However, the results show clear deficits regarding grammar, spelling and style issues of the current state UIT. Finally, a case insensitive, spelling variation tolerant match between the translation suggested by the users and the UIT entries occurred in 299 of 600 cases.

Discussion

Is it worthwhile investing in terminology acquisition as well as UIT maintenance, in the light of ever increasing performance of machine translation tools? Already in 2013, we had found rather surprising translation results proposed by Google Translate [8]. We suggested a combination of translations done by medical students and done by machine translations; the method we have been using in the described UIT project. Still, machine translation systems have problems with the generation of compound nouns, as well as with the production of sufficient numbers of term variants and combinations. Also, their power depends on the amount of training data they are fed; good results can therefore not be expected for rarely used terms or languages that have less content on the Web.

We are currently supporting the construction of an interface terminology for Portuguese, using the Spanish version from SNOMED International as source language. For this language pair, the power of web-based machine translation systems seems to be much poorer than for German / English.

The advantage of interface terminologies is obvious when re-structuring narrative content of the EHR in terms of SNOMED CT. This opens up the ability to be linked with aggregation terminologies, such as the International Classification of Diseases (ICD-10 or the upcoming ICD-11) [9]. Apart from classification systems of diseases and procedures, such semantic standards are rarely used in routine documentation. As long as interface terminologies act in the background, where they are only used by text mining software, their size and their tendency to over-generate term variants (most of which are never found in any text) do not constitute a problem. This changes if interface terms are to be used by humans; here, uncommon terms should be filtered out before use in order to limit the variety of terms per concept. Such filter criteria could use large clinical corpora together with a general corpus such as Wikipedia and require that a generated term was used at least once. This may, however, obviate the translation of long artificial terms like the one cited above, which are unlikely to be found in any clinical document.

Extracting them from a narrative can be thought of as a result of fuzzy matching between the narrative and multiple term candidates, an approach to be exploited, especially with new, powerful word embedding methods and deep learning.

Another route for translating medical terms capitalises on the fact that large parts of them derive from Latin or Greek roots and share a regular morphology. Such terms can be automatically translated by rewriting rules [10,11]. A machine learning process can learn from two language sets, identify letter-based patterns and apply these rules for the automatic translation of new terms. The word "bronchoscopy" ending with "-oscopy" would be such an example that could be detected by this method and automatically translated by inferring rewriting rules ("Bronchoskopie"). The accuracy of this method depends on how many pairs of terms get compared.

Conclusions

We presented an approach to acquire interface terms from SNOMED CT reference term translations and data from various clinical corpora and map these interface terms onto SNOMED CT reference terms. During the process we focused on frequent words both in SNOMED CT and in clinical texts with the goal to cover all commonly used medical terms and their synonyms, integrating them into our system. An analysis of the current quality of the UIT by blinded human assessment terminology states equivalence regarding term understandability compared to a fully automated Web-based translator. This tool, however, yields much less synonyms, so that there are good reasons to further develop our semi-automated technology and recommend it for other language pairs.

References

[1] S. Schulz, Building an experimental German user interface terminology linked to SNOMED CT, *SNOMED Expo*, Bratislava, Slovakia, October 2017. https://goo.gl/cHgvmo [accessed on March 31st, 2019]

[2] S. Schulz, "Using language technology for SNOMED CT localization (poster), *SNOMED Expo*, Montevideo, Uruguay, October 2015. https://goo.gl/z2aCX1 [accessed on March 31st, 2019]

[3] D. Kalra, S. Schulz, D. Karlsson, R.V. Stichele, R. Cornet, K.R. Gøeg, G. Cangioli, C. Chronaki, R. Thiel, S. Thun,and V. Stroetmann, *ASSESS-CT recommendations*, 2016, *http://assess-ct.eu/fileadmin/assess_ct/final_brochure/assessct_final_brochure.pdf* [accessed on March 31st, 2019]

[4] SNOMED International. *SNOMED CT Starter Guide*, 2018,

[5] S. Schulz, J.M. Rodrigues, A. Rector, and C. Chute, Interface terminologies, reference terminologies and aggregation terminologies: a strategy for better integration, *Studies in Health Technology and Informatics* **245** (2017), 940-944.

[6] M. Kreuzthaler, C. Martínez-Costa, P. Kaiser, and S. Schulz, Semantic technologies for re-use of clinical routine data, *Studies in Health Technology and Informatics* **245** (2017), 539-543.

[7] DeepL translator, https://www.deepl.com/en/translator [accessed on March 31st, 2019]

[8] S. Schulz, J. Bernhardt-Melischnig, M. Kreuzthaler, P. Daumke, and M. Boeker, Machine vs. human translation of SNOMED CT terms, *Studies in Health Technology and Informatics* **192** (2013), 581-584.

[9] M. Mamou, A. Rector, S. Schulz, J. Campbell, H. Solbrig, and J.M. Rodrigues, ICD-11 (JLMMS) and SCT Inter-Operation, *Studies in Health Technology and Informatics* **223** (2016), 267-272.

[10] V. Claveau, Translation of biomedical terms by inferring rewriting rules. *Information Retrieval in Biomedicine: Natural Language Processing for Knowledge Integration*, IGI - Global. 2009, Chap 6.

[11] J.A. Miñarro-Giménez, J. Hellrich J, and S. Schulz, Acquisition of character translation rules for supporting SNOMED CT localizations. *Studies in Health Technology and Informatics* **210** (2015), 597-601.

Address for correspondence

David Hashemian Nik, david.hashemian-nik@stud.medunigraz.at

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190203

Characterizing Frequent Flyers of an Emergency Department Using Cluster Analysis

Emile Ramez Shehada[a], Lu He[a], Elizabeth V. Eikey, PhD[a], Maxwell Jen, MD[b], Andrew Wong, MD[c], Sean D. Young, PhD[d,e], Kai Zheng, PhD[a]

[a] *Department of Informatics, University of California, Irvine, Irvine, CA, USA*
[b] *Department of Emergency Medicine University of California, Irvine, Irvine, CA, USA*
[c] *Emergency Department, University of California, Davis, Davis, CA, USA*
[d] *Department of Family Medicine, University of California, Los Angeles, Los Angeles, CA, USA*
[e] *University of California Institute for Prediction Technology, Los Angeles, CA, USA*

Abstract

Emergency department (ED) overcrowding has been a pain point in hospitals across the globe. "Frequent flyers," who visited the ED at a much higher rate than average, account for almost one third of ED visits even though they represent only a small proportion of all ED patients. In this study, we used data-mining methods to cluster ED frequent flyers at a large academic medical center in the US. The objective was to identify distinct types of frequent flyers, and the common characteristics associated with each type. The results show that the frequent flyers at the ED have three subgroups each exhibiting distinct characteristics: (1) the elderly with chronic health conditions, (2) middle-aged males with unhealthy behavior, and (3) adult females who are generally healthy. These findings may inform targeted interventional strategies for patients of each subgroup, who likely have distinct reasons for visiting the ED frequently, to reduce ED overcrowding.

Keywords:
Data Mining; Hospital Emergency Service; Cluster Analysis

Introduction

ED overcrowding has long been an issue in hospitals across the world [1–3]. Frequent flyers, which are generally defined as patients who visit the ED four or more times per year [4,5], represent as little as 4.5% to 8% of all ED patients but account for as much as 28% of all ED visits [6]. Accordingly, hospitals have a vested interest in characterizing their populations of ED in order to determine how to reduce their impact on ED operations.

Researchers in the field of emergency medicine have written extensively about the typical characteristics of ED frequent flyers. However, there is still too much variation; frequent flyers simply are not a homogenous group. Within the literature, there is a general consensus that three factors are associated with the number of times a frequent flyer visits the ED per year: mental illness, substance abuse and dependence, and alcohol abuse [7–9]. Additionally, employment status and government insurance are often correlated with high ED use [6,10,11]. For instance, younger age groups were also implicated in ED overuse due to the characteristics of the neighborhood surrounding the hospital or the type of hospital from which the data was acquired [6,7]. In some cases, researchers observed a bimodal age distribution with increased risks in frequents users who were either 25 to 44 years old or over 64 years old [7]. Similarly, research has generally found that males and minorities tend to use the ED more frequently, but findings are sometimes inconsistent. For example, a 2005 study conducted by Blank et al. concluded that the sex and race composition of their ED frequent flyer population did not differ substantially from that of their general ED population, which was 51% female and 57% white [10]. Another study conducted by Milbrett and Halm indicated that the frequent flyer population at their "large Midwestern urban hospital" was "commonly" female, middle-aged and white [11].

Thus, there is no "one-size-fits-all" list of characteristics that apply to all hospitals. In order to support hospital-level decision making, a different approach to uncovering these subgroups is necessary. There is a need to identify hospital-specific subgroups and characteristics of frequent flyers. Clustering, or the process of organizing objects or measurements into groups, is a promising method. It differs from mere classification in that the groups used in classification are pre-defined groups, whereas the groups exposed by clustering do not exist beforehand [12]. As an automated unsupervised method, it is also less labor-intensive than manually identifying subgroups in a population and therefore is realistic to be implemented in real-life settings. In previous work, clustering techniques have been used on ED data to both predict patient outcomes or characterize ED usage. In 2007, Huang et al. used the K-means algorithm to cluster patients according to medical utilization, discovering that their population of frequent flyers also tended to utilize other medical services more often [13]. In a similar vein as Huang, Hastings et al. used latent cluster analysis (LCA) to reveal medical utilization clusters among elderly ED patients [14].

In this paper, we performed cluster analysis on 1748 unique patients who have visited the ED more than 4 times in a year at an UC medical center to identify characteristics of subgroups. We applied two commonly used cluster methods that are able to handle mixed data types: the K-prototypes algorithm and Partition-Around-Medoid (PAM). Finally, we discuss the three subgroups of frequent flyers that emerged from our analysis and the implications for using subgroup level interventions to reduce ED misuse.

Methods

Data description and processing

The data received from UCI Medical Center consisted of records of 1748 unique patients. Some data was derived from other data elements; for instance, blood pressure and height were used to yield a computed hypertension, which was cross-referenced against patients' medical history. Computed hypertension is used in the final analysis in lieu of both blood pressure and height. Patient medical history was also used to determine whether the patient had any history of substance abuse, mental illness, or cancer. The data contains both categorical attributes and continuous attributes. Categorical attributes include: hypertension, Diabetes, Binge drinking, tobacco, gender, employment, insurance, substance abuse, other mental illness, and cancer. Continuous attributes include: age, Body Mass Index (BMI), average emergency severity index (ESI), heart rate (HR), blood pressure (RR), temperature (TempC) and oxygen saturation (SpO2). Table 1 provides descriptive statistics of our dataset. Decimals in the table indicate percentages for categorical variables.

Table 1-Descriptive statistics

Variables	Mean (SD)
Demographic	
Gender (Male=1)	.52
Age	46 (18)
Race (Hispanic)	.47
Race (White)	.36
Race (Black)	.05
Race (Asian)	.08
Race (Others)	.04
Medical history	
Cancer (Yes=1)	.18
Diabetes (Yes=1)	.16
Other mental illness (Yes=1)	.33
Health indicators	
BMI	28 (7.3)
HR	91 (16)
RR	18 (2.2)
TempC	37 (.27)
SpO2	98 (1.9)
Hypertension (Normal)	.21
Hypertension (Prehypertension)	.44
Hypertension (Stage1)	.26
Hypertension (Stage2)	.07
Hypertension (ZHT Crisis)	.02
Binge Drinking (Daily or almost daily)	.21
Binge Drinking (Weekly)	.09
Binge Drinking (Monthly)	.07
Binge Drinking (Less than monthly)	.17
Binge Drinking (Never)	.57
Substance Abuse (Yes=1)	.12

Cluster methods

While K-means is often regarded as the most widely used cluster method, it can only handle numerical data [15]. Though categorical data can be transformed through techniques such as one-hot encoding, important information may be lost in the process. Another commonly used cluster method, K-modes, can only handle categorical data, and similarly, while we can condense numerical attributes into categorical ones, how the intervals are defined is also in question. Therefore, to handle the mixed data type we have, we choose the K-prototypes algorithm. To validate the resulting clusters produced by the K-prototypes algorithm, we used PAM on the same dataset. For each attribute, we performed the Kruskal-Wallace test to see if there exists any statistically significant difference across the three clusters. To sketch an overview of the subgroup characteristics, we combined several sub-categories, for instance, for tobacco use, "never" and "former" are combined to be "no". Since cluster analysis is unsupervised, we do not know how much natural clusters exist in the dataset. Therefore, we computed the Silhouette coefficient as well as the gap statistics to decide how many clusters we should expect from our dataset.

K-prototypes

K-prototype was first proposed by Huang in 1998 [16]. It differs from k-modes in that k-prototypes allow clustering of datasets with mixed data attributes, utilizing a combined dissimilarity measure. Another feature of K-prototypes is that for each resulting cluster, the algorithm returns a real data point from the original dataset as the "prototype" or the representation of the cluster. Cluster methods such as K-means usually return an averaged centroid instead of a real data point.

PAM

PAM, also known as the k-medoids algorithm, clusters data according to medoids [17]. The user inputs the desired number of clusters and the algorithm selects a random set of k items to be the set of starting medoids. Then, the algorithm constructs clusters by iterating through the remaining observations and assigning them to the "closest" medoids. "Closeness" in a PAM analysis is determined by a distance function. Euclidean distance is frequently used in k-means and similar clustering techniques where the data being clustered is entirely continuous. However, because our dataset contains several categorical variables, we used Gower distance, which is capable of handling dichotomous, continuous, and categorical data [18].

Optimal number of clusters: The Silhouette Coefficient & The Gap Statistic

To select the optimal number of clusters to feed into PAM, we used silhouette width. Silhouettes allow users to visualize the quality of a specified number of clusters [19]. When the data has been partitioned in a manner that closely tracks the "natural" cluster, silhouette width is generally high. As Figure 1 shows, when there are 3 clusters, the Silhouette coefficient reaches the peak, which indicates that more natural clusters can be produced. The gap statistic is another way to decide the optimal number of clusters, which is illustrated in much detail in Tibshirani et al.'s paper [20]. Simply put, the higher the gap statistic is, the more natural the resulting clusters will be. Figure 2 shows the gap statistic where there are local peaks at $k = 3$ and $k = 5$. However, the Silhouette coefficient hits a local minima at $k = 5$ but peak at $k = 3$. Balancing the Silhouette coefficient and the Gap statistic, we decided to use $k = 3$ as the number of clusters for the following cluster analysis.

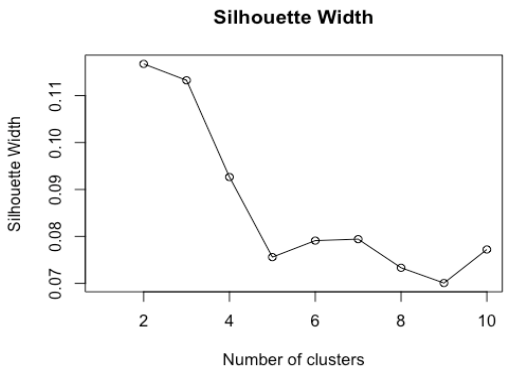


Figure 1 – The Silhouette Coefficient

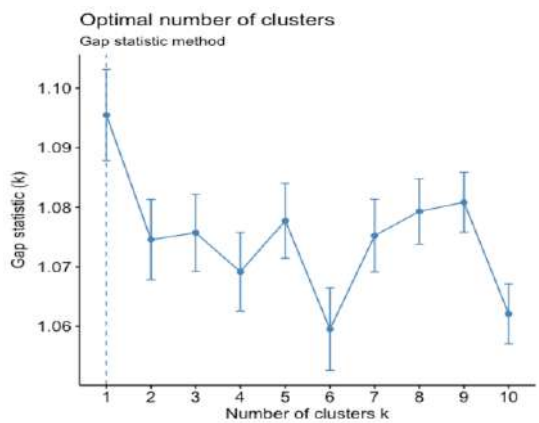


Figure 2 – The Gap statistic

Results

In Table 2, we present the demographical characteristics of the three clusters produced by K-prototypes and PAM. In Table 3, we present the medical history features of the subgroups. In Table 4, we present the health indicator features of the subgroups. The p-values of the Kruskal-Wallace test are marked next to each attribute. The results of K-prototypes are listed on the left of each column, and the results of PAM are on the right. For attributes that contain multiple categories, we only report the categories that are of the majority within the clusters.

Table 2 – Demographical features of subgroups (: p< 0.001)*

Attributes	Cluster 1		Cluster 2		Cluster 3	
Size	379	399	685	520	684	829
Gender* (Male=1)	.44	.64	.62	.73	.36	.24
Age*	72.9	67.1	49.7	43.8	30.3	39.6
Race* (Hispanic)					.55	.72
Race (White)	.48	.55	.33	.58		
Employment* (Retired)	.69	.61				
Employment* (Unemployed)			.76	.76	.71	.72

Table 3 – Medical history features of subgroups

Attributes	Cluster 1		Cluster 2		Cluster 3	
Size	379	399	685	520	684	829
Cancer* (Yes=1)	.38	.37	.16	.09	.09	.14
Diabetes* (Yes=1)	.28	.24	.18	.13	.07	.14
Mental illness* (Yes=1)	.39	.37	.31	.37	.32	.29

Table 4 – Health indicators of subgroups

Attributes	Cluster 1		Cluster 2		Cluster 3	
Size	379	399	685	520	684	829
BMI	26.1	26.0	28.4	27.8	28.0	28.6
HR*	84.3	87.8	87.4	92.7	98.9	92.0
RR*	17.9	17.9	17.8	17.8	17.9	17.9
TempC*	36.8	36.8	36.8	36.8	36.9	36.9
SpO2*	97.3	97.4	98.0	98.0	98.4	98.3
Binge Drinking (Yes=1)*	.27	.23	.49	.83	.30	.28
Tobacco Use (Yes=1)*	.10	.07	.37	.70	.32	.12
Substance Abuse (Yes=1)*	.06	.07	.15	.25	.13	.07

Next, we summarize and describe the distinguishable characteristics of the three subgroups that have emerged from the cluster analysis. The two algorithms seem to diverge in deciding the Binge Drinking and Tobacco Use categories for cluster 2, but if we compare them across different subgroups, they still remain meaningful. Overall we observed differences that are statistically significant across the three groups through the Kruskal-Wallace test.

The Elderly Subgroup

The two cluster algorithms all clearly identified the elderly subgroup, whose members have an averaged age of 67.1~72.9. More than 62% of this subgroups are retired and primarily white. Members in the elderly subgroup also have relatively healthy habits and normal BMI, with the highest proportion (73%~77%) that has never engaging in binge drinking, the highest proportion (90%~93%) of currently non-smokers, and the highest proportion (93%~94%) without substance abuse problems. However, the elderly subgroup has a larger proportion that have cancer (38%) and diabetes (28%), compared to the middle-aged subgroup where the proportion of having cancer is 16% and diabetes is 18%, and the adult subgroup (where the proportion of having cancer is 9% and diabetes 7%).

The Middle-aged Subgroup

The largest proportion of the middle-aged subgroup are unemployed (76%) and male (62%~73%) with an averaged age of 44~49 years old. Compared to the elderly subgroup, the middle-aged subgroup has a lower proportion with cancer (16%) and diabetes (18%). However, the middle-aged subgroup has less healthy habits in general – 30%~66% are current tobacco smokers and 15%~25% have substance abuse issues, which are higher than the elderly subgroup (6%~7%). The

middle-aged subgroup also has higher BMI are considered as overweight concerning their age groups.

The Adult Subgroup

This subgroup consists of primarily unemployed (72%) females (64%~76%) that are generally below 40 years old (younger than the middle-aged subgroup). They have the least proportion of having cancer (9%) and diabetes (7%). Compared to the middle-aged subgroup, they have healthier habits – less proportion are current smokers (12%~32%) and only 7%~13% have substance abuse issues. This subgroup also has higher than average BMI and are considered obese concerning their age group.

Discussion

While the two algorithms have some slight discrepancies in deciding certain characteristics, many of their results are consistent. Comparing across the three subgroups, our findings are consistent with previous literature that older population, unemployed population and patients with mental illnesses are more likely to become frequent flyers. However, one significant contribution of our study is that the frequent flyer population consists of subgroups that are substantially different from each other. If only looking at a macro level, these subgroup characteristics may be easily obfuscated. For instance, the general characteristics such as chronic health conditions simply do not apply to all subgroups – the middle-aged subgroup and the adult subgroup seem to have other concerns that lead to their overuse of the ED resource. In addition, the adult subgroup that has the lowest proportion of having chronic health conditions and unhealthy behaviors also takes up a large part of the frequent flyer population, and reasons behind this are rarely explored in existing literature. More investigation needs to be done for the adult female subgroup in order to understand their concerns. Unemployment and burdens of child raising could be factors that contribute to this subgroup's overuse of ED resource. The elderly group who need to cope with chronic health conditions may have difficulty commuting to the right medical facility, and ED may be the most feasible choice in such situations. For the middle-aged male subgroup, less healthy behaviors such as binge drinking and substance abuse may play a bigger role in overusing the ED resource. Thus, simply implementing interventions that are built upon the general frequent flyer characteristics is not able to address the concerns faced by the those different subgroups. Hospitals may consider first investigate what subgroups of frequent flyers constitute the patient populations at their EDs, why certain subgroups are more likely to make repeated visits to EDs, and devise separate intervention programs that target at those subgroups to address their varying needs.

The frequent flyer population at different locations may differ significantly, and our study has suggested cluster analysis as a feasible and promising exploratory stage to understand which subgroups constitute their frequent flyer population at a particular site. Simply borrowing intervention programs from hospitals at different locations or of different settings may fail to suit the needs of the particular frequent flyer subgroup at one hospital.

Limitations and Future Work

Our analysis was based on the structured Electronic Health Records of patients at a single academic medical center, and therefore the results may not be generalizable to hospitals at other locations or of other settings. In addition, while our analysis was able to reveal the subgroups of the frequent flyer population of the study site, the data we used lacks the explanatory power to uncover why these three groups become frequent flyers of ED. It is our future work to extend the analysis to clinical notes that contain richer information of patients' visits.

Conclusions

We performed cluster analysis on 1748 frequent flyers at an academic Medical Center. Three subgroups with substantially different characteristics have emerged from our analysis: the elderly subgroup with chronic health conditions, the middle-aged male subgroup with unhealthy behaviors and the adult female subgroup that are generally healthy. Our findings suggested that the commonly used general characteristics may not apply to all subgroups and different subgroups may face varying challenges in reaching out the right medical resource. Our study sets the stage for more tailored subgroup-level interventions to reduce ED overuse in hospitals across the world.

Acknowledgements

The project described was supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institute of Health, USA, through Grant UL1 TR001414. The content is solely the responsibility of the authors and does not necessarily represnet the official views of the NIH.

References

[1] R.W. Schafermeyer and B.R. Asplin, Hospital and emergency department crowding in the United States, *Emerg Med* **15** (2003), 22–27.
[2] R.A. Barish, P.L. Mcgauly, and T.C. Arnold, Emergency room crowding: a marker of hospital health, *Trans A. Clin Climatol Assoc* **123** (2012), 304–311.
[3] P. Bodenmann, V.-S. Velonaki, J.L. Griffin, S. Baggio, K. Iglesias, K. Moschetti, O. Ruggeri, B. Burnand, J.-B. Wasserfallen, F. Vu, J. Schupbach, O. Hugli, and J.-B. Daeppen, Case management may reduce emergency department frequent use in a universal health coverage system: a randomized controlled trial, *J Gen Intern Med* **32** (2017), 508–515.
[4] K.A. Hunt, E.J. Weber, J.A. Showstack, D.C. Colby, and M.L. Callaham, Characteristics of frequent users of emergency departments, *Ann Emerg Med* **48** (2006), 1–8.
[5] C.A. Grover and R.J. Close, Frequent users of the emergency department: risky business, *West J Emer. Med* **10** (2009), 193–194.
[6] E. LaCalle and E. Rabin, Frequent users of emergency departments: the myths, the data, and the policy implications, *Ann Emerg Med* **56** (2010), 42–48.
[7] Y. Ledoux and P. Minner, Occasional and frequent repeaters in a psychiatric emergency room, *Soc Psychiatry Psychiatr Epidemiol* **41** (2006), 115–121.
[8] L.E. Mehl-Madrona, Prevalence of psychiatric diagnoses among frequent users of rural emergency medical services, *Can J Rural Med* **13** (2008), 9.
[9] J. Billings and M.C. Raven, Dispelling an urban legend: frequent emergency department users have substantial

burden of disease, *Health Aff Proj Hope* **32** (2013), 2099–2108.
[10] F.S.J. Blank, H. Li, P.L. Henneman, H.A. Smithline, J.S. Santoro, D. Provost, and A.M. Maynard, A descriptive study of heavy emergency department users at an academic emergency department reveals heavy ED users have better access to care than average users, *J Emerg Nur.* **31** (2005), 139–144.
[11] P. Milbrett and M. Halm, Characteristics and predictors of frequent utilization of emergency services, *J Emerg Nurs* **35** (2009), 191–198.
[12] R.A. Meyers, ed., *Encyclopedia of Analytical Chemistry: Applications, Theory, and Instrumentation*, Wiley, Chichester, NY, 2000.
[13] J.-A. Huang, R.-H. Weng, C.-S. Lai, and J.-S. Hu, Exploring medical utilization patterns of emergency department users, *J Formos Med Assoc* **107** (2008), 119–128.
[14] S.N. Hastings, C. Horney, L.R. Landerman, L.L. Sanders, M.B. Hocker, and K.E. Schmader, Exploring patterns of health service use in older emergency department patients, *Acad Emerg Med* **17** (2010), 1086–1092.
[15] R.O. Duda, *Pattern Classification and Scene Analysis*, Wiley, Chichester, NY, 1973.
[16] Z. Huang, Extensions to the k-means algorithm for clustering large data sets with categorical values, *Data Min Knowl Discov* **2** (1998), 283–304.
[17] L. Kaufman and P.J. Rousseeuw, *Finding Groups in Data, An Introduction to Cluster Analysis*, Wiley, Chichester, NY, 1990.
[18] J.C. Gower, A general coefficient of similarity and some of its properties, *Biometrics* **27** (1971), 857–871.
[19] P.J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *J Comput Appl Math* **20** (1987), 53–65.
[20] R. Tibshirani, G. Walther, and T. Hastie, Estimating the number of clusters in a data set via the gap statistic, *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63** (2001) 411–423.

Address for correspondence

Lu He
Department of Informatics,
Donald Bren School of Information and Computer Sciences,
University of California, Irvine
E-mail: lhe11@uci.edu

Kai Zheng, PhD
Department of Informatics,
Donald Bren School of Information and Computer Sciences,
University of California, Irvine
E-mail: kai.zheng@uci.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190204

An Ensemble Deep Learning Model for Drug Abuse Detection in Sparse Twitter-Sphere

Han Hu[a], NhatHai Phan[a], James Geller[a], Stephen Iezzi[a], Huy Vo[b], Dejing Dou[c], Soon Ae Chun[d]

[a] *Ying Wu College of Computing, New Jersey Institute of Technology, Newark, NJ, USA*
[b] *Department of Computer Science, The City College of New York, New York, NY, USA*
[c] *Computer and Information Science, University of Oregon, Eugene, OR, USA*
[d] *Information Systems & Informatics, City University of New York, Staten Island, NY, USA*

Abstract

As the problem of drug abuse intensifies in the U.S., many studies that primarily utilize social media data, such as postings on Twitter, to study drug abuse-related activities use machine learning as a powerful tool for text classification and filtering. However, given the wide range of topics of Twitter users, tweets related to drug abuse are rare in most of the datasets. This imbalanced data remains a major issue in building effective tweet classifiers, and is especially obvious for studies that include abuse-related slang terms. In this study, we approach this problem by designing an ensemble deep learning model that leverages both word-level and character-level features to classify abuse-related tweets. Experiments are reported on a Twitter dataset, where we can configure the percentages of the two classes (abuse vs. non abuse) to simulate the data imbalance with different amplitudes. Results show that our ensemble deep learning models exhibit better performance than ensembles of traditional machine learning models, especially on heavily imbalanced datasets.

Keywords:

Machine Learning, Social Media, Substance-Related Disorders

Introduction

Misuse and abuse of prescription drugs and of illicit drugs have been major public health problems in the United States for decades. A "Public Health Emergency" declared in 2016 [1] and several official surveys [2] all show that the problem has been getting worse in recent years. For example, the most recent reports from the National Survey on Drug Use and Health (NSDUH) [2] estimate that 10.6% of the total population of people ages 12 years and older (i.e., about 28.6 million people) have misused illicit drugs in 2016, which represents an increase of 0.5% over 2015. According to the Centers for Disease Control and Prevention (CDC), opioid drugs were involved in 42,249 known deaths in 2016 nationwide [3] . In addition, the number of heroin-related deaths has been increasing sharply over five years and has surpassed the number of firearm homicides in 2015 [4] . The emerging new problems, such as the epidemic of illicitly manufactured fentanyl (IMF) [5] , marijuana-related traffic accidents [6] , and marijuana use among adolescents [7] are posing further increasing threats to public health.

To fight this epidemic of drug abuse, methods of social media monitoring with wider scope and shorter response time are needed. Social media, such as Twitter, have been proven to be sufficient and reasonably reliable data sources for social-level detection and monitoring tasks [8] . Twitter is a popular social media platform that has 100 million daily active users and 500 million daily tweets [10] (messages posted by Twitter users), most of which are publicly accessible, on a wide range of topics.

We are using algorithms for filtering and classification for acquiring abuse-related tweets for analysis and monitoring. Filtering is the very first and most basic step toward extracting potentially useful tweets from the large number posted every day. Filtering, by itself, even with standard drug names (e.g. heroin), generally does not suffice to produce a dataset pure enough for practical use. Thus, machine learning classifiers have to be trained to further identify tweets that are related to drug abuse. However, most abuse-related Twitter datasets have the problem of imbalanced class distributions. Typical datasets, collected with only the names of drugs, may have 5% to 30% of positive (abuse-related) tweets, due to the topic diversity and language irregularity of tweets. The percentage of positive tweets decreases sharply when more keywords, especially slang names for drugs (e.g., snow) and abuse behavior keywords (e.g., snorting), are included in a tweet dataset. The imbalanced class distribution and the noisy nature of the Twitter data make it hard to train a classifier with good performance.

In this paper, we propose an ensemble of two types of deep learning-based methods as better options, among classifiers, for situations in which the collected data is inevitably imbalanced, because they are more robust than traditional machine learning models. Our ensemble deep learning model combines word-level CNN models and character-level CNN models to perform classification. We compare our models with baseline models on a dataset we collected, where we can configure the class distribution of positive versus negative tweets in the training data and test data. By changing the percentage of positively and negatively labeled data in the dataset, we can simulate the imbalanced datasets that were collected by different means. We validate the performance of different models in a variety of settings to gain a clearer picture of how imbalanced data affect classification performance.

Related Works

Large scale surveys, such as NSDUH [2] , Monitoring the Future [11] , the MedWatch program [12] , and the results derived from these surveys [13] , clearly show that there is an epidemic of drug abuse across the United States. However, a recent report [14] states that the estimated number of deaths due to prescription drugs could be inflated due to the difficulties in determining whether a drug is obtained by prescription or not. We assert that the ambiguities highlighted in this new report raise questions about the reliability of the earlier surveys,

and thus, such a report illustrates the potential value of social media-based studies.

In fact, several studies found positive correlations between Twitter data and real world data. Chary et al. [15] performed semantic analysis on 3.6 million tweets with 5% labeled and found significant agreement with the NSDUH data. Hanson et al. [16] conducted a quantitative analysis on 213,633 tweets discussing Adderall, and found positive geo-temporal correlations. Furthermore, Shutler et al. [17] performed a qualitative analysis of prescription opioid-related tweets and found that indications of abuse were common. On the other hand, several studies focused on designing machine learning models to preform tweet classification. Mahata et al. [18] performed a comprehensive study on using deep learning models to identify mentions of drug intake in tweets. Katsuki et al. [20] trained support vector machine (SVM) on a dataset of 1,000 tweets for classification of tweets for relevance and favorability of online drug sales. Hu et al. [19] showed the potential of applying deep learning models in a drug abuse monitoring system to detect abuse-related tweets. Sarker et al. [9] proposed an ensemble of traditional machine learning models to classify drug abuse tweets and non-abuse tweets of certain drugs. Other studies focus on social media users, such as Fan's work [26] utilizes user interaction networks to identify opioid users on Twitter. In this paper, we will be developing ensemble deep learning models to expand the classification of tweets to a border scope of drugs and their abuse behaviors with better performance in the unbalanced class distribution settings.

Methods

In this section, we present the definition of the *drug abuse-related risk behavior detection problem*, our methods for collecting tweets, our methods for labeling tweets, and our ensemble deep learning approach.

Problem Definition

In this paper, our first goal was to build a Twitter dataset consisting of tweets that are related to drug abuse risk behaviors (**positive** tweets), and tweets that are not (**negative** tweets). The "drugs" in the term "drug abuse risk behaviors" in this study include Schedule 1 and Schedule 2 drugs and their derivatives [21] , including marijuana, heroin, cocaine, fentanyl, etc. The reasons we included *marijuana* even though it is legalized in several states are that: **(1)** Marijuana is still a controlled substance in the federal law, whether for medical use or recreational use; and **(2)** Marijuana can still cause harm to adolescents [7] , can cause "use disorder" [13] , and is related to traffic fatalities [6] . The term "abuse risk behavior" can be defined as "The existence of likely abusive activities, consequences, and endorsements of drugs." Tweets that contain links to or summarize news and reports related to drug abuse, and tweets that merely express opinions about drug abuse, are counted as negative in this study. Our main goal in this paper is to train a model that can accurately classify positive and negative tweets in a highly imbalanced (drug abuse) dataset.

Data Collection

Although there are human-labeled drug abuse Twitter datasets (e.g. Sarker's dataset [9]) available, due to Twitter's data policy, which prohibits the direct sharing of tweet contents, by the time we access the tweets in that dataset, more than 40% of tweets are either removed or hidden from the public. This significantly affects the quality and integrity of existing publicly available datasets. Therefore, we need to build a new dataset from scratch. In our framework, raw tweets are collected through a set of Application Programming Interfaces (Twitter APIs) via keyword filtering. By defining a set of keywords, the API will fetch tweets that contain any of the keywords from either the real-time stream of tweets or from archived tweets. For a more complete coverage of drug-related topics, we selected three types of keywords: **(1)** Full and official names of drugs, e.g. marijuana, cocaine, OxyContin, fentanyl, etc.; **(2)** Slang terms for drugs, e.g. pot, blunt, coke, crack, smack, etc.; and **(3)** Drug abuse-related behaviors and symptoms, e.g., high, amped, addicted, headache, dizzy, etc. The number of keywords we used is limited to 400 by the Twitter APIs.

Data Annotation

We built a comprehensive guide, accessable at **https://goo.gl/tqWddS**, based on Sarker's guide [9] . Each one of the three members in our research team with experience in health informatics annotated the 1,794 tweets from Hu et al.'s study [19] independently following the guide. A final label for each seed tweet is determined by majority voting from the three labels.

To acquire annotated tweets rapidly, at low cost, and with increased percentage of positive tweets, we **(1)** used these labeled tweets as *"seed"* tweets to train a SVM classifier; **(2)** ran the SVM classifier on the unlabeled dataset, and randomly sampled 5,000 machine labeled tweets that have a prediction probability (esitmated with Platt scaling) > 0.8; and **(3)** posted the 5,000 tweets (without identification information) onto the Amazon Mechanical Turk (AMT) crowdsourcing platform for annotation. AMT is a well-known crowdsourcing platform where Posters can post Human Intelligence Tasks (HITs) and Workers finish HITs for micro-payments. A literature study [22] evaluated AMT as a thrustworthy platform to obtain human labeled data. The same guide is used to guide the Workers on how to annotate the tweets. Each tweet is posted as one HIT that requires the Worker to label it as positive or negative following the guide. Each HIT is replicated as three *assignments* to be completed by three individual Workers. We set the price of each assignment to be $0.05, a very generous price compared to what was reported in Buhrmester's work [22] All HITs are completed within hours after being posted. The final label of each tweet is aggregated from the three labels by majority voting. We also label 1,000 tweets randomly sampled from the 5,000 tweets with our annotator as a measure of quality check.

Feature Extraction

Machine learning models require numerical features to work. Feature extraction transforms text features into numerical features in the form of vectors. To cover the content ambiguity in drug abuse-related tweets, a variety of feature extraction methods are used in this study. In our word-level, CNN models, we use pre-trained word embedding models that were trained on large corpora to transform words into dense vectors. We tested several pre-trained models as Mahata's work [18] suggested. With our word-level CNN model, the *Drug Chatter* embedding had the best average performance on our dataset; thus, it was chosen as the pre-trained, word embedding model for this study. The details of the tested, word embedding models are shown in Table 1. Each tweet is converted to a sequence of 400-dimensional vectors. Considering that the length limit of each tweet nowadays is 280 characters, the sequence length is set to 40. In our char-level CNN, the preprocessing step only turns all characters to lower case as suggested by [23] . Each char. is then converted into a 128-dimensional trainable

randomly-initialized vector. Instead of being fixed, the character embeddings are trained along with other layers in the model.

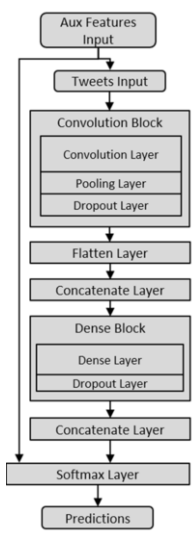
Table 1. Details of Pre-Trained Word Embedding

Name	Model	Corpus	Dimension
GoogleNews	Word2vec	~100 billion words	300
Glove Common	Glove	~42 billion words	300
Godin	Word2vec	~400 million tweets	400
Drug Chatter	Word2vec	~1 billion tweets	400

We also replicated the features extracted in Sarker et al. study [9] , including: **(1)** The tokenization process; **(2)** The abuse-indicating term features, consisting of the presence and the counts of abuse-indicating terms obtained from Hanson et al. [16] ; **(3)** The drug-slang lexicon features, consisting of the presence and the counts of terms longer than five characters found in an online drug abuse dictionary [24] ; **(4)** The word cluster features, represented by 150-dimensional one-hot vectors, were constructed by identifying words that belong to certain word clusters in a dataset [9] that contains 150 drug-related word clusters; and **(5)** The synonym expansion features, accomplished by identifying all synonyms of all nouns, verbs, and adjectives in the tokenized tweets using WordNet [25] .

An Ensemble Deep Learning Model for Drug Abuse Detection in Sparse Twitter-Sphere

In this section, we present our novel, ensemble deep learning model for drug abuse risk behavior detection by integrating extracted features from tweets into CNN models. Our ensemble model takes the outputs of multiple prediction models, word-level CNN (**W-CNN**) and char-level CNN (**C-CNN**) [29] in our case, and feed them to a meta-learner that provides the final predictions. We designed W-CNN and C-CNN for this task. In fact, both the W-CNN and the C-CNN share a similar structure as shown in Figure 1.



Input Parameters

Parameter	W-CNN	C-CNN
tweet input dimension	[40,400]	[280,128]
aux feature dimension	[154,]	[154,]

Convolution Block Parameters

Parameter	W-CNN	C-CNN
kernel width	{2,3,4,5}	{3,5,7,10}
# kernels	128 for each kernel width	256 for each kernel width
activation function	Relu	Tanh
pooling	MaxPooling	GloablMax-Pooling
dropout	Uniform-Dropout with rate=0.1	Alpha-Dropout with rate=0.1
# blocks	2	1

Dense Block Parameters

Parameter	W-CNN	C-CNN
# hidden units	1024	1024
activation function	Relu	Selu
dropout	Uniform-Dropout with rate=0.5	Uniform-Dropout with rate=0.5
# blocks	2	2

Figure 1. Ensemble CNN Model Structures

The inputs of our W-CNN are vectors of shape [40, 400] where 40 is the maximum sequence length (number of words allowed) in an input tweet, and 400 is the length of the pre-trained word embeddings. The input of our C-CNN is shaped as [280, 128] where 280 is the maximum possible length of a tweet, and 128 is the length of the vector representation of each character in the charset. The auxiliary features in the input include: **(1)** The synonym expansion features in the form of synonymous words are directly concatenated with the input tweets (before they are transformed into vectors); and **(2)** The remaining auxiliary features, in the form of 154-dimensional vectors, are concatenated to the last hidden layer of the dense layers. For each convolution kernel size, the W-CNN model has two convolution layers with ReLU activation functions stacked together. Each is followed by a max-pooling layer.

The C-CNN model has one convolution layer for each convolution kernel size with Tanh activation function, followed by a global-max-pooling layer that performs max-pooling over all outputs of convolution layers with different kernel sizes. Both models have one dense layer block, consisting of two dense layers with 1,024 hidden units each, and one Softmax output layer with two units. The activation functions are slightly different, as the W-CNN model uses ReLU, while the C-CNN model uses SELU. The output of the last hidden layer is concatenated with vectors of abuse-indicating term features, drug-slang lexicon features, and word cluster features, before being fed into the output layer.

Finally, a number of independently trained CNN models of both types are ensembled together by using majority. Model ensembles were also used in Sarker et al. study [9] to reduce variability and bias, in order to improve prediction performance. We apply the same ensemble strategy to both our deep learning models and the baseline models.

Experimental Design

Our main objective in this experiment was to directly compare the performances of the ensemble traditional machine learning model and the ensemble deep learning model. For the ensemble traditional machine learning model, two of each type of baseline models, six in total, are trained and ensembled together. For the ensemble deep learning model, six models of three types (two for each type) are used. The three types are denoted as follows. **(1)** "*char_aux*" is the char-level CNN model with auxiliary features. **(2)** "*char_cnn*" is the plain char-level CNN without any auxiliary features. **(3)** "*word_aux*" is the word-level CNN model with all auxiliary features. For deep learning models, it is extremely easy to overfit, due to the rather small number of training and test data elements; thus, the model is saved at each training epoch, and the best epoch is found among the saved models. For each class distribution scenario, each model is trained with the same six sets of training data and tested on the corresponding test data. All results reported are averaged results from the 6-fold cross-validation.

Experimental Results

Data Annotation Results

From Jan 2017 to Feb 2017, we collected 3,265,153 tweets in total. The *"seed"* dataset that we annotated to be used to train the pre-filter consisted of 1,794 tweets, including 280 positive labels and 1,514 negative labels. Our annotator achieved the agreement score of 0.414, measured by Krippendoff's Alpha. For the AMT labeled dataset, we removed duplicate tweets from it, with a resulting dataset containing 4,736 tweets with 2,657 positive labels and 2,079 negative labels. The agreement score is 0.456 measured by Alpha, which can be considered as a reliable result in our study as demostrated in [27] , since: **(1)** We are performing data annotation with data aggregation to reduce variability, instead of typical content analysis [28]; **(2)**

The Krippendoff's Alpha is sensitive to data imbalance; and **(3)** We focus on sparse and imbalanced data distributions. For the 1,000 tweets for quality check, we received the Kappa score of 0.910 between our final labels and the labels we obtained from AMT. This is showing that our annotation guide was followed consistently by both our annotators and AMT Workers.

To simulate the data imbalance scenarios, we configured the class distribution and pre-sampled the dataset into six blocks for each distribution scenario, for 6-fold cross-validation. Each model was trained and tested on the same sets of training and test data to ensure a fair comparison. The number of data points included in each distribution scenario was maximized, but it was inevitably different between scenarios. Table 2 shows the dataset in each class distribution scenario.

Table 2. Dataset Variants

Class distribution (positive: negative)	# of training data items	# of test data items
50:50 split	3450	690
40:60 split	2850	570
30:70 split	2450	490
20:80 split	2150	430
10:90 split	1900	380

Drug Abuse Detection Results

Table 3 shows the results for all individual models and two ensemble models. The ensemble model results are separated from the individual models for easier viewing. The highest value of each measure is marked in bold font. There is an interesting trend in the results of the ensemble models. When the data is balanced or nearly balanced, the traditional ensemble machine learning model has a better performance than the ensemble deep learning model. At 50:50 and 40:60 splits, the ensemble machine learning model is superior over the ensemble deep learning model for most of the criteria. This is partially due to the relatively small dataset size. When the data becomes more imbalanced, e.g., at a 30:70 split, the ensemble deep learning model becomes better and has a higher F1-score for positive labels, compared with the traditional ensemble machine learning model. At 20:80 and 10:90 splits, the ensemble deep learning model takes the lead, most significantly in each measure for positive labels. The larger model capacity and the ability of the deep learning models to learn more complex non-linear functions, can better distinguish the semantic differences between positive tweets and negative tweets, when the distribution of classes is heavily imbalanced.

Looking at individual machine learning models, Random Forest and SVM showed a strong performance on all datasets, and they perform well when the dataset is balanced. Naïve Bayes also has a good performance on a balanced dataset, but on an imbalanced dataset, it is heavily biased towards negative labels and has a poor performance for positive labels. Deep learning models generally have more stable performance, compared to traditional machine learning models, across all datasets, and a smaller difference between precision and recall, but their peak performances are not as good. Comparing between deep learning models, auxiliary features do not provide C-CNN significant performance boost, and W-CNN performed worse than the C-CNN model. However, in additional results that are not shown in this paper due to space limitations, auxiliary features give the plain W-CNN model a performance boost.

By investigating the performance of each individual model and the ensemble model that includes it, we can see that our ensemble strategy works well for deep learning models, as most

Table 3. Experimental Results

Class Distribution: 50:50 split								
Measure	**Ensemble CNN**	**Ensemble ML**	**char_aux**	**char_cnn**	**word_aux**	**SVM**	**Random Forest**	**Naive Bayes**
Accuracy	0.8510	**0.8575**	0.8506	0.8477	0.8466	0.8415	**0.8586**	0.8384
Precision_p	**0.8468**	0.8350	0.8315	0.8240	0.8198	0.8063	**0.8404**	0.8319
Recall_p	0.8575	**0.8918**	0.8797	0.8845	0.8894	**0.9000**	0.8860	0.8493
F1_score_p	0.8520	**0.8623**	0.8549	0.8531	0.8529	0.8504	**0.8624**	0.8402
Class Distribution: 40:60 split								
Measure	**Ensemble CNN**	**Ensemble ML**	**char_aux**	**char_cnn**	**word_aux**	**SVM**	**Random Forest**	**Naive Bayes**
Accuracy	0.8567	**0.8582**	0.8528	**0.8563**	0.8430	0.8444	0.8494	0.8427
Precision_p	**0.8079**	0.8047	0.8007	0.8055	0.7818	**0.8104**	0.7770	0.7862
Recall_p	0.8428	**0.8531**	0.8421	0.8454	0.8443	0.7982	**0.8746**	0.8341
F1_score_p	0.8249	**0.8280**	0.8207	**0.8248**	0.8113	0.8041	0.8229	0.8093
Class Distribution: 30:70 split								
Measure	**Ensemble CNN**	**Ensemble ML**	**char_aux**	**char_cnn**	**word_aux**	**SVM**	**Random Forest**	**Naive Bayes**
Accuracy	**0.8599**	0.8595	0.8522	0.8507	0.8483	0.8429	**0.8537**	0.8452
Precision_p	0.7402	**0.7426**	0.7253	**0.8223**	0.718	0.7467	0.7137	0.7218
Recall_p	**0.8231**	0.8163	0.8209	0.8180	0.8158	0.7234	**0.8583**	0.7914
F1_score_p	**0.7792**	0.7771	0.7695	0.7666	0.7635	0.7336	**0.7789**	0.7538
Class Distribution: 20:80 split								
Measure	**Ensemble CNN**	**Ensemble ML**	**char_aux**	**char_cnn**	**word_aux**	**SVM**	**Random Forest**	**Naive Bayes**
Accuracy	**0.8674**	0.8508	**0.8624**	0.8568	0.8506	0.8384	0.8475	0.8527
Precision_p	**0.6416**	0.5908	**0.6325**	0.6128	0.5965	0.5640	0.5838	0.6261
Recall_p	0.7713	**0.8295**	0.7558	0.7868	0.8023	**0.8547**	0.8295	0.6609
F1_score_p	**0.7001**	0.6900	**0.6878**	**0.6878**	0.6823	0.6792	0.6850	0.6425
Class Distribution: 10:90 split								
Measure	**Ensemble CNN**	**Ensemble ML**	**char_aux**	**char_cnn**	**word_aux**	**SVM**	**Random Forest**	**Naive Bayes**
Accuracy	**0.8728**	0.8636	0.8638	0.8664	0.8445	0.8355	0.8592	**0.8961**
Precision_p	**0.4338**	0.3975	0.4112	0.4153	0.3760	0.3609	0.3875	**0.4762**
Recall_p	**0.7281**	0.6754	0.7368	0.7346	0.7171	**0.8114**	0.6776	0.2939
F1_score_p	**0.5389**	0.4999	0.5243	**0.5275**	0.4882	0.4990	0.4925	0.3611

of the measures for the ensemble model are higher than for any of its components corresponding measures. This effect was only observed a few times for traditional machine learning models. We expect that, by using more complicated ensemble strategies, deep learning has the potential to reach an even better performance level.

Conclusions

In this study, we investigated how the data imbalance issue influences the performance of classifiers that are trained for identifying tweets that are related to drug abuse. We first collected a dataset with a broad selection of drug abuse-related keywords and slang terms. We explored the use of the Amazon Mechanical Turk platform as a reliable source for acquiring human-labeled tweets, and we obtained a solid dataset. We designed an ensemble deep learning classification model with both word-level and char-level CNNs, and we conducted a direct comparison with traditional machine learning models on our dataset, with simulated class imbalance. Experimental results show that our ensemble deep learning models have better performance than traditional machine learning models when the data is off-balance. Results also show that the ensemble strategy we used is effective for improving deep learning models. Finally, our analysis of the collected three million tweets, labeled by our model, shows an interesting temporal pattern that agrees with our intuition.

Acknowledgements

The authors gratefully acknowledge the support from the National Science Foundation (NSF) grants CNS-1650587, CNS-1747798, CNS-1624503, and CNS-1850094.

References

[1] H.P. Office, HHS Acting Secretary Declares Public Health Emergency to Address National Opioid Crisis, (2017). https://www.hhs.gov/about/news/2017/10/26/hhs-acting-secretary-declares-public-health-emergency-address-national-opioid-crisis.html.

[2] T.S. Abuse, and M.H.D. Archive, National Survey on Drug Use and Health, (2018). https://www.samhsa.gov/data/data-we-collect/nsduh-national-survey-drug-use-and-health.

[3] N.I. on Drug Abuse, Overdose Death Rates (2018). https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates.

[4] G.V. Archive, 2015 Gun Violence Archive (2018). http://www.gunviolencearchive.org/past-tolls.

[5] J.K. O'Donnell, J. Halpin, C.L. Mattson, B.A. Goldberger, and R.M. Gladden, Deaths Involving Fentanyl, Fentanyl Analogs, and U-47700-10 States, July-December 2016, *MMWR Morb Mortal Wkly Rep*. **66** (2017), 1197–1202.

[6] B. Hansen, K.S. Miller, and C. Weber, Early Evidence on Recreational Marijuana Legalization and Traffic Fatalities, Nat'l Bu. of Econ. Res., 2018.

[7] E.J. D'Amico, A. Rodriguez, J.S. Tucker, E.R. Pedersen, and R.A. Shih, Planting the seed for marijuana use: Changes in exposure to medical marijuana advertising and subsequent adolescent marijuana use, cognitions, and consequences over seven years, *Drug Alcohol Depend*. **188** (2018), 385–391.

[8] T. Sakaki, M. Okazaki, and Y. Matsuo, Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors, in: Proc. 19th Int. Conf. WWW, Raleigh, North Carolina, USA, 2010: pp. 851–860.

[9] A. Sarker, and others, Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from Twitter, *Drug Saf*. **39** (2016), 231–240.

[10] S. Aslam, Twitter by the Numbers, (2018). https://www.omnicoreagency.com/twitter-statistics/.

[11] A. Arbor, Monitoring The Future, (2018). http://www.monitoringthefuture.org/.

[12] U. FDA, MedWatch: The FDA Safety Information and Adverse Event Reporting Program, (2018). https://www.fda.gov/safety/medwatch/.

[13] D.S. Hasin, and others, US adult illicit cannabis use, cannabis use disorder, and medical marijuana laws: 1991-1992 to 2012-2013, *Jama Psychiatry* **74** (2017), 579–588.

[14] P. Seth, R.A. Rudd, R.K. Noonan, and T.M. Haegerich, Quantifying the Epidemic of Prescription Opioid Overdose Deaths, *Am J Public Health Res*. **108** (2018), 500–502.

[15] M. Chary, N. Genes, C. Giraud-Carrier, C.L. Hanson, L.S. Nelson, and A.F. Manini, Epidemiology from Tweets: Estimating Misuse of Prescription Opioids in the USA from Social Media, *J Med Toxicol*. **13** (2017), 278–286.

[16] C.L. Hanson, S.H. Burton, C. Giraud-Carrier, J.H. West, M.D. Barnes, and B. Hansen, Tweaking and tweeting: exploring Twitter for nonmedical use of a psychostimulant drug (Adderall) among college students, *J Med Internet Res*. **15** (2013).

[17] L. Shutler, L.S. Nelson, I. Portelli, C. Blachford, and J. Perrone, Drug use in the Twittersphere: a qualitative contextual analysis of tweets about prescription drugs, *J Addict Dis*. **34** (2015) 303–310.

[18] D. Mahata, J. Friedrichs, Hitkul, and R.R. Shah, #phramacovigilance -Exploring Deep Learning Techniques for Identifying Mentions of Medication Intake from Twitter, *CoRR*. (2018). http://arxiv.org/abs/1805.06375.

[19] H. Hu, and others, Deep Learning Model for Classifying Drug Abuse Risk Behavior in Tweets, in: 2018 IEEE Int. Conf. Healthcare Inform. (ICHI), IEEE, 2018: pp. 386–387.

[20] T. Katsuki, T.K. Mackey, and R. Cuomo, Establishing a link between prescription drug abuse and illicit online pharmacies: analysis of Twitter data, *J Med Internet Res*. **17** (2015).

[21] O. of the Law Revision Counsel, Drug Abuse Prevention and Control. Definitions. 21 U.S.C Sect. 802, (2018).

[22] M. Buhrmester, T. Kwang, and S.D. Gosling, Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspect Psychol Sci*. **6** (2011), 3–5.

[23] X. Zhang, J.J. Zhao, and Y. LeCun, Character-level Convolutional Networks for Text Classification, *CoRR*. (2015). http://arxiv.org/abs/1509.01626.

[24] NoSlang.com, Drug Slang Translator, (2018). https://www.noslang.com/drugs/dictionary.php.

[25] G.A. Miller, WordNet: a lexical database for English, *Commun ACM*. **38** (1995), 39–41.

[26] Y. Fan, Y. Zhang, Y. Ye, and X. Li, Automatic Opioid User Detection from Twitter: Transductive Ensemble Built on Different Meta-graph Based Similarities over Heterogeneous Information Network., in: IJCAI, 2018: pp. 3357–3363.

[27] L.A. Jeni, J.F. Cohn, and F. De La Torre, Facing Imbalanced Data–Recommendations for the Use of Performance Metrics, in: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, IEEE, 2013: pp. 245–251.

[28] K.A. Hallgren, Computing inter-rater reliability for observational data: an overview and tutorial, *Tutorials in Quantitative Methods for Psychology* **8** (2012), 23.

[29] X. Zhang, J.J. Zhao, and Y. LeCun, Character-level Convolutional Networks for Text Classification, CoRR. (2015). http://arxiv.org/abs/1509.01626.

Address for correspondence

Corresponding author: NhatHai Phan, New Jersey Institute of Technology, NJ, USA, Email: phan@njit.edu.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190205

Handwriting Features of Multiple Drawing Tests for Early Detection of Alzheimer's Disease: A Preliminary Result

Tatsuya Ishikawa[a], Miyuki Nemoto[b], Kiyotaka Nemoto[b], Tomoko Takeuchi[b], Yuriko Numata[b], Ryohei Watanabe[b], Eriko Tsukada[b], Miho Ota[b], Shinji Higashi[b], Tetsuaki Arai[b], Yasunori Yamada[a]

[a] IBM Research, Tokyo, Japan
[b] University of Tsukuba, Ibaraki, Japan

Abstract

Early detection of Alzheimer's disease (AD) has become increasingly important. Healthy monitoring technology focusing on behavioral changes is a promising approach in this vein. Among such technologies, handwriting features measured by digital tablet devices have attracted attention as potential indicators for detecting AD and mild cognitive impairment (MCI). However, previous studies have mainly investigated features in single tasks, and it remains unclear whether combining the features of multiple tasks could improve the performance of detecting AD and MCI. In this study, we investigated features in five representative tasks used in neuropsychological tests collected from 71 seniors including some diagnosed with MCI and AD. We found that our three-class classification model improved diagnosis accuracy by up to 11.3% by combining features of multiple tasks, for a final accuracy of 74.6%. We also suggested that drawing behaviors during multiple tasks might be useful for estimating disease progression simply by utilizing the labels of disease groups.

Keywords:

Dementia, Handwriting, Classification

Introduction

As the world's elderly population increases, the number of people living with dementia is rising rapidly, making dementia an increasingly serious health and social problem. According to a previous survey, around 47 million people globally were living with dementia as of 2015, corresponding to about 7.6% of the world's over-65-year-olds [1]. At the same time, diagnostic coverage worldwide remains low and dementia is often undiagnosed. Even in high-income countries, only 40–50% of dementia sufferers have received a diagnosis [2, 3]. The low diagnosis coverage makes it difficult for many patients and their families to receive appropriate support and care. In addition, while dementia affects the individuals with the disease, it also affects their supporters—including relatives and the wider society—because people with dementia require constant and costly care for years. In fact, healthcare costs have risen significantly, reaching over $818 billion USD in 2015, and this figure is estimated to rise to $2 trillion USD by 2030 [4]. One strategy to reduce some of this cost is early intervention at the mild cognitive impairment (MCI) or preclinical stages. In fact, longitudinal studies suggest the possibility of early intervention at the MCI stage to reduce the progression to dementia [5]. An intervention that could delay the onset of Alzheimer's disease (AD) by five years is estimated to result in a 57% reduction in the number of AD patients and to reduce 45% of the projected Medicare costs [6].

Health monitoring technology focusing on behavioral features is expected to help improve diagnosis coverage and to detect AD at an earlier stage by expanding opportunities for receiving assessment from the clinical setting to more varied situations including everyday situations. For example, previous lab studies have suggested that behavioral features in gait, speech, and eye movement can be useful indicators for identifying AD and MCI [7]. Being able to infer AD and MCI from these behaviors with better accuracy and wider applicability would be tremendously useful.

One promising behavior to explore is drawing behavior. An advantage of this approach is ease of data collection brought about by the popularization of portable devices such as tablets and smartphones. Drawing behavior assessments such as the Clock Drawing Test (CDT) [8] and the Trail Making Test (TMT) [9] have proven useful for measuring cognitive decline and detecting AD and are commonly used in conventional in-clinic neuropsychological tests. Recent research on drawing behaviors using tablet devices has shown the possibility of automatic detection of patients with cognitive or motor impairments [10–22]. For example, [10] extracted pressure and kinematic features during several tasks including the CDT, while [11] investigated frequencies, velocities, and temporal features during a variant of the TMT. These features have been shown as differentiating healthy subjects and AD patients. Although previous studies have demonstrated how we can build a model for detecting AD and/or MCI by using behavioral features during individual drawing tasks, whether and how we can improve the model performance by combining behavioral features during multiple tasks remains unexplored. In addition, most of these studies focused on developing a classification model for differentiating patients with MCI and AD. Being capable of inferring disease progression on a scalar or ordinal scale defined by in-clinic cognitive assessment scores or biomarkers such as amyloid beta and tau deposition would extend the scope of application, for example, through visualization of the effects of intervention and prevention.

In this study, we investigated drawing behaviors during five representative tasks used in in-clinic neuropsychological tests collected from 71 Japanese seniors including some diganosed with MCI and AD. We extracted a series of drawing behavioral features including pressure, velocity, acceleration, jerk, and in-air and on-screen durations and then built a three-class classification model to distinguish healthy controls (HCs), patients with MCI, and patients with AD. Through

Table 1 – Demographics of participants.

Status	No. of participants (Female)	Mean age (SD)	Mean MMSE score (SD)
HC	36 (21)	70.0 (5.0)	28.3 (1.5)
MCI	25 (15)	75.9 (5.3)	26.8 (3.1)
AD	10 (7)	76.7 (6.0)	18.8 (3.9)

comparison, we found that the model using features in all five tasks could improve accuracy by up to 11.3% by combining the features of multiple tasks, achieving a final accuracy of 74.6% (chance rate 40%). We next investigated using the model for inferring disease progression on a continuous scale. Specifically, we first trained our model to differentiate three groups using only drawing features and disease labels (HC, MCI, and AD), and then we investigated whether the model could estimate in-clinic cognitive assessment scores. The results showed that the scores estimated by the model were significantly correlated with in-clinic cognitive assessment scores, even though we did not use the assessment scores themselves. These results indicate that our approach focusing on drawing behaviors during multiple tasks might be useful for inferring AD progression.

Materials and Methods

Participants

A total of 71 participants were enrolled by the University of Tsukuba Hospital. Ten participants were patients with AD, 25 were patients with MCI, and 36 were HCs. All participants were evaluated with the Mini-Mental State Examination (MMSE), a screening measure of global cognitive functioning [23]. Table 1 shows the number of participants (female), mean age, and mean MMSE score for the HC, MCI, and AD groups.

None of the participants in the HC group were diagnosed as having MCI or dementia before the experiment. The definitions of the MCI and AD groups were based on diagnosis by psychiatrists through medical examinations including structural magnetic resonance imaging, blood tests, and neuropsychological tests. More specifically, the doctors followed the guidelines and criteria in [24] for MCI and [25] for AD. Informed consent was obtained from all participants in accordance with a procedure by the ethics committee, the University of Tsukuba Hospital (H29-65).

Apparatus

A digital drawing tablet (Wacom Cintiq Pro 16) was used to acquire handwriting movements. The detailed specifications of the tablet are as follows: external dimensions (width × depth × height) 410 × 265 × 17.5 mm, spatial resolution 3840 × 2160 dots, pixel size 0.090 × 0.090 mm, temporal resolution 30 ms, and pressure levels 8,192.

The data include 3D coordinates (x, y, z) and the pressure of the pen-tip, altitude and azimuth of the pen, a binary variable (1 for writing state and 2 for erasing state), and timestamp.

When the stylus touches the start button, the software starts acquiring the data, and a black line reproducing the written trace appears. Thus, participants can monitor in real-time what they are writing.

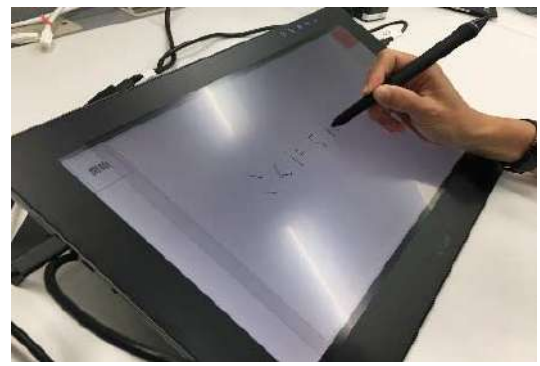


Figure 1 – Example of writing a spontaneous sentence.

Experimental procedure

Participants were seated on a chair with the digital tablet placed on a desk and could freely adjust the position of the device. Each participant was asked to perform five tasks:

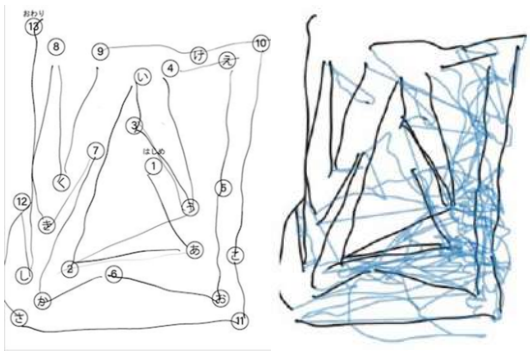
1. Writing spontaneous sentences. This task is included in the Mini-Mental State Examination (MMSE) [23]. Paticipants are asked to write any complete sentences on a sheet.
2. Drawing crossed pentagons. This is also a part of the MMSE. At first, participants see a figure of two intersecting pengagons. Then they are asked to draw the same figure as shown on a sheet.
3. Trail Making Test (TMT) part A. This task requires participants to draw lines connecting consecutive numbers randomly distributed on a sheet [9].
4. TMT part B. This is similar to the TMT part A, but instead of just linking numbers, participants are required to draw lines connecting numbers and letters alternately in their respective sequence [9].
5. Clock Drawing Test. This test asks participants to draw an analog clock-face showing 10 minutes after 10 on a blank sheet [8].

Figure 1 shows an example of writing a spontanous sentence on our tablet device.

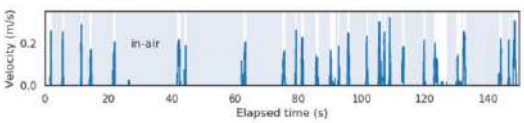
Data analysis

Pressure profiles were obtained from the apparatus as a raw dataset. Numbers of segments were counted to characterize the handwriting behaviors. As kinematic parameters, velocity (m/s), acceleration (m/s^2), and jerk (m/s^3) of the pen-tip movements over the 2D coordinates (x, y) on the tablet surface were computed. The durations (s) of on-screen and in-air stylus pen movements were considered as timing parameters. Figures 2 and 3 show examples of the handwriting behavioral features.

After completion of all five tasks, a three-class classification analysis using the handwriting parameters calculated to distinguish HC, MCI, and AD was performed. To investigate whether the model can be used to estimate in-clinic cognitive assessment scores, correlation analysis between MMSE scores and our model scores was also conducted.

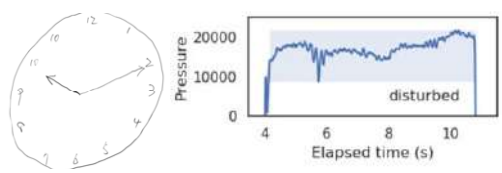


(a) Left: final outcome and right: extracted on-screen (black) and in-air (blue) trajectories

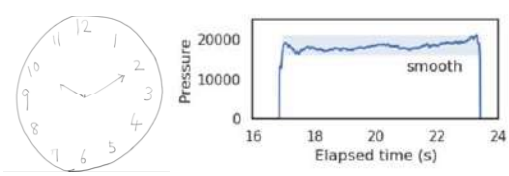


(b) On-screen pen-tip velocity and in-air duration (shaded area)

Figure 2 – Example of handwriting behavioral features of an MCI patient in TMT part A.



(a) Disturbed pressure history of MCI



(b) Smooth pressure history of HC

Figure 3– Example of a difference of pressure features during circle drawing in CDT between MCI and HC regardless of the similarity of the final outcomes.

Results

Three-class classification models

On the basis of the extracted handwriting features, classification analysis using a generalized linear model with a logit link function was performed. In order to differentiate HC, MCI, and AD, we assigned 0, 1, and 2 as targeted values, respectively. Explanatory variables were selected by means of a stepwise method to optimize the Akaike's Information Criterion (AIC). We evaluated three-class classification performances of the models using features of single tasks and all tasks.

Tables 2 and 3 show the accuracy and the selected features of the resultant classification models. As shown, the model that used features of all tasks improved the diagnosis accuracy by up to 11.3% and achieved an accuracy of 74.6% (chance rate 40%). The classification performance of this best model is presented in Table 4.

Table 2 – Classification performance.

Tasks	Accuracy (%)
MMSE sentences	67.6
MMSE pentagons	66.2
TMT part A	69.0
TMT part B	63.3
CDT	67.6
All tasks	74.6

Table 3 – Selected variables in classification models.

Tasks	Selected variables
MMSE sentences	Age, mean velocity, and mean pressure
MMSE pentagons	Age, mean pressure, CV of pressure, in-air duration, and mean velocity
TMT part A	Age, in-air duration, and CV of acceleration
TMT part B	Age, CV of pressure, CV of jerk, and on-screren duration
CDT	Age, number of segments, in-air duration, and CV of jerk
All tasks	Mean velocity and mean pressure (MMSE sentences), mean pressure (MMSE pentagons), on-screen duration (TMT part A), CV of pressure (TMT part B), and number of segments, in-air duration, and mean pressure (CDT)

Table 4 – Classification performance of the model using features from all tasks.

		Actual		
		HC	MCI	AD
Estimated	HC	35	11	1
	MCI	1	10	1
	AD	0	4	8

Estimation of in-clinic cognitive assessment scores from the classification models

We next investigated whether the three-class classification model could estimate in-clinic cognitive assessment scores. Specifically, we explored the relationship between a parameter in the classification model and the MMSE score.

The generalized linear model with the logit link function estimates a continuous and ordinal score that can be interpreted as a probability parameter of a Bernoulli trial. Here, we may regard this as a predicted disease progression score of AD. A correlation analysis was performed between the disease progression scores predicted with the models and the MMSE scores. According to the Pearson's correlation coefficients, our predicted disease progression score obtained from the model combining features of multiple tasks was most strongly correlated with the MMSE scores among all models ($r = -0.70$, Table 5).

Figure 4 shows the best predicted scores according to the labels of the groups. Figure 5 indicates the relationship between the best predicted scores and MMSE scores.

Table 5 – Pearson's correlation coefficients between predicted disease progression scores and MMSE scores.

Tasks	*r*	*p*-value
MMSE sentences	–0.38	$<1.0 \times 10^{-3}$
MMSE pentagons	–0.51	$<1.0 \times 10^{-5}$
TMT part A	–0.59	$<1.0 \times 10^{-7}$
TMT part B	–0.60	$<1.0 \times 10^{-7}$
CDT	–0.65	$<1.0 \times 10^{-9}$
All tasks	–0.70	$<1.0 \times 10^{-11}$

Discussion

As stated in the introduction, healthy monitoring technology focusing on behavioral changes has shown promise for early detection of AD. Among these technologies, handwriting behaviors measured by digital tablet devices are attracting attention as potentially useful indicators. The objectives of the present study were to investigate the effectiveness of combining behavioral features of multiple drawing tasks for detecting AD and MCI and the possibility of estimating disease progression by using such features. We collected and analyzed a series of behavioral features of five representative handwriting tasks used in MMSE, TMT, and CDT from 71 participants including some diagnosed with MCI and AD.

First, although the selected predictive features in our analysis were different depending on the tasks, we found tendencies that features relating to fine motor controls such as velocity, acceleration, and pressure were selected for the MMSE tasks and features relating to attention or cognitive performances such as durations and number of segments were selected for the TMT and the CDT. These results may reflect the characteristics of the different tasks and suggest that exploiting the combined features of multiple tasks may improve the performance of predictive models. We demonstrated that our three-class classification model combining features of multiple tasks improved the accuracy.

Next, we demonstrated that the classification model can also estimate in-clinic cognitive assessment scores via correlation analysis between the model estimated score that can be regarded as a predicted disease progression and the MMSE assessment score. It should be noted that the predicted disease progression scores were obtained just from the labels of disease groups and behavioral features without any progression measures. A more accurate validation of disease progression might require biomarkers such as amyloid beta in the brain and tau in cerebrospinal fluid (CSF), especially when we attempt to build a model for detecting patients at the preclinical AD stage. It would be difficult or impossible to use these biomarkers for training a model, considering the cost and/or invasiveness for assessing positron emission tomography (PET) and CSF. Our approach, which attempts to predict the level of disease progression by using just disease stage information instead of these biomarkers, should reduce the amount of required sample data (e.g., just validation data would be needed).

Some limitations exist in this study. First, since the number of participants was relatively small, generalizability of our results may be limited. Further study with a larger number of participants is our future work. Second, we evaluated the handwriting features using just disease labels and neuropsychological assessment scores. Further study is needed to investigate whether our model for estimating scores can predict known biomarkers for AD such as amyloid beta and tau deposition.

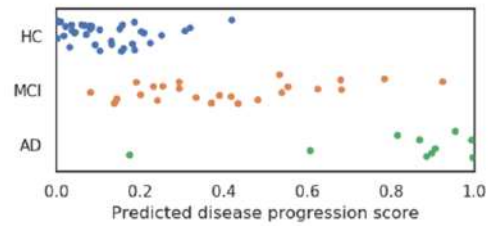


Figure 4 – Strip plot of predicted disease progression scores obtained from model combining multiple tasks.

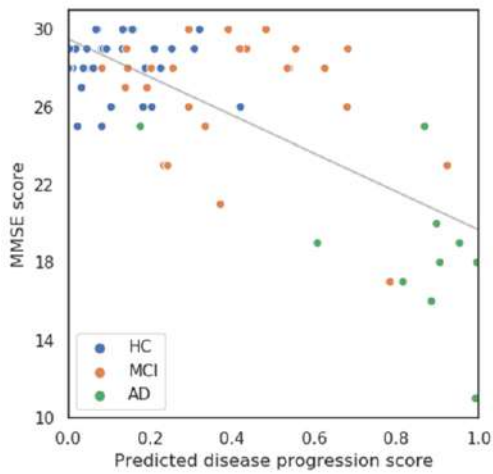


Figure 5 – Scatter plot with the line of best fit between the predicted disease progression scores obtained from the model combining multiple tasks and MMSE score.

References

[1] M. Prince, *et al*., *The global impact of dementia 2013–2050*. Alzheimer's Disease International, 2013.

[2] M. Prince, *et al*., *World Alzheimer report 2016: Improving healthcare for people living with dementia: Coverage, quality, and costs now and in the future*. Alzheimer's Disease International, 2016.

[3] V. Kotagal, *et al*., Factors associated with cognitive evaluations in the United States. *Neurology* **81** (2015), 64–71.

[4] A. Wimo *et al*., The world costs of dementia 2015 and comparisons with 2010. *Alzheimer's Dement*, **13** (2017), 1–7.

[5] T. Horr, *et al*., Systematic review of strengths and limitations of randomized controlled trials for non-pharmacological interventions in mild cognitive impairment: Focus on Alzheimer's disease. *The Journal of Nutrition, Health & Aging,* **19.2** (2015), 141–153.

[6] R.A. Sperling, *et al*. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia* **7.3** (2011), 280–292.

[7] A. Alberdi, A. Aztiria, and A. Basarab. On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey. *Artificial Intelligence in Medicine* **71** (2016), 1–29.

[8] M.D. Lezak, D.B. Howieson, and D.W. Loring. *Neuropsychological Assessment (4th ed.)*. Oxford University Press, 2004.

[9] C.R. Bowie, and P.D. Harvey, Administration and interpretation of the Trail Making Test, *Nature Protocols* **1** (2006), 2277–2281.
[10] J. Garre-Olmo *et al*., Kinematic and pressure features of handwriting and drawing: Preliminary results between patients with mild cognitive impairment, Alzheimer's disease, and healthy controls, *Current Alzheimer Research* **14** (2017), 960–968.
[11] P. Gulde, *et al*., Step by Step: Kinematics of the reciprocal trail making task predict slowness of activities of daily living performance in Alzheimer's disease. *Front Neurol* **9** (2018), 140.
[12] C. Kahindo *et al*., Characterizing early-stage Alzheimer's through spatiotemporal dynamics of handwriting. *IEEE Signal Processing Letters* **25** (2018), 1136–1140.
[13] K. Lopez-de-Ipina, *et al*., Automatic analysis of Archimedes' spiral for characterization of genetic essential tremor based on Shannon's entropy and fractal dimension. *Entropy* **20** (2018), 531–545.
[14] G. Pirlo *et al*., Early diagnosis of neurodegenerative diseases by handwriting signature analysis. *Proceedings of International Conference on Image Analysis and Processing 2015*, 290–297.
[15] R. Camicioli *et al*., Handwriting and pre-frailty in the Lausanne cohort 65+ (Lc65+) study. *Archives of Gerontology and Geriatrics* **61** (2015), 8–15.
[16] A. Bisio *et al*., The kinematics of handwriting movements as expression of cognitive and sensorimotor impairments in people with multiple sclerosis. *Scientific Report* **7** (2017), 17730.
[17] M. Mitolo, *et al*., The new qualitative scoring MMSE pentagon test (QSPT) as a valid screening tool between autopsy-confirmed dementia with Lewy bodies and Alzheimer's disease. *J Alzheimer Dis* **39** (2014), 823–832.
[18] J. Dehmen *et al*., An analysis of a digital variant of the trail making test using machine learning techniques. *Technol Health Care* **25** (2017), 251–264.
[19] R. Fellows, *et al*., Multicomponent analysis of a digital trail making test. *Clin Neuropsychol* **31** (2017), 154–167.
[20] J. Heinik, *et al*., Computerized kinematic analysis of the clock drawing task in elderly people with mild major depressive disorder: An exploratory study. *Int Psychogeriatr* **22** (2010), 479–488.
[21] Z. Harbi *et al*., Segmentation of clock drawing based on spatial and temporal features. *Proceedia Computer Science* **60** (2015), 1640–1648.
[22] W. Soullard-Mandar, *et al*., Learning classification models of cognitive conditions from subtle behaviors in the digital clock drawing test. *Mach Learn* **102** (2016), 393–441.
[23] M.F. Folstein, S.E. Folstein, and P.R. McHugh, "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12** (1975), 189–198.
[24] R.C. Petersen *et al*., Aging, memory, and mild cognitive impairment. *International psychogeriatrics* **9** (1997), 65–69.
[25] G. McKhann, *et al*., The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7** (2011), 263–269.

Address for correspondence

Tatsuya Ishikawa: ISIKAWA@jp.ibm.com
Yasunori Yamada: YSNR@jp.ibm.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190206

Development of a Machine Learning Model Predicting an ICU Admission for Patients with Elective Surgery and Its Prospective Validation in Clinical Practice

Stefanie Jauk[a,c], Diether Kramer[b], Günther Stark[b], Karl Hasiba[b], Werner Leodolter[b], Stefan Schulz[c], and Johann Kainz[b]

[a] *CBmed, Graz, Austria*
[b] *Steiermärkische Krankenanstaltengesellschaft m.b.H. (KAGes), Graz, Austria*
[c] *Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria*

Abstract

Frequent utilization of the Intensive Care Unit (ICU) is associated with higher costs and decreased availability for patients who urgently need it. Common risk assessment tool, like the ASA score, lack objectivity and do account only for some influencing parameters. The aim of our study was (1) to develop a reliable machine learning model predicting ICU admission risk after elective surgery, and (2) to implement it in a clinical workflow. We used electronic medical records from more than 61,000 patients for modelling. A random forest model outperformed other methods with an area under the curve of 0.91 in the retrospective test set. In the prospective implementation, the model achieved a sensitivity of 73.3% and a specificity of 80.8%. Further research is essential to determine physicians' attitudes to machine learning models and assess the long term improvement of ICU management.

Keywords:

Electronic health records; Intensive care units; Machine learning

Introduction

In the last decade there has been a distinct increase in life expectancy among the older population that is accompanied by a progression in morbidity rate. For this reason many patients need the specialised health care facility that is offered by an Intensive Care Unit (ICU). Intensive care beds in Austria comprise a higher proportion of total hospital beds in comparison to other European countries. As the availability of this resource varies throughout time, a constant analysis of appropriateness of ICU use is needed [1]. Moreover, frequent utilization of the ICU does not decrease hospital mortality, but is associated with higher costs and decreases the access for those patients who may profit more from the intensivists' expertize [2]. Therefore, patients likely to benefit from ICU admission and those who are at risk for severe postoperative illness have to be identified preventively.

The most common method used to stratify preoperative risk in anaesthesia is the "American Society of Anaesthesiologists – Physical Status (ASA) tool". This world-wide used instrument allows classifying and grading the physical condition of a patient before surgery. It allows estimating possible intraoperative and postoperative anaesthetic complications associated with the patient's preoperative condition. However, one of the disadvantages of the ASA tool is its subjectivity that may lead to only moderate inter-rater reliability when comparing ASA scores estimated by two physicians for the same patient [2].

In a recent published paper, Chan et al. combined the ASA score with electronic medical records (EMR) into the CARES model to predict the need for ICU admission postoperatively [3]. They extracted routinely collected EMR data of more than 90,000 patients and developed a multivariable regression model. The CARES model outperformed the ASA score in the prediction results with an area under the receiver-operating characteristic curve (AUROC) of 0.84 (0.81 - 0.87). However, their final model was based on seven variables only, including the ASA score and a surgical risk classification parameter.

Considering the increasing amount of available health care data, models based on machine learning (ML) are becoming an alternative way for risk prediction. Through ML methods, more variables can be evaluated for the purpose of generating newer, more precise models.

Recently, an ML model predicting major complications and death after surgery was published by Bihorac et al. [4]. One of the analysed complications was an ICU admission greater than 48 hours. The authors used a generalized additive method (GAM) for model development, as this method had outperformed other prior ML methods for anlayses (e.g. logistic regression and support vector machine) [5]. The AUROC for the prediction of ICU admission with GAM was 0.88 (0.87 - 0.89), with a sensitivity of 0.76 and a specificity of 0.83. Utilisation of this prediction model could be beneficial for managing ICU capacities in general, but in routine work it is more crucial to accurately estimate the risk of short term ICU patients.

Several ML-based risk prediction models have been published during the last few years, but only few of them made their way to clinical practice. In order to ensure highest performance of a prediction model and patient safety respectively, research in real clinical settings is needed. Prediction results might differ between validation on retrospectively collected data and prospective validation with real-time prediction.

KAGes (Steiermärkische Krankenanstaltengesellschaft) is the biggest health care provider in Styria, a federal state of Austria. Due to the coverage of more than 90 % of all hospital beds in Styria, KAGes has access to approximately two million longitudinal patient histories. This large amount of clinical data forms a highly valuable basis for modelling and the use of ML methods. KAGes runs several subsidiary hospitals, one of them being the hospital Hochsteiermark – Bruck an der Mur (abbreviated BRU). Hospital BRU has a total of 330 inpatients beds, including twelve beds in the ICU with mechanical ventilation, and eight beds in the intensive monitoring unit without mechanical ventilation. Prior to surgery, patients undergo at least one assessment by an

anaesthesiologist. In 2017, the anaesthesiologits conducted 4,606 preoperative consultations on 3,881 patients. Of these preoperative assessments, 4,370 (94.9 %) cases underwent a surgery, of which 462 (10.6 %) cases were admitted to ICU postoperatively.

During preoperative assessment, anaesthesiologists estimate the patients' risk of ICU admission after surgery, based on patient histories and ASA score classification. This risk estimation is crucial for preoperative anaesthetic management, and ICU bed management respectively. However, due to the subjectivity of the ASA score the accuracy of this assessment may vary between patients. Another problem concerning the accuracy of risk estimation is the large amount of information in some patients' medical histories and having to review all diagnostic findings that were collected in the past.

To overcome this challenge, objective assessment methods are required that support anaesthesiologists in their risk estimation. In clinical practice, such methods will only be helpful when many more variables are integrated than current common surgery risk tools offer. Most of the prediction models published in recent years have shown good results on validation of test data, but little of them have been validated during a real-time scenario. However, a model's usefulness can only be determined through a validation on real-time data. Therefore, our goal in the underlying study was to develop a reliable ML-based model predicting the risk of an ICU admission after an elective surgery and implement the model within a user-friendly interface in the routine clinical workflow.

Methods

The development and implementation of the study received approval from the Ethics Committee of the Medical University of Graz (30-146 ex 17/18). We used the TRIPOD statement [6] as a guideline for the development, validation and reporting. The data process and modelling design of the study are shown in Figure 1.

Data Set

The routine clinical data of KAGes was stored in the Hospital Information System (HIS) openMEDOCS, which was based on IS-H/i.s.h.med information systems, and implemented on SAP platforms. For data analysis, the data warehouse platform SAP HANA was used for controlling and quality issues. All analyses including pre-processing and modelling were performed using the R software environment.

The predicted outcome of our classification model was the admission to ICU within five days after an elective surgery. The time of prediction was defined as the last preoperative assessment of a patient before undergoing the elective surgery.

Cohort Building and Feature Set

For the first step, we extracted a retrospective cohort of hospitalised patients from the hospital BRU and a second KAGes-hospital, the hospital LEO. In this cohort, we included all patients that had undergoe a preoperative assessment in one of the two hospitals between 1st of January, 2010, and 31st of March, 2018. We excluded all patients for whom the elective surgery was not executed after the assessment, as well as patients younger than 18. The final cohort resulted in 61,864 patients.

For the second step, we extracted routinely stored EMR data for the whole cohort. These records included structured information, such as demographic data, ICD-10 diagnosis codes, current Austrian procedures codes, lab data, patient transfer data and nursing protocols. We considered information in lab data from the last 30 days and ICD-10 codes from the last three years. ICD-10 codes for chronic diseases like diabetes, chronic kidney disease, asthma etc. were used without any exclusion. In order to handle rather infrequent ICD-10 codes, we computed groups of related chapters, e.g. the feature ICD_E10_E14 that covered any diabetes code for the patient. This strategy also helped us to handle coding errors for ICD-10 codes which we had detected in our previous research [7]. In addition, we calculated several indices like the Charlson Comorbidity Index [8], length of last inpatient stay or total number of stays. The data extraction and pre-processing resulted in 630 features summarised in Table 1.

In previous studies we had examined the impact of data imputation for EMRs used for modelling [9]. As those studies had shown no improvement in model performance when using imputation methods, we decided against the imputation of values based on population measurements. Instead, we solely performed last-observation-forward imputation for lab data over the last 30 days and for nursing protocols from the last hospital stays.

Table 1 – Modelling Features (n = 630) Extracted from Electronic Medical Records for Prediction of an ICU Admission.

Feature group	Examples	n
Demographic Data	Age, sex, mother tongue, additional private insurance	30
Diagnosis Codes	ICD-10 codes, groups of ICD-10 codes	345
Procedure Codes	X-ray, MRI, physiotherapy, CT, endoprosthesis implantation	103
Laboratory Data	CRP, cholesterol, gamma-GT, haemoglobin,	46
Nursing Protocols	Respiratory problems, sleeping disorder, body mass index	96
Administrative Data, Indices	Transfers, hospital admissions, Charlson comorbidity index	10

Modelling and Calibration

In order to choose the best model for our prediction, we compared various ML methods. We computed all pre-processing and modelling steps in R, using the *caret* package [10] and associated packages for modelling. For a deep learning model we used the *mxnet* package [11]. We split the data set into 75 % training data (n = 46,460) and 25 % test data (n = 15,404). After splitting, we trained all models on the same training data with a five times repeated 10-fold cross validation and used AUROC as a metric to select the optimal model. To overcome the problem of imbalanced data, we applied up- and down-sampling during training. Finally, we calculated model performance parameters for each model (accuracy, sensitivity, specificity and precision) on the remaining test data using the "closest topleft" threshold of the ROC.

The compared models included linear discriminant analysis (LDA), logistic regression (GLM), feed-forward neural networks with single hidden layer (NN), deep neural networks (MXNet), stochastic gradient boosting (GBM) and random forest (RF).

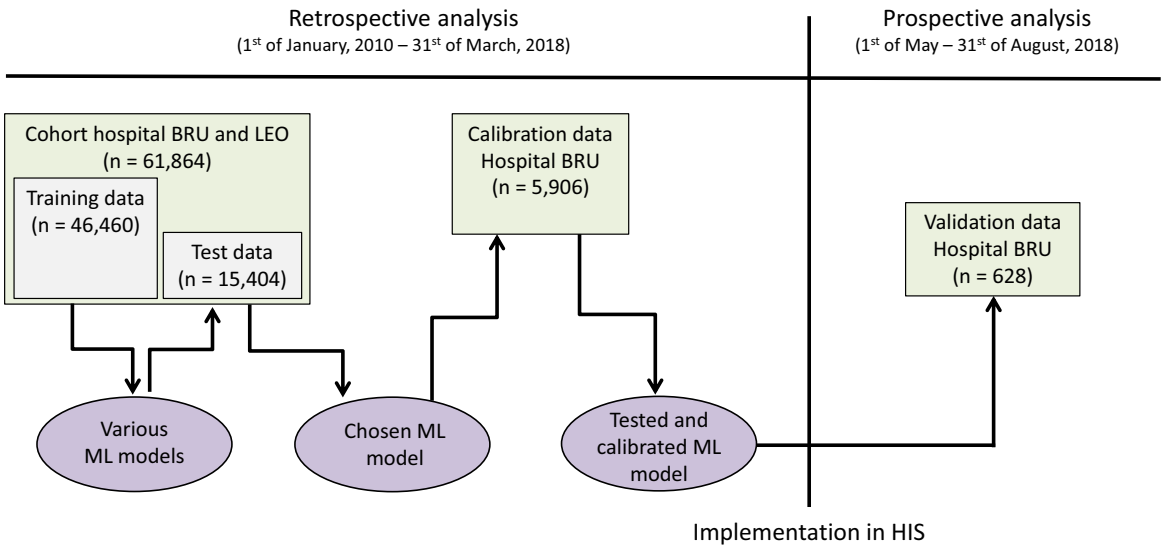


Figure 1 – Data process design for the development of a machine learning (ML) based prediction model for the risk of an ICU admission after surgery, and its prospective validation after an implementation in the HIS.

Based on the results of the area under the receiver operating characteristic curve (AUROC) and accuracy in the test data, we selected the best performing model for further calibration. We chose the thresholds with respect to the characteristics of the department. The aim of the calibration was to avoid over-alerting and to map the patients' risk probabilities to three categories: low risk, high risk and very high risk. In order to determine the best thresholds for these categories, we evaluated the number of available ICU beds and discussed the distributions with the hospital's senior physicians.

Prospective Validation in Clinical Practice

The best performing prediction model was implemented in the HIS during a pilot phase starting on 1st of May, 2018, and ending on 31st of August, 2018. For every patient who underwent a preoperative assessment by an anaesthesiologist, an individual risk of ICU admission was calculated in real-time.

During those four months, the anaesthesiologists were able to review the latest predicted risk for the patients. A traffic light symbol integrated in the HIS represented the patients' risk: high risk presented in yellow and very high risk in red. A click on the symbol revealed details on the prediction. The information integrated in a user interface was designed to support clinical reasoning and highlight the underlying health records that influenced the model's outcome. To increase usability, the presented features were ordered using the calculated variable importance of the model.

After the implementation phase, we analysed the performance of the model in real-time prediction and compared it to the performance in the test data. To calculate the sensitivity and specificity of the validation data, we combined the two categories high risk and very high risk, and defined them as a positive prediction of ICU admission, For every patient we considered the prediction result just before the preoperative assessment. If more than one preoperative assessment was available, only the latest prediction was evaluated.

In addition, the anaesthesiologists in our group examined all patients who were classified into a wrong category. Patients with an ICU stay and a prediction of low risk, as well as patients without ICU stay and a prediction of high risk were defined as incorrectly classified. We analysed their patient histories and further information of the hospital stay including surgical procedures, lab data etc.

Results

The demographic data of the three data sets are presented in Table 2. The percentage of ICU admissions was 7.7 % in the training set and 7.3 % in the test set. For ICU patients, the length of stay was comparable between the two test sets with a median stay of 50 hours and 49 hours, respectively. Distributions of sex and age were comparable between training and test data set, too.

Table 2 – Descriptive Statistics for the Training, Test and Validation Data. (Values of categorical variables are presented in absolute frequencies and column percentages).

	Training (n = 46,460)	**Test (n = 15,404)**	**Validation (n = 628)**
ICU	3,573 (7.7 %)	1,131 (7.3 %)	60 (9.6 %)
Non-ICU	42,887 (92.3 %)	14,273 (92.7 %)	568 (90.4 %)
Length of ICU stay in hours[a]	50.3 (91.9)	49.0 (73.6)	24.5 (24.6)
Male	22,733 (48.9 %)	7,533 (48.9 %)	343 (55.1 %)
Female	23,727 (51.1 %)	7,871 (51.1 %)	280 (44.9 %)
Age[b]	56.8 (18.3)	56.8 (18.4)	63.2 (15.1)

[a] reported in median (IQR) and for ICU patients only
[b] reported in mean (standard deviation)

Table 3 – Modelling Results for Different Machine Learning Methods Predicting an ICU Admission after Surgery (If not otherwise indicated, down-sampling was used to account for imbalanced data).

Model	R-package	Parameters	AUROC[a]	Accuracy	Sensitivity	Specificity	Prec.
GBM	gbm	n.trees=500, maxTreeDepth=9, shrinkage=0.075, minNodeSize=10	0.883 (0.875 - 0.892)	0.801	0.815	0.801	0.223
GLM	glm	type=classification	0.823 (0.811 - 0.835)	0.757	0.771	0.756	0.182
LDA	mass	tol = 1.0e-12	0.830 (0.818 - 0.842)	0.762	0.771	0.762	0.185
NN	nnet	size(hiddenUnits)=1, weightDecay=2	0.842 (0.831 - 0.853)	0.779	0.781	0.779	0.199
NN-up[b]	nnet	size(hiddenUnits)=1, weightDecay=2	0.846 (0.835 - 0.857)	0.770	0.801	0.768	0.195
MXNet	mxnet	nLayers=3, activation=softrelu	0.793 (0.781 - 0.805)	0.718	0.763	0.714	0.158
RF	randomForest	ntree=1000, mtry=√630	0.888 (0.879 - 0.897)	0.799	0.823	0.797	0.222
RF-up[b]	randomForest	ntree=500, mtry=√630	**0.911 (0.903 - 0.919)**	**0.828**	**0.833**	**0.827**	**0.253**

[a] Area under the receiver operating curve; values in parenthesis present the 95 % confidence intervals; [b] upsampling was used instead of downsampling. Note: Prec.: precision; mtry: number of randomly chosen predictors for each split.

Modelling and Calibration

The model performance of the different ML-based methods on the test data is summarized in Table 3. Comparing the methods, RF with up-sampling performed better than all other models with the highest AUROC (0.911), and accuracy (0.828).

The calibration of the RF model on the test data resulted in two thresholds dividing all predictions into three risk categories. The lowest 82 % of the predicted risk scores were classified to the low risk category, the following 11 % to high risk, and the upper 7 % of all predictions to the very high risk category.

Prospective Validation in Clinical Practice

During the four months of implementation in clinical workflow, 628 patients had an elective surgery with an antecedent preoperative assessment (see Table 4).

The thresholds set during calibration split the predictions accurately into the three risk categories with the category of high risk being slightly bigger for validation than for calibration (17 % compared to 11 %). The AUROC for the validation data was high with 0.843 (0.798 - 0.887), as well as the accuracy with 80.3 %. The sensitivity of the prediction in the validation data was 73.3 %, and the specificity 80.8 %.

We performed qualitative checks on 42 wrongly classified patients. The false negative group consisted of 16 ICU patients with a low risk prediction at the preoperative assessment. For nine patients out of the 16, little information of the patient histories was stored in the HIS. In some cases, there were no ICD-10 codes available at all.

Table 4 – Real-time Prediction of an ICU Admission for All Patients with Preoperative Aassessments

ICU admission	Predicted Risk Category			
	Low	High	Very high	Total
No	459 (80.8)	83 (14.6)	26 (4.6)	568
Yes	16 (26.7)	26 (43.3)	18 (30.0)	60
Total	475 (75.6)	109 (17.4)	44 (7.0)	628

Note: The numbers in parentheses show rounded row percentages.

The false positive group included 26 non-ICU patients with a very high risk prediction. All 26 patients underwent non-severe surgeries, e.g. percutaneous transluminal angioplasty (PTA), shunt procedures, cataract surgeries, or lipoma surgeries. Due to the performed procedures those patients were highly unlikely to be admitted to ICU, although 21 of them were classified as ASA 3, and three of them as ASA 4.

Discussion

In this study, we developed a ML-based classification model to predict an ICU admission after an elective surgery. We used the EMRs of more than 61,000 patients for the modelling process and validated the model on more than 600 patients prospectively during a real-time prediction. In retrospective test data, our model achieved a higher performance than prior published comparable models. Our best performing model, a random forest method, resulted in an AUROC of 0.91, while the previously published multivariable regression model CARES reported an AUROC of 0.84 only [3]. Our literature review revealed only one published ML model predicting a postoperative ICU stay [4]. This model predicted an ICU admission with a stay greater than 48 hours and resulted in an AUROC of 0.88. However, these results are only to some extent comparable to our model, as we focused on any ICU admission independent of the length of stay.

In addition, we offered new insights into the implementation of ML-based prediction models and their performance in prospective settings. During a period of four months, we predicted the risk of ICU admission in real-time just before the preoperative assessment for all patients of one KAGes hospital. The risk for each patient was presented to the anaesthesiologists within the HIS. As expected, in the prospective evaluation the performance of the model decreased compared to the performance in the test data. Nevertheless, the result was satisfying with an AUROC of 0.84 and an accuracy of 80.3 %.

With the goal to improve the prediction model and to find possible deficits in the development, we checked all wrongly classified patients. Interestingly, all patients in the false positive group had surgeries with quite simple procedures, e.g. cataract or lipoma surgery. To overcome this bias in future model development, two additional features should be added: (1) the information of the severity of an elective surgery assessed by an anaesthesiologist, and (2) the type of anaesthesia. This information might improve the prediction for certain groups of patients. One group that might benefit from such an extension is are patients that undergo a severe elective surgery only due to a vital indication. The model assumes that these patients undergo non-severe surgeries only, and therefore have no need for an ICU bed. However, in the

validation data we could not distinguish well between elective surgeries and emergency ones due to missing real-time information, and, thus, might have underestimated the risk for emergency patients with highly severe surgeries. In future work we will need to address this limitation in order to provide the best prediction.

Some other limitations of our study should be considered when interpreting the results. First, all analyses are based on data from a HIS. Although most of the patient histories in the HIS of KAGes are quite comprehensive, the availability of structured information caused some problems in our study. For some of the wrongly classified patients, no coded diagnoses were found in their history, although free-text diagnoses were available in the discharge letters. Future data extraction might include natural language processing methods in order to make use of unstructured information.

Second, we did not consider the length of an ICU stay in our risk prediction, but rather focused on any admission to ICU after surgery. Bihorac et al. [4] had predicted ICU admissions with a stay greater than 48 hours. This prediction is not enough to ensure a precise ICU bed management, as almost 50 % of our ICU patients have a stay with less than 48 hours. It might be useful to combine various prediction models for ICU admission: First, predict any admission to ICU for all patients, and then, predict the length of ICU stay for patients with a previous positive result.

In this study, we focused on the quantitative results of the prediction model. Future work on the model should include an evaluation of the users' experience, especially attitudes and fears towards the use of ML models for prognosis in health care as well as the satisfaction with the results in real-time prediction. In addition, the usability of the implementation in the HIS could be improved through the users' feedback.

Conclusions

We developed a machine learning model that predicts the risk of ICU admission after elective surgery. There is a high need for precise ICU bed management in order to avoid unnecessary costs, improve healthcare for patients, and to support decision makers. We implemented our ML-based prediction model in a clinical workflow to support anaesthesiologists in their clinical reasoning, and prospectively evaluated the model's performance. Further research is essential to determine physicians' attitudes to machine learning based on risk prediction, and to assess the long-term improvement of ICU management and patient outcomes when using ML-based prediction models.

Acknowledgements

This work has been carried out with the K1 COMET Competence Centre CBmed, which is funded by the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT); the Federal Ministry of Science, Research and Economy (BMWFW); Land Steiermark (Department 12, Business and Innovation); the Styrian Business Promotion Agency (SFG); and the Vienna Business Agency. The COMET program is executed by the FFG. KAGes and SAP provided significant resources, manpower and data as basis for research and innovation.

References

[1] C. Wild, and M. Narath, Evaluating and planning ICUs: Methods and approaches to differentiate between need and demand., *Health Policy Amst. Neth.* **71** (2005) 289–301. doi:10.1016/j.healthpol.2003.12.020.

[2] I. Kose, C. Zİncircioglu, M. Çakmak, G. Cabbaroglu, N. Senoglu, and M. Gonullu, Postoperative patients in the intensive care unit: Identifying those who do not really need it, *J Crit Care* **30** (2015) 1295–1298. doi:10.1016/j.jcrc.2015.08.012.

[3] D.X.H. Chan, Y.E. Sim, Y.H. Chan, R. Poopalalingam, and H.R. Abdullah, Development of the Combined Assessment of Risk Encountered in Surgery (CARES) surgical risk calculator for prediction of postsurgical mortality and need for intensive care unit admission risk: A single-center retrospective study, *BMJ Open* **8** (2018) e019427. doi:10.1136/bmjopen-2017-019427.

[4] A. Bihorac, T. Ozrazgat-Baslanti, A. Ebadi, A. Motaei, M. Madkour, P.M. Pardalos, G. Lipori, W.R. Hogan, P.A. Efron, F. Moore, L.L. Moldawer, D.Z. Wang, C.E. Hobson, P. Rashidi, X. Li, and P. Momcilovic, MySurgeryRisk: Development and Validation of a Machine-learning Risk Algorithm for Major Complications and Death After Surgery, *Ann Surg* (2018) 1. doi:10.1097/SLA.0000000000002706.

[5] P. Thottakkara, T. Ozrazgat-Baslanti, B.B. Hupf, P. Rashidi, P. Pardalos, P. Momcilovic, and A. Bihorac, Application of machine learning techniques to high-dimensional clinical data to forecast postoperative complications, *PLOS ONE* **11** (2016) e0155705. doi:10.1371/journal.pone.0155705.

[6] G.S. Collins, J.B. Reitsma, D.G. Altman, and K. Moons, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement, *BMC Med.* **13** (2015) 1. doi:10.1186/s12916-014-0241-z.

[7] S. Jauk, D. Kramer, S. Schulz, and W. Leodolter, Evaluating the impact of incorrect diabetes coding on the performance of multivariable prediction models., *Stud. Health Technol Inform* **251** (2018) 249–252.

[8] M.E. Charlson, P. Pompei, K.L. Ales, and C.R. MacKenzie, A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation, *J. Chronic Dis.* **40** (1987) 373–383.

[9] S. Jauk, D. Kramer, and W. Leodolter, Cleansing and imputation of body mass index fata and its impact on a machine learning based prediction model, *Stud Health Technol Inform* (2018) 116–123. doi:10.3233/978-1-61499-858-7-116.

[10] M. Kuhn, caret: Classification and Regression Training. R package version 6.0-78., 2017.

[11] T. Chen, Q. Kou, and T. He, MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems, 2017.

Address for correspondence

Diether Kramer, KAGes – Steiermärkische Krankenanstaltengesellschaft m.b.H., Billrothgasse 18a , 8010 Graz, Austria;
E-mail: diether.kramer@kages.at.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190207

Mapping Korean EDI Medical Procedure Code to SNOMED CT

Eun Jung Hwang[a], Hyeoun-Ae Park[b], Seung Kook Sohn[a], Hong Bock Lee[a], Hee Kyoung Choi[a], Sangmi Ha[a], Hak Jun Kim[c], Tae Wan Kim[d], Wook Youm[a]

[a] Health Insurance Review and Assessment Service, Wonju, Korea,
[b] College of Nursing, Seoul National University, Seoul, Korea,
[c] Korea University Guro Hospital, Seoul, Korea,
[d] Seoul National University Boramae Medical Center, Seoul, Korea

Abstract

The Electronic Data Interchange (EDI) medical procedure code is the code used for health insurance claims in Korea. We mapped Korean EDI codes to SNOMED CT to explore the global interoperability of health insurance claims data. We developed rules for mapping based on the mapping guideline provided by SNOMED CT International. The first and second authors mapped 726 EDI codes used to claim reimbursement in five specialty areas to SNOMED CT. Eight subject matter experts reviewed the mapping results. Out of 726 procedure codes, 82.5% were exactly or partially mapped to SNOMED CT. An EDI code was mapped to an average of 2.04 SNOMED CT concepts. Twenty-one attributes were identified in the EDI codes mapped to SNOMED CT concepts. We identified strategies to improve the EDI code in this study. They include introducing hierarchical structures, adding inclusion and exclusion criteria for procedure codes, and improving EDI code labels.

Keywords:

Systematized Nomenclature of Medicine, Health insurance reimbursement, Health information interoperability

Introduction

Korea introduced universal health care coverage in 1989. As of 2016, 97.1% of the population is covered by the National Health Insurance (NHI) and 2.9% by Medical Aid [1]. The NHI is managed comprehensively in a form of social insurance, and financed by contributions from the insured and their employers (84.3%), government subsidies (9.2%), and tobacco surcharges and others (6.5%) [2]. A single payer-system was introduced in 2000 and all healthcare providers in the country are required to enroll in the National Health Insurance System (NHIS) to get reimbursed.

The NHIS serves as the insurer, and the Health Insurance Review and Assessment Service (HIRA) as the reviewer of the claims and quality of healthcare services [3]. The reimbursement process starts with the healthcare institutions filing claims for medical fees to HIRA. Most claims (99.7%) are submitted electronically using the Medical Claim Portal Service (MCPoS) or Electronic Data Interchange (EDI) system. HIRA reviews and assesses the claims; it notifies the NHIS of the result for the medical cost payment, and notifies healthcare institutions for verification. Fee-for-service is the main payment system in Korea, covering 99.7% of outpatient care and 80.7% of inpatient care [4].

There are two categories of medical services in the NHIS: 'covered' services in which patients are responsible for only part of the cost and 'uncovered' services in which patients are responsible for the entire cost. Lists of covered and uncovered services are determined by the Ministry of Health and Welfare and maintained by the HIRA. The medical fee schedule is determined by multiplying the resource-based relative value scale (RBRVS) of each medical service by the unit price of the service [3].

As of August 2017, there are 9,252 items of medical services including treatments and surgical procedures, tests, long-term care, medical images, basic treatments, dental services, hospice care, emergency services, anesthesia, traditional medicine, meals, and pharmacy [3]. Out of 9,252 items, 8,759 are covered by the National Health Insurance. The national standard code, also known as the 'EDI code,' was assigned to each medical service. The EDI code is the basic structure for reimbursement claims of medical services in the fee-for-service system, and is used as a major component of patient classification for the DRG-based payment system covering seven groups of diseases or operations.

As use of ICT in healthcare increases, opportunities of sharing and exchanging health information between different healthcare institutions and countries have increased. This led to interest in standard healthcare terminologies and classifications. The Korean Classification of Diseases (KCD), which was modified from the ICD, has been used in Korea for the classification of medical diagnoses since 1952. However, there is no standard classification of medical procedures used in Korea, just as there is no global standard for medical procedures as widely used as the ICD. Even though WHO has been developing the International Classification for Health Interventions (ICHI) for reporting and analysing health interventions for statistical purposes since 2007, there are concerns around some aspects of its operationalization. [5,6].

Even though the EDI medical procedure code has been used for more than 20 years in Korea to file claims for reimbursement, it has a few limitations to be qualified as a classification. For example, the code has no consistent hierarchical structure, even though it contains some medical procedures organized by anatomy. The levels of granularity of the EDI procedure code vary by domain, and sometimes even within the same domain. Thus, it is not easy to use the current EDI code as a means to exchange and share information among the stakeholders. There have been efforts to improve the EDI code as a standardized medical procedure classification interoperable with the global classification by benchmarking ICHI, and ICD-10-PCS in

Korea [7,8]. However, the results of these efforts have not been fully implemented in the EDI procedure code.

SNOMED CT is one of the most widely used standard reference terminologies for information exchange across EHR systems globally [9]. Clinical databases differ in both purpose and design. EHRs are aimed at supporting clinical practice at the point of care, while claims data are built for the insurance reimbursement processes. SNOMED CT as a standard terminology is used to describe clinical conditions and medical procedures in EHRs. SNOMED CT has one of the most extensive coverages among biomedical vocabularies in the world with 340,659 active concepts contained in the July 2018 release [10]. SNOMED CT concepts are defined with the rules called the SNOMED CT concept model with permitted attributes and values that may be applied to each concept. SNOMED CT is used in more than 50 countries and managed by SNOMED International, a non-profit international organization with 35 member countries [11]. Mapping the procedure codes such as CPT and ICD-10-PCS toward SNOMED CT is a useful and rather common exercise [12,13].

In this study, we mapped the EDI medical procedure code to SNOMED CT concepts in order to explore ways to improve the global interoperability of Korean health insurance claims data.

Methods

We developed methods and rules for mapping based on the mapping guideline provided by SNOMED International [14].

Mapping Materials

The source of mapping is the Korean EDI procedure code updated in 2017, and the target of mapping is the SNOMED CT concept within the 'Procedure' top-level hierarchy released by SNOMED International in July 2017. A total of 726 medical procedure codes used to claim reimbursement for treatment and surgery in five specialty areas -- Internal Medicine, Colorectal Surgery, Ophthalmology, Orthopedics and Pediatrics -- were mapped. These specialties were selected due to their broad content coverage and the active participation of the subject matter experts in the use of clinical terminology.

Mapping Procedure

Using the SNOMED International SNOMED CT Browser, the first and second authors searched and identified the SNOMED CT concept that has a synonymous relationship with an EDI medical procedure within the 'Procedure' top-level hierarchy [15]. The identified SNOMED CT concept was used as a lead term for the browse-up and down approaches. In the browse-up approach, parent concept(s) of the identified concept were examined. This helped us to examine if any other children concepts of the parent concept were also relevant and should be included.

For instance, 'Anal sphincteroplasty for obstetrical laceration' was first identified for the EDI medical procedure code 'Sphincteroplasty without levatorplasty'. Browsing up from the SNOMED CT concept 'Anal sphincteroplasty for obstetrical laceration' yields a parent 'Repair of anal sphincter'. Examination of the siblings of 'Anal sphincteroplasty for obstetrical laceration' under 'Repair of anal sphincter' reveals some siblings can be included as a relevant concept of 'Sphincteroplasty limited to anal sphincter', such as 'Post-anal repair', while other siblings such as 'Construction of anal neosphincter' cannot be included.

In the browse-down approach, descendant concepts of the identified concept were examined to see if any of the descendant concepts should be excluded. For instance, for the EDI medical procedure code 'Primary repair of mesentery', the child concepts of 'Repair of mesentery' such as 'Mesentery closure', 'Repair of mesentery of colon' or 'Repair of mesentery of small intestine' can be included. However, 'Mesentery reconstruction' and 'Plication of mesentery' cannot be included.

The browse-up and down sequence of the search was repeated as needed to look for other relevant SNOMED CT concepts. Then, included concepts were numbered from 1 to n. The Boolean logic of the map was provided if necessary. We mapped to concepts with the same meaning, considering use cases. Definition of the EDI medical procedure code, as well as the formal definition, fully specified name and synonyms of the SNOMED CT concept were examined during each mapping.

During the mapping, we identified the attribute relationships of the source concept and the values of the attributes. The attribute relationships will contribute to the definition of the source concept by associating them with the values. We also identified strategies to improve the structure and label of the EDI medical procedure code to improve the global interoperability of the claims data submitted using the EDI code.

Classification of Map

Each map was classified as either an 'exact' or 'partial' map according to the level of correspondence between the EDI medical procedure code and SNOMED CT concept(s). Since the EDI code is used for reimbursement purpose, a single EDI code can have multiple anatomical sites or multiple procedures. An EDI code with more than one site or more than one procedure was dissected to express different meaning of the code and then mapped to multiple SNOMED CT concepts.

For a EDI medical procedure with a single anatomical site and a single procedure, if the identified SNOMED CT concept was equivalent with the EDI medical procedure code, it was classified as an exact map. If the identified SNOMED CT concept was less specific than the EDI code, it was classified as a broad map. For a broad map, post-coordination of SNOMED CT concept can be introduced to represent the EDI code. Otherwise, it was classified as a narrow map.

For a EDI medical procedure with more than one anatomical site or more than one procedure, all of the identified SNOMED CT concepts were equivalent with the dissected EDI medical procedure code, it was classified as an exact map. Otherwise it was classified as a partial map. Partial maps were further classified as a broad, a narrow or a mixed map. If all of the identified SNOMED CT concepts was less specific than the dissected EDI medical procedure code, it was classified as a broad map. For a broad map, post-coordination of SNOMED CT concepts can be introduced to represent the EDI code. If all of the identified SNOMED CT concepts was more specific than the dissected EDI medical procedure code, it was classified as a narrow map. If the identified SNOMED CT concepts were less specific, more specific or equivalent with the dissected EDI medical procedure code, it was classified as a mixed map.

Table 1– Type of Map

Type		Description
Exact		The identified SNOMED CT concept(s) is equivalent with the EDI medical procedure code.
Partial	Broad	The identified SNOMED CT concept(s) is less specific than the EDI medical procedure code.
	Narrow	The identified SNOMED CT concept(s) is more specific than the EDI medical procedure code.
	Mixed	The identified SNOMED CT concepts are either less specific, more specific or equivalent with the dissected EDI medical procedure code.
No map		There is no appropriate SNOMED CT concept to represent the EDI medical procedure code.

For each map, the cardinality of the map between the EDI medical procedure code and SNOMED CT concept(s) was counted. If all descendant concepts of an identified SNOMED CT concept were included as relavant concepts of an EDI medical procedure code, only the parent concept was counted.

Validation of Mapping

Eight subject matter experts from five specialty areas reviewed mapping results to see if SNOMED CT concepts validly represent EDI medical procedure codes. They received five hours of training and education on SNOMED CT before the review. We posed questions to the reviewers based on SNOMED CT concepts and descendants. An example question is 'Should 'Excision of diverticulum from large intestine' or 'Excision of diverticulum of small intestine' be included in 'Diverticulectomy' according to the definition of the EDI code?' If the experts do not agree with the mapping, a consensus building process was introduced. The reviewers were also asked to explore ways to improve global interoperability of the claims data collected using the EDI procedure code during the validation.

Results

Mapping examples of EDI medical procedure codes with SNOMED CT concepts are presented in Table 2.

The mapping rates of 726 EDI medical procedure codes to SNOMED CT concepts are presented in Table 3. Overall, 82.5% of the EDI codes were exactly or partially mapped, ranging from 76.9% to 100% by specialty. The mapping rate of the EDI medical procedure codes in Colorectal surgery was higher than those of other specialties in exact maps (73.8%).

We did not attempt to map 79 EDI codes to SNOMED CT since they are not the codes describing medical procedures. Such examples are EDI codes used to claim reimbursement for objects (dialyser, tubing set, guide wire, IV set, syringe, needle, and catheter) and substances (heparin and normal saline) for extracorporeal dialysis of ascitic fluid.

Out of 726 EDI procedure codes, 48 codes were not mapped since there was no SNOMED CT concept describing these medical procedure codes. For example, the EDI code 'Ultra-Low anterior resection with Intersphincteric Resection' was not mapped. Another example is 'Transaortic approach for transcatheter aortic valve implantation'. SNOMED CT does not have any concept describing this procedure even though there were two SNOMED CT concepts describing transcatheter aortic valve implantation - 'Transapical approach for transcatheter aortic valve implantation' and 'Transfemoral approach for transcatheter aortic value implantation'.

Table 2– Examples of EDI medical procedure code mapped to SNOMED CT concepts with boolean logic

EDI Code	SNOMED CT concepts	#of SNOMED CT concepts	Boolean logic	Type of map
Radio-frequency Ablation of Supra-ventricular Arrhythmia through 3D Mapping	1. Radiofrequency ablation operation for arrhythmia (procedure) 2. Percutaneous transluminal three dimensional electroanatomic mapping of conducting system of heart (procedure)	2	1 AND 2	Broad
Diverti-culectomy	1. Excision of diverticulum from large intestine (procedure) 1.1 Excision of diverticula of colon (procedure) 2. Excision of diverticulum of small intestine (procedure) 2.1 Diverticulectomy of duodenum (procedure) 2.1.1 Excision of diverticulum of ampulla of Vater (procedure) 2.2 Excision of Meckel's diverticulum (procedure)	2	1 OR 2	Exact
Primary Repair of Mesentery	1. Mesentery closure (procedure) 1.1 Suture of mesentery (procedure) 2. Repair of mesentery of colon (procedure) 3. Repair of mesentery of small intestine (procedure)	3	1 OR 2 OR 3	Exact

Table 3– Result of mapping by specialty

Type of map	Internal Medicine	Colo-rectal Surgery	Ophthal-mology	Ortho-pedics	Pedia-trics	Total
Exact	72 (50.3%)	76 (73.8%)	61 (50.4%)	91 (25.9%)	4 (50.0%)	304 (41.9%)
Broad	35 (24.5%)	19 (18.4%)	14 (11.6%)	39 (11.1%)	4 (50.0%)	111 (15.3%)
Narrow	4 (2.8%)	-	25 (20.7%)	140 (39.9%)	-	169 (23.3%)
Mixed	7 (4.9%)	4 (3.9%)	4 (3.3%)	-	-	15 (2.1%)
Not mapped	19 (13.3%)	4 (3.9%)	17 (14.0%)	8 (2.3%)	-	48 (6.6%)
Did not map	6 (4.2%)	-	-	73 (20.8%)	-	79 (10.9%)
Total	143 (100%)	103 (100%)	121 (100%)	351 (100%)	8 (100%)	726 (100%)

In total, 726 EDI codes mapped to 1,481 SNOMED CT concepts with an average of 2.04 SNOMED CT concepts per EDI code. The cardinality of the maps is provided in Table 4. There were 352 EDI codes mapped to one SNOMED CT concept, and 247 EDI codes mapped to more than one SNOMED CT concept with a maximum of 19 concepts. Some of EDI codes do not have a separate code describing surgical procedure with laparoscopic approach such as 'Laparoscopic appendectomy'. Healthcare providers in Korea claim laparoscopic appendectomy using the EDI code for 'Appendectomy with open approach'. Thus, we mapped an EDI surgical procedure which can be performed with an open or laparoscopic approach to two SNOMED CT concepts, one for the open approach and the other for the laparoscopic approach.

Table 4– Cardinality of maps by specialty

Specialty / Cardinality	Internal Medicine	Colorectal Surgery	Ophthal-mology	Ortho-pedics	Pedia-trics	Total
0	25	4	17	81	0	127 (17.5%)
1	70	28	79	167	8	352 (48.5%)
2	23	12	12	62	0	109 (15.0%)
3	5	24	6	19	0	54 (7.4%)
4	9	9	2	14	0	34 (4.7%)
5	4	9	1	5	0	19 (2.6%)
6	0	3	0	1	0	4 (0.6%)
7	0	3	0	1	0	4 (0.6%)
8	7	1	1	1	0	10 (1.4%)
9	0	1	0	0	0	1 (0.1%)
≥10	0	9	3	0	0	12 (1.7%)
Total	143	103	121	351	8	726 (100%)

In total, 21 attribute relationships were identified in Table 5. They are Procedure site, Procedure site - Direct, Procedure site - Indirect, Procedure morphology, Direct morphology, Indirect morphology, Method, Procedure device, Direct device, Indirect device, Using device, Using access device, Access, Surgical approach, Direct substance, Using substance, Priority, Has focus, Has intent, Revision status, Route of administration, and Using energy. For example, the EDI code 'Fusion of joint of lumbar spine with internal fixation by anterior approach' has a 'procedure site - direct' relationship to the concept 'lumbar spine joint structure', 'method' relationship to the concept 'fusion - action', 'surgical approach' relationship to the concept 'anterior approach', and 'using device' relationship to the concept 'Orthopedic internal fixation system, device'. Attribute relationships of the source code identified in this study can be used in the data retrival process where the meaning of a concept is needed to determine whether the EHR record matches the query criteria and advanced analytics.

Table 5– A list of attributes identified in SNOMED CT concepts mapped to EDI codes

Specialty / Cardinality	Internal Medicine	Colorectal Surgery	Ophthal-mology	Ortho-pedics	Pedia-trics
Procedure site	10/3	39/10	20/12	17/8	-
Procedure site - Direct	150/27	515/58	170/40	444/170	4/3
Procedure site - Indirect	121/22	37/22	47/20	92/56	1/1
Procedure morphology	-	6/2	4/1	-	-
Direct morphology	64/12	117/26	66/12	149/26	2/2
Indirect morphology	3/2	14/1	1/1	4/2	-
Method	315/36	590/42	238/51	565/47	5/4
Procedure device	2/2	-	-	-	-
Direct device	110/22	5/4	11/7	81/32	1/1
Using device	70/21	32/5	21/7	66/9	1/1
Using access device	61/14	35/3	-	11/2	-
Access	33/1	5/3	1/1	21/2	2/1
Surgical approach	6/3	63/7	10/5	4/2	-
Direct substance	3/2	19/4	6/5	28/6	-
Using substance	11/3	5/3	2/2	-	-
Priority	-	7/4	-	-	2/1
Has focus	22/2	7/4	34/4	7/2	2/1
Has intent	23/1	-	-	-	3/1
Revision status	2/1	7/2	2/1	53/2	-
Route of administration	-	-	2/1	-	-
Using energy	-	-	1/1	-	-

Table 5 shows the number of attribute relationships identified in the SNOMED CT concepts mapped to EDI codes with the number of unique values for each attribute in each specialty. 'Method' was the most popular attribute, followed by the 'Procedure site - Direct', 'Direct morphology' and 'Procedure site – Indirect' attributes.

Discussion

We attempted to map 726 medical procedure codes in five specialty areas by identifying relevant SNOMED CT concepts using the browser-up and down search sequence. With this approach, we were able to identify all relevant SNOMED CT concepts for eligible medical procedures to claim with a specific EDI code. Thus, it was possible to achieve the similar mapping results to the map of SNOMED CT concepts to EDI medical procedure codes. Overall, 82.5% of the EDI codes were exactly or partially mapped to the SNOMED CT concepts, which is a slightly lower than that of ICD-9-CM Procedure Codes to SNOMED CT map [12].

Out of 726 EDI medical procedure codes, 79 codes used to claim reimbursement for medical procedures with additional comorbidity or accompanying surgeries for higher payment, and materials or drugs used in medical procedures were not mapped. There were 48 medical procedures for which we could not identify appropriate SNOMED CT concepts. They fall broadly into two categories; one category includes EDI medical procedure codes with mismatching body sites or approaches. An example of this is 'Intravenous catheterization for hemodialysis in the subclavian vein and internal jugular vein with cut-down method.' A cut-down approach may be used when inserting an implantable port into the cephalic vein or external jugular vein. The second category includes medical procedures not modelled in SNOMED CT yet. An example of this is 'Repair of nonunion or malunion in fingers or toes'. This procedure can be modelled in SNOMED CT with two concepts: 'Repair of nonunion of phalanges' and 'Repiar of malunion of phalanges'.

During the mapping process, we identified strategies to improve the interoperability of the claims data collected using the EDI medical procedure code. First, we would like to suggest introducing a hierarchical structure to the EDI medical procedure code. We mapped the EDI medical procedure 'Hemorrhoidectomy' to the SNOMED CT concepts 'Closed hemorrhoidectomy', 'Open hemorrhoidectomy' and 'Submucosal hemorrhoidectomy'. However, the EDI code 'Hemorrhoidectomy' is used for a claim for the simultaneous removal of three primary hemorrhoids. There is no EDI code to claim the removal of less than three hemorrhoids. Physicians use the EDI code of 'Excision of thrombosed hemorrhoids' for this claim. This issue can be resolved by introducing a hierarchical structure to distinguish between different types of hemorrhoidectomy. EDI codes used to claim for materials, medication or medical procedures with additional qualifiers such as acuity of patient, co-morbidity or co-treatment can be better organized with the introduction of a hierarchical structure to the EDI code. A list of attributes identified can be used to move the current enumerated EDI medical procedure code to a structured classification.

Second, we would like to suggest revising labels of EDI medical procedure codes to reflect the meaning of the procedure as accurately as possible. 'Enucleation' -- a general surgical technique that removes a mass without cutting into or dissecting it -- was used to describe the removal of an eyeball. 'Enucleation of eyeball' could be a better term to describe this procedure.

Third, we would like to suggest removing EDI medical procedure codes being used rarely for claims. 'Extracoporeal Ascites Dialysis' is being used seven times since 2013 in Korea. This procedure can be removed from the EDI medical procedure code list.

Fourth, we would like to suggest adding examples or synonyms to each EDI medical procedure code as an additional guide to help healthcare providers. For instance, 'Laparoscopic appendectomy' can be added as an example of the EDI code 'Appendectomy'.

We recognize the following limitations in our study. We only focused on EDI medical codes used to claim treatments and surgical procedures in five specialty areas. The EDI code includes other types of medical procedures such as laboratory tests, and imaging studies. Thus, the results may not be generalizable to the medical procedures in other specialties and other type of medical procedures.

Conclusions

The findings of this study will be used to improve the Korean EDI medical procedure code so that global interoperability of health insurance claims data can be improved. The EDI procedure code can be improved by introducing a hierarchical structure, standardizing EDI procedure code labels, and providing better coding reproducibility by clarifying inclusion and exclusion criteria for each procedure code. The global interoperability of health insurance claims data will be improved with the mapping between the EDI medical procedure code and SNOMED CT.

We will extend the mapping project to the EDI codes used in other specialty areas so that claims data can be used for national and international comparison and healthcare research. We hope that this work will trigger the use of SNOMED CT in the EHR system in Korea just as the EDI system had triggered the introduction of health information systems throughout Korea in the early 1990s. If healthcare providers use SNOMED CT when they collect clinical information at different stages of patient care, automated generation of reimbursement claims will be possible with data documented in the EHR system. This will make it possible to monitor the utilization of resources.

In the future, the findings of the mapping between the EDI procedure code and SNOMED CT concepts can be used for linking the EDI procedure code to the ICHI developed by WHO so that we can produce high quality, valid, comparable and sharable health insurance claims data meeting global standards.

Acknowledgements

The research being reported in this paper was supported by the Health Insurance Review and Assessment Service in Korea. The authors of this paper serve as staff or advisors for the Health Insurance Review and Assessment Service.

References

[1] National Insurance Review and Assessment Service and National Health Insurance Service, *National Health Insurance Statistical Yearbook*, National Insurance Review and Assessment Service and National Health Insurance Service, Wonju, 2017.

[2] National Health Insurance System, *National health Insurance and long term care insurance system in Republic of Korea,* National Health Insurance System, Wonju, 2018. http://www.nhis.or.kr/static/html/wbd/g/a/wbdga0704.html [accessed on October 30, 2018]

[3] Health Insurance Review and Assessment Service, *2018 Health Insurance Review and Assessment Service's functions and Roles – Part 2 Covered services and payment system,* Health Insurance Review and Assessment Service, Wonju, 2018.

[4] Health Insurance Review and Assessment Service, *Claim Statistic 2017,* Health Insurance Review and Assessment Service, Wonju, 2017.

[5] University of Udine, *International Classification of Health Interventions: Introduction*. http://mitel.dimi.uniud.it/ichi/docs/#introduction [accessed on October 30, 2018]

[6] M. Cerniauskaite, R. Quintas, C. Boldt, A. Raggi, A. Cieza, J.E. Bickenbach, and M. Leonardi, Systematic literature review on ICF from 2001 to 2009: its use, implementation and operationalisation, *J Disability rehabilitation* **33** (2011), 281-309.

[7] Health Insurance Review and Assessment Service and Korean Medical Association, *Development of Korean Medical Procedures Classification and resource-based relative value scale for Interoperability with International Classification (Medicine),* 2013.

[8] Health Insurance Review and Assessment Service and Seoul National University Hospital, *Research on improvement of Korean Standardized Medical Procedure Classification System,* 2016.

[9] European Union, Assess CT: *Assessing SNOMED CT for large scale eHealth deployment in the EU, European Union*, December 2016. http://assess-ct.eu/fileadmin/assess_ct/final_brochure/assessct_final_brochure.pdf. [accessed on October 30, 2018]

[10] SNOMED International, *SNOMED CT Worldwide*, SNOMED International, https://www.snomed.org/snomed-ct/snomed-ct-worldwide. [accessed on November 1, 2017]

[11] SNOMED International, *SNOMED International Members*, SNOMED International. http://www.snomed.org/members. [accessed on November 1, 2017]

[12] US National Library of Medicine, *ICD-9-CM Procedure concepts to SNOMED CT Map*. https://www.nlm.nih.gov/research/umls/mapping_projects/icd9cmv3_to_snomedct.html [accessed on November 1, 2017]

[13] A. Romero-Gutiérrez, *Mapping SNOMED CT procedure concepts to ICD-10-PCS*. https://confluence.ihtsdotools.org/pages/viewpage.action?pageId=29952418 [accessed on October 30, 2018]

[14] SNOMED International, *Mapping to SNOMED CT: Discussion and Guidance on Mapping Based Solutions Design for Migration or Transformation,* SNOMED International, 2013. https://doc.ihtsdo.org/download/xdoc_MappingToSnomedCt_Current-en-US_INT_21030402.pdf. [accessed on October 30, 2018]

[15] SNOMED International, *SNOMED CT Browser,* SNOMED International, http://browser.ihtsdotools.org/ [accessed on November 1, 2017]

Address for correspondence

Hyeoun-Ae Park, address: College of Nursing, Seoul National University, 103 Dahakro, Jongno-gu, Seoul, 03080, South Korea. email: hapark@snu.ac.kr. phone: (82)10-3786-3284. fax: (82)2-765-4102.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190208

Automatic Identification of Individual Drugs in Death Certificates

Soon Jye Kho[a], Amit Sheth[a], Olivier Bodenreider[b]

a Kno.e.sis Center, Department of Computer Science and Engineering, Wright State University, Dayton, Ohio, USA
b *National Library of Medicine, Bethesda, Maryland, USA*

Abstract

Background. *Establishing trends of drug overdoses requires the identification of individual drugs in death certificates, not supported by coding with the International Classification of Diseases. However, identifying drug mentions from the literal portion of death certificates remains challenging due to the variability of drug names.* ***Objectives***. *To automatically identify individual drugs in death certificates.* ***Methods***. *We use RxNorm to collect variants for drug names (generic names, synonyms, brand names) and we algorithmically generate common misspellings. We use this automatically compiled list to identify drug mentions from 703,106 death certificates and compare the performance of our automated approach to that of a manually curated list of drug names.* ***Results***. *Our automated approach shows a slight loss in recall (4.3%) compared to the manual approach (for individual drugs), due in part to acronyms.* ***Conclusions***. *Maintenance of a manually curated list of drugs is not sustainable and our approach offers a viable alternative.*

Keywords:
Death Certificates; Drug Overdose; RxNorm

Introduction

Recent mortality trends show a substantial increase in drug overdose death rates in the United States. From 2010 to 2015, the rate of drug overdose death has increased from 12.3 to 16.3 per 100,000 in the U.S. population [1]. Researchers have devoted a significant amount of effort to describing drug overdose trends and to identifying the population at risk, as attempts to address this public health crisis [1–3].

Mortality data are a valuable source of information for establishing drug overdose trends. Causes of death are classified in accordance with the International Classification of Disease, Tenth Revision (ICD-10). ICD-10 codes X40-X49 identify unintentional drug poisoning or overdose deaths, while drug-specific overdose deaths are identified by the contributory causes of death indicated by "T" codes (e.g., T40.1 indicates death due to poisoning by, adverse effect of and underdosing of heroin). ICD-10 codes have been used to facilitate the drug overdose trends analysis [4–6].

While ICD-10 codes support consistent coding of the underlying causes of death, they do not provide enough granularity especially when it comes to reporting drug overdose at the level of individual drugs. While some drugs are assigned a unique ICD-10 code, most of them are not. For example, drug overdose cases caused by heroin and methadone are assigned distinct ICD-10 codes (T40.1 and T40.3 respectively), but drug overdose cases caused by fentanyl and tramadol are clustered together and assigned the same code (T40.4 - poisoning by synthetic narcotics). This issue also affects all the other opioids (T40.2), as well as barbiturates (T42.3) and benzodiazepines (T42.4).

To mitigate the granularity issue, researchers have utilized the literal portion of death certificates (i.e., a short textual description of the cause of death) to identify the contribution of a specific drug to drug overdose cases [7]. The death certificates in which a specific drug is mentioned can be retrieved and used for establishing specific drug overdose trends. Yet, identifying drug mentions in death certificates is not a trivial task as drug entities are often denoted by different terminology variants, including generic names, brand names and synonyms. Moreover, drug names are sometimes misspelled in death certificates. Therefore, it is important to identify these variants in order to have complete and accurate retrieval of death certificates in which drugs are mentioned.

Trinidad et. al. [8] manually inspected death certificates over a 5-year period (2010-2014) and identified a list of search terms for drugs. These search terms were partially populated from Substance Abuse and Mental Health Services Administration's (SAMHSA) Drug Abuse Warning Network (DAWN) Drug Reference Vocabulary (DRV) and complemented by drug mentions identified manually from the literal portion of death certificates. These search terms consist of various terminology variants, including synonyms, abbreviations, brand names, and misspellings. A study team trained in pharmacy and pharmaco-epidemiology then organized these search terms into their corresponding drug entities (around a "principal variant"), as illustrated in Figure 1. We refer to this list as the Manually Curated List (MCL).

This list covers a large number of drugs involved in the death of decedents but requires a significant amount of manual effort from domain experts for its curation. Manual curation is time-consuming and does not constitute a scalable and sustainable approach as new drugs are marketed and new variants appear in death certificates. Moreover, while terminology variants for a given drug entity were grouped together, the type of variant (e.g., brand name, misspelling) was not documented, making it impossible to study the specific contribution of each type of terminology variant.

The objective of this study is to explore an automated approach to generating a list of search terms for drugs to support the identification of individual drugs in death certificates. We assess the performance of our Automatically Compiled List (ACL) against the Manually Curated List (MCL) for the retrieval of death certificates. The main contribution of our work is to address the scalability and sustainability of drug identification in death certificates, by proposing an automated approach to generating the drug list.

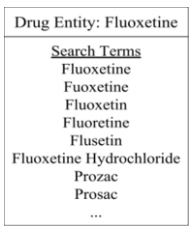


Figure 1 –Example of drug entity in the Manually Curated List with the "principal variant" (top), and the set of terminology variants.

Methods

We use RxNorm to collect variants for drug names (generic names, synonyms, brand names) and we algorithmically generate common misspellings. We use this automatically compiled list (ACL) to identify drug mentions from death certificates and compare the performance of our automated approach to that of the manually curated list (MCL) of drug names. (Figure 2).

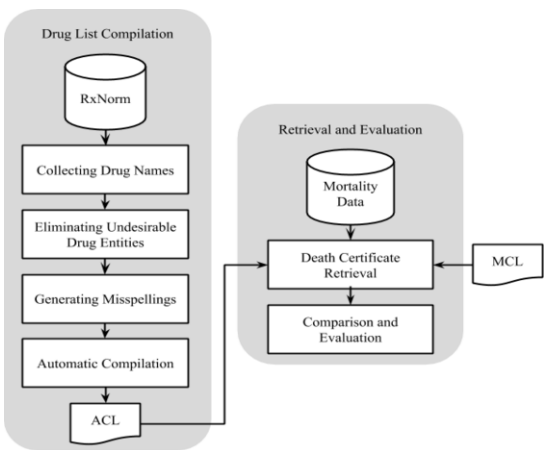


Figure 2 – Overview of the methods.

Drug List Compilation

The first stage focuses on compiling the list of search terms for drugs from RxNorm.

Collecting Drug Names from RxNorm

RxNorm is a normalized naming system for generic and branded drugs from a collection of commonly used public and private drug vocabularies [9]. The present study leverages RxNorm for populating drug names due to its comprehensive coverage of clinical drugs. We collected all names for drug entities from the following three RxNorm categories:

- Ingredients (IN) - a compound or moiety that gives the drug its distinctive clinical properties. Ingredients generally use the United States Adopted Name (USAN), e.g., *Fluoxetine*.
- Precise Ingredient (PIN) - a specified form of the ingredient that may or may not be clinically active. Most precise ingredients are salt or isomer forms, e.g., *Fluoxetine Hydrochloride*.
- Brand Name (BN) - a proprietary name for a family of products containing a specific active ingredient, e.g., *Prozac*.

The rationale for focusing on these three categories is that they are sufficient for identifying drug mentions in death certificates. While brand names generally correspond to prescription drugs (e.g., *oxycontin*), generic names for ingredients in RxNorm may also include substances for which no prescription drugs are available (e.g., *rofecoxib*). Other RxNorm categories, such as 'Semantic Clinical Drug' contain additional information (e.g., dose form and strength), usually not mentioned in death certificates and not essential for the retrieval of death certificates. We collect all synonyms from RxNorm for every drug name.

Eliminating Undesirable Drug Entities

RxNorm includes drug entities that are not of interest for identifying drugs in death certificates and will possibly generate false positives in retrieval, e.g., micro-organisms found in vaccines or used for allergy testing. We utilized semantic categorization (semantic type information) in the Unified Medical Language System (UMLS) [10] for filtering undesirable drug entities. All drug concepts from RxNorm are part of the UMLS Metathesaurus. For the purpose of retrieving death certificates, we only consider RxNorm concepts with the following UMLS semantic types:

- T109 - Organic Chemical
- T116 - Amino acid, peptide or protein
- T121 - Pharmacologic Substance
- T126 - Enzyme

In practice, we search in the UMLS Metathesaurus every RxNorm drug concept collected in the previous step and filter out those for which the semantic type is not one of the four listed above.

Another practical issue in identifying drug mentions in death certificates is that some drugs have brand names that correspond to frequently used English terms (e.g., *Prevail*, *Thrive*, *Today*). When found in death certificates, these terms generally do not denote a drug and will generate false positives in retrieval. In practice, we used the list of the top-5000 words from Word Frequency Data[1] computed from the Corpus of Contemporary American English[2] to filter out from our drug list any brand name that is present in this list.

Generating Misspellings

Some drug names are misspelled in death certificates. These misspellings indicate the mentions of drug entities but are often missed out in retrieval processes. We generated potential misspellings using an algorithm inspired by Pimpalkhute et. al. [11]. There are two phases in the algorithm: generation phase and filtering phase.

We first generated all variants with an edit distance of 1 (i.e., differing from the original by one character through insertion [*bupropion* / *bupropr*ion], deletion [*fluoxetine* / *fuoxetine*] or substitution [*Prozac* / *Prosac*]). We did not generate misspellings for chemical names (e.g., *1,1,2,2-tetrafluoroethane*) or drug names smaller than 5 characters (e.g.

[1] https://www.wordfrequency.info/free.asp

[2] https://corpus.byu.edu/coca/

Agar, *Urea*, *Tums*) to avoid proliferation and ambiguity, respectively.

Misspelling generation has the potential to create a large number of variants. Further filtering (phoneme, lexical and semantic) is needed to select the most relevant misspellings. Phoneme filtering helps to reduce the spelling variants generated to a manageable number. Lexical and semantic filtering helps to avoid generating ambiguous variants that would likely generate false positives in retrieval.

1. ***Phoneme filtering*** helps select misspellings that sound like the original term, which are the most likely to be found in text. We use the metaphone algorithm to identify spelling variants of the original drug name with similar pronunciation (see [11] for details). The other spelling variants generated in previous steps are discarded.
2. ***Lexical filtering*** eliminates existing or potentially ambiguous variants. Short variants are discarded (variants of 4 characters or less) because short words tend to be ambiguous. Variants that correspond to existing English words are discarded because their mention in text most likely denotes an entity other than the drug. Finally, variants that correspond to existing drug names are discarded, because they are already covered by the main drug list.
3. ***Semantic filtering*** eliminates variants that correspond to existing biomedical concepts outside the drug domain for the same reason we eliminate variants that correspond to existing English words. In practice, we use the filter developed for eliminating undesirable drug entities (see above).

Automatically Compiled List of Drug Names

For the purpose of identifying drug mentions in death certificates, we organized the Automatically Compiled List (ACL) of drug names around RxNorm ingredients (IN). In practice, we grouped RxNorm precise ingredients (PIN) and brand names (BN), along with their synonyms and spelling variants, together with their corresponding ingredient. For example, mentions of *Prozac* (brand name for *fluoxetine*) and *fluoxetine hydrochloride* (precise ingredient, salt form of *fluoxetine*) are counted as mentions of *fluoxetine*. However, we keep track of the specific type of terminology variant (e.g., brand name) for each variant.

Evaluation

We use this automatically compiled list to identify drug mentions from death certificates and compare the performance of our automated approach to that of the manually curated list of drug names. We use a corpus of death certificates from Washington State. This corpus spans a period of 14 years (from the year 2003 until 2016) and comprises a total of 703,106 death certificates. After indexing the death certificates with the *Elasticsearch* search engine, we used the automatically compiled list (ACL) and the manually curated list (MCL) to query the corpus of death certificates and retrieved one set for each list. Counts of death certificates are aggregated by ingredient (ACL query) or by principal variant (MCL query). For the purpose of comparing ACL and MCL, we normalized terms from the MCL to RxNorm using the RxNorm API.

Results

Drug List Compilation

Collecting and Filtering Drug Names from RxNorm

A total of 22,161 drug names are collected from RxNorm. After eliminating undesirable drug entities, 21,459 drug names remain (11,396 ingredients, 2714 precise ingredients, 5887 brand names and 1462 synonyms).

The 702 variants eliminated are described below.

- UMLS filtering: 689 RxNorm concepts were eliminated because their UMLS semantic type was outside the drug domain. Examples include *Adenine* (T114 Nucleic acid), *Air* (T167 Substance), *Apple Juice* (T168 Food), *Candida albicans* (T004 Fungus) and *Human poliovirus* (T005 Virus). Additionally, five obsolete drug names could not be mapped to the UMLS and were eliminated.
- Eight brand names corresponding to frequently used English terms (*Legend*, *Prevail*, *React*, *RID*, *Thrive*, *Today*, *Tomorrow* and *Triumph*).

Generating Misspellings

Our algorithm generated a total of 3,255,198 spelling variants, but most of them were eliminated during the filtering phase (Table 1). A total of 903,831 spellings variants was retained.

Table 1 - Number of variants eliminated at each step

Filtering step	Number of variants eliminated	Number of remaining variants	Examples of variants eliminated
Phoneme filtering	2,343,923	911,275	*azciximab, prozaw*
Short variant filtering	7132	904,143	*born, corhd, parox, dylan*
Existing English word filtering	242	903,901	*captain, watery, concert*
Existing drug name filtering	63	903,838	*butabarbital, nexavar, protamine*
Semantic types filtering	7	903,831	*phosphide, ostium, ocular*

Automatically Compiled List of Drug Names

The Automatically Compiled List (ACL) of drug names, after eliminating undesirable drug entities and generating misspellings, comprises 925,290 variants organized around 11,396 main drug entities. Figure 3 shows the terminology variants for the drug entity *diphenhydramine*.

Evaluation

Overall Retrieval of Death Certificates

A total of 37,215 unique death certificates were retrieved with the ACL query and 49,163 with the MCL query. Of these, 35,822 death certificates were retrieved by both queries, leaving 1,393 death certificates retrieved only with the ACL query and 13,341 retrieved only with the MCL query. The total number of death certificates retrieved with any query is 50,556, of which the ACL query retrieved 73.6% and the MCL query 97.2%. The difference in recall between ACL and MCL (23.6%) was expected, since MCL contains non-drug substances and drug classes, while ACL is restricted to drug names by construction.

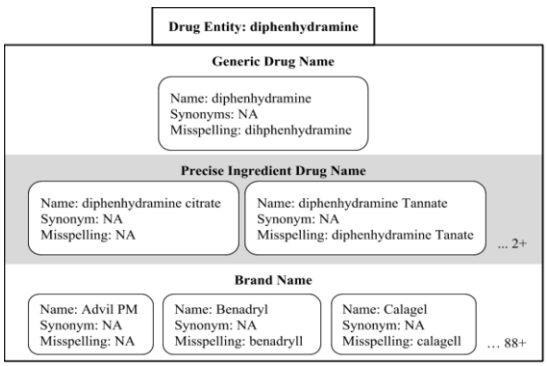


Figure 3 –Example of a drug entity from ACL with different types of variants.

Retrieval of Death Certificates for Individual Drugs

Quantitative evaluation. There are 1654 drug entities present in both MCL and ACL. To assess variant generation in ACL, we specifically compared the number of death certificates retrieved for those drug entities in common between the two lists. A total of 35,674 unique death certificates were retrieved with the ACL query and 37,278 with the MCL query. Of these, 35,633 death certificates were retrieved by both queries, leaving 41 death certificates retrieved only with the ACL query and 1,645 retrieved only with the MCL query. The total number of death certificates retrieved with any query is 37,319, of which the ACL query retrieved 95.6% and the MCL query 99.9%. The difference in recall between both lists, when restricted to individual drugs, is only 4.3% (i.e., much lower than the loss in recall observed overall). Given the significant amount of manual effort involved in curating MCL, the slight loss in recall is indicative of a strong performance for the ACL query. Moreover, the fact that the ACL query retrieved 41 death certificates not retrieved with the MCL query (for variants including *alteplace*, *remicaide*, and *gentamycin*) demonstrates the benefit of a systematic, algorithmic approach to collecting drug names and spelling variants.

Qualitative evaluation. To assess whether ACL and MCL would identify similar trends in drug-related mortality, we compared the top-20 drug entities retrieved by MCL and ACL in death certificates. Table 2 shows that both lists essentially identify the same top-20 drugs. Among the top-20 drug entities in both lists, 19 overlaps. (*Cholesterol* is inappropriately identified as a drug in ACL, but not MCL.) The ranking of these drug entities is also identical in both list, except for the permutation of *Citalopram* and *Alprazolam*, whose frequencies are very close.

Discussion

Specific Contribution of Variant Types

Table 3 shows the overall frequency of death certificates retrieved by each type of terminology variant from the ACL. While ingredient names account for the vast majority of drug mentions, brand names and misspellings also contribute to the retrieval of death certificates.

Table 4 presents the top-5 brand names mentioned in death certificates. Overall, the usage of brand names constitutes 3.6% (1323/37,215) of all drug mentions. One exception is *Coumadin* (926 mentions) , the brand name of *Warfarin* (788 mentions).

Table 2 - Top-20 drug entities identified by MCL and ACL.

MCL		ACL	
Drug entity	**Freq.**	**Drug entity**	**Freq.**
Ethanol	4561	Ethanol	3588
Methadone	3377	Methadone	3375
Methamphetamine	2964	Metamphetamine	2946
Heroin	2456	Heroin	2446
Cocaine	2171	Cocaine	2166
Oxycodone	2100	Oxycodone	2093
Warfarin	1712	Warfarin	1709
Morphine	1412	Morphine	1404
Hydrocodone	1043	Hydrocodone	1037
		Cholesterol	*970*
Diphenhydramine	955	Diphenhydramine	944
Alprazolam	*886*	*Citalopram*	*879*
Citalopram	*884*	*Alprazolam*	*876*
Diazepam	828	Diazepam	827
Oxygen	814	Oxygen	814
Nicotine	685	Nicotine	685
Amitriptyline	623	Amitriptyline	622
Acetaminophen	555	Acetaminophen	552
Iron	533	Iron	533
Hydromorphone	508	Hydromorphone	497
Fentanyl	469	Fentanyl	468

Table 3 - Overall frequency of death certificates retrieved by each type of terminology variant.

Drug category	Variant type	Aggregated number of death certificates*		
Ingredient	Drug Name	35,916	36,007	Total = 37,215
	Synonyms	48		
	Misspellings	398		
Precise Ingredient	Drug Name	122	122	
	Synonyms	0		
	Misspellings	0		
Brand Name	Drug Name	1,299	1,323	
	Synonyms	1		
	Misspellings	25		

** Some death certificates are counted multiple times here when more than one drug is mentioned.*

Table 4 - Top-5 brand names mentioned in death certificates.

Brand name	Generic name	Number of death certificates retrieved
Coumadin	Warfarin	926
Plavix	Clopidogrel	57
Tylenol	Acetaminophen	35
Adriamycin	Doxorubicin	33
Lovenox	Enoxaparin	16

Error Analysis

The error analysis reveals the search terms that lead to loss in recall. Table 5 shows the top-10 drug entities which have the highest difference in retrieval performance. These drug entities explain 61.2% (1007/1645) of the loss in recall. Most of the search terms responsible for the loss in recall are abbreviations (e.g., *ETOH* for *ethanol*). This particular term is responsible for 52.9% (870/1645) of the loss in recall. Adding this one term into ACL would decrease the loss in recall from 4.30% to 1.97%.

Upon inspection of the top-20 drug entities, we observed that some drug entities (from both ACL and MCL) tend to generate

false positive results, i.e., retrieve death certificates for which the drug mentioned is not the cause of death (e.g., *Iron*, mentioned in the context of 'iron deficiency anemia', not iron-related overdose). Most of these drug entities have the semantic type of 'T196 Element' (e.g. *iron*, *oxygen*, *gold*, *helium*) and could easily be eliminated if it is confirmed that they yield false positives.

Table 5 - Top-10 drug entities which have the highest difference in retrieval performance. The search terms are only present in MCL and responsible for the loss in recall. Only the most significant search terms are listed, along with their terminology variant types.

Drug entity	Number of death certificate retrieved		Difference	Search terms (Number of death certificates not retrieved with ACL query)
	MCL	ACL		
Ethanol	4561	3588	973	Abbreviation - *ETOH* (870)
Cannabis Sativa Subsp. Indica Top	107	0	107	Synonym - *Marijuana* (29)
1,1-difluoroethane	78	0	78	Abbreviation - *Difluoroethane* (51)
Alteplase	37	5	32	Abbreviation - *TPA* (32)
Dronabinol	31	3	28	Abbreviation - *THC* (5)
Isopropyl Alcohol	33	7	26	Synonym - *Isopropanol* (10)
Bupropion	271	248	23	Misspelling - *Buproprion* (1)
Cyclobenzaprine	345	326	19	Misspelling - *Cyclobenzapine* (1)
Methamphetamine	2964	2946	18	Abbreviation - *Meth* (7)
Dextromethorphan	213	196	17	Misspelling - *Nyquil* (1)

Sustainability

RxNorm is updated monthly. Since ACL is built programmatically, it can be easily updated when new versions of RxNorm are released. Unlike MCL, the updated ACL will identify mentions of new drugs in death certificates. The algorithm used for generating misspellings is fast and could be run easily on new versions of RxNorm.

Limitations

Our study only focused on drug identification. Assessing whether drug mentions in death certificates actually correspond to drug-related deaths is beyond the scope of this investigation.

Substances mentioned in death certificates are sometimes non-drugs (e.g., *fumes*, *cigarette*, *asbestos*) or drug class names (e.g., *opiates*, *narcotics*, *antipsychotic*, *steroid*) and are not covered by RxNorm. If ICD-10 is not sufficient for coding these substances, other sources can be investigated (e.g., drug classifications systems).

In this preliminary retrieval study, we only included the 184 misspellings that appear in our corpus of death certificates. Restricting misspellings to this small subset would not support the identification of different spelling variants in another corpus of death certificates. However, given the performance of the search engine, using the complete list of misspellings would not be a major issue.

Conclusions

In this study, we explored an automated approach to compiling a list of search terms for identifying drug mentions in death certificates. We showed that the automatically compiled list (ACL) has a competitive retrieval performance for individual drugs compared to the manually curated list (MCL), with only a slight loss in recall, and can reproduce similar drug overdose trend analysis results. Importantly, unlike the manual approach, our automated approach is dynamic, scalable and sustainable.

Acknowledgements

This work was supported by the Intramural Research Program of the NIH, National Library of Medicine. The authors would like to thank Margaret Warner and Kate Brett from the National Center for Health Statistics, Centers for Disease Control and Prevention (CDC), for providing insightful discussions.

References

[1] R.A. Rudd, P. Seth, F. David, and L. Scholl, Increases in Drug and Opioid-Involved Overdose Deaths - United States, 2010-2015, MMWR Morb. Mortal. Wkly. Rep. **65** (2016), 1445–1452.

[2] L.H. Chen, H. Hedegaard, and M. Warner, Drug-poisoning Deaths Involving Opioid Analgesics: United States, 1999-2011, NCHS Data Brief (2014), 1–8.

[3] H. Hedegaard, L.-H. Chen, and M. Warner, Drug-poisoning deaths involving heroin: United States, 2000-2013, NCHS Data Brief (2015), 1–8.

[4] C.J. Ruhm, Drug poisoning deaths in the United States, 1999-2012: a statistical adjustment analysis, Popul. Health Metr **14** (2016), 2.

[5] C.J. Ruhm, Drug involvement in fatal overdoses, SSM Popul Health **3** (2017), 219–226.

[6] M. Warner, L.H. Chen, D.M. Makuc, R.N. Anderson, and A.M. Miniño, Drug poisoning deaths in the United States, 1980-2008, NCHS Data Brief (2011), 1–8.

[7] E.M. Ossiander, Using textual cause-of-death data to study drug poisoning deaths, Am. J. Epidemiol **179** (2014), 884–894.

[8] J.P. Trinidad, M. Warner, B.A. Bastian, A.M. Minino, and H. Hedegaard, Using Literal Portion from the Death Certificate to Enhance Mortality Statistics: Characterizing Drug Involvement in Deaths, Natl. Vital Stat. Rep. **65** (2016), 1–15.

[9] S. Liu, W. Ma, R. Moore, V. Ganesan, and S. Nelson, RxNorm: prescription for electronic drug information exchange, IT Prof **7** (2005), 17–23.

[10] Bodenreider, The Unified Medical Language System (UMLS): integrating biomedical terminology, Nucleic Acids Res **32** (2004), D267–70.

[11] P. Pimpalkhute, A. Patki, A. Nikfarjam, and G. Gonzalez, Phonetic spelling filter for keyword selection in drug mention mining from social media, AMIA Jt Summits Transl Sci Proc 2014 (2014), 90–95.

Address for Correspondence

Olivier Bodenreider, National Library of Medicine, email: olivier@nlm.nih.gov

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190209

Pretraining to Recognize PICO Elements from Randomized Controlled Trial Literature

Tian Kang[a], Shirui Zou[b], Chunhua Weng[a]

[a] *Department of Biomedical Informatics, Columbia University, New York, NY, United States*
[b] *Longstar Healthpro, Inc., Los Angeles, CA, United States*

Abstract

*PICO (**P**opulation/problem, **I**ntervention, **C**omparison, and **O**utcome) is widely adopted for formulating clinical questions to retrieve evidence from the literature. It plays a crucial role in Evidence-Based Medicine (EBM). This paper contributes a scalable deep learning method to extract PICO statements from RCT articles. It was trained on a small set of richly annotated PubMed abstracts using an LSTM-CRF model. By initializing our model with pretrained parameters from a large related corpus, we improved the model performance significantly with a minimal feature set. Our method has advantages in minimizing the need for laborious feature handcrafting and in avoiding the need for large shared annotated data by reusing related corpora in pretraining with a deep neural network.*

Keywords:

Natural Language Processing, Evidence-Based Medicine, Randomized Controlled Trial

Introduction

Evidence-Based Medicine (EBM) is the conscientious, explicit, judicious and reasonable use of modern, best evidence in making decisions about the care of individual patients [1]. However, the evidence base has been growing exponentially. It is practically impossible to catch up with the explosion of the biomedical scientific literature and realize fast and effective evidence retrieval and decision making for EBM practitioners [2; 3]. Evidence adoption at clinical practice remains suboptimal due to poorly formulated clinical questions, ineffective evidence search strategies, and disconnected databases preventing access to the best evidence.

Successfully retrieving relevant evidence begins with a well-structured question. Thus, the ability to question formulation is fundamental to locate and synthesize related resources . PICO (Population/problem, Intervention, Comparison, and Outcome) is widely adopted for formulating clinical questions to retrieve evidence from the literature. PICO stands for :

P *– Population/Problem. What are the most critical characteristics of the enrolled population? What is the primary disease?*
I *– Intervention. What is the primary intervention considered?*
C *– Comparison. To what the intervention is compared?*
O *– Outcome. What are the anticipated measures, improvements or effects?*

The PICO framework is specialized to help break down the need for evidence into searchable keywords and to formulate answerable research questions [4]. A prior study has shown that utilization of the PICO framework can improve evidence search against PubMed [5]. However, due to high demands for technical skills and medical domain knowledge for using PICO, practitioners and the general public who require searching evidence may find it either time consuming to incorporate into their busy clinical workflow, or difficult to learn. Automatic extraction of PICO statements in the biomedical literature is desired to facilitate evidence retrieval, appraisal and synthesis by clinicians and the public [6; 7].

Natural language processing (NLP) in particular promises to help us achieve this goal. Previous work has explored the use of NLP techniques to identify PICO elements in biomedical text. During the last decade, the primary solutions have evolved from knowledge-based to statistical-based such as Support Vector Machine (SVM) and Conditional Random Field (CRF) [8-11]. However, this area has attracted less attention than it should have from the NLP community, primarily caused by the lack of publicly available, annotated corpora [12], and systems almost all heavily rely on laborious handcrafted features including those specifically designed to incorporate domain knowledge.

In practice, there also lacks modularized fundamental NLP tools to support different aspect of evidence synthesis and EBM, such as tools for Named Entity Recognition (NER) to recognize PICO elements and their attributes in literature for indexing, information extraction (IE) systems for parsing and structuring study design and results from free-text literature, as well as information retrieval (IR) tools based on the PICO framework to support effective searching in literature.

With rapid advances in neural network and deep learning, recent state-of-the-art NLP systems have been developed using neural models, including some for the biomedical domain. For the Named Entity Recognition (NER) task, the best performance is achieved by biLSTM-CRF [13-15]. And transfer learning attracts increasing attention to solve high demand of large data for training neural networks [16; 17]. Recently a corpus of 5000 RCT abstracts with multi-level annotations of Patient, Intervention, and Outcomes was published, enabling new NLP application development for EBM research [12].

Compared to prior work, our PICO extraction method makes the following three significant and innovative contributions. First, it is the initial publicly available open-source NLP system for recognizing PICO elements and their attributes/measures in RCT abstracts. PICO elements are normalized with UMLS CUIs (https://github.com/Tian312/PICO_Parser). Second, this tool is developed with the minimum human labor but achieves comparable and even better performance in some categories: only a small size of gold standards is created with high inter-annotator agreement; only word feature, and no laborious handcrafted features, is used. Third, we contribute a method to reuse a large related corpus [12] under annotation guidelines different from ours to improve our model performance.

Methods

Our PICO statement extraction tool processes RCT literature following these steps: 1) Named Entity Recognition for PICO

elements and attributes; 2) UMLS encoding; 3) XML output formatting. An overview of our workflow to develop the model and tool is shown in Figure 1.

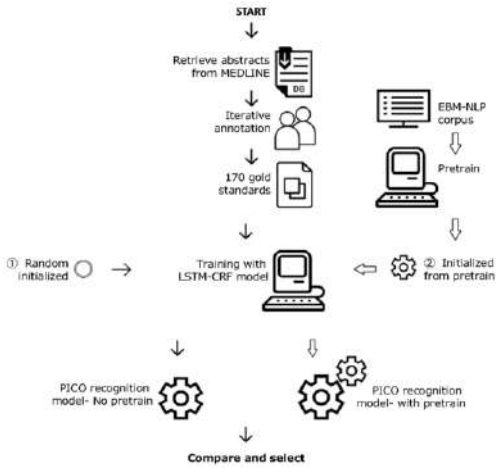


Figure 1 - Overview of the PICO recognition tool development. We compared two optional ways for training the LSTM-CRF model (in blue): 1) with random initialization of model parameters (green, on the left); 2) "pretrain" the model with the same architecture on EBM-NLP corpus, resulting in a better parameter initialization.

Data Collection

Small Size of Gold Standards from Manual Annotation

We randomly retrieved 170 RCT publications using indexed metadata from the MEDLINE database. Abstracts were retrieved from the articles and prepared in brat, a web server based collaborative annotation tool [18]. One medical professional (ZS) and one informatic researcher (TK) designed the annotation guideline for entity and attribution using an iterative process. Entity classes included in the annotation: Population, Intervention (Comparison is merged with Intervention as a subclass), and Outcome, each strictly following standard definition from the PICO framework [4]. The context for PICO elements consists of 2 types of attributes: Qualifier, a qualitative description of PICO elements (e.g., *"difference", "similar", "higher"*), and Measure, a quantitative description of PICO elements (e.g., *"138 +/- 13 mg daily"*). During the annotation process, both annotators followed the guideline for asking answerable research question [19]: each RCT abstract is first classified into one of 5 common clinical question types: Treatment, Prevention. Diagnosis, Prognosis, and Etiology, then annotated with PICO elements based on research type. Attributes are also identified in order to form PICO statements in entity-operator-value triplets. Each abstract is annotated at least twice by two annotators in order to ensure it strictly follows the guideline. As a rule, annotators skip annotating background and implication sections in abstracts since those parts do not usually describe study design or report objective results. An example annotation interface in brat is shown in Figure 2. This step is aimed to create a small size of annotation with a high inter-annotator agreement and high quality, serving as gold standards and core training set.

A Related, Large Publicly Available Corpus

A corpus of 5000 abstracts with multi-level annotation of PIO (C is categorized as a subclass in I as well) has recently been published by a group of EBM researchers [12] (referred as EBM-NLP corpus later in this paper). The annotation was generated primarily by laypersons from Amazon Mechanical Turk (AMT) and a small part by medical professionals. The average inter-annotator agreement is measured by F1 score as 0.3, 0.18, 0.1 in span annotation and 0.5, 0.6, 0.69 in the hierarchical annotation for P, I, O classes, respectively. After reviewing this corpus, we decided that this annotated corpus cannot be directly used for our task for two major reasons. First, the annotation guidelines, primarily in part of defining element boundary and granularity, are different. In the EBM-NLP corpus, identified PICO tend to be the longest description within a sentence. While our guideline is designed to break down abstracts into the most basic elements, which can be used as "building block" for PICO statements and encoded to represent study design and results of each RCT article. For example, "*Seventy-two consecutive anti-HBe-positive chronic hepatitis B patients (59 male and 13 female, median age 41 yr)*" (PMID 10235220) was annotated as one Population element in EBM-NLP. While in our annotation, we recognize "*consecutive anti-HBe-positive chronic hepatitis B*", "*male*", "*female*" and "*median age*" as 4 independent Population entities, and "59", "13", "41yr" as measures. Second, as aforementioned, the EBM-NLP corpus combines measured values descriptive statistics with the PICO terms, while we have separated classes. Instead of directly training on this corpus, we believe it can be helpful for modeling a similar context as a pretrain and guide the next model training on our small gold standards.

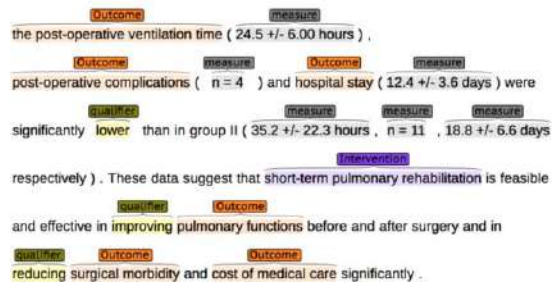


Figure 2 – Example of our annotation in brat

Base model learning and pretraining

We model the task to identify PICO statements, which comprise PICO elements and their attributes in the biomedical literature, as a sequence labeling task for Named Entity Recognition (NER). NER is fundamental in general text mining, as well as in biomedical domain, e.g., recognizing problems, drug names in clinical notes, or protein, gene names in literature. As deep learning based approaches to this tasks have been gaining attention in recent years, NLP researchers now tend to prefer those methods over traditional models alone such as Support Vector Machine (SVM) or Condition Random Field (CRF) since the parameters can be learned end-to-end without the need for hand-engineered features [15]. This is particularly true in biomedical domain where traditional biomedical NLP systems heavily rely on hand-made rules and ontologies in order to reach a good performance. Deep learning methods also start attracting biomedical NLP researchers. However, these approaches are usually built upon large, high-quality labeled data, which is expensive to obtain especially in the biomedical domain because labeling biomedical text requires special medical training.

To address the lack of training corpus, recent researches start focusing on training multi-task models [20], and conducting data augmentation or transfer learning [16; 17]. Inspired by their work, we explore the feasibility and the potential way to overcome such two challenges (i.e., hand-engineered features

and large, high-quality data) in a small training set and simple feature with the help of the public data. We adopt the bidirectional Long short-term memory (LSTM), a kind of recurrent neural network as our base model, and decode with a linear chain CRF in the output layer (biLSTM-CRF). This architecture now achieves state-of-the-art performance in NER tasks. The model details are illustrated in Figure 3. We use classical "BIO" tags to represent the boundary of terms of interest: "O" means it is outside the target terms. "B" represents the beginning word, and "I" tags all the inside words. We compare the results trained with raw tagging to BIO tagging methods. Tagged output for the example in Figure 3 is:

- ***Pre-operative****/B-Intervention* ***short-term****/I-Intervention* ***pulmonary****/I-Intervention* ***rehabilitation****/I-Intervention for patients …*

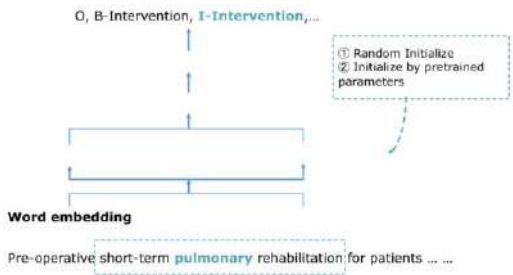


Figure 3 - Base model detailed architecture. It's used to train both PICO recognition model and EBM-NLP corpus.

The base model is similar to [14], and is also used in EBM-NLP corpus paper to generate task baseline for identifying PICO span. The dark green in Fig. 3 represents modules learned during training. While the light green, namely word embedding is pretrained on an entire collection of abstracts on PubMed from 1990-2018 using word2vec toolkit [21]. During the learning phase, the model first generates a character-based representation of each word and concatenate it with pretrained word vectors. In other words, each input word is represented by two concatenated vectors in both word level and character level. In the next step, each sentence as sequenced word vectors is then fed into a bidirectional LSTM to extract the contextual representation of each word.

At this step, we can get a likelihood at word level through a decoder layer. A significant drawback of optimizing by word-level likelihood is that it doesn't consider dependencies between neighboring in the sentence. Thus, a CRF layer is added to model the entire sentence structure. CRF is a log-linear graphical model that additionally considers the transition score from one tag to the next. This characteristic makes it a classic model in traditional NER tasks. After decoding by CRF, the log likelihood is maximized for the entire sentence in order to select the best tag for the target word. Like in Figure 3, when the target word is "pulmonary", all the neighbor words in a window are considered to generate the tag. Only word features, i.e., word vectors and character vectors, are used in this model, without any feature engineering.

Without pretraining, biLSTM in the base model is randomly initialized and then optimized using the Adam optimizer. For pretraining, we use the entire 5000 abstracts in EBM-NLP corpus with "starting span" annotation (only in PICO level, no further hierarchical labeling) in the same model architecture, and then transfer the learned weights to the PICO recognition model. Next, we fine-tune the PICO recognition model to reach the best performance. All models are trained using TensorFlow (https://www.tensorflow.org/).

Concept Normalization and Output Structuring

We select the best model as the backend support of our PICO recognition tool. Given one or a set of free text abstracts as input, the tool automatically recognizes Patient, Intervention (including Comparison), Outcome elements and corresponding attributes. In order to support further computational tasks, the recognized PICO elements are encoded with the Unified Medical Language System (UMLS, https://www.nlm.nih.gov/research/umls/), an integrated biomedical terminology, by applying a UMLS concept extraction tool QuickUMLS [22]. Extracted semantics are further organized into a structured format. The default output format is XML, while users can also choose JSON, as more recently published APIs use JSON as standard data format.

Results

Descriptive Statistics of the Annotated Corpus

We created a sharable, finely annotated corpus for PICO extraction with its descriptive statistics provided in Table 1.

Table 1. Descriptive statistics of the annotated corpora

	Entity class			Attribute class	
	P.	**I. (+C.)**	**O.**	**Qualifier**	**Measure**
Count	1185	2027	2140	766	904
Agreement	0.916	0.844	0.727	0.955	0.954

The inter-annotator agreement is evaluated by cohen's κ statistic. The overall agreement between the two annotators for 5 categories is 0.83. The category-specific κ measures is reported in Table 1. Our goal in this step is to create a corpus of high-quality annotation with high agreement. Thus, our annotation team spent much time on iterative annotation guideline design and test run in sample corpus for multiple rounds to resolve discrepancies between annotators and arrive at consensus understandings for each class and required granularity. Compared to the related NLP work, we have a relatively small corpus to minimize human labors. We plan to achieve satisfactory performance with such small corpus.

Model Performance

For evaluation purpose, 6-fold cross-validation is applied. 170 abstracts are equally divided into 6 groups. Among each run of model learning and testing, 4/6 of the data used as training set, 1/6 as validation set and 1/6 as test set. We report the performance on test sets.

Classic evaluation metrics are generated for evaluating NER tasks: precision, recall and, F1 score, to evaluate model performance in two different levels: word level and token level and use represent the two by *span* and *trunk*. In word level or span evaluation, the basic unit is the word, while in trunk evaluation, basic unit is a token. Using an intervention element with BIO tagging as an example, "*short-term/B-Intervention pulmonary/I-Intervention rehabilitation/I-Intervention*", in span evaluation, there are 3 predictions for each word. A true positive is counted when both BI tag and class are predicted correctly. There are 3 true positives at most. While in trunk evaluation, "short-term pulmonary rehabilitation" is counted as one token, 1 true positive is counted only if both boundary and class of this token are correctly predicted.

The model performance is reported based on the test set in Table 2. We test performance with different tagging methods (raw/BIO tagging) and pretrain or not. For each model setting, we report the best evaluation among 6 sets for cross-validation

and also averaged measures. In summary, using BIO tagging and pretrain can both improve model performance. In word-level, span evaluation, the best performance comes from the model using pretrain and raw tagging with 0.78 in averaged F1 score, and the best F1 in the subset is 0.89. Compared to the model setting with raw tagging as well but using no pretrain, the F1 score has been improved about 10%. With pretrain and BIO tagging, the performance is also improved, but not as much as using raw tagging. It's a reasonable result as in BIO tagging a word is counted as true positive requiring both BI tags and class are correct while raw tagging only require class prediction. On the other hand, BIO tagging provides more information for learning to help identify the boundary of each element. This is reflected by the evaluation in trunks/token level. The best performance in token level is generated by model with pretrain and BIO tagging (average F1 score 0.62, best 0.64). Compared to the two model only trained on 170 abstracts (0.52/0.54 for average F1), pretraining on EBM-NLP corpus and transferring learned parameters also help improve the model performance significantly by 10%. Therefore, applying pretrain and BIO tagging can best improve the recognition of PICO elements boundaries and predicting PICO classes.

Table 2. Model performance in different training settings.

		No Pre + Raw		No Pre + BIO		Pre + Raw		Pre + BIO	
		Best	**Ave.**	**Best**	**Ave.**	**Best**	**Ave.**	**Best**	**Ave.**
Test set (Span)	**Precision**	0.78	0.76	0.84	0.85	0.93	0.87	0.86	0.83
	Recall	0.62	0.63	0.68	0.66	0.80	0.70	0.71	0.7
	F1 score	0.69	0.66	0.75	0.74	**0.89**	**0.78**	0.78	0.73
Test set (Trunk)	**Precision**	0.54	0.52	0.58	0.53	0.74	0.61	0.63	0.63
	Recall	0.53	0.51	0.56	0.52	0.74	0.56	0.64	0.61
	F1 score	0.53	0.52	0.57	0.54	0.74	0.58	**0.64**	**0.62**

We further analyzed the individual performance from one of the 6 sets using the best model setting (Pre+BIO). The details are shown in Table 3. As evaluated by F1 score, entities with B tags are generally better than I tags, indicating the model is better at predicting if there is an entity, but need to be improved to predict the span of it. I tag prediction is especially poor in evaluation for Modifier, with F1 score only 0.25. We retrieved raw prediction results for Modifier class. We found due to the fact that we have a small gold standard set, and only 1/6 used for testing, there are only 166 Modifier tokens in the test set, among which only 6 are not unigram (have I tags). Modifiers are usually one-word token such as "higher", "rise", and "similar". Among 6 modifiers with I tags in the test set, the model predicts 2 I-modifier tags and 1 of them is correct. Thus, precision/recall is 0.5/0.16 and F1 is calculated as low as 0.25, but actually caused by its small total number in entire corpus.

Table 3. Detailed evaluation for one set in PICO/attribute

	B-Pop.	**I-Pop.**	**B-Int.**	**I-Int.**	**B-Out.**	**I-Out.**
Precision	0.82	0.84	0.82	0.78	0.88	0.85
Recall	0.68	0.65	0.70	0.50	0.75	0.42
F1 score	0.75	0.74	0.75	0.61	0.81	0.56
	B-Mea.	**I-Mea.**	**B-Qua.**	**I-Qua.**		
Precision	0.77	0.85	0.91	0.5		
Recall	0.65	0.65	0.60	0.17		
F1 score	0.71	0.74	0.72	0.25		

Sample output

The models are trained with following parameters: mini batch (size of 5) and Adam optimizer are selected for training; the dimensions of the word and character vectors are 200 and 100; the learning rate is set as 0.001with a decay of 0.9. Pretrain converges within 50 epochs and the training on 170 abstracts within 10 epochs. A sample recognition result in XML format is shown in Figure 4 It contains rich parsed semantic and positional information that can support further computational tasks such as relation extraction and information retrieval. Sample JSON output can be found in our github repository.

```
<abstract pmid="43164">
  <sent section="OBJECTIVES">...</sent>
  <sent section="METHODS">
    <text>Considerable differences in dose ( atenolol 138
    +/- 13 mg daily ; labetalol 308 +/- 34 mg daily ;
    metoprolol 234 +/- 22 mg daily ; and pindolol 24 +/-2 mg
    daily were required to produce similar antihypertensive
    effects .</text>
    <attribute class="qualifier" index="T94" start="1">
    differences</attribute>
    <entity class="Outcome" UMLS="" index="T95" start="3">
    dose</entity>
    <entity class="Intervention" UMLS="C0004147:atenolol"
    index="T96" start="5">atenolol</entity>
    <attribute class="measure" index="T97" start="6">138 +/-
    13 mg daily</attribute>
    <entity class="Intervention" UMLS="C0022860:labetalol"
    index="T98" start="12">labetalol</entity>
    <attribute class="measure" index="T99" start="13">308
    +/- 34 mg daily</attribute>
    <entity class="Intervention" UMLS="C0025859:metoprolol"
    index="T100" start="19">metoprolol</entity>
    <attribute class="measure" index="T101" start="20">234
    +/- 22 mg daily</attribute>
    <entity class="Intervention" UMLS="C0031937:pindolol"
    index="T102" start="27">pindolol</entity>
    <attribute class="measure" index="T103" start="28">24
    +/-2 mg daily</attribute>
    <entity class="Outcome" UMLS="C0003364:antihypertensive"
    index="T104" start="37">antihypertensive effects</entity>
  </sent>
</abstract>
```

Figure 4 - Sample output for our PICO extraction method

Discussion

Error Analysis

The most common error happens when multiple PICO terms appear in conjunction. For example, an RCT paper titled "Perioperative enteral nutrition and quality of life of severely malnourished head and neck cancer patients: a randomized clinical trial", and one Population entity recognition result is:

<entity class=Population UMLS='C0278996:head and neck cancer,C0162429:malnourished,C0205082:severely' index='T3' start='8'> severely malnourished head and neck cancer </entity>

Although we define the PICO elements to be the annotated as the most basic concepts (should be "*malnourished*" and "*head and neck cancer*" the two P entities in this case), there was

variance in what annotators considered, also causing inconsistency when calculating inter-annotator agreement.

Comparative Performance Evaluation Results

In the EBM-NLP corpus, the best performance for the baseline model trained on 5000 abstracts for P, I, O classes are 0.71, 0.65 and 0.63 by F1 score respectively (mathematical mean: 0.66) at the word level. In contrast, with a small gold standard set (170 abstracts) and without any hand-engineered features, our model reaches 0.78 for the best average F1 score at word level and 0.62 at the token level. Our results prove the effectiveness of pretraining in minimizing human efforts in annotation and features engineering while reaching satisfactory performance.

Future Work

We have not yet related the attributes to their PICO elements nor distinguished PICO elements by arms. To further complete the structured information, negation and semantic relations need to be identified. We will progressively complete the functions of this tool, and eventually turn it to comprehensive information extraction system to computationally represent abstracts describing RCTs.

Conclusions

In this study, we demonstrate the early promise of pretraining to improve model performance tuned on a small training set, with only word feature, and we achieve better performance than conventional machine learning models trained on a larger corpus. This result is significant in showing the feasibility of overcoming the challenges in the dearth of annotated data and laborious feature handcrafts in biomedical NLP. We also contribute an open source NLP tool to automatically recognize PICO elements and their attributes from RCT abstracts. This tool, can be used to structure study design and results and can further enhance evidence retrieval and synthesis from biomedical literature to facilitate evidence-based medicine.

Acknowledgments

This project was supported by USA NIH grant R01LM009886-08A1 (Bridging the semantic gap between research eligibility criteria and clinical data; PI: Weng).

References

[1] I. Masic, M. Miokovic, and B. Muhamedagic, Evidence based medicine - new approaches and challenges, *Acta Inform Med* **16** (2008), 219-225.

[2] H. Bastian, P. Glasziou, and I. Chalmers, Seventy-five trials and eleven systematic reviews a day: how will we ever keep up?, *PLoS medicine* **7** (2010), e1000326.

[3] A.G. Fraser and F.D. Dunstan, On the impossibility of being expert, *Bmj* **341** (2010), c6815.

[4] W.S. Richardson, M.C. Wilson, J. Nishikawa, and R.S. Hayward, The well-built clinical question: a key to evidence-based decisions, *ACP journal club* **123** (1995), A12-A12.

[5] C. Schardt, M.B. Adams, T. Owens, S. Keitz, and P. Fontelo, Utilization of the PICO framework to improve searching PubMed for clinical questions, *BMC medical informatics and decision making* **7** (2007), 16.

[6] P. Fontelo, F. Liu, and M. Ackerman, ask MEDLINE: a free-text, natural language query tool for MEDLINE/PubMed, *BMC medical informatics and decision making* **5** (2005), 5.

[7] B.C. Wallace, I.J. Dahabreh, C.H. Schmid, J. Lau, and T.A. Trikalinos, Modernizing the systematic review process to inform comparative effectiveness: tools and methods, *Journal of comparative effectiveness research* **2** (2013), 273-282.

[8] D. Demner-Fushman and J. Lin, Answering clinical questions with knowledge-based and statistical techniques, *Computational Linguistics* **33** (2007), 63-103.

[9] S.N. Kim, D. Martinez, L. Cavedon, and L. Yencken, Automatic classification of sentences to support evidence based medicine, in: *BMC bioinformatics*, BioMed Central, 2011, p. S5.

[10] I.J. Marshall, J. Kuiper, E. Banner, and B.C. Wallace, Automating biomedical evidence synthesis: RobotReviewer, in: *Proceedings of the conference. Association for Computational Linguistics. Meeting*, NIH Public Access, 2017, p. 7.

[11] B.C. Wallace, J. Kuiper, A. Sharma, M. Zhu, and I.J. Marshall, Extracting PICO sentences from clinical trial reports using supervised distant supervision, *The Journal of Machine Learning Research* **17** (2016), 4572-4596.

[12] B. Nye, J.J. Li, R. Patel, Y. Yang, I.J. Marshall, A. Nenkova, and B.C. Wallace, A Corpus with Multi-Level Annotations of Patients, Interventions and Outcomes to Support Language Processing for Medical Literature, *arXiv preprint arXiv:1806.04185* (2018).

[13] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, Neural architectures for named entity recognition, *arXiv preprint arXiv:1603.01360* (2016).

[14] X. Ma and E. Hovy, End-to-end sequence labeling via bi-directional lstm-cnns-crf, *arXiv preprint arXiv:1603.01354* (2016).

[15] D.S. Sachan, P. Xie, M. Sachan, and E.P. Xing, Effective Use of Bidirectional Language Modeling for Transfer Learning in Biomedical Named Entity Recognition.

[16] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C.H. So, and J. Kang, BioBERT: pre-trained biomedical language representation model for biomedical text mining, *arXiv preprint arXiv:1901.08746* (2019).

[17] J.M. Giorgi and G.D. Bader, Transfer learning for biomedical named entity recognition with neural networks, *Bioinformatics* **34** (2018), 4087-4094.

[18] P. Stenetorp, S. Pyysalo, G. Topić, T. Ohta, S. Ananiadou, and J.i. Tsujii, BRAT: a web-based tool for NLP-assisted text annotation, in: *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, 2012, pp. 102-107.

[19] E. Fineout-Overholt and L. Johnston, Teaching EBP: Asking searchable, answerable clinical questions, *Worldviews on Evidence-Based Nursing* **2** (2005), 157-160.

[20] X. Wang, Y. Zhang, X. Ren, Y. Zhang, M. Zitnik, J. Shang, C. Langlotz, and J. Han, Cross-type biomedical named entity recognition with deep multi-task learning, *arXiv preprint arXiv:1801.09851* (2018).

[21] T. Mikolov, K. Chen, G. Corrado, and J. Dean, Efficient estimation of word representations in vector space, *arXiv preprint arXiv:1301.3781* (2013).

[22] L. Soldaini and N. Goharian, Quickumls: a fast, unsupervised approach for medical concept extraction, in: *MedIR workshop, sigir*, 2016.

Address for correspondence

Chunhua Weng, Department of Biomedical Informatics, Columbia University, 622 W 168th Street, PH-20 room 407, New York, NY 10032, USA. E-mail: chunhua@columbia.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190210

A Study of Medical Problem Extraction for Better Disease Management

Youngjun Kim[a], Stéphane M. Meystre[a,b]

[a] *Medical University of South Carolina, Charleston, South Carolina, USA*
[b] *Clinacuity, Inc., Charleston, South Carolina, USA*

Abstract

This study focuses on the extraction of medical problems mentioned in electric health records to support disease management. We experimented with a variety of information extraction methods based on rules, on knowledge bases, and on machine learning, and combined them in an ensemble method approach. A new dataset drawn from cancer patient medical records at the University of Utah Healthcare was manually annotated for all mentions of a selection of the most frequent medical problems in this institution. Our experimental results show that a medical knowledge base can improve shallow and deep learning-based sequence labeling methods. The voting ensemble method combining information extraction models outperformed individual models and yielded more precise extraction of medical problems. As an example of applications benefiting from acurate medical problems extraction, we compared document-level cancer type classifiers and demonstrated that using only medical concepts yielded more accurate classification than using all the words in a clinical note.

Keywords:
Medical Informatics, Natural Language Processing, Neural Networks

Introduction

Computerized physician order-entry systems with decision support have been implemented as a solution to reduce medical errors, but these systems rely on structured and coded information in the electronic health record (EHR). Unfortunately, a substantial proportion of the information available in the EHR is only mentioned in narrative clinical notes. Electronic lists of problems are available in most EHRs, but they require manual management and consequently are often incomplete, inaccurate, and out of date.

To improve the completeness, correctness, and timeliness of electronic lists of medical problems and allergies, we are building on earlier efforts [26] and developing a system based on natural language processing (NLP) to continually extract relevant information from EHR notes with high accuracy [27]. This system is integrated with the EHR system and helps ensure the majority of medical problems and allergies of a patient are known by their healthcare providers, and are available for decision support and quality improvement.

In this study, we focus on medical problem extraction (MPE) to identify medical problem concepts, such as disease, sign, symptom, and other abnormalities mentioned in EHR text notes. MPE makes relevant patient information available to improve patient data quality and disease management. The challenge of this task is in limited access to shared data, unstructured narrative text format, and heterogeneous contents across various healthcare providers.

For this research, we employ a variety of information extraction (IE) methods including rule-based and machine-learning based methods. Early clinical NLP applications used rule-based approaches combining syntax analysis and semantic lexicons to identify and encode medical information [6; 10; 30]. MetaMap [1] is a prominent rule-based system example we also used for this research. It maps terms found in biomedical or clinical text with Unified Medical Language System (UMLS) [2; 21] Metathesaurus concepts.

Most current clinical IE systems use statistical machine learning approaches that often achieve better performance than rule-based approaches. The MPE task is closely related to named entity recognition (NER). Over the past two decades, NER has applied a variety of machine learning algorithms [8; 17; 24]. Clinical NER has advanced from general text or biomedical literature NER by sharing the algorithms and features. More recently, NER based on deep neural networks [7; 18; 22] has reached even higher accuracy than previous machine learning-based approaches. We also adopted these recent advances and trained long-short-term memory (LSTM) recurrent neural networks (RNN) [12] with the output of rule-based systems.

Machine learning-based approaches have been effective in information extraction because of their ability to automatically capture lexical, syntactic, or morphological patterns in the terms of interest and their local surroundings. However, when focused only on certain types of information with small amounts of labeled text for training, supervised learning methods tend to only allow for reduced accuracy. They tend to miss information that does not appear in training data, or extract false positive information that is not of interest.

We investigate information extraction methods to identify only commonly recorded medical problems in EHRs. We present a new clinical text corpus from the EHR of patients with breast, lung, or gastrointestinal cancer. We focused on the 168 medical problems most frequently recorded at the University of Utah Healthcare in 2013. We annotated all mentions of the selected medical problems and of all directly related medical problems (i.e., parent or children such as "*Hypertensive crisis*" or "*Systolic hypertension*" with "*Hypertensive disease*").

First, we create four different MPE applications implementing different methods and evaluate performance with our new corpus. Next, we show that augmenting features derived from a medical knowledge-base allows for better performance, and using a combination of various MPE methods can yield more precise extraction of medical problems.

As another contribution of our study, we examine whether the MPE methods can help with document classification. We trained multi-class classifiers using the words contained in the documents and medical problem concepts identified by the MPE applications. We demonstrate that cancer type classifiers using only medical concepts can lead to more accurate classification.

Methods

In the following paragraphs, we present an NLP system that extracts medical problems from clinical notes. We describe how we created a new clinical text corpus and examine how we developed four different MPE applications.

Data Description

For this study, we randomly selected a cohort of 154 patients treated at the Huntsman Cancer Institute (Salt Lake City, Utah) for breast, lung, or gastrointestinal cancer. For each patient, five most recent clinical notes of select types were used. To create a reference standard, nine medical experts annotated all mentions of the selected medical problems (i.e., 168 most frequently recorded medical problems at the University of Utah Healthcare in 2013) found in the clinical notes. The inter-annotator agreement was good with an F_1 score of 70% or even 90% if not considering detailed modifiers (e.g., negation, temporality, subject), allowing for a good quality reference standard.

Our corpus was randomly split in a training subset of 495 notes and a test subset of 275 notes. The training subset contains 9,337 annotated medical problem concepts and the test subset contains 5,360 concepts. Table 1 shows the ten most prevalent medical problems in each text collection categorized by cancer type. Some medical concepts, such as '*dyspneas*' and '*swollen lymph nodes*', are common across all texts. However, the prevalence of most medical concepts varies greatly. For example, '*colon cancers*' is far more common in the medical records of gastrointestinal cancer patients (not surprisingly).

Medical Problem Extraction Applications

We describe how medical problems were recognized using a variety of information extraction methods, including knowledge base, rule-based, and machine learning models.

MetaMap

We used MetaMapLite [9], a knowledge-based system that identifies UMLS Metathesaurus concepts in free text. We used the CUIs (Concept Unique Identifiers) assigned by MetaMap to select concepts of interest for our task. Once we collected all of MetaMap's findings, we chose the MetaMap concepts with CUIs listed in our terminology subset of 16,384 UMLS concepts described below.

Rule-based Application

The rule-based system used dictionary lookup with a custom medical problem terminology subset. We used a modified version of ConceptMapper [31], an easily configurable dictionary annotator. We identified the longest phrase of contiguous word tokens that matched one of the dictionary entries.

We created a medical problem terminology subset based on the Clinical Observations Recordings and Encoding (CORE) Problem List Subset of SNOMED CT [11]. For the 168 most frequently recorded problems, the parent and all children (i.e., more detailed concepts) were also included. As a result, our dictionary included 57,624 medical problem concepts. For each concept, the corresponding UMLS CUI was recorded. A CUI was often linked with multiple terms. For example, '*renal fibrosis*', '*kidney fibrosis*' and '*fibrosis kidney*' were all linked with '*C0151650*'.

Conditional Random Fields (CRF) Application

We created sequential models using linear chain CRFs [17]. We applied the Stanford CoreNLP tool [23] to our datasets for tokenization, sentence annotation, lemmatization, part-of-speech (POS) tagging, and named entity recognition (NER). To train the CRF models, we used Wapiti [20], a simple and fast discriminative sequence labeling software.

The training data was reformatted with BIO token tags (B: at the beginning, I: inside, or O: outside of a concept). The feature set was based on our previous medical concept extraction studies [14; 15]. We defined features for the targeted word's lexical string, lemma, POS tag, affix(es), orthographic features (e.g., alphanumeric, InitCaps), named entity tag (e.g., PERSON, LOCATION, DATE, TIME), word skip-grams, word embedding features, and the predictions of *MetaMap* and the *Rule-based application*. We used the orthographic features defined in McDonald and Pereira [25] with a slight modification. For example, "`[A-Z].*`" is a regular expression to capture whether the first letter of a word is capitalized (InitCaps). We performed K-means clustering for word embedding using word2Vec [28]. We created 1,000 clusters of semantically related words in the 2014 English Wikipedia data with word2Vec default parameters. We used the cluster identifier of each word in a sentence as a feature.

We implemented four different CRF variations: without external system output (CRF), with MetaMap output (CRF+M), with the Rule-based application output (CRF+R), and with the outputs of both systems (CRF+M+R) as features. To fine tune

Table 1– Ten Most Frequent Medical Problems in the Training Set

CUI	Examples	Count			
		Total	brCA	giCA	lgCA
C0006142	*breast cancer, malignant tumor of breast*	360	350	4	6
C0027497	*nausea, feeling queasy*	353	90	192	71
C0007102	*colon cancers, malignant colon tumor*	326	17	300	9
C0013404	*dyspneas, short of breath*	309	102	99	108
C0015967	*high body temperature, fevers*	296	93	149	54
C0000737	*abdominal pains, bellyache*	288	69	202	17
C0042963	*vomiting symptom, throwing up*	227	60	128	39
C0011991	*diarrheas, loose stools*	223	38	119	66
C0497156	*lymphadenopathy, swollen lymph nodes*	216	62	85	69
C0015672	*fatigue, tiredness*	194	37	80	77

brCA: breast cancer, giCA: gastrointestinal cancer, lgCA: lung cancer

the parameters of machine learning-based classifiers, we randomly selected 100 documents (about 20% of the training set) from the training set as held-out data. Parameters were tuned to maximize the F_1 score with the held-out data. We observed that CRF models performed very well with the parameter settings used in our medical concept extraction studies [14; 15]. We set the size of the interval for the stopping criterion to be $e = 0.001$. For regularization, *L1* and *L2* penalties were set to 0.005 and 0.4, respectively.

RNN Application

We trained RNN-based deep neural network models proposed by Ma and Hovy [22]. To build BLSTM-CNN-CRF models, the combination of bidirectional LSTM network, character-level CNN (convolutional neural network) and CRF decoding, we employed the NER system of Reimers and Gurevych [29]. Similarly to CRF models, we trained RNN models with or without the outputs of two external systems. Words and the output of the MetaMap and Rule-based application were used as inputs. We used *RMSProp* gradient descent method to produce more stable results than other optimizers. The RMSProp gradient descent is an adaptive learning rate method proposed by Tieleman and Hinton [32] and keeps the gradient relatively small and consistent during iterations. After experimenting with different values on the development data, we set the learning rate at 0.0001 and the models were trained within fifty epochs.

As demonstrated by the experimental results of Reimers and Gurevych [29], non-deterministic approaches, such as LSTM network with random weight initialization and a random shuffling of the training data, can lead to statistically significant differences between multiple runs, even with the same hyperparameters. To obtain the most representative or consistent model among the various models, we applied the model selection method based on the pairwise comparison. We trained five different RNN models with the same hyperparameters and extracted medical concepts from the test set. We then calculated the similarity (i.e., pairwise F_1 score) between the results of five trials and selected the model with the highest F_1 score on average.

Voting Ensemble Method

A voting ensemble can provide a convenient and effective way to combine multiple applications without retraining the new model. It has been applied to several NLP tasks and has shown that it can outperform individual applications [4; 5; 16]. We created an ensemble method combining all MPE applications with voting from the predictions of the individual application. We used the ensemble method with voting that we successfully applied to our previous research on medical concept extraction [14; 15] or text de-identification [13]. The ensemble method collects all concept predictions from the individual MPE applications.

The voting ensemble selects a concept candidate when having one vote for each component to determine whether the candidate is a medical concept. When two different concepts have overlapping text spans, the concept that receives more votes is selected. For overlapping concepts with identical vote counts, the one produced by the highest ranked MPE application is selected. To determine the ranking of the applications, we measured how an MPE application agreed with the others for medical concepts extraction. Similarly to the model selection method applied to the RNN, an application ranking was based on the F_1 scores measured with the other applications. The higher the average F_1 score, the higher the application was ranked.

Cancer Type Classification

To confirm that our MPE applications could benefit other high-level tasks, we tackled the cancer type classification problem. This task involves recognizing three types of cancer: *breast cancer*, *gastrointestinal cancer*, and *lung cancer*. The classifier assigns one of the three cancer categories to a clinical note using its textual content.

We created a multi-class OGD (online gradient descent) [3] classifier with words and medical problem concepts. We used the Vowpal Wabbit [19] online learning library to train one-against-all models. For each word contained in the document, we defined 3-skip-trigrams that allow for words to be skipped. These lexical features were also captured for each word contained in the medical problem term. The number of training passes was set to 100 after trials with default Vowpal Wabbit parameters.

Results

We present experimental results for each MPE application and for the voting ensemble method. We measured recall, precision, and the F_1 score (harmonic mean of recall and precision with equal weight). We present the results of exact and partial text span matching. With exact matching, the text span must exactly match the reference annotation. With partial matching, any overlap between the reference standard text span and the concept detected by the system is considered a match. We also assess the classification of the document for patient cancer types. We report the accuracy of how our multi-class model correctly identifies the type of cancer in each patient note.

Evaluation of MPE Methods

Table 2 shows the results measured with individual MPE applications and the voting ensemble method. MetaMap showed moderate performance. Filtering with the list of CUIs of interest allowed for a precision of 94.3%, but a relatively low recall of 79.7% with partial matches. Our dictionary lookup (Rule-based application) reached good performance with a recall of 87.8% and 91.6% with exact and partial matching respectively. It produced higher recall and F_1 score than the CRF and RNN MPE applications.

Table 2– Problem Extraction Result of Each Method

Methods	Exact			Partial		
	P	**R**	**F**	**P**	**R**	**F**
MetaMap (M)	87.4	73.8	80.0	94.3	79.7	86.4
Rules (R)	88.8	87.8	88.3	92.7	91.6	92.2
CRF	90.3	81.7	85.8	95.3	86.3	90.6
CRF+M	91.5	86.8	89.1	95.4	90.5	92.9
CRF+R	93.0	91.8	92.4	95.5	94.2	94.8
CRF+M+R	93.4	91.8	92.6	95.9	94.2	95.0
RNN	86.3	85.3	85.8	91.5	90.5	91.0
RNN+M	89.6	88.7	89.2	93.8	92.9	93.3
RNN+R	92.9	91.4	92.1	95.5	94.0	94.7
RNN+M+R	92.3	**92.7**	92.5	94.7	**95.0**	94.9
Vote	**93.8**	92.1	**92.9**	**96.0**	94.3	**95.1**

P = Precision, R = Recall, and F = F_1 score

CRF with MetaMap (CRF+M) and Rule-based (CRF+R) outputs increased F_1 scores by 3.3 (= 89.1% − 85.8%) and 6.6 (= 92.4% − 85.8%) respectively. Among the four CRF models, the CRF+M+R system obtained the best results with an F_1 score of 92.6% with exact matching and 95.0% with partial matching.

In general, RNN models achieved higher recall than corresponding CRF models. The RNN model with MetaMap and Rule-based outputs (RNN+M+R) produced the best recall at 92.9% and 95.0% with exact and partial matching, respectively.

We evaluated the performance of the ensemble method with voting combining ten MPE application versions. The last row in Table 2 shows the results with a voting threshold of five (i.e. five votes are needed to annotate a concept). This voting ensemble method achieved higher precision than any of the individual classifiers, with 96.0% precision, 94.3% recall, and 95.1% F_1 score with partial matching. Using a paired t-test to compare F_1 scores demonstrated that the voting ensemble method reached significantly higher accuracy than all other approaches at the 95% significance level with exact matching, but not significantly better than the CRF+R, CRF+M+R, and RNN+M+R models with partial matching.

Evaluation of Cancer Type Classification

We examined the impact of the various MPE approaches and how they contributed to cancer type classification. Table 3 shows the results produced with the multi-class models using the combination of all words and medical problem concepts extracted from the clinical notes. For comparison, we also created two additional models with reference standard medical concepts.

Table 3– Cancer Type Classification at Note Level

Features	Accuracy (%)
Words	75.6
Problems (given)	90.2
Problems (system)	88.7
Words + Problems (given)	87.9
Words + Problems (system)	87.3

The model trained with only words (first row in Table 3) achieved 75.6% accuracy. The models with only problem concepts outperformed the models with the combined features from words and problems. The model trained with system detected problems (third row) reached an accuracy of 88.7%. This demonstrates that our MPE task has the potential to efficiently acquire relevant patient information for improved data quality and decision making.

Discussion

Figure 1 shows the recall, precision and F_1 score measured with the voting ensemble method and 10 different MPE application versions. Note that the y-axis scale in each graph does not start at zero to focus on the value ranges of interest.

Partial matching results with voting thresholds ranging from one to ten are presented. The voting threshold had a significant impact on performance. The curves show that the higher the threshold, the higher the precision but the lower the recall. When the voting threshold was set to four or five, the voting ensemble achieved the highest F_1 score at 95.1%.

We analyzed false negatives from the voting ensemble method. Table 4 shows false negative examples with the number of occurrences in the training data. When contextual differences in the words surrounding the medical concept terms were found, they were misclassified, although they often appeared in the training data. We also witnessed some false negatives written in a similar way to medical concept terms that are not our interest. Some abbreviated concepts were missed when they never or rarely appeared in the training data. For example, CVA (cerebrovascular accident) occurred only two times and SVT (supraventricular tachycardia) never appeared in the training data.

Figure 1– Voting Ensemble Method Results

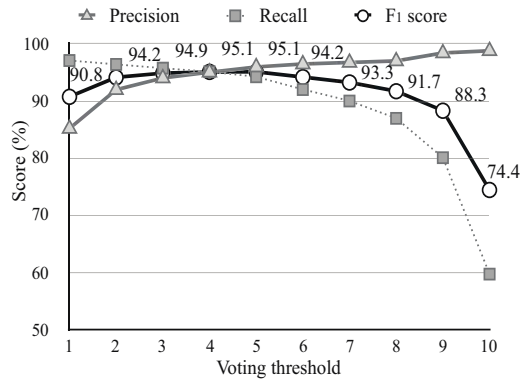


Table 4– False Negative Examples

Occurrences	Examples
More than 50 times	*vomiting, chest pain, JVD*
Less than 5 times	*CVA, N&V*
Unseen	*T2DM, SVT*

Conclusions

We demonstrated that a medical knowledge-base can improve information extraction when labeled data is not sufficient. Both CRF and RNN MPE applications benefited from a list of medical problems compiled from the UMLS Metathesaurus. Our voting ensemble outperformed individual models and its outputs contributed most to classifying patient cancer types.

One avenue for further study would be applying semi-supervised learning that exploits large amounts of unlabeled clinical texts. New models can be trained through the acquisition of clinical notes where there are medical problems of our interest but not mentioned in the labeled data.

Acknowledgments

This research was supported in part by the National Cancer Institute (R41CA180190). We thank Paul Heider for organizing the corpus.

References

[1] A.R. Aronson and F.-M. Lang, An overview of MetaMap: historical perspective and recent advances, *JAMIA* 17 (2010), 229–236.

[2] O. Bodenreider, The unified medical language system (UMLS): integrating biomedical terminology, *Nucleic Acids Res.* 32 (2004), D267–270.

[3] L. Bottou, Online learning and stochastic approximations, *On-line learning in neural networks*, 1998, pp. 9–42.

[4] E. Brill and J. Wu, Classifier Combination for Improved Lexical Disambiguation, *Proceedings of the 36th ACL and 17th ICML*, 1998, pp. 191–195.

[5] S. Brody, R. Navigli, and M. Lapata, Ensemble Methods for Unsupervised WSD, *Proceedings of the 21st COLING and the 44th ACL*, 2006, pp. 97–104.

[6] L.M. Christensen, P.J. Haug, and M. Fiszman, MPLUS: A Probabilistic Medical Language Understanding System, *Proceedings of the ACL-02 Workshop on Natural Language Processing in the Biomedical Domain*, 2002, pp. 29–36.

[7] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, Natural language processing (almost) from scratch, *J Mach Learn Res* 12 (2011), 2493–2537.

[8] K. Crammer and Y. Singer, Ultraconservative online algorithms for multiclass problems, *J Mach Learn Res* 3 (2003), 951–991.

[9] D. Demner-Fushman, W.J. Rogers, and A.R. Aronson, MetaMap Lite: an evaluation of a new Java implementation of MetaMap, *JAMIA* 24 (2017), 841–844.

[10] C. Friedman, P.O. Alderson, J.H. Austin, J.J. Cimino, and S.B. Johnson, A general natural-language text processor for clinical radiology, *JAMIA* 1 (1994), 161–174.

[11] K.W. Fung, C. McDonald, and S. Srinivasan, The UMLS-CORE project: a study of the problem list terminologies used in large healthcare institutions, *JAMIA* 17 (2010), 675–680.

[12] S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural computation* 9 (1997), 1735–1780.

[13] Y. Kim, P. Heider, and S.M. Meystre, Ensemble-based Methods to Improve De-identification of Electronic Health Record Narratives, *AMIA Annu Symp*, 2018, pp. 663–672.

[14] Y. Kim and E. Riloff, Stacked Generalization for Medical Concept Extraction from Clinical Notes, *Proceedings of the BioNLP 2015*, 2015, pp. 61–70.

[15] Y. Kim, E. Riloff, and J.F. Hurdle, A Study of Concept Extraction Across Different Types of Clinical Notes, *AMIA Annual Symposium Proceedings*, 2015, pp. 737–746.

[16] T. Kudo and Y. Matsumoto, Chunking with Support Vector Machines, *Proceedings of the 2nd NAACL*, 2001, pp. 1–8.

[17] J.D. Lafferty, A. McCallum, and F.C.N. Pereira, Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, *Proceedings of the 18th ICML*, 2001, pp. 282–289.

[18] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, Neural Architectures for Named Entity Recognition, *Proceedings of the 2016 Conference of the NAACL: HLT*, 2016, pp. 260–270.

[19] J. Langford, L. Li, and A. Strehl, Vowpal wabbit online learning project, Technical report, http://hunch.net, 2007.

[20] T. Lavergne, O. Cappé, and F.c. Yvon, Practical Very Large Scale CRFs, *Proceedings of the 48th ACL*, 2010, pp. 504–513.

[21] D.A.B. Lindberg, B.L. Humphreys, and A.T. McCray, The unified medical language system, *Yearb. Med. Inform.* (1993), 41–51.

[22] X. Ma and E. Hovy, End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF, *Proceedings of the 54th ACL*, 2016, pp. 1064–1074.

[23] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky, The Stanford CoreNLP Natural Language Processing Toolkit, *Proceedings of 52nd ACL: System Demonstrations*, 2014, pp. 55–60.

[24] A. McCallum, D. Freitag, and F.C. Pereira, Maximum Entropy Markov Models for Information Extraction and Segmentation, *Proceedings of the 17th ICML*, 2000, pp. 591–598.

[25] R. McDonald and F. Pereira, Identifying gene and protein mentions in text using conditional random fields, *BMC. Bioinformatics* 6 (2005), S6.

[26] S.M. Meystre and P.J. Haug, Randomized controlled trial of an automated problem list with improved sensitivity, *Int J Med Inform* 77 (2008), 602–612.

[27] S.M. Meystre, J. Shao, and G. Jones, Automated Dynamic Problem and Allergy Lists for Efficient Electronic Health Record Management, *AMIA Annu Symp*, 2016, p. 1512.

[28] T. Mikolov, K. Chen, G. Corrado, and J. Dean, Efficient estimation of word representations in vector space, *arXiv preprint arXiv:1301.3781* (2013).

[29] N. Reimers and I. Gurevych, Reporting Score Distributions Makes a Difference: Performance Study of LSTM-networks for Sequence Tagging, *Proceedings of the 2017 EMNLP*, 2017, pp. 338–348.

[30] G.K. Savova, J.J. Masanz, P.V. Ogren, J. Zheng, S. Sohn, K.C. Kipper-Schuler, and C.G. Chute, Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications, *JAMIA* 17 (2010), 507–513.

[31] M.A. Tanenblatt, A. Coden, and I.L. Sominsky, The ConceptMapper Approach to Named Entity Recognition, *LREC*, 2010, pp. 546–551.

[32] T. Tieleman and G. Hinton, Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude, *COURSERA: Neural networks for machine learning* 4 (2012), 26–31.

Address for correspondence

Youngjun Kim
Translational Biomedical Informatics
Medical University of South Carolina,
135 Cannon Street, 1st Floor,
Charleston, SC 29425, USA;
kimy@musc.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190211

Recurrent Deep Network Models for Clinical NLP Tasks: Use Case with Sentence Boundary Disambiguation

Benjamin C. Knoll[a], Elizabeth A. Lindemann[b], Arian L. Albert[a], Genevieve B. Melton[a,b], Serguei V.S. Pakhomov[a,c]

[a] *Institute for Health Informatics, University of Minnesota, Minneapolis, Minnesota, USA*
[b] *Department of Surgery, University of Minnesota, Minneapolis, Minnesota, USA*
[c] *College of Pharmacy, University of Minnesota, Minneapolis, Minnesota, USA*

Abstract

Although a number of foundational natural language processing (NLP) tasks like text segmentation are considered a simple problem in the general English domain dominated by well-formed text, complexities of clinical documentation lead to poor performance of existing solutions designed for the general English domain. We present an alternative solution that relies on a convolutional neural network layer followed by a bidirectional long short-term memory layer (CNN-Bi-LSTM) for the task of sentence boundary disambiguation and describe an ensemble approach for domain adaptation using two training corpora. Implementations using the Keras neural-networks API are available at https://github.com/NLPIE/clinical-sentences.

Keywords:

Natural Language Processing, Machine Learning, Neural Networks (Computer)

Introduction

In contrast to general English, clinical notes have significant differences in structure and content. For instance, clinical text often contains units of thought that fit the technical definition of sentences that are not terminated by the standard sentence boundary symbols or any symbols in many cases. Structures such as labels, section headers, text arranged in tables, and lists are examples of clinical text that do not follow general English rules for sentence termination. Furthermore, clinical text contains a disproportionately high number of acronyms, abbreviations, and ordinal numbers frequently decorated with punctuation symbols and containing variable capitalization. Segmentation errors caused by these ambiguities are magnified in downstream processing.

Previous research has shown that transfer learning in deep networks can improve generalization to tasks of related problems with small data sets [1]. Ensemble methods that engage in meta-learning through weighted voting models such as boosting, bagging, and stacking also reduce the generalization error over standard models [2]. We utilize transfer learning both in the use of word embeddings, and in our method for domain adaptation of models trained on one corpus to a different, but related corpus of clinical text.

Sentence boundary disambiguation (SBD), also known as sentence segmentation or sentence boundary detection, is a well-understood and explored problem in the domain of general English text. In well-formed general English text, most sentences are terminated by sentence boundary symbols. Ambiguities caused by acronyms, abbreviations, quotations, and ordinal numbers are handled by further rules, or by statistical methods such as maximum entropy classification of boundaries. Using these methods on general English text results in accuracy performance above 95% [3]. Relying on sentence boundary symbols for SBD on the entire text of a clinical document leads to errors in detecting non-terminated sentence boundaries [4]. Although deep learning is the prevailing approach for many machine learning problems, it remains underutilized in clinical applications, and the generalizability of clinical applications using current approaches is limited [5].

In this paper we report on applying deep neural network methods with the use case of sequence labeling to the SBD problem on the entire text of clinical notes with no preprocessing or cleaning. We show that an architecture consisting of word embeddings enriched with character information run through a bi-LSTM have high accuracy in detecting sentence boundaries. We also explore the generalization problem of using a trained model for SBD on previously unseen text using several implementations, including combining training data, resuming training with data from the new corpus, and a stacking method where the hidden layer results of two models trained separately on both corpora are summed before prediction using a shared prediction layer. We show that the stacking method has the lowest generalization error with 96% F1 score for beginning of sentence tags.

Methods

Sentence Segmentation

SBD is often the first step in solving any problem using natural language processing (NLP). Availability of sentence boundaries is necessary both for many general language tasks, such as part-of-speech tagging and parsing, and for domain-specific analytical tasks such as document classification. Errors in sentence detection tend to propagate to many other areas in a system making sentence accuracy critical for any downstream tasks in a text analysis system.

In Table 1, some examples of text from Fairview Medical Services notes where sentences are not terminated with sentence boundary symbols are shown.

Table 1 – Examples of sentences without termination

Text
RECOMMENDATIONS FOR MDs/PROVIDERS TO ORDER:
Recommendations already ordered by Registered Dietitian (RD): Calorie counts reordered
Diet: dysphagia diet level 2 mechanical, thin liquids, magic cup between meals, Nepro between meals
Pt reported his appetite is getting better, he likes the supplements
(+) No chance of pregnancy C-spine cleared: N/A, no H/O Chronic pain,no other significant disability

There are several existing commonly used implementations of SBD that rely on or expect sentence boundary symbols and perform poorly in their absence. Stanford CoreNLP [6] provides a rule-based algorithm, which makes decisions based on the results of a tokenizer to disambiguate whether sentence boundary symbols indicate sentence splits. Natural Language Toolkit (NLTK) [7] implements SBD using a method that combines rules for sentence boundaries with an unsupervised algorithm for the detection of acronyms and abbreviations, a common source of errors in SBD [8]. The Apache OpenNLP toolkit [9] and Apache cTAKES [10] provide SBD based on the maximum entropy method described in Reynar, et. al [3]. Previous evaluations have looked at the performance of SBD and have noted the difficulty of the task in the domain of clinical notes and have noted the performance issues on non-terminated sentences [4].

Approaches to the NLP problem of sequence tagging are well-suited for the SBD problem—which can be expressed as a tagging task where words beginning sentences are tagged 'B' and words internal to sentences are tagged 'I'. The architecture for sequence tagging involving recurrent neural networks (RNNs) has shown good results when applied to the tasks of part-of-speech (PoS) tagging and named entity recognition (NER) [11], and improvements were shown when character information is combined with word embeddings via a convolutional neural network (CNN) [12].

Dataset

Source Corpora

We created two source corpora for the sentence detection task. The first dataset was drawn from the MIMIC-III (Medical Information Mart for Intensive Care) corpus [13], which is a de-identified corpus of notes associated with 40,000 intensive care unit patients at the Beth Israel Deaconess Medical Center between 2001 and 2012. Our MIMIC corpus consisted of 749 randomly sampled notes. A second, target dataset was drawn from the Fairview Health Services (FV) EHR system. We used a stratified sampling strategy, in which we created batches of 56 notes made up of 16 inpatient notes and 40 outpatient notes. The inpatient notes were selected proportional to the distribution of note type (Table 2) and the outpatient notes were selected proportional to the distribution of department (Table 3). A total of 952 notes from FV were used, 17 complete batches. MIMIC notes were in plaintext while FV notes were converted from RTF to plaintext using the BioMedICUS system [14]. The MIMIC corpus used in this study contains a total of 315,797 tokens, and the FV corpus contained 415,112 tokens.

Table 2 – Inpatient note types per FV batch

Note Type	Number
Progress Note	3
Plan of Care	3
ED Notes	2
8 other note types	1 each

Table 3 – Outpatient note departments per FV batch

Note Type	Number
Family Medicine	5
Internal Medicine	4
Pediatrics	3
Obstetrics and Gynecology	3
Hematology and Oncology	3
Urgent Care	3
Physical Therapy	3
Cardiovascular Disease	3
13 other departments	1 each

Manual Annotation of Sentences

The manual annotation of sentences was performed in the BRAT Rapid Annotation Tool [15] by a pair of trained annotators. Annotators were instructed to label all complete thoughts, section headers, item labels, list items, and fragments using a "Sentence" annotation. For any data that was not groupable into sentence-like units (e.g., purely numeric tables, lists of laboratory data, lines of vital signs measures, and metadata tables such as those in header information), or for any other areas of text for which annotators had low confidence in their ability to correctly label sentences, annotators were instructed to use an "Unsure" annotation. After sentences were manually annotated, the documents were tokenized and converted to tagged sequences where 'B' was applied to the first token in every sentence, 'I' was applied to the rest of the tokens in the sentence, and 'O' was applied to all the tokens in the "Unsure" category.

Cross-validation structure

To evaluate generalization error, a cross-validation structure was used where 100% of the MIMIC data was used for cross-validation with 80% as a training split and 20% as a validation split; for the FV data 50% was used for cross-validation (again using an 80-20 training-validation split) and 50% was held out as an unseen test corpus. For architecture and hyper-parameter tuning, the MIMIC validation split was used. The FV cross-validation set was used for training models alone or augmenting MIMIC-trained models. During training, the validation data was used to provide validation loss as an estimation of generalization error to determine when the model has stopped improving and training could be halted.

Model Architecture

Words were tokenized according to rules that split whenever any whitespace, any symbols, or any digits are encountered. Words were represented using a 300-dimension word embedding trained using the Facebook fastText software package [16] on the entire MIMIC-III corpus preprocessed to replace any symbols with spaces, to replace digits with their English names in separate words, i.e. "1.23" to "one two three", and to lowercase all letters. These word embeddings are enriched by summing with the results of a convolutional neural network (CNN) on 30-dimension character embeddings which are learned during training on the SBD tagging task. The CNN is made up of one convolutional layer with 300 filters, each looking at the sequences of the embeddings of four characters, followed by global max pooling. The results of the CNN func-

tion as an adjustment vector to the original word vector for the sentence tagging task. The input of the character CNN is all characters of the word (including symbols and whitespace) along with a context of up to seven characters between the previous word and the word; and up to seven characters between the word and the next word. Special characters were inserted for the end of the previous word, the beginning of the next word, the beginning and end of the word, as well as the beginning and end of the document if those fell into the context. Using all the original characters, including whitespace, allows structural information about the document's formatting to be used for SBD decisions. After the word representation is constructed the results are batch-normalized before passing to the next layer. The architecture of the word representation layer is shown in Figure 1.

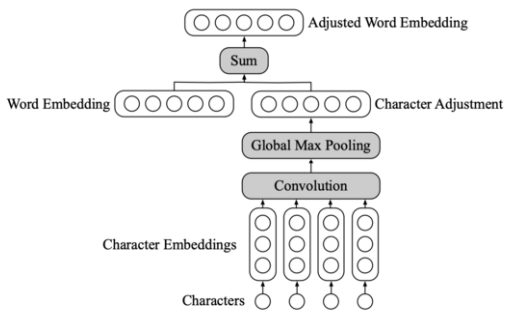


Figure 1 – Word Representation Layer

To encode contextual word representations, we use a bi-directional long short-term memory (LSTM) layer. LSTM units are iteratively run on time-series data (sequences of words in the case of text), maintaining an internal cell state as it moves from one input to the next. LSTMs are optimized during training to learn what information is important to remember from previous inputs to the cell. A LSTM layer is parameterized by the number of LSTM units, each contributing one output dimension. In a bi-directional LSTM layer, the inputs are run both ways through the layer, with one set of LSTM units responsible for seeing the data in order and one set responsible for seeing the data in reverse. Dropout and recurrent dropout [17] were used to provide regularization of the learned weights and prevent overfitting. The results of the bi-LSTM layer are batch-normalized before being passed to the inference layer. In Figure 2, computation of a bi-LSTM on time series data is shown, each node labeled LSTM-F and LSTM-R is the same set of units at different points in the time series, and the lines drawn between nodes represent the propagation of internal memory states to the next item in the time series. The outputs from the forward and backward LSTMs are concatenated to a single contextual word representation, an embedding of the word and surrounding words.

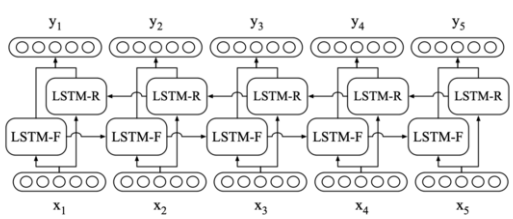


Figure 2 – Bi-directional LSTM

After the bi-directional LSTM layer, a sigmoid-activated dense NN prediction layer is used on each contextual word embedding to output the log-probability that the word is the beginning of a sentence. Lasso or L1-norm regularization was used on the weights of the prediction layer to prevent overfitting. The complete graph of our architecture is shown in Figure 3.

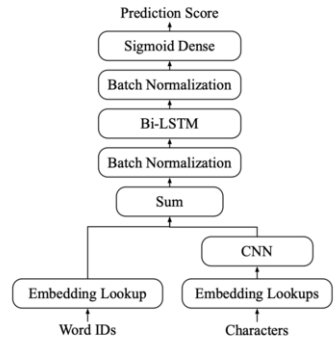


Figure 3 – Complete Model Graph

Domain Adaptation

In addition to using models trained on each individual corpus, we evaluated three methods for domain adaptation of models trained on MIMIC to the FV hold-out test set. First, we looked at merging the cross-validation data from both corpora. Second, we looked at resuming training of the network trained on the MIMIC cross-validation data with the FV cross-validation data. Third, we looked at using an ensemble stacking method for transfer learning where the hidden-layer contextual word representations of the network trained on MIMIC were summed with the contextual word representations of a new network before the sigmoid dense NN prediction layer. In this architecture the output of the second network functions as corrections to the first network for the FV training data. This stacked network architecture is shown in Figure 4.

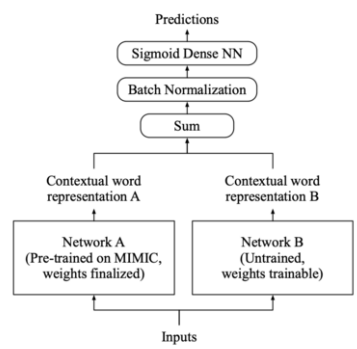


Figure 4 – Ensemble of Two Networks

Training

Based on the results of tuning using CV on the MIMIC corpus, we selected the gradient descent variant ADAM (Adaptive Moment Estimation) [18] as the optimizer of network weights. During training, only models that were improvements on validation loss were saved, and after 5 epochs with no im-

provement training was terminated. Binary cross-entropy loss was used for training, and loss values were weighted by the ratio between the target tag probability and an equal distribution of tags, shown in the equation in Figure 5. Mini-batching was used for training, sequences of 32 words were batched into groups of 32 for gradient optimization.

$$weight_{class} = \frac{samples_{total}}{n_{classes} \times samples_{class}}$$

Figure 5 – Weighting of classes

Results

Manually Annotated Corpora

On an overlap of 100 MIMIC notes annotated by both annotators, ignoring "Unsure" annotations, the Cohen's kappa of Sentence annotations was computed as 0.957 using the irr library in R version 3.4.4. The agreement between annotators on "Unsure" annotations was 0.646. After conversion, on the subset of 100 notes labeled by both annotators Cohen's kappa was 0.71 for all tags and 0.95 after excluding items labeled as 'O' by either annotator. Tables 4 and 5 describe the distribution of 'B', 'I', and 'O' tags after this conversion.

Table 4 – Distribution of Tags in MIMIC

Tag	Count	Percentage
B	23,648	7.5%
I	200,272	63.4%
O	91,877	29.1%

Table 5 – Distribution of Tags in FV

Tag	Count	Percentage
B	43,636	10.5%
I	336,018	80.9%
O	35,458	8.5%

In both source corpora, sentences not terminated by sentence boundary symbols are highly prevalent. Section headers and text labels were common and often ended by the colon sentence boundary symbol. Table 6 shows the quantity of sentences terminated by each symbol.

Table 6 – Sentence Termination Type

Type	MIMIC	FV
Period	12,698 (53.7%)	13,619 (31.2%)
Exclamation Point	4	19
Question Mark	24 (0.1%)	261 (0.6%)
Semi-colon	48 (0.2%)	10
Colon	4,855 (20.5%)	6,180 (14.2%)
Quotation	4	58 (0.1%)
No symbol	6,018 (25.4%)	23,506 (53.8%)

Evaluation of SBD Approaches

For our evaluation we ignored all tags that were labeled as 'O' both during training and during evaluation. Thus, the recall, precision, and F1 for 'B' and 'I' are symmetric, every false positive 'B' is a false negative 'I' and every false negative 'B' is a false positive 'I'. We've reported only the 'B' scores as they are directly proportional to the overall accuracy of detected sentences. The best architecture and hyper-parameter tuned models from cross validation achieved 98.6% F1 on both 'B' and 'I' tags on the MIMIC validation set and 99.2% F1 on 'B' and 'I' tags in the FV validation set.

We evaluated our implementation of SBD against the 50% FV hold-out data set (476 notes). In addition to the architecture described above, we evaluated a maximum entropy / logistic regression classifier (listed as LR) using optimization of a sigmoid-activated dense NN layer on an input of 7 characters before, at the beginning, at the end, and following every word. This is an approach like, but not as tuned as individual implementations of maximum entropy SBD. The primary metrics used for evaluation were the precision, recall, and F1-score for the beginning of sentence class tag.

In addition to models trained on MIMIC and FV individually, we evaluated three methods to test generalizability against the FV test corpus. The models trained solely against one corpus are listed as "MIMIC" and "FV." The results of a model trained on a both corpora's cross-validation set combined are listed as "MIMIC+FV." The continued training of one network is listed as "MIMIC then FV" and the ensemble model is listed as "Ensemble." Results of these evaluations are show in Table 7 with best results in bold.

Table 7 – 'B' Tag Accuracy Against FV Hold-out

Method	Precision	Recall	F1
LR-MIMIC	0.511	0.840	0.636
LR-FV	0.650	0.948	0.771
MIMIC	0.829	0.971	0.895
FV	0.923	0.991	0.956
MIMIC+FV	0.919	**0.995**	0.956
MIMIC then FV	0.910	0.992	0.949
Ensemble	**0.933**	0.989	**0.96**

Discussion

The complexity, grammatic idiosyncrasies, and domain variability of clinical text lead to significant hurdles in designing and training generalizable models for NLP tasks. This common challenge necessitates the use high-capacity, complex machine learning models such as the deep neural network approach described here. Leveraging transfer learning and domain adaptation, such as the ensemble method used here, is an important tool to regularize models created from smaller domain-specific corpora with data from external corpora. In the SBD task, the clinical-specific structuring of sentences in our target corpus led us to applying these approaches.

In all experiments, recall was higher than precision, which can be explained by the class weighting structure. Models are penalized much higher for missing a 'B' tag than for replacing an 'I' with a 'B' tag, leading to models being overeager in splitting sentences. Adaptation of models trained on one corpus to another corpus of text show clear but relatively small losses in performance—we can see that the MIMIC trained model has an F1-score approximately 0.06 lower than the FV trained model.

The ensemble method slightly improves the F1 score against the other domain adaptation methods, increasing precision at a slight cost to recall. The ensemble method was the best performing overall with a 0.96 F1 score on B-tags. This F1-score is on par with the 0.957 Cohen's Kappa inter-rater agreement on the MIMIC data that represents a "ceiling" for performance of SBD algorithms. As shown in the information about our corpora, the FV corpus has different distribution of sentence 'B', 'I', and 'O' tags than the MIMIC corpus, demonstrating that these methods are successful in adapting to a corpus with significant syntactic differences.

There is a loss in performance in the continued training method versus the MIMIC+FV method. In this method, weights may not be able to recover from sub-optimal positions for predicting FV data from the training on MIMIC data. The gradient descent may not be able to find a path from the current position of the weights to the more optimal position of weights found by the FV-only and the MIMIC+FV trained models.

Conclusions

Our study shows that there are improvements in SBD using deep networks over using traditional classification methods, and that these networks can perform well even against different corpora and against corpora with large proportions of sentences that are not terminated by sentence boundary symbols. We've also shown that transfer learning approaches for domain adaptation such as the ensemble model have lower generalization error than combining training sets or continued training.

Further Work

The generalization performance gains from using a two-network ensemble indicate further exploration into meta-learning and ensemble approaches may be fruitful. Furthermore, usage of this or other transfer learning ensemble methods with general-domain English corpora included as training data for base models remains an unexplored possibility.

The accuracy of automatically detected sentences can have substantial consequences on downstream components in a processing pipeline. These benefits are significant on face but have not been formally quantified, and these effects are a potential target for future research.

Acknowledgements

We would like to acknowledge Michael Hietpas, one of the annotators for our data. This research was supported in part by NIH/NCATS UL1TR002494, NIH/NCATS U01TR002062, NIH/NIGMS R01GM102282, and AHRQ R01HS022085. The content is solely the responsibility of the authors and does not necessary represent the official views of the National Institutes of Health or the Agency for Healthcare Research and Quality.

References

[1] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition, *CoRR*. **abs/1310.1** (2013). http://arxiv.org/abs/1310.1531.

[2] E. Bauer, and R. Kohavi, An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants, *Machine Learning* **36** (1999), 105–139. doi:10.1023/A:1007515423169.

[3] J.C. Reynar, and A. Ratnaparkhi, A maximum entropy approach to identifying sentence boundaries, in: Proceedings of the Fifth Conference on Applied Natural Language Processing, 1997, 16–19.

[4] D. Griffis, C. Shivade, E. Fosler-Lussier, and A.M. Lai, A Quantitative and Qualitative Evaluation of Sentence Boundary Detection for the Clinical Domain., *AMIA Joint Summits on Translational Science Proceedings. AMIA Joint Summits on Translational Science*. **2016** (2016), 88–97. http://www.ncbi.nlm.nih.gov/pubmed/27570656 (accessed November 15, 2018).

[5] Y. Wang, L. Wang, M. Rastegar-Mojarad, S. Moon, F. Shen, N. Afzal, S. Liu, Y. Zeng, S. Mehrabi, S. Sohn, and H. Liu, Clinical information extraction applications: A literature review, *Journal of Biomedical Informatics* **77** (2018), 34–49. doi:10.1016/J.JBI.2017.11.011.

[6] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky, The Stanford CoreNLP Natural Language Processing Toolkit, *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. (2014) 55–60. doi:10.3115/v1/P14-5010.

[7] S. Bird, E. Loper, and E. Klein, Natural Language Processing with Python, O'Reilly Media Inc, 2009.

[8] T. Kiss, and J. Strunk, Unsupervised Multilingual Sentence Boundary Detection, *Computational Linguistics* **32** (2006), 485–525. doi:10.1162/coli.2006.32.4.485.

[9] Apache OpenNLP, (n.d.). https://opennlp.apache.org/ (accessed November 15, 2018).

[10] Apache cTAKES™ - clinical Text Analysis Knowledge Extraction System, (n.d.). http://ctakes.apache.org/ (accessed November 15, 2018).

[11] Z. Huang, W. Xu, and K. Yu, Bidirectional LSTM-CRF Models for Sequence Tagging, *CoRR*. **abs/1508.0** (2015). doi:10.18653/v1/P16-1101.

[12] X. Ma, and E.H. Hovy, End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF, *CoRR*. **abs/1603.0** (2016). http://arxiv.org/abs/1603.01354.

[13] A.E.W. Johnson, T.J. Pollard, L. Shen, L.H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R.G. Mark, MIMIC-III, a freely accessible critical care database, *Scientific Data*. **3** (2016) 160035. doi:10.1038/sdata.2016.35.

[14] BioMedICUS, (n.d.). https://nlpie.github.io/biomedicus/ (accessed November 17, 2018).

[15] brat rapid annotation tool, (n.d.). http://brat.nlplab.org/ (accessed November 15, 2018).

[16] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, Enriching Word Vectors with Subword Information, *CoRR*. **abs/1607.0** (2016). http://arxiv.org/abs/1607.04606.

[17] S. Semeniuta, A. Severyn, and E. Barth, Recurrent Dropout without Memory Loss, *CoRR*. **abs/1603.0** (2016). http://arxiv.org/abs/1603.05118.

[18] D.P. Kingma, and J. Ba, Adam: A Method for Stochastic Optimization, *CoRR*. **abs/1412.6** (2014). http://arxiv.org/abs/1412.6980.

Address for correspondence

Serguei Pakhomov, PhD
Department of Pharmaceutical Care and Health Systems
College of Pharmacy
University of Minnesota
5-130 WDH 1332A
308 Harvard St SE
Minneapolis, MN 55455
Pakh0002@umn.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190212

Annotating German Clinical Documents for De-Identification

Tobias Kolditz,[a] Christina Lohr,[a] Johannes Hellrich,[a] Luise Modersohn,[a]
Boris Betz,[b] Michael Kiehntopf,[b] Udo Hahn[a]

[a] Jena University Language & Information Engineering (JULIE) Lab, Friedrich Schiller University Jena, Jena, Germany
[b] Institute of Clinical Chemistry and Laboratory Diagnostics, Jena University Hospital, Jena, Germany

Abstract

We devised annotation guidelines for the de-identification of German clinical documents and assembled a corpus of 1,106 discharge summaries and transfer letters with 44K annotated protected health information (PHI) items. After three iteration rounds, our annotation team finally reached an inter-annotator agreement of 0.96 on the instance level and 0.97 on the token level of annotation (averaged pair-wise F1 score). To establish a baseline for automatic de-identification on our corpus, we trained a recurrent neural network (RNN) and achieved F1 scores greater than 0.9 on most major PHI categories.

Keywords:

Data Anonymization, Confidentiality, Natural Language Processing

Introduction

Natural Language Processing (NLP) research, in general, has greatly benefited from sharing software resources (tools) and datasets (corpora, lexical repositories, etc.), as well as organizing challenge competitions to evaluate this research infrastructure [1]. This role model has also been adopted by biomedical NLP [2], including clinical NLP [3]. The i2b2 series of competitions[1] mark the cornerstone for the development of numerous clinical datasets[2] and the offspring of specialized software solutions[3] that shape the current state-of-the-art in clinical NLP–yet, for the English clinical language only. Still, almost all of these (de-identified) datasets are accessible to the international research community under the legal conditions of Data Use Agreements (DUA) building on less restrictive data privacy law in the US than in many other parts of the world.

National privacy law in Europe and the General Data Protection Regulation (EU) 2016/679 pose a number of obstacles on the way to extramural sharing clinical corpora. Yet, the minimal prerequisite for all shared clinical corpora is de-identification, i.e., the removal of any patient-identifying information from clinical data. In Germany, one of the top representatives of a particularly restrictive data privacy legislation, the heavily funded Medical Informatics research Initative (MII)[4] was launched recently whose goal is, among others, to empower interoperability of and transparent access to data (including clinical document corpora) and software resources within and across different clinical sites at the national level, in conformance with valid data privacy regulations. Our team heads the NLP acitivities within one of the four funded consortia, *Smart Medical Information Technology for Healthcare* (SMITH)[5] [4]. One of the early outcomes of these efforts is 3000PA [5], a German-language clinical corpus made of 3,000 discharge summaries and transfer notes from three national university hospitals located in Jena, Leipzig and Aachen. Moreover, we are also involved in developing automated workflows for *Healthcare Integrated Biobanking* (STAKI^{2}B^2) where de-identification of clinical corpora is a prerequisite when it comes to workflow validation across different biobanks. In this paper, we report on the de-identification of a list of protected health information (PHI) categories reflecting US law criteria (Health Insurance Portability and Accountability Act (HIPAA)),[6] and their adaptation to national requirements and clinical particularities in German hospitals.

Related Work

Many systems for automatic de-identification have been developed, especially for the English language, but only a few large de-identification gold-standard corpora are reported in the literature–even fewer are publicly available, typically protected by DUAs. In this paper, we report on our efforts in creating a de-identification gold-standard corpus for the German language. To put this work in perspective, we focus on previous work on the creation of corpora annotated with PHI categories.

For clinical English, several corpora have been compiled and annotated for the purpose of de-identification. Building on manual de-identification work by Douglass et al. [6; 7], Neamatullah et al. [8] assembled an automatically de-identified gold-standard corpus of 2,434 nursing progress notes of patients from intensive care units collected in the MIMIC II project [9]. The final corpus consists of almost 340K tokens, contains 1,779 instances of PHI and is available on PhysioNet.[7]

Two different gold-standard corpora for de-identification were assembled for the i2b2 de-identification challenges in 2006 and 2014. For the first challenge [10], 889 medical discharge summaries consisting of almost 550K tokens were annotated semi-automatically, with almost 20K instances of PHI. For the second challenge [11], a corpus of 1,304 longitudinal medical records consisting of more than 800K tokens was annotated, with more than 28K PHI items. The authors report an averaged token-based F1 score of 0.927 between the annotators and the gold standard. For the CEGS N-GRID shared task on de-identification [12], a corpus of 1,000 psychiatric intake records

[1] https://www.i2b2.org/
[2] Accessible via http://dbmi.hms.harvard.edu/programs/health-care-data-science-program/clinical-nlp-research-data-sets
[3] Most notably, the cTAKES system: http://ctakes.apache.org/
[4] http://www.medizininformatik-initiative.de/
[5] http://www.smith.care/?lang=en
[6] https://www.hhs.gov/hipaa/index.html
[7] https://www.physionet.org/

consisting of more than 1,8M tokens was annotated, with more than 34K PHI instances. Token-level agreement of 0.91 F1 score between the annotators and the gold standard was slightly worse than for the i2b2 annotation campaign in 2014. After surrogate replacement, all three challenge corpora were made publicly available via DUAs. The 2006 i2b2 corpus (training set) and a carefully genre-balanced corpus from the Veterans Health Administration (800 documents, with 5.5K PHI items; test set) formed the basis for an "out-of-the-box" evaluation of five operational text de-identification systems [13].

Deléger et al. [14; 15] annotated 3,503 clinical notes of more than 22 different types (more than 1M tokens) with about 30K PHI. They report an inter-annotator agreement of 0.92 F1 score. In a later publication [16] they modify the corpus such that it can be shared for research purposes. Similar annotated corpora (albeit not always publicly available) exist for a few other languages, such as French [17] and Swedish [18; 19]; synthetic (fictional) data were created for Japanese [20].

Unfortunately, no shareable de-identified corpus exists for the German language–in general, there has been only little work on the de-identification of German clinical data. Seuss et al. evaluate a commercial de-identification tool on German clinical reports [21] and to this end assemble a large semi-automatically annotated corpus. They manually correct the output of the tool on 1,400 documents consisting of 5M tokens, identifying more than 23K PHI items (in nine categories pre-defined by the tool), and report high recall values for the system. However, the manual annotation process is tightly coupled with the tool and there is no evaluation of inter-annotator agreement between the three human annotators. Recently, Richter-Pechanski et al. [22] presented a de-identification tool that combines an off-the-shelf named entity recognizer (German Stanford NER tool, retrained on non-medical data), regular expressions and gazetteers with spelling variation detection. For evaluation purposes, they annotated a set of 15 notes (~ 14K tokens) with ten different categories, identifying 680 PHI entities overall.

Manual Annotation

In this section, we describe in detail the manual annotation campaign we conducted: our data, annotation guidelines, evaluation measures, the rationale for several annotation iterations, and the final annotated corpus. All annotations were created using the *Brat Rapid Annotation Tool* (BRAT) [23].

Data

Annotations were based on the Jena slice of the 3000PA corpus [5], a collection of 1,106 discharge summaries, short summaries and transfer letters. All documents were extracted from EPRs of deceased patients (due to privacy concerns) who were treated in either internistic or ICU units for at least five days between 2010 and 2015. The extraction included a conversion from a proprietary data format to plain text (for details see [24]). Our work on this data was approved by the local ethics committee (4639-12/15) and the Data Protection Officer of the Jena University Hospital.

Annotation Guidelines

Our annotation guidelines are based on the 18 PHI categories defined in HIPAA and subsequent manual and automatic annotations of these PHI types, most notably as part of the i2b2 de-identification challenges [10-12; 25]. We iteratively developed and updated our guidelines in three preliminary annotation rounds, adjusting them to our data, common clinical requirements and the particularities of the German language. The final guidelines define eight broad categories, three of which are further divided into more specific annotation types:

- **Age:** age of patient or relative (any age)
- **Contact**: URL, IP address, email, phone or fax number
- **Date:** any date (excluding single days of the week and times of the day)
  - **Birthdate:** date of birth
- **ID:** any ID or code (patient id, medical record number, codes with unknown semantics that might be PHI)
  - **Typist:** shorthand symbols for medical typists
- **Location:** place names (mainly addresses, including street, house number, zip code, city, district, state)
- **MedicalUnit:** medical units (hospital names, names of hospital departments and ambulant medical units)
- **Person:** names of persons
  - **Patient:** patient names
  - **Relative:** names of relatives
  - **Staff:** names of medical staff
- **Other:** any remaining PHI

Each category is also an annotation type by default; thus, we come up with 13 annotation types. Annotators were required to select only the most specific applicable type for each entity. For instance, if they found the name of a physician, they were instructed to assign it the type *Staff*, not *Person*. Only if the context was unclear about whether a name belongs to a patient, a relative or a member of staff, or if the person denoted by a name belongs to neither of these three groups, the annotators were instructed to assign it the generic *Person* type.

Our scheme collapses several of the HIPAA-defined types of PHI because they are infrequent or indistinguishable in our data: *ID* comprises social security, medical record, account, certificate/license, patient-related serial numbers, and any numbers or codes that may directly or indirectly identify a patient or member of staff. Since we work with plain texts, we dropped the category for photographic images. On the other hand, we introduced some more fine-grained distinctions, e.g., between names of patients, relatives and staff. The *Typist* category was introduced after the second pre-iteration (see below) to resolve systematic disagreement among annotators.

Evaluation Measures

In the following, we report inter-annotator agreement (IAA) as pair-wise averaged F1 score [26], both on instance level and on token level. An instance is a single annotation, possibly spanning multiple tokens, e.g., "*Jane Smith"* denotes one instance of a patient's name comprised of two tokens. Tokens are based on the tokenization script of the TreeTagger[8] with minor post-processing. To obtain token-based agreement, we annotate each token overlapping with an annotation with the annotation's type, sub-token annotations are expanded to full tokens. Due to missing spaces, we got two annotation types per token in 13 cases–we kept both annotation types for each of these tokens when calculating agreement, but used only one as input to the classifier described in the second part of this paper.

Annotation Iterations

We ran three preliminary annotation iterations (summarized in Table 1) before we started the final annotation project. Iterated annotation rounds for building de-identification gold standard corpora have also been suggested by Browne et al. [27] as a strategy to cope with the complexities of this task that are

[8] http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/

meticulously described in their paper. The number of annotations in this table includes redundant annotations from different annotators on the same documents. In the first iteration, eight medical students and two physicians worked on the full set of 1,106 documents, covering all protected health information with generic annotations according to a first set of guidelines based on the list of PHI items defined in HIPAA and the i2b2 de-identification challenges [11; 25]. We automatically pre-annotated dates with a regular expression covering standard numerical date representations. Annotators at this stage could focus on finding PHI items in the text and were not supposed to categorize them. These first annotations were intended to serve as pre-annotations for later iterations (ensuring that each document was checked for false negatives at least twice) and formed the empirical basis for a refined set of annotation guidelines. After a qualitative analysis of the results, we further specified the extent of each annotation type, defined a list of 12 types of PHI (the types listed in the previous section, excluding *Typist*) and added illustrative examples for each category.

Table 1 – Annotation iteration setups and results

Annotation	Pre-1	Pre-2	Pre-3	Final
# documents	1,106	25	12	1,106
# agr. docs	0	25	12	50
# annotators	10	4	4	5
# ann. types	1	12	13	13
# annotations	40,664	4,347	2,556	51,814
Avg. F1 (token)	-	0.91	0.96	0.97
Avg. F1 (inst.)	-	0.87	0.92	0.96

Using simple patterns and keywords, the generic annotations from the first iteration were automatically categorized according to the new guidelines. These data served as pre-annotations for the second iteration in which four medical students each worked on the same set of 25 pre-annotated documents. Their task was to identify missing PHI items and correct both extent and type of the existing annotations. They achieved an inter-annotator agreement of 0.87 ($\sigma = 0.02$) on instance level and 0.91 ($\sigma = 0.02$) on token level.

To further train our annotators and achieve even higher agreement values, a third pre-iteration was conducted with four annotators (one of the previous annotators had to leave the team temporarily and was replaced by a new annotator). We defined an additional category, *Typist*, for shorthand symbols identifying medical typists to resolve the ambiguity between *ID* and *Person*[9] and added more examples to the annotation guidelines to resolve cases of systematic disagreement among annotators. Each annotator was provided with six documents with annotations he or she had created in the previous iteration[10] and six documents with semi-automatically generated pre-annotations. Inter-annotator agreement improved on both the old and the new documents. For the full set of documents, the averaged pair-wise F1 score reached 0.92 ($\sigma = 0.04$) on instance level and 0.96 ($\sigma = 0.02$) on token level.

For the main annotation, five annotators worked on the 1,106 Jena documents of 3000PA. A subset of 50 documents (cf. Table 1, '# agr. docs') was annotated by all five annotators and served as the basis for the computation of IAA which improved further to 0.97 ($\sigma = 0.01$) on token level and 0.96 ($\sigma = 0.01$) on instance level. We also noticed an increase in IAA within the final annotation hinting towards a training effect over time.

Final Corpus

The final corpus consists of roughly 1.4M tokens for which our annotators created more than 44K annotations, excluding redundant agreement annotations.[11] Table 2 gives an overview of instance and token frequencies by annotation type for the entire corpus, as well as instance- and token-level agreement by annotation type on 50 multiply annotated agreement documents. Looking at the instance frequencies, we notice that almost half of the PHI items in our corpus are dates, followed by medical units, locations and staff names, each accounting for more than ten percent of all annotations. The large share of dates is partially a consequence of our broad definition of *Date* which only excludes days of the week without month or year and times of the day. At the other end of the frequency spectrum, we find the generic types *Person* and *Relative*. The fact that only few instances fall under the *Person* category indicates that most names could be assigned unambiguously to either *Patient, Relative* or *Staff*. The low frequency of *Relative* is probably an artifact of our data sampling criteria–most of the deceased patients were elderly people for which family history is less important, hence names of relatives occur rarely. The frequency distribution of tokens mirrors that of instances and gives information about the average length of each annotation type. The share of *Date*s is lower as they usually consist of only one token, whereas instances of the types *Location*, *MedicalUnit* and *Staff* often span multiple tokens, raising their relative frequency at the token level.

Table 2 – Number of instances and tokens per annotation type in the final corpus. Instance-level and token-level agreement (Avg. F1) on 50 multiply annotated agreement documents.

Category	Type	Instances		Tokens	
		Frequency	Avg. F1	Frequency	Avg. F1
Age	Age	498	1.00	500	1.00
Contact	Contact	613	0.97	2,009	0.98
Date	Date	20,603	0.98	24,277	0.99
	Birthdate	1,103	1.00	1,103	1.00
ID	ID	398	0.81	424	0.82
	Typist	655	0.86	1,418	0.93
Location	Location	5,429	0.98	11,286	0.99
MedicalUnit	MedicalUnit	6,189	0.90	12,499	0.95
Person	Person	14	-	23	-
	Patient	3,180	0.99	5,167	1.00
	Relative	36	0.80	62	0.88
	Staff	5,231	0.95	10,003	0.97
Other	Other	218	0.28	271	0.26
Total	*	44,167	0.96	69,042	0.97

[9] Shorthand symbols or initials denote persons but are different from typical person names.

[10] The fourth annotator received 12 new documents.

[11] Of all annotations created for the 50 multiply annotated agreement documents, the final corpus only contains those created by the annotator who achieved the highest IAA.

Instance and token-level agreement lie between 0.8 and 1 for most annotation types. However, for *Relative, Other*, and *ID* each annotator found 11 or less instances on average. These values thus cannot be considered reliable. *Person* occurred 14 times in the whole corpus, none of these occurrences are in the 50 agreement documents. Low agreement for *Other* does not come as a surprise (independent of IAA reliability problems), as annotators were supposed to assign this category whenever they found information that they considered confidential, but could not match with any other annotation type. This vague definition led to a number of varying subjective judgements. For many annotation types (especially *MedicalUnit, Typist,* and *Staff*), token-level agreement is higher than instance-level agreement, indicating that there is disagreement about the exact extent of some annotations: partially overlapping annotations raise agreement on token level, but not on instance level. We also observe cases where one annotator decided to create one long annotation while another one created two short ones covering the same tokens. For the task of de-identification, the latter kind of disagreement is less severe, as we are mainly interested in covering all sensitive information.

Automatic De-Identification

In order to provide a first baseline for automatic de-identification on our annotated data and to submit the quality of our annotations to an empirical test, we trained and evaluated a neural network for named entity recognition (NER) designed by Sterbak [28] on the final corpus.

Methods

We randomly sampled 80% of the documents for training and used the remaining 20% as test set. Splitting the corpus on documents rather than sentences ensures a representative share of out-of-vocabulary (OOV) tokens. We provide only tokens and characters as input. The network employs a unidirectional LSTM [29] to learn word representations based on character embeddings. These character-based representations are concatenated with word embeddings (all embeddings were initialized randomly). The resulting complex word representations serve as input to a bidirectional LSTM. At each timestep, the output of the bidirectional LSTM is fed into a densely connected layer of units with a softmax activation function. Each unit of the last layer represents one possible classification result (the annotation types and an additional outside tag, *O*, for non-PHI data). The softmax function provides a probability distribution over these possible results. For each token, we predict the type with maximal probability.

To obtain reliable results despite randomly initialized weights, we train five different models on the same train-test split. We compute standard NER evaluation measures for each model and report averages over all five models. Additionally, we determine the percentage of OOV tokens for an annotation type *T*, that is, tokens that occur in the test set labeled as *T,* but do not occur in the training set with the same annotation type *T*.

Results

Table 3 yields the classifier's precision, recall and F1 score on the test set, as well as the support (the number of expected occurrences) and the percentage of OOV tokens for each annotation type and the outside tag (*O*). Totals are averages over all types weighted by support. For most annotation types, we achieve F1 scores between 0.88 and 0.98. 14% of all PHI tokens are OOV, whereas only 4% of all tokens are OOV. Disregarding confusions between different PHI types and only considering the binary distinction between PHI and non-PHI, the classifier achieves a precision of 0.98 and a recall of 0.96.

Table 3 – Automatic de-identification results on the test set: token-level precision, recall, and F1 score, support and percentage of out-of-vocabulary tokens per annotation type.

Type	Prec	Rec	F1	Supp	OOV
Age	0.885	0.910	0.897	89	25%
Birthdate	0.959	0.992	0.975	222	68%
Contact	0.966	0.945	0.956	435	7%
Date	0.959	0.970	0.964	4,969	15%
ID	0.693	0.385	0.480	95	91%
Location	0.977	0.940	0.958	2,188	11%
MedicalUnit	0.967	0.938	0.952	2,424	4%
O	0.998	0.999	0.999	269,337	3%
Other	0.592	0.384	0.463	63	46%
Patient	0.952	0.962	0.957	990	35%
Person	0.000	0.000	0.000	5	80%
Relative	0.000	0.000	0.000	16	100%
Staff	0.975	0.961	0.968	2,015	7%
Typist	0.898	0.862	0.879	245	12%
Total	0.996	0.996	0.996	283,093	4%
Total w/o O	0.959	0.947	0.952	13,756	14%

Discussion

We observe a clear drop in performance for smaller classes, especially *ID*, *Other, Person*, and *Relative*. The latter two together contribute only 85 tokens in the whole corpus, 21 of which are in the test set. So, we cannot expect the network to learn these rather difficult types. The confusion matrix reveals that, on average, around 43 out of 95 *ID* tokens were misclassified as *Date*s which suggests that the internal representations learned by the network are rather shallow–combinations of digits and punctuation marks may be *ID*s or *Date*s depending on the context–and to a large extent explains the low recall for *ID*. Furthermore, we find that low IAA scores are reflected in the classifier's performance for *Other* and *ID.* The overall F1 score (without the outside tag *O*) is slightly below the token-level results reported by Stubbs et al. [25] where the best system achieves an F1 score of 0.961 for the i2b2-PHI categories and 0.976 for the HIPAA-PHI categories. However, this comparison has to be taken with a grain of salt since both our annotation types and the data are different. The percentage of OOV tokens in the last column of Table 3 gives an idea about the recall a lazy learner might achieve by just remembering tokens occurring in the training set. Since our dataset stems from two units of a single hospital, the percentage of OOV tokens for *MedicalUnit* and *Staff* is low which puts the results for these types into perspective. For *Patient*, we observe a high percentage of OOV tokens, but for this type of PHI we sometimes find clear contextual cues (e.g., German equivalents of *Mr.* or *Ms./Mrs.*) which facilitate automatic recognition.

Conclusions

We annotated 13 PHI categories on the Jena slice of the German-language 3000PA corpus (1,106 documents) based on annotation guidelines that evolved over three annotation pre-iterations. The annotation process ensured that each document was checked twice for PHI and high inter-annotator agreement scores suggest that the resulting annotations are of very good quality. The neural baseline model we trained on the final annotated corpus achieved promising results, indicating that the annotation types defined in our guidelines lend themselves well to a data-driven approach to automatic de-identification. Future work will be directed at incorporating the 1,000 document

slices from the other two clinical sites, Leipzig and Aachen. With these more heterogeneous data, the classifier will be further tuned to maximize recall while ensuring generalizability to provide reasonable input for the planned pseudonymization engine as a resilient basis for the DUA-controlled distribution of a large corpus of German clinical documents.

Acknowledgements

This work was supported by the German Federal Ministry of Education and Research (BMBF) within the SMITH project under grant 01ZZ1803G and the Deutsche Forschungsgemeinschaft (DFG) under grants HA 2079/8-1 and KI 564/2-1 within the STAKI^{2}B^2 project.

References

[1] M. Nissim, L. Abzianidze, K. Evang, R. van der Goot, H. Haagsma, B. Plank, and M. Wieling, Sharing is caring: the future of shared tasks, *Comput Linguistics* **43** (2017), 897-904.

[2] C.-C. Huang and Z. Lu, Community challenges in biomedical text mining over 10 years: success, failure and the future, *Brief Bioinform* **17** (2016), 132-144.

[3] W.W. Chapman, P.M. Nadkarni, L. Hirschman, L.W. D'Avolio, G.K. Savova, and Ö. Uzuner, Overcoming barriers to NLP for clinical text: the role of shared tasks and the need for additional creative solutions, *JAMIA* **18** (2011), 540–543.

[4] A. Winter *et al.*, Smart Medical Information Technology for Healthcare (SMITH). Data integration based on interoperability standards, *Methods Inf Med* **57** (2018), e92-e105.

[5] U. Hahn, F. Matthies, C. Lohr, and M. Löffler, 3000PA: towards a national reference corpus of German clinical language, in: *MIE 2018*, Gothenburg, Sweden, 2018, pp. 26-30.

[6] M. Douglass, G. Clifford, A. Reisner, G. Moody, and R. Mark, Computer-assisted de-identification of free text in the MIMIC II database, in: *Comput Cardiol 2004*, Chicago, IL, USA, 2004, pp. 341-344.

[7] M. Douglass, G. Cliffford, A. Reisner, W. Long, G. Moody, and R. Mark, De-identification algorithm for free-text nursing notes, in: *Comput Cardiol 2005*, Lyon, France, 2005, pp. 331-334.

[8] I. Neamatullah, M. Douglass, L.-w. Lehman, A. Reisner, M. Villarroel, W. Long, P. Szolovits, G. Moody, R. Mark, and G. Clifford, Automated de-identification of free-text medical records, *BMC Med Inform Decis Mak* **8** (2008), 32-47.

[9] M. Saeed, M. Villarroel, A. Reisner, G. Clifford, L.-w. Lehman, G. Moody, T. Heldt, T. Kyaw, G. Moody, and R. Mark, Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC II): a public-access intensive care unit database, *Crit Care Med* **39** (2011), 952-960.

[10] Ö. Uzuner, Y. Luo, and P. Szolovits, Evaluating the state-of-the-art in automatic de-identification, *JAMIA* **14** (2007), 550-563.

[11] A. Stubbs and Ö. Uzuner, Annotating longitudinal clinical narratives for de-identification: the 2014 i2b2/UT-Health corpus, *J Biomed Inform* **58 Suppl** (2015), S20-S29.

[12] A. Stubbs, M. Filannino, and Ö. Uzuner, De-identifica-tion of psychiatric intake records: overview of 2016 CEGS N-GRID Shared Tasks Track 1, *J Biomed Inform* **75 Suppl** (2017), S4-S18.

[13] Ó. Ferrández, B. South, S. Shen, F. Friedlin, M. Samore, and S. Meystre, Evaluating current automatic de-identification methods with Veteran's Health Administration clinical docu-ments, *BMC Med Res Methodol* **12** (2012), #109.

[14] L. Deléger, K. Molnar, G. Savova, F. Xia, T. Lingren, Q. Li, K. Marsolo, A. Jegga, M. Kaiser, L. Stoutenborough, and I. Solti, Large-scale evaluation of automated clinical note de-identification and its impact on information extraction, *JAMIA* **20** (2013), 84-94.

[15] L. Deléger, Q. Li, T. Lingren, M. Kaiser, K. Molnar, L. Stoutenborough, M. Kouril, K. Marsolo, and I. Solti, Building gold standard corpora for medical natural language processing tasks, in: *AMIA 2012*, Chicago, IL, USA, 2012, pp. 144-153.

[16] L. Deleger, T. Lingren, Y. Ni, M. Kaiser, L. Stoutenborough, K. Marsolo, M. Kouril, K. Molnar, and I. Solti, Preparing an annotated gold standard corpus to share with extramural investigators for de-identification research, *J Biomed Inform* **50** (2014), 173-183.

[17] C. Grouin and A. Névéol, De-identification of clinical notes in French: towards a protocol for reference corpus development, *J Biomed Inform* **50** (2014), 151-161.

[18] S. Velupillai, H. Dalianis, M. Hassel, and G.H. Nilsson, Developing a standard for de-identifying electronic patient records written in Swedish: precision, recall and F-measure in a manual and computerized annotation trial, *Int J Med Inform* **78** (2009), e19-e26.

[19] H. Dalianis and S. Velupillai, De-identifying Swedish clinical text: refinement of a gold standard and experiments with conditional random fields, *J Biomed Semant* **1** (2010), #6.

[20] M. Morita, Y. Kano, T. Ohkuma, M. Miyabe, and E. Aramaki, Overview of the NTCIR-10 MedNLP Task, in: *NTCIR-10*, Tokyo, Japan, 2013, pp. 696-701.

[21] H. Seuss, P. Dankerl, M. Ihle, A. Grandjean, R. Ham-mon, N. Kaestle, P.A. Fasching, C. Maier, J. Christoph, M. Sedlmayr, M. Uder, A. Cavallaro, and M. Hammon, Semi-automated de-identification of German content sensitive reports for big data analytics, *Röfo* **189** (2017), 661-671.

[22] P. Richter-Pechanski, S. Riezler, and C. Dieterich, De-identification of German medical admission notes, in: *GMDS 2018*, Osnabrück, Germany, 2018, pp. 165-169.

[23] P. Stenetorp, S. Pyysalo, G. Topić, T. Ohta, S. Anania-dou, and J.i. Tsujii, BRAT : a web-based tool for NLP-assisted text annotation, in: *EACL 2012: Demonstrations*, Avignon, France, 2012, pp. 102–107.

[24] C. Lohr, S. Luther, F. Matthies, L. Modersohn, D. Ammon, K. Saleh, A. Henkel, M. Kiehntopf, and U. Hahn, CDA-compliant section annotation of German-language discharge summaries: guideline development, annotation campaign, section classification, in: *AMIA 2018*, San Francisco, CA, USA, 2018, pp. 770-779.

[25] A. Stubbs, C. Kotfila, and Ö. Uzuner, Automated sys-tems for the de-identification of longitudinal clinical narra-tives: overview of 2014 i2b2/UTHealth Shared Task Track 1, *J Biomed Inform* **58 Suppl** (2015), S11-S19.

[26] G. Hripcsak and A.S. Rothschild, Agreement, the F-measure, and reliability in information retrieval, *JAMIA* **12** (2005), 296-298.

[27] A. Browne, M. Kayaalp, Z. Dodd, P. Sagan, and C. McDonald, The challenges of creating a gold standard for de-identification research, in: *AMIA 2014*, Washington, D.C., USA, 2014, pp. 353-358.

[28] T. Sterbak, Enhancing LSTMs with character embed-dings for named entity recognition, in: *Depends on the defini-tion, URL: https://www.depends-on-the-definition.com/lstm-with-char-embeddings-for-ner/*, Last Access: Nov 3, 2018.

[29] S. Hochreiter and H.J. Schmidhuber, Long short-term memory, *Neural computation* **9** (1997), 1735-1780.

Address for correspondence

Tobias Kolditz, Jena University Language & Information Engineering (JULIE) Lab, Friedrich-Schiller-Universität Jena, Fürstengraben 27, 07743 Jena. Email: tobias.kolditz@uni-jena.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190213

Creating a Queer Ontology: The Gender, Sex, and Sexual Orientation (GSSO) Ontology

Clair Kronk[a], Giao Q. Tran, PhD[b], Danny T.Y. Wu, PhD, MSI[a]

[a] Department of Biomedical Informatics, University of Cincinnati, Cincinnati, Ohio, United States,
[b] Department of Women's, Gender, and Sexuality Studies, University of Cincinnati, Cincinnati, Ohio, United States

Abstract

Cultural attitudes, linguistic variation, and historical pathology have led to a plethora of terms concerning gender, sex, and sexual orientation that have caused confusion and uneasiness among both lay people and relevant professionals. For members of the LGBTQIA+ community, these negative reactions are compounded by identities which have historically and contemporarily been mistreated by medical professionals. In an effort to provide a reliable resource for patients and clinicians, we have created the Gender, Sex, and Sexual Orientation (GSSO) ontology, which currently includes over 4,000 entities from multiple disciplines. The GSSO is a manually curated resource utilizing related glossaries from biology, medicine, psychology, sociology, and gender studies. With links to over 20 other ontology resources such as SNOMED-CT and MedDRA, the GSSO aims for accessibility and interoperability with existing systems. It is also open-source and features an easy-to-use web interface (https://github.com/Superraptor/GSSO). Future work involves multiple language support efforts and empirical evaluation.

Keywords:
Linguistic, Sexual and Gender Minorities; Social Determinants of Health

Introduction

An ontology is a description of a common, shared, or otherwise controlled vocabulary used to explain information in a particular domain, such as linguistics or medicine. More often than not, due in part to advances in biomedical informatics and the rise of big data, ontologies are usually designed primarily to be machine-interpretable, rather than strictly human-readable [1].

Noy and McGuinness point out several reasons why ontology development is crucial in a general sense, including reuse of domain knowledge, analysis of that knowledge, and creating explicitness within that domain. This explicitness is especially important for scientific communication, namely helping the public understand the importance of such communication, especially when such communication concerns health outcomes. As Markides put so well in 2014 "[c]ommunication is the most important component of our work with patients. It is the cornerstone of our interaction with people" [2].

Clear communication is even more crucial when it concerns people belonging to groups that have historically and contemporarily experienced discrimination in healthcare and mistreatment by medical practitioners. If communication is unclear, disparaging, or otherwise offensive, patient outcomes can be affected [3–6]. Various forms of discrimination or perceived discrimination, many of which include linguistic components, have been strongly linked to suicide and other areas of mental health [7–15].

One such group that experiences such discrimination are LGBTQIA+ (lesbian, gay, bisexual, transgender, queer/questioning, intersex, agender/asexual, and other umbrella gender and sexual identity minorities) populations. For example, it is well documented that LGBTQIA+ populations have significantly higher risk of suicide than general population. The differences in suicidality have been variously attributed to social determinants of health (SDOH) including anti-LGBTQIA+ government policies, medicalization of gender and sexual nonconformity, and unemployment linked to LGBTQIA+ status [16–18].

Unfortunately, the chasm between healthcare providers and the LGBTQIA+ community persists, with one study suggesting 23% of transgender respondents did not see a doctor when they needed due to fear of mistreatment [19]. Even well-intentioned providers can find reliable information difficult to access. As one clinician explained [20]:

> [D]espite trying to find ways to improve my expertise, I just didn't know where to go or who to talk to, or where to get the information, and I felt really bad because some of my initial attempts to help these people—I sent them to people I wish I hadn't sent them to.

Here, we present preliminary efforts to structure information about gender, sex, and sexual orientation into a comprehensive ontology for use by healthcare professionals and patients alike, with potential to help mitigate communication issues and provide reliable, trustworthy information on LGBTQIA+ terminology and medical needs.

Methods

Ontology Development

An iterative approach was used to manually create and expand the Gender, Sex, and Sexual Orientation (GSSO) ontology, following the procedures outlined by Corcho et al. for annotation [21]. These include four main steps: 1) marking, 2) exploring, 3) mapping, and 4) abstracting.

First, the subject area was clearly marked as including gender-, sex-, and sexual orientation-related knowledge, helping to exclude materials in psychiatry, psychology, and other fields outside the purview of those fields. Second, current lists of terminology and vocabulary regarding these content areas were explored primarily using PubMed, Google Books and Scholar,

and the HathiTrust Digital Library, supplemented by wiki-based sites featuring specific slang and media styleguides (such as those provided by LGBTQIA+ organizations like GLAAD and PFLAG) to provide necessary context. Examples of exploratory sources included *The International Encyclopedia of Human Sexuality* (2015), *Encyclopedia of Gender and Society* (2009), *The Wiley Blackwell Encyclopedia of Gender and Sexuality Studies* (2016), and several editions of the *Diagnostic and Statistical Manual of Mental Disorders* (1952, 1968, 1980, 1987, 1994, 2013).

Third, once a group of terms was identified, they were mapped onto existing terminologies available via the NCBO (National Center for Biomedical Ontology) BioPortal where appropriate (a summary of these mappings is shown in Table 1). Originally, terms mappings were performed automatically using full-string matching or near-string matching, but this process introduced significant false positive matching and addition of inaccurate, incomplete, and entirely missing annotations. Therefore, mapping was supplemented by a manual process to approve any automatic matching. Finally, these terms were "mapped up" using existing ontologies, moving toward higher abstractions (eventually running into the basic formal ontology's [BFO] *entity* concept) [22]. Often when searching resources, web-related documents and print periodicals would be consulted to approximate leveraging clinical definitions and more colloquial definitions. During this process, new terms appeared, necessitating beginning the approach from the first step.

Table 1. Cross-references to various other ontologies. Despite over 1,200 mappings, over 2,000 terms are entirely new.

Acronym	Full Name	# Terms
BFO	Basic Formal Ontology	19
ChEBI	Chemical Entities and Biological Interest	4
DO	Disease Ontology	62
EFO	Experimental Factor Ontology	35
FMA	Foundational Model of Anatomy	43
GO	Gene Ontology	32
GOLD	General Ontology for Linguistic Description	13
HPO	Human Phenotype Ontology	30
IAO	Information Artifact Ontology	10
ICD-9-CM	International Classification of Diseases, 9th Edition	30
ICD-10-CM	International Classification of Diseases, 10th Edition	28
LifO	Life Ontology	6
MedDRA	Medical Dictionary for Regulatory Activities	129
MeSH	Medical Subject Headings	261
NCBI Taxon	NCBI Taxonomy	11
NCIT	NCI Thesaurus	261
OBI	Ontology for Biomedical Investigations	6
RO	Relations Ontology	6
SCTID	SNOMED-CT	241
SIO	Semanticscience Integrated Ontology	116
STY	Semantic Types Ontology	16
TA	Terminologia Anatomica	3
TE	Terminologia Embryologica	30
TH	Terminologia Histologica	2
UBERON	Uber-anatomy Ontology	22
	Total:	**1416**

Construction of the GSSO is ongoing using a combination of this methodology and user feedback concerning necessary additions and terms considered missing or in need of editing.

Ontology Evaluation

The GSSO was evaluated using three techniques derived from Brank et al. [23]. These methods included comparisons and mappings to other ontologies (which was folded into the "mapping up" done during construction), usage of the MIMIC-III database for finding patients based on GSSO terminology versus ICD-9-CM (automated matching/classification as part of an application), and preliminary feedback from experts in different fields affected by the GSSO. Several more comprehensive validation techniques are in development and outside the scope of the current study.

Results

The GSSO includes over 4,000 classes, with over 1,000 definitions and more than 1,400 external biomedical ontology references (not including links to wiki-based sources, which would bring the total of external links to over 3,300). Over 70 scholarly articles and 80 books are fully cited and sourced, covering a range of topics, from Perry M. Lichtenstein's infamous "The 'Fairy' and the Lady Lover" (1922) to "Induced Lactation in a Transgender Woman" (2018). A list of all sources (as of 11 November 2018) is included as a Microsoft Excel spreadsheet on the project's GitHub (https://github.com/Superraptor/GSSO/blob/master/docs/gsso_sources_11.15.2018_v1.xlsx). In addition, as of the time of writing, the GSSO is the only known ontology to include its sources as classes in the ontology itself, a unique feature which allows one to analyze concept shift over time, and to drill down to original definitions. Generally speaking, highly connected relationships were prioritized, meaning terms that appeared more often in source materials usually ended up with more detailed annotations (such as *transgender*, visualized as a map in Figure 1).

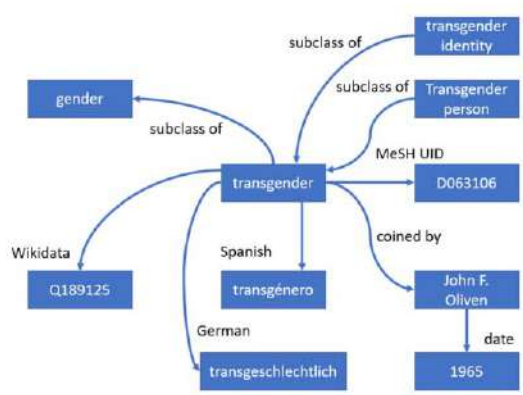


Figure 1. Visualization of Highly-Connected Relationships in the GSSO (emphasizing database cross-references and translations) Highly-connected networks allow for more complicated annotations and queries which are more reflective of complicated relationships in reality.

More than 200 usage notes are included, discussing reclaimed terms (such as *dyke*), terms that are considered problematic in certain contexts (*transsexual*), and terms that are obsolete

(*transgenderism*). An sample entry, the term *homosexual*, is shown with such notes in Figure 2.

Because some of these terms are slang, taboo, or only used within select groups, online sources played a large part in tracing word origins and sentiment. All of these sources were archived using the Wayback Machine provided by the Internet Archive, ensuring that they will be available in their original state for the foreseeable future (this includes nearly 100 online sources captured from the early days of Usenet to present).

Many terms did not appear in any of the 600+ available biomedical ontologies, especially those relating to transgender health care: gender expression, non-binary gender identities, gender passing, and cultural gender roles (*hijra*, *māhū*, two-spirit, etc.) were all unaccounted for.

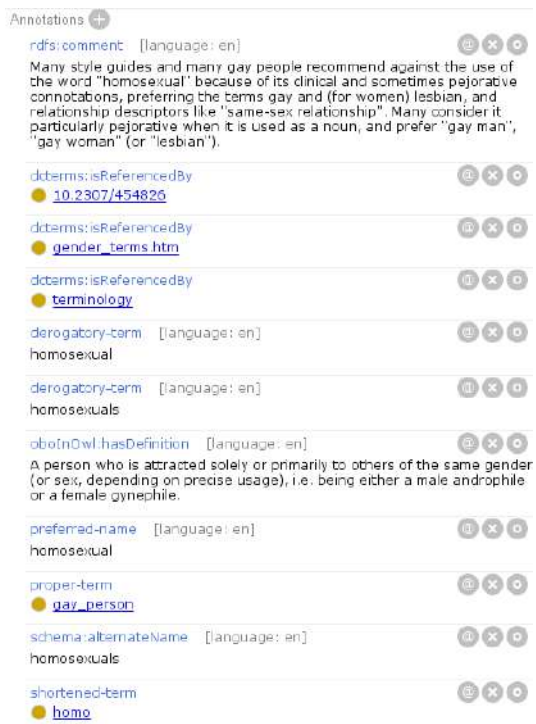


Figure 2. Example Annotation Showcasing Usage Notes. Displayed using Protégé, a free, open-source ontology editing framework.

Validation efforts are ongoing, as the ontology itself is still growing. Exploratory efforts showed promising results. Preliminary feedback from experts regarding the ontology's content, subject area, and resourcing has been positive based on discussion at the University of Cincinnati Graduate Student Research Forum (GSRF), the American Medical Informatics Association (AMIA) 2018 Annual Symposium, and the GLBT Museum and Archives in San Francisco, California. A mailing list sent by the GLBT Archives to various archivists and librarians across the United States yielded similar positive feedback. High facial validity was also shown in the comparison and mapping of GSSO to other ontologies. When identifying transgender patients in the MIMIC-III clinical notes database [24], only six verified transgender patients could be identified using ICD-9 codes alone. However, when using the ontology within unstructured clinical notes, the GSSO allowed for detection of those six, along with seven others. Additionally, six of these patients had an incorrect sex assigned at birth noted, which could only be detected after using the ontology to analyze the clinician's notes specifically. Use in a wider clinical dataset is underway.

Discussion

In this project, we developed the GSSO ontology and demonstrated preliminary evaluation results. As previously stated, over 2,000 terms within the ontology are entirely new, having been unaccounted for in other biomedical ontologies. Ultimately, we hope that this ontology can be layered in with existing electronic health record technologies to educate medical professionals, without requiring additional separate trainings (problematic terms could be underlined and indicated as such without being interruptive, providing extra information about why a certain term is derogatory, for instance).

Using such a system could help orient and educate patients and clinicians approaches biomedical ontologies from an uncommon perspective, building from a human readability standpoint first. Of course, the ontology could be used in machine learning or natural language processing applications, but having a very human readable focus is not particularly common among other ontology systems. For this reason, the paired automatic and manual approaches towards validation are considered especially crucial.

LGBTQIA+ specific forums present on Reddit (https://www.reddit.com/) and Usenet (archived here: https://groups.google.com) are being analyzed to provide online contexts, and to see if similar results can be obtained on more chaotic, unstructured text present on the internet. Such an approach parallels that of the previously constructed Vaccine Misinformation Ontology (VAXMO), which aimed to collect misunderstandings about vaccines [25]. In the VAXMO paper, suggested use cases included producing datasets for machine learning and semantic-driven approach for misinformation detection, both of which the GSSO could handle in its own respective domain. For example, the GSSO could be used to identify positive and negative connotations in the LGBTQIA+ entity space as well as looking for information that could be considered medically misinformed (such as incorrect information regarding hormone replacement therapy or sexually transmitted infections). The GSSO has great potential to produce information that could be used to better inform clinicians and thereby facilitate effective patient communication, which is a critical first step to solve social and political issues faced by this underserved patient population.

In addition to these automated approaches, we are also designing inclusive protocols for both patients and providers to evaluate the ontology in a more subjective manner, including the ontology's presentation, usefulness, and potential missing entries. As such, the ontology is publicly available for comment at https://github.com/Superraptor/GSSO, with a simple search interface available at http://homepages.uc.edu/~kronkcj/gsso/ (Figure 3).

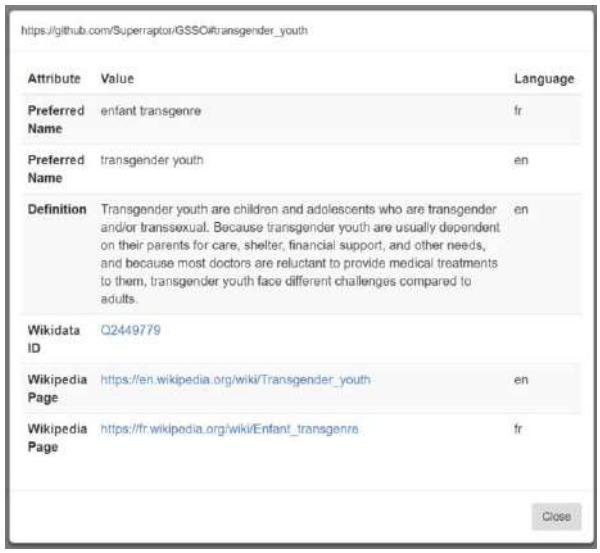


Figure 3. Modal Displaying Annotations for the Term, Transgender Youth*. Interface was built entirely using JavaScript and provides accessibility to those not necessarily familiar with ontologies.*

Conclusions

Language has more power in medicine and outside of medical environments than people realize. As one activist put so well: "To omit specific labels… is to deny the existence of the labelled" [26]. Since Ralph Werther became the first transgender person to publish an autobiography in the United States in 1918, it was clear that abusive language affected mental well-being [27,28]:

> Mean-spirited boys would call me a girl in derision, and twit me about my conduct of early childhood, thus awakening a violent desire to commit suicide.

One hundred years later, despite over coming many significant hurdles, the National Transgender Discrimination Survey still reported that 50% of transgender patients sampled had to teach their medical provider about transgender care—and 19% reported being refused care because of their gender status [29]. Using ontologies as a baseline to educate both patients and providers helps attack the problem at both ends, potentially leading to better health outcomes for LGBTQIA+ patients in general as communication and understanding between groups improves.

Acknowledgements

In addition to thanking my mentors Drs. Danny T. Y. Wu and Giao Q. Tran, I would like to thank Jonathan Ned Katz, Randall Sell, and Susan Stryker for their help in determining historical contextualization which helped facilitate the development of the GSSO. I would also like to thank Nalini Elias and Patricia Delara of the GLBT Historical Society Museum & Archives.

References

[1] R. Hoehndorf, P.N. Schofield, and G.V. Gkoutos, The role of ontologies in biological and biomedical research: a functional perspective, *Briefings in Bioinformatics*. **16** (2015) 1069–1080. doi:10.1093/bib/bbv011.

[2] M. Markides, The Importance of Good Communication Between Patient and Health Professionals:, *Journal of Pediatric Hematology/Oncology*. **33** (2011) S123–S125. doi:10.1097/MPH.0b013e318230e1e5.

[3] N. Abdulhadi, M. Al Shafaee, S. Freudenthal, C.-G. Östenson, and R. Wahlström, Patient-provider interaction from the perspectives of type 2 diabetes patients in Muscat, Oman: a qualitative study, *BMC Health Services Research*. **7** (2007). doi:10.1186/1472-6963-7-162.

[4] D.C. Dugdale, R. Epstein, and S.Z. Pantilat, Time and the patient-physician relationship, *J Gen Intern Med*. **14 Suppl 1** (1999) S34-40.

[5] V. Okunrintemi, E.S. Spatz, P. Di Capua, J.A. Salami, J. Valero-Elizondo, H. Warraich, S.S. Virani, M.J. Blaha, R. Blankstein, A.A. Butt, W.B. Borden, K. Dharmarajan, H. Ting, H.M. Krumholz, and K. Nasir, Patient–Provider Communication and Health Outcomes Among Individuals With Atherosclerotic Cardiovascular Disease in the United States: Medical Expenditure Panel Survey 2010 to 2013, *Circulation: Cardiovascular Quality and Outcomes*. **10** (2017). doi:10.1161/CIRCOUTCOMES.117.003635.

[6] M.A. Stewart, Effective physician-patient communication and health outcomes: a review, *CMAJ*. **152** (1995) 1423–1433.

[7] L.W. Li, G.C. Gee, and X. Dong, Association of Self-Reported Discrimination and Suicide Ideation in Older Chinese Americans, *The American Journal of Geriatric Psychiatry*. **26** (2018) 42–51. doi:10.1016/j.jagp.2017.08.006.

[8] S. Assari, M. Moghani Lankarani, and C. Caldwell, Discrimination Increases Suicidal Ideation in Black Adolescents Regardless of Ethnicity and Gender, *Behavioral Sciences*. **7** (2017) 75. doi:10.3390/bs7040075.

[9] A. Ahuja, LGBT adolescents in America: Depression, discrimination and suicide, *European Psychiatry*. **33** (2016) S70. doi:10.1016/j.eurpsy.2016.01.981.

[10] A.P. Haas, M. Eliason, V.M. Mays, R.M. Mathy, S.D. Cochran, A.R. D'Augelli, M.M. Silverman, P.W. Fisher, T. Hughes, M. Rosario, S.T. Russell, E. Malley, J. Reed, D.A. Litts, E. Haller, R.L. Sell, G. Remafedi, J. Bradford, A.L. Beautrais, G.K. Brown, G.M. Diamond, M.S. Friedman, R. Garofalo, M.S. Turner, A. Hollibaugh, and P.J. Clayton, Suicide and Suicide Risk in Lesbian, Gay, Bisexual, and Transgender Populations: Review and Recommendations, *Journal of Homosexuality*. **58** (2010) 10–51. doi:10.1080/00918369.2011.534038.

[11] a Committee on Improving the Health, Safety, and Well-Being of Young Adults, ed., Investing in the health and well-being of young adults, The National Academies Press, Washington, D.C, 2014.

[12] M.L. Halim, K.H. Moy, and H. Yoshikawa, Perceived ethnic and language-based discrimination and Latina immigrant women's health, *Journal of Health Psychology*. **22** (2017) 68–78. doi:10.1177/1359105315595121.

[13] L.R.M. Hausmann, K. Jeong, J.E. Bost, and S.A. Ibrahim, Perceived Discrimination in Health Care and Health Status in a Racially Diverse Sample:, *Medical*

Care. **46** (2008) 905–914. doi:10.1097/MLR.0b013e3181792562.
[14] D.H. Sorkin, Q. Ngo-Metzger, and I. De Alba, Racial/Ethnic Discrimination in Health Care: Impact on Perceived Quality of Care, *Journal of General Internal Medicine*. **25** (2010) 390–396. doi:10.1007/s11606-010-1257-5.
[15] Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care (with CD), National Academies Press, Washington, D.C., 2003. doi:10.17226/12875.
[16] J.M. White Hughto, S.L. Reisner, and J.E. Pachankis, Transgender stigma and health: A critical review of stigma determinants, mechanisms, and interventions, *Social Science & Medicine*. **147** (2015) 222–231. doi:10.1016/j.socscimed.2015.11.010.
[17] J.A. Bauermeister, How Statewide LGB Policies Go From "Under Our Skin" to "Into Our Hearts": Fatherhood Aspirations and Psychological Well-Being Among Emerging Adult Sexual Minority Men, *Journal of Youth and Adolescence*. **43** (2014) 1295–1305. doi:10.1007/s10964-013-0059-6.
[18] M.L. Hatzenbuehler, K.A. McLaughlin, K.M. Keyes, and D.S. Hasin, The Impact of Institutional Discrimination on Psychiatric Disorders in Lesbian, Gay, and Bisexual Populations: A Prospective Study, *American Journal of Public Health*. **100** (2010) 452–459. doi:10.2105/AJPH.2009.168815.
[19] S.E. James, J.L. Herman, S. Rankin, M. Keisling, L. Mottet, and M. Anafi, The Report of the 2015 U.S. Transgender Survey, (2016). https://transequality.org/sites/default/files/docs/usts/USTS-Full-Report-Dec17.pdf (accessed March 12, 2019).
[20] J.W. Snelgrove, A.M. Jasudavisius, B.W. Rowe, E.M. Head, and G.R. Bauer, "Completely out-at-sea" with "two-gender medicine": A qualitative analysis of physician-side barriers to providing healthcare for transgender patients, *BMC Health Services Research*. **12** (2012). doi:10.1186/1472-6963-12-110.
[21] O. Corcho, M. Fernández-López, and A. Gómez-Pérez, Methodologies, tools and languages for building ontologies. Where is their meeting point?, *Data & Knowledge Engineering*. **46** (2003) 41–64. doi:10.1016/S0169-023X(02)00195-7.
[22] R. Arp, B. Smith, and A.D. Spear, Building ontologies with Basic Formal Ontology, Massachusetts Institute of Technology, Cambridge, Massachusetts, 2015.
[23] J. Brank, M. Grobelnik, and D. Mladenić, A Survey of Ontology Evaluation Techniques, (n.d.). http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.4788&rep=rep1&type=pdf.
[24] A.E.W. Johnson, T.J. Pollard, L. Shen, L.H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R.G. Mark, MIMIC-III, a freely accessible critical care database, *Scientific Data*. **3** (2016) 160035. doi:10.1038/sdata.2016.35.
[25] M. Amith, and C. Tao, Representing vaccine misinformation using ontologies, *Journal of Biomedical Semantics*. **9** (2018). doi:10.1186/s13326-018-0190-0.
[26] J. Kennedy, Ottawa Council wimps out with generic "Pride Week" proclamation, *The Ottawa Citizen*. (1995) 11.
[27] S. Stryker, Transgender history, Seal Press : Distributed by Publishers Group West, Berkeley, CA, 2008.
[28] E. Lind, R. Werther, and R. Werther, Autobiography of an androgyne, Fredonia Books, Amsterdam, The Netherlands, 2005.
[29] J.M. Grant, L.A. Mottet, and J. Tanis, National Transgender Discrimination Survey Report on Health and Health Care, (2010). https://web.archive.org/web/20181022191420/https://cancer-network.org/wp-content/uploads/2017/02/National_Transgender_Discrimination_Survey_Report_on_health_and_health_care.pdf (accessed October 21, 2018).

Address for correspondence

Contact Clair Kronk via email at kronkcj@mail.uc.edu and/or clair.kronk@gmail.com. Specific issues or discussion related to the GSSO should be posted at the GitHub address, i.e. https://github.com/Superraptor/GSSO.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190214

An Iconic Approach to the Browsing of Medical Terminologies

Jean-Baptiste Lamy[a], Van Bui Thuy[a], Agnès Lillo-Le Louët[b], Cédric Bousquet[a,c]

[a] *LIMICS, Université Paris 13, Sorbonne Paris Cité, 93017 Bobigny, INSERM UMRS 1142, Sorbonne Université, Paris, France*
[b] *Centre Régional de Pharmacovigilance HEGP, AP-HP, Paris, France.*
[c] *SSPIM, Unit of Public Health and Medical Information, CHU University Hospital of Saint Etienne, Saint-Étienne, France*

Abstract

Medical terminologies are the basis of interoperability in medicine. They allow connecting the various systems and data and facilitate searches in databases. An example is the MedDRA terminology, used in particular for coding drug adverse events. However, these terminologies are often complex and involve a huge number of terms. Consequently, it is difficult to browse them or find the desired terms. Traditional approaches consist of lexical search, with the problems of synonymy and polysemy, or tree-based navigation, but the user often gets "lost" in the tree.

Here, we propose a new approach for browsing medical terminologies: the use of pictograms and icons, for formulating the query in complement to a textual search box, and for displaying the search results. We applied this approach to the MedDRA terminology. We present both the methods and search algorithms and the resulting browsing interface, as well as the qualitative opinions of two pharmacovigilance experts.

Keywords:
Terminology as Topic, Pharmacovigilance, Nonverbal Communication.

Introduction

Medical terminologies are the basis of interoperability in medicine [1]. They allow connecting the various systems and data, and facilitate searches in databases [2]. An example is the MedDRA terminology, used in particular for coding drug adverse events. Traditional search approaches are either based on lexical search or on the navigation in a hierarchy [3]. However, lexical searches suffer from synonymy, polysemy and metonymy, e.g. a search with "auricular" return terms related to both heart and ear. Navigation in a tree is usually complex and the user often gets "lost" in the tree [4]. Brown [5] described the various difficulties encountered when performing searches in MedDRA-coded pharmacovigilance database.

In the literature, several authors suggest the use of pre-coordination for defining terms in terminologies, i.e. the decomposition of disorders into anatomic location and morphology, e.g. "nephritis," defined as an inflammation of the kidney, and post-coordination for selecting terms [6]. Bakhshi-Raiez et al. [7] proposed to take advantage of the compositional aspect of the SNOMED CT terminology to facilitate the searches in the terminology. Lee et al. [8] proposed the use of post-coordinated expressions to retrieve ICD codes. Third-generation, knowledge-based terminologies, aim at formalizing semantic relations and decompositions between terms, which can help with searches [9]. The use of semantic web standards such as RDF has also been proposed to facilitate data retrieval from MedDRA [10].

In previous works [11], we proposed an iconic compositional language for medical concepts. In this paper, we propose an iconic approach to the browsing of medical terminologies. Our approach uses icons at two levels: (1) for formulating the query, in combination with a traditional textual search box, and (2) for displaying the search results, by grouping similar terms using icons. This work is part of the Pegase project, which aims at helping pharmacovigilance experts with the use of the MedDRA terminology.

Methods

MedDRA is a medical classification often used for coding drug adverse effects and events. It has 5 levels, from the more general to the more specific: system organ class (SOC), high-level group terms (HLGT), high-level terms (HLT), preferred terms (PT) and low-level terms (LLT). The classification is multi-axial: a term can have more than one parent. We worked on the French translation of MedDRA 19.

VCM [11] is an iconic language for representing the patient main clinical conditions, including symptoms, diseases, physiological states (*e.g.* age class or pregnancy), risks and history of diseases, drug and non-drug treatments, lab tests and follow-up procedures. VCM includes a set of graphical primitives (colors and pictograms) and a grammar to combine these elements for creating icons.

For representing clinical signs and disorders, a VCM icon is made of a color, a basic shape, a central pictogram and a set of modifier pictograms. The color indicates the temporal aspect of the icon: red for current states of the patient, orange for risk of future states, and brown for past states. The basic shape is a circle for physiological states or a square for pathological states (diseases or symptoms). The central pictogram indicates the anatomico-functional location (e.g. endocrine system) or the patient characteristic (*e.g.* pregnancy) involved; and special pictograms are available for a few specific disorders associated with a specific anatomico-functional location (*e.g.*

```
function lexico_icono_hierarchical_search(keywords, pictograms, selected_levels):
    if only keywords (i.e. pictograms = ∅ ):
        terms = { t such that MedDRA(t) and match(t.label, keyword) and t.levels ∈ selected_levels }
        icons = { i such that Icon(i) and ∃ t ∈ terms with has_icon(t, i) }
    else if only pictograms (i.e. keywords = ∅ ):
        icons = { i such that Icon(i) and ∀ p ∈ pictograms we have has_part(i, p) }
        terms = { t such that MedDRA(t) and t.levels ∈ selected_levels and ∃ i ∈ icons with has_icon(t, i) }
        icons = { i such that i ∈ icons and ∃ t ∈ terms with has_icon(t, I)}
    else (both keywords and pictograms):
        terms = { t such that MedDRA(t) and match(t.label, keyword) and t.levels ∈ selected_levels }
        icons = { i such that Icon(i) and ∀ p ∈ pictograms we have has_part(i, p) }
        terms = { t such that t ∈ terms and ∃ i ∈ icons with has_icon(t, i) }
        icons = { i such that i ∈ icons and ∃ t ∈ terms with has_icon(t, I)}
    return (icons, terms)
```

Figure 2– Algorithm performing the lexico-icono-hierarchical search in pseudo-code. The search takes three input parameters: a set of lexical keywords, a set of pictograms, and a set of selected levels in the hierarchy (e.g. {PT, LLT}). It returns two values: the set of icons and the set of MedDRA terms to display. We used a first-order logic syntax in the algorithm, i.e. classes are unary predicates and properties are two-argument predicates, e.g. "MedDRA(t)" means that it is a MedDRA term and "has_icon(t, i)" means that it has icon i. Finally, match() is a function that performs keyword-based string comparisons.

diabetes for the endocrine system). Modifier pictograms can be added to specify (a) a general pathological process (*e.g.* tumor or infection), and (b) a "transversal" anatomical structure that can be present in many anatomico-functional locations (*e.g.* blood vessels and nerves are present in most organs).

The VCM ontology has been designed for formalizing the semantics of VCM icons [12]. This ontology includes 3 parts: (1) graphical concepts corresponding to VCM graphical primitives (*i.e.* the various shapes, colors and pictograms), (2) medical concepts (*i.e.* the main anatomical structures, biological functions, pathological processes, *e.g.* liver, hepatic function and inflammation, but *not* the various disorders, such as hepatitis), and (3) relations between the graphical and the medical concepts (*e.g.* the "liver" central pictogram is associated with both the liver anatomic structure and the hepatic biological function). In a previous work [13], we started to design a mapping between MedDRA and VCM.

Model

We designed an OWL ontology including: (1) all MedDRA terms with their codes, labels and parent-child relations, (2) MedDRA to VCM mapping and (3) decomposition of VCM icons in pictograms. Figure 1 shows the general model of the resulting ontology.

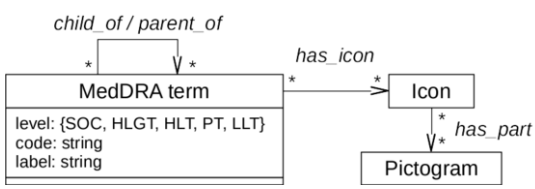


Figure 1– General model of the ontology, as a UML class diagram.

We used PyMedTermino [14] to extract the MedDRA terminology and to deal with VCM icons. The resulting OWL ontology includes 531,095 RDF triples (46 Mb) and belongs to the $\mathcal{ALH}(D)$ family of description logics. The ontology was stored in an optimized quadstore based on a SQLite3 database, using the Owlready ontology-oriented programming module [15, 16] for Python. Owlready was used because it allows an object-oriented access to ontologies with performances similar to SQL databases. It also supports a full-text search.

Search Approaches

The proposed interface combines three approaches for searching terms; they are detailed in the following paragraphs.

The **full-text search** approach consists of searching for one or more keywords in the labels of the MedDRA terms. Each keyword has to match the beginning of one word of the term, *e.g.* keyword "Abs" match term "Renal abscess" but not "Bile acid malabsorption". We used Owlready2 to perform the search, which relies on the FTS5 implementation of Sqlite3. In the interface, keywords are entered into a search bar.

The **iconic approach** consists of selecting one or more pictograms and searching for all terms that are associated with icons that include all the selected pictograms. To facilitate the selection of pictograms, we organized them on a schematic anatomical sketch, called "Mister VCM". It shows 37 pictograms, grouped in 6 regions: the head, the hat (for social medicine), the thoughts (for psychology and addictions), the trunk, the arm (only one limb is detailed) and the etiologies. The entire VCM language has more than 150 pictograms, however, many of them are specialized versions of a more general pictogram, and "Mister VCM" shows only the general pictograms. The user can click on a pictogram to select it, and the gray pictogram is replaced by a red icon. By clicking the same pictogram again, the user can deselect it. When several pictograms are selected, we consider their intersection, *i.e.* the terms have to match all the selected pictograms.

The **hierarchical approach** considers the hierarchical structure of medical terminologies. In the present work, we limited this approach to the ability to filter MedDRA terms according to their depth level in the terminology, *e.g.* search only for PT, or for both PT and LLT, ... The interface includes 5 checkboxes, one per MedDRA level; they allow the user to select the desired levels.

The output of the search is a set of MedDRA terms, but also a set of VCM icons. The set of icons are be used to display the search results in a synthetic and visual way. Figure 2 gives the algorithm that combines the three search approaches. If only keywords are present in the query, we search for the terms of

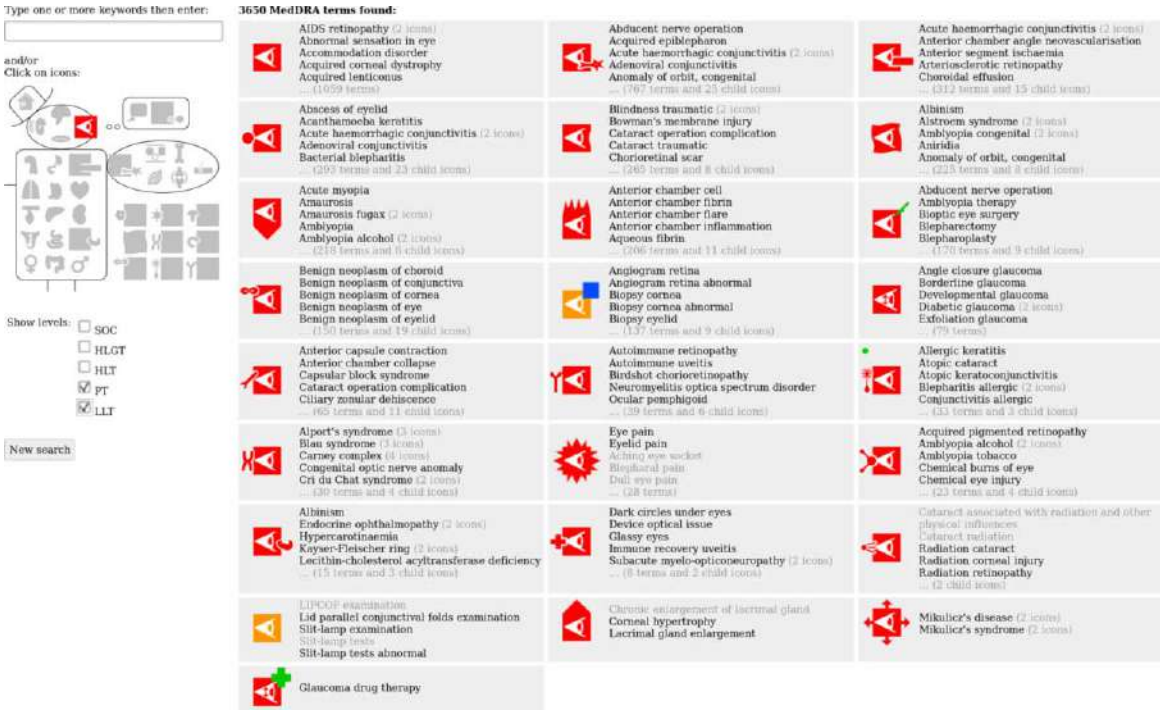


Figure 3– Screenshot of the results of an iconic search for PT and LLT with the "eye" pictogram.

the desired levels whose label matches the keywords; then we search for their icons. If only pictograms are present, we search for the icons having these pictograms; then we search for the terms of the desired levels that have at least one of these icons; finally, we filter the set of icons to those associated with at least one term in the resulting set of terms (due to the filtering on hierarchical levels, some icons can be associated with no terms). When both keywords and icons are present, we search for terms (taking into account keywords and levels) and for icons (taking into account pictograms) independently, and then we restrict both resulting sets to the elements associated with at least one element in the other set. In addition, the proposed algorithm is designed so that each step involves at the most two classes out of the three in the model (*i.e.* MedDRA terms, Icons, and Pictogram). This improves performances of the search.

Displaying the Search Results

Search results can be numerous; therefore, we use icons to organize them. Icons are sorted by the number of associated terms. For each icon, we display a panel with the icon on the left and the list of the corresponding MedDRA terms on the right. LLT are displayed in gray and other levels in black. When more than 6 terms are present, we display only the first 5 ones (in alphabetical order) and the total number of terms; the entire list of terms can be obtained by clicking on the panel. Whenever a given term is associated with more than one icon in the resulting set of icons, the term appears in several panels (one per icon), with a gray label indicating the number of icons (*e.g.* "2 icons"). By mouse-hovering this label, the user can see a popup bubble showing all the icons.

Moreover, when there are more than 10 icons, we reduce the list of icons by removing all icons that are more specific than another. In fact, the VCM ontology defines is-a relation between icons. For example, the icon "renal infection" is more specific than "renal disorder" and "infection". The terms associated with the removed icons are re-associated with their parent icon (if several parents are present, the terms are associated with all of them and will be displayed in several panels, as previously). When the panel of the parent icon is clicked, the terms initially associated with the parent icon are displayed, as well as other panels for the child icons.

Preliminary Evaluation

As a preliminary evaluation, we showed the proposed interface to two pharmacovigilance experts working daily with the MedDRA terminology. We gathered their opinions and their comments using a qualitative form including the following 7 general questions: Is the interface useful for finding MedDRA terms when coding a new case? when gathering codes for performing an exhaustive search in a pharmacovigilance database? Do you search mainly at the PT level or the LLT level? Is the iconic search interesting? alone? in combination with the lexical search? Is a training necessary for using VCM icons?

Results

Iconic Interface for Browsing MedDRA

Figure 3 shows a screenshot of the proposed interface. On the left, the user can enter keywords, click pictograms on "Mister VCM" and select the desired MedDRA levels. These three

operations can be done in the order of his choice. The proposed system has a good performance: even when searching over the largest part of MedDRA (the tumor pictogram), the response time does not exceed 0.6 seconds.

Figure 3 shows the results of an iconic search with the "eye" pictogram. 3,650 terms were found and grouped according to 25 icons (more specific, children, icons exist but are not shown). When clicking on an icon panel, the entire list of associated terms is shown, as well as more specific icons. Figure 4 shows an example of such a panel, obtained after clicking on the "allergic inflammation of eye" icon on Figure 3 (it is the icon marked with a green dot). The panel shows 2 more specific child icons: "allergic inflammation of the optic nerve" and "allergic edema of eye".

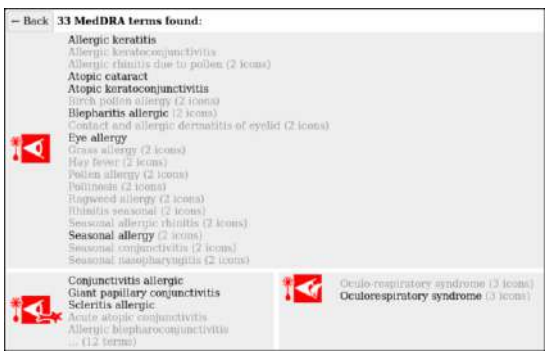


Figure 4– Screenshot of the panel shown when clicking the "allergic inflammation of eye" icon in Figure 3 (the one marked with a green dot).

Figure 5 shows an example of a search mixing lexical and iconic queries. We wanted to search for abscess in the urinary tract. "Abscess" is a clearly defined and specific keyword, so we typed it into the lexical search box. On the contrary, "urinary tract" is more general: many synonyms exist, and it also includes several subparts (*e.g.* ureter). Thus, we clicked on the "kidney" pictogram that represents the urinary system. The results show 2 icons with 17 terms.

When a term is clicked, it is displayed on a page with its label, its MedDRA code, its icons and its position in the MedDRA hierarchy, including ancestors and children terms. Ancestors and children can be clicked. As a consequence, the proposed interface also allows a classical tree-based navigation in the terminology.

Expert Opinions

The purely iconic search was not found very interesting by the two experts. On the contrary, the combination of the lexical and iconic search was described by them as very interesting, for example when dealing with synonymy (*e.g.* "auricular" keyword + heart pictogram to avoid auricular terms referring to ear) or for increasing the sensibility of the search, since most pictograms are rather general. This second point makes our interface particularly appealing to them when gathering codes for performing an exhaustive search. Moreover, the ability to select the depth level allows starting at a high level and then going down to lower levels if there is no result on the higher level.

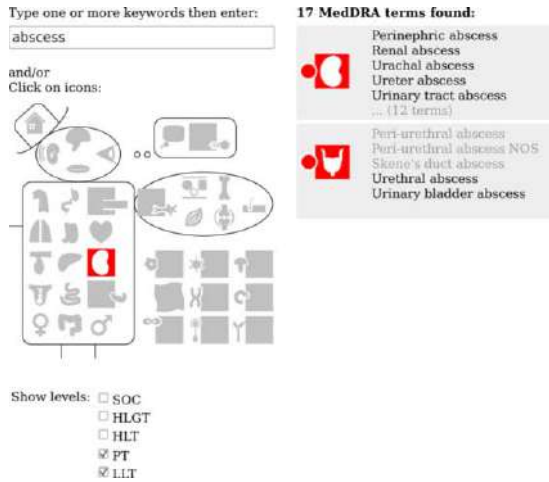


Figure 5– Screenshot of a lexico-iconic search: we searched for PT and LLT in MedDRA with the keyword "abscess" and pictogram "urinary".

On the contrary, when coding a new case, experts have to "stick" to the diagnostic mentioned in the case, and thus a lexical search through LLT is more appropriate. The iconic interface is thus less interesting for them. However, one expert qualified VCM icons as an "Esperanto of medical language". She considered that the iconic interface would be very useful for clinical research associates in the industry, because they are often non-specialists (*i.e.* not a physician or pharmacist) and could benefit from the guidance of icons: they may help to clearly identify the anatomic location or the morphology of a complex pathology, *e.g.* VCM icons explicitly represent "glomerulonephritis" as an inflammation of a part of the kidney while it may not be obvious to a non-specialist. In addition, many of them have to use MedDRA in English, even if it is not their native language.

Finally, both experts considered that training is needed before using VCM icons.

Discussion

In this paper, we proposed an original approach for browsing and searching medical terminologies, using pictograms and icons in addition to lexical keywords and terms. We applied this new approach to MedDRA. The resulting tool was found interesting by pharmacovigilance experts, in particular for gathering terms before performing an exhaustive search in a pharmacovigilance database, and for non-specialist users.

"Mister VCM" was initially designed to gather and organize a small set (5-50) of VCM icons. Here, we used it in a very different way, for allowing the selection of one or more pictograms in order to formulate a query in a visual way. This is the first time we use VCM for a querying purpose.

The icon enables a semantic search in the terminology. For example, the heart pictogram allows searching for all cardiac-related terms, without having to deal with the problems of synonymy or polysemy, while a lexical search would have to consider the many existing synonyms: "heart", "cardiac",

"card-", but also parts of the heart, such as "ventricular" or "coronary", and polysemy, *e.g.* "cardia" is not related to the heart but to the stomach, despite its short lexical distance which "cardiac" would suggest.

Some medical terminologies have an ontological structure that could be used for performing a semantic search. For instance, SNOMED CT relates all disorders to their finding site. These relations could be used to find all disorders related to the heart (including the heart subparts). However, the huge number of concepts in SNOMED CT makes this approach difficult and tedious to use. Moreover, ontological definitions are usually stricter than a human expert would expect, especially when performing broad search: *e.g.* in SNOMED CT, pericardium is not included in the heart because, strictly speaking, it is not a part of the heart. On the contrary, VCM pictograms are far less numerous, and they have a broader definition: *e.g.* the heart pictogram also covers the heart-related structure such as pericardium.

In the literature, Massari et al. [17] proposed meta-terms based on medical specialties to enrich terminologies and facilitate searches. Meta-terms can be combined with other types of searches. This approach is similar to the one we proposed here, but neither iconic nor visual.

Conclusions

Icons are a new, promising approach for browsing and searching medical terminologies. The perspectives of this work are twofold. First, the proposed interface will be properly evaluated in pharmacovigilance settings, using quantitative methods. Second, it could also be adapted to other medical terminologies such as ICD10 or SNOMED CT and used for helping physicians associate coded terms to their patients. Applications in medical education are also possible.

Acknowledgements

This work was supported by the French National Research Agency (ANR) through the Pegase project [grant number ANR-16-CE23-0011].

References

[1] Cimino J J. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med. 1998;37(4-5):394–403.

[2] Degoulet P, Sauquet D, Jaulent M C, Zapletal E, Lavril M. Rationale and design considerations for a semantic mediator in health information systems. Methods Inf Med. 1998;37(4-5):518–26.

[3] Bousquet C, Trombert B, Souvignet J, Sadou E, Rodrigues JM. Evaluation of the CCAM Hierarchy and Semi Structured Code for Retrieving Relevant Procedures in a Hospital Case Mix Database. In: AMIA Symposium. vol. 2010; 2010. p. 61–5.

[4] Sundvall E, Nyström M, Petersson H, Åhlfeldt H. Interactive visualization and navigation of complex terminology systems, exemplified by SNOMED CT. Stud Health Technol Inform. 2006;124:851–856.

[5] Brown EG. Methods and pitfalls in searching drug safety databases utilising the Medical Dictionary for Regulatory Activities (MedDRA). Drug Saf. 2003;26(3):145–58.

[6] Cornet R, Nyström M, Karlsson D. User-directed coordination in SNOMED CT. Stud Health Technol Inform. 2013;192:72–6.

[7] Bakhshi-Raiez F, de Keizer NF, Cornet R, Dorrepaal M, Dongelmans D, Jaspers MWM. A usability evaluation of a SNOMED CT based compositional interface terminology for intensive care. IJMI. 2012;81:351–362.

[8] Lee D, Cornet R, Lau F, de Keizer N. A survey of SNOMED CT implementations. J Biomed Inform. 2013;46(1):87–96.

[9] Souvignet J, Declerck G, Asfari H, Jaulent MC, Bousquet C. OntoADR a semantic resource describing adverse drug reactions to support searching, coding, and information retrieval. J Biomed Inform. 2016;63:100–107.

[10] Alecu I, Bousquet C, Degoulet P, Jaulent MC. PharmARTS: terminology web services for drug safety data coding and retrieval. Stud Health Technol Inform. 2007;129(Pt 1):699–704.

[11] Lamy JB, Duclos C, Bar-Hen A, Ouvrard P, Venot A. An iconic language for the graphical representation of medical concepts. BMC Medical Informatics and Decision Making. 2008;8:16.

[12] Lamy JB, Soualmia LF. Formalization of the semantics of iconic languages: An ontology-based method and four semantic-powered applications. Knowledge-Based System. 2017;135:159–179.

[13] Lamy JB, Tsopra R. Combining semantic and lexical methods for mapping MedDRA to VCM icons. In: Stud Health Technol Inform. vol. 247; 2018. p. 905–909.

[14] Lamy JB, Venot A, Duclos C. PyMedTermino: an open-source generic API for advanced terminology services. Stud Health Technol Inform. 2015;210:924–928.

[15] Lamy JB. Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies. Artif Intell Med. 2017;80:11–28.

[16] Lamy JB. Ontology-Oriented Programming for Biomedical Informatics. Stud Health Technol Inform. 2016;221:64–68.

[17] Massari P, Pereira S, Thirion B, Derville A, Darmoni S J. Use of super-concepts to customize electronic medical records data display. Stud Health Technol Inform. 2008;136:845–50.

Address for Correspondence
Jean-Baptiste Lamy <jibalamy@free.fr>
Bureau 149, LIMICS, UFR SMBH, Université Paris 13
74 rue Marcel Cachin
93017 Bobigny, France

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190215

An Empirical Test of GRUs and Deep Contextualized Word Representations on De-Identification

Kahyun Lee, Michele Filannino, Özlem Uzuner

Department of Information Science and Technology, George Mason University, Fairfax, VA, United States

Abstract

De-identification aims to remove 18 categories of protected health information from electronic health records. Ideally, de-identification systems should be reliable and generalizable. Previous research has focused on improving performance but has not examined generalizability. This paper investigates both performance and generalizability. To improve current state-of-the-art performance based on long short-term memory (LSTM) units, we introduce a system that uses gated recurrent units (GRUs) and deep contextualized word representations, both of which have never been applied to de-identification. We measure performance and generalizability of each system using the 2014 i2b2/UTHealth and 2016 CEGS N-GRID de-identification datasets. We show that deep contextualized word representations improve state-of-the-art performance, while the benefit of switching LSTM units with GRUs is not significant. The generalizability of de-identification system significantly improved with deep contextualized word representations; in addition, LSTM units-based system is more generalizable than the GRUs-based system.

Keywords:

Data Anonymization, Natural Language Processing, Machine Learning

Introduction

Removing identifiable information from personal data is important for protecting privacy. The General Data Protection Regulation (GDPR) in the European Union (EU) prohibits processing of personal data without obtaining explicit consent unless identifiable information is first removed from the data. In the United States, the Health Insurance Portability and Accountability Act (HIPAA) requires removal of 18 categories of protected health information (PHI) from electronic health records (EHRs) in order to protect the privacy of patients. The task of removing such information is often referred to as de-identification. This task is often the first step that must be performed before EHRs can be shared. In this paper, we focus on the de-identification of EHRs for HIPAA compliance. This task can be accomplished both manually and automatically. However, manual de-identification is non-reusable, time-consuming, and costly. For example, Dorr et al.[1] assessed the difficulty and time costs of manual de-identification in clinical notes. On average, manual de-identification of a clinical note, with an average note length of 261 ± 352 words, required 87.3 ± 61 seconds. Douglass et al.[2] found that each human annotator could read 16,000 to 20,000 words hourly and was paid $50 per hour. Given the vast amount of EHRs and the time and financial cost of manual de-identification, fully manual de-identification is not feasible and must benefit from automated methods.

The earliest automated de-identification systems were primarily rule-based [3–5]. Researchers utilized external knowledge resources, such as lists of personal names and addresses, as well as medical term dictionaries. They also implemented regular expressions to capture morpho-lexical features of PHI.

Recently, machine-learning systems have made significant contributions to de-identification. Most machine learning models interpret de-identification as a sequence-labeling problem [6–11]. These models can handle sequences of characters, words, or sentences. For the 2014 i2b2/UTHealth shared task on de-identification [12], Yang and Garibaldi [7] earned the highest F1-score among participants using a sequence-labeling classifier based on conditional random fields (CRFs) which were complemented with customized rules. All the systems that ranked among the top five in this shared task were based on CRFs. More recently, artificial neural networks (ANNs), which are also sequence labelers, have been used for de-identification. Among various architectures of ANNs, recurrent neural networks (RNNs) have been the most common. A type of RNNs, referred to as long short-term memory (LSTM) units combined with CRFs along with manually encoded rules [11] have displayed the best performance in the 2016 CEGS N-GRID de-identification shared task [13]. Interestingly, the best systems in both shared tasks complemented their machine learning solutions with hand-crafted rules that were optimized to their datasets. Independently of these shared tasks, Dernoncourt et al. [14] built NeuroNER, and achieved state-of-the-art performance on the 2014 i2b2/UTHealth dataset. NeuroNER is also based on bi-directional LSTM (bi-LSTM) units and CRFs but includes no hand-crafted features or external resources other than GloVe word embeddings [15]. NeuroNER implements the bi-LSTM units on a character-level input to get character embeddings and concatenates character embeddings with pre-trained token embeddings. The resulting character-enhanced token embeddings are fed into the bi-LSTM units again and the sequence of probabilities are then tuned with the CRF sequence optimizer to produce the system output.

The success of NeuroNER prompts some follow up questions. The first concerns the use of different recurrent units. Although RNN with traditional neural units could handle sequence-based problems such as de-identification, they cannot effectively handle tasks that contain long-term dependencies. This is often referred to as the vanishing gradient problem. To mitigate the vanishing gradient problem, researchers have devised either enhanced learning algorithms [16] or more sophisticated activation functions [17, 18]. Among these activation functions, LSTM units are the most popular. However, in 2014, Cho et al. [18] proposed a different type of gating mechanism, namely the gated recurrent units (GRUs). Like LSTM units, GRUs can address the vanishing gradient problem, but they also have simpler structure relative to LSTM units. While LSTM units contain “forget,” “update,” and “output” gates, GRUs merely have “reset” and “update” gates. Despite this structural advantage and comparable performance [19], GRUs have not

been evaluated in de-identification.

The second question entails the benefit of incorporating different combinations of embeddings. Although basic word representations such as GloVe [15] and Word2Vec [20] have become the core of most recent natural language-processing (NLP) research, they face challenges when handling out-of-vocabulary (OOV) words and ambiguity, i.e., words that do not have an embedding (such as misspellings, dates, telephone numbers, and ID numbers) and words that are ambiguous (such as the word 'can' as a modal in the sentence "I can do it" and the same word as a noun in the sentence "this is a trash can"). Character embeddings could compensate for some of the weaknesses of these representations providing character-level information that can encapsulate the semantics of words. In their ablation test to investigate the contribution of each component to the performance of NeuroNER, Dernoncourt et al. [14] determined that their system performed reasonably well using only character-level embeddings (i.e., without word embeddings). However, both word embeddings and character embeddings contain limited or no information about context, which can be pivotal for capturing the true semantics of words. Recently, Peters et al. [21] devised word representations that can incorporate context into token embeddings. These "deep contextualized word embeddings" are calculated from a bi-directional language model (biLM) and are evaluated in our experiments.

Finally, despite the high performance of existing machine-learning-based de-identification systems, these systems do not generalize well to new types of data. The result of the 2016 CEGS N-GRID shared task on "sight unseen" data supports this observation. This task tested the performance of existing, unmodified de-identification systems on new types of data consisting of psychiatry records and was the first shared task of its type that utilized psychiatry data. Nine teams participated in this task. Each team could train their system on any data of their choosing. All systems were evaluated on the same previously unseen test data. The highest-performing team on this task could not reach 0.80 in terms of strict entity-based F1-score [13].

To reliably remove PHI from EHRs in practice, however, de-identification systems should generalize to new types of data that are different from what they trained on. Systems tested on the 2016 CEGS N-GRID shared task on "sight unseen" data were each trained on different sets of data, which rendered comparison of generalizability of their models impossible. Here, we present a systematic evaluation of generalizability of our systems. To the best of our knowledge, this is the first use of GRUs and deep contextualized word representations in de-identification and provides a systematic assessment of generalizability of discussed solutions.

Methods

State-of-the-art in de-identification was developed by Dernoncourt et al. [14]. This system, NeuroNER, consists of four layers, including a character-embedding layer, a token-embedding layer, a label-prediction layer, and a label-sequence-optimization layer. NeuroNER utilizes bi-LSTM units on character-level input to get character embeddings and concatenates character embeddings with pre-trained token embeddings. The resulting character-enhanced token embeddings are fed into bi-LSTM units again, and the resulting sequence of probabilities are then tuned with a CRF label sequence optimizer.

We selected NeuroNER as our baseline model. We maintained its overall pipeline, and modified its components (as described in the following experiments) to evaluate GRUs and deep contextualized word representations in de-identification, and to assess generalizability of the resulting models.

Datasets

We evaluated de-identification performance on the 2014 i2b2/UTHealth and the 2016 CEGS N-GRID de-identification data. The 2014 i2b2/UTHealth de-identification dataset (henceforth, the 2014 dataset) consists of 1,304 medical records, including discharge summaries, progress notes, doctor's notes, doctor-patient communications from 296 diabetic patients. These records contain 28,872 PHI entities in 28 types (including subtypes for some HIPAA PHI categories) [12]. The 2016 CEGS N-GRID de-identification dataset (henceforth, the 2016 dataset) includes 1,000 psychiatric intake records containing 34,364 PHI entities with the same types [13]. Table 1 summarizes the statistics of each dataset. As indicated in the table, the 2016 dataset contains more tokens and PHI per record. Considering the ratio between the average token per document and the average PHI per document, the 2014 dataset is denser than the 2016 dataset.

Both the 2014 and 2016 datasets consisted of training (60%) and test (40%) sets. We divided each of the training sets into training (40%) and validation (20%) sets. We used the training and validation sets for system development. We report results on the test sets.

Table 1 – Overview of the 2014 and 2016 Datasets

	2014 Dataset	2016 Dataset
Records	1,304	1,000
Total # of tokens	805,118	1,862,452
Average # of tokens per record	617	1,862
Total # of PHI entities	28,872	34,364
Average # of PHI per record	22	34

Evaluation Metrics

For evaluation, we used micro-averaged precision, recall, and F1-score computed using the following equations:

- Precision (P) = true positives / (true positives + false positives)
- Recall (R) = true positives / (true positives + false negatives)
- F1-score (F1) = 2*P*R / (P+R)

Organizers of the 2014 and 2016 de-identification shared tasks measured multiple versions of these metrics (strict/relaxed/overlap with token-based/entity-based matching) over two sets of PHI types (i2b2 PHI vs. HIPAA PHI). Strict, relaxed, and overlap measurements differ in terms of their acceptance of exact or inexact textual spans corresponding to PHI. Token-based versus entity-based evaluation refers to whether the complete string or individual words in the string are considered for de-identification. i2b2 PHI and HIPAA PHI differ in terms of granularity of PHI. i2b2 categories classify the 18 HIPAA categories into sub-types (e.g., a location HIPAA PHI type can be classified into street, city, state, zip i2b2 types). In this paper, we evaluate performance with the strictest settings—strict entity-based matching over i2b2 PHI.

System Structure

NeuroNER serves as both the starting point and the benchmark for the experiments presented in this paper. Our goal is to improve performance of this state-of-the-art system on de-identification and to assess its generalizability under different conditions. We experiment with different RNNs and

embeddings in order to evaluate the contribution of each to system performance and generalizability.

Our system accepts tokenized narrative text as input. For this, the narratives are first pre-processed using spaCy [22] for sentence detection and tokenization. Given the pre-processed narratives and NeuroNER, we experimented with bi-LSTM units and bi-GRUs. LSTM units and GRUs are alike in that they both incorporate history into their predictions. But as previously mentioned, GRUs have simpler structure than LSTM units. As a result, GRUs have fewer parameters than LSTM units, and can converge faster [19]. In terms of their performance, the superiority or inferiority of either LSTM units or GRUs remains a matter of debate [19, 23], while the merits of both units relative to the traditional RNNs units (for example, tanh) are undeniable [24,25].

In order to evaluate the performance of these units, we tested three different types of embeddings with them. Pre-trained word embeddings [12,17] are widely used and have become a core element in recent machine-learning-based NLP research; however, they can fall short on OOV and ambiguous words. One way of resolving these cases would entail the use of character-enhanced token embeddings or deriving from deep contextualized word representations. Rather than training a word vector for each token, deep contextualized word representations train a context function that is based on two-layer biLMs with character convolutions and that returns a flexible word vector [21]. This function assigns a different vector to each word based on its morphological features and context; even the same two words in the same sentence can have different word vectors. Deep contextualized word representations can mitigate both OOV and ambiguous word problems. We used 1,024-dimensional deep contextualized word representations in addition to character embeddings and word embeddings. We processed our datasets (using only the training and validation portions) with a pre-trained model (which is a context function trained on the 1B Word Benchmark [26]) to extract deep contextualized word representations.

To investigate the contribution of different embeddings to different RNNs on de-identification, we experimented with all possible combinations of items within the four layers of NeuroNER architecture. The layers are summarized in Table 2.

Table 2 – Experiment Items in Each Architectural Layer of NeuroNER

Location	Experiment items
4th layer	CRF label sequence optimizer
3rd layer	bi-LSTM units vs. bi-GRUs
2nd layer	word embeddings (WEs) and/or deep contextualized word representations (DEs)
1st layer	character embeddings (CEs)

Training and Hyperparameters

We tuned our hyperparameters using the training set of each dataset. NeuroNER accomplished its optimal performance on both datasets with these parameters:

- character-embedding dimension: 25
- character-based token-embedding dimension: 25
- token-embedding dimension: 100
- label-prediction dimension: 100
- dropout probability: 0.5

In addition to the above hyperparameters, the dimension of the deep contextualized word representations was 1,024. Therefore, the inputs to the label-prediction layer entail a minimum of 25 dimensions (in the case that only character embeddings are tested) and a maximum of 1,149 dimensions (in the case that all three embeddings are tested). When training, iteration was stopped if there was no improvement in the F1-score on the validation set for 10 epochs.

Results

Performances of both LSTM units-based and GRUs-based RNNs are presented in Tables 3 and 4. The statistical significance of F1-score difference from the baseline model was tested using approximate randomization [27] with 9,999 shuffles. Statistically significant F1-score differences at the level of 0.01 are asterisked.

Table 3 indicates the performance of LSTM units-based RNNs with different combinations of embeddings. The baseline performance of NeuroNER is underlined, and the precision, recall, and F1-score of the highest-performing combination for each dataset is in bold font. In the 2014 dataset, the role of character embeddings was pivotal. Without word embeddings, LSTM units-based RNNs accomplished 89.70 for F1-score. Any combinations that included deep contextualized word representations significantly outperformed the combination of character embeddings and word embeddings, which were our baseline, in F1-score. Even the system that used deep contextualized word representations alone could outperform this baseline in F1-score. The highest F1-score achieved with LSTM units-based RNNs was 92.82, which is 1.15 higher than the baseline. In the 2016 dataset, the system with only character embeddings underperformed relative to the system with only word embeddings. Deep contextualized word representations worked effectively on the 2016 dataset as well, but it did not surpass the baseline system until it was combined with the other two embeddings. The highest F1-score was 89.00. In both the 2014 and 2016 datasets, the optimal performances were observed when all three embeddings were combined.

Table 3 – Performance of LSTM Units-Based RNNs

Embeddings	2014 Dataset			2016 Dataset		
	P	R	F1	P	R	F1
CEs	92.83	86.77	89.70*	84.96	78.58	81.65*
WEs	85.46	79.09	82.15*	85.92	79.99	82.85*
DEs	93.50	90.79	92.13*	88.23	86.56	87.39*
CEs+WEs	92.39	90.97	91.67	89.15	87.09	88.11
CEs+DEs	93.14	90.84	91.98*	88.65	86.95	87.79
WEs+DEs	93.34	90.91	92.11*	87.99	87.92	87.95
All three	94.33	91.36	92.82*	90.88	87.20	89.00*

Table 4 – Performance of GRUs-Based RNNs

Embeddings	2014 Dataset			2016 Dataset		
	P	R	F1	P	R	F1
CEs	91.98	87.30	89.58*	84.77	77.20	80.81*
WEs	86.86	77.81	82.09*	84.66	78.73	81.59*
DEs	93.04	91.23	92.13*	88.48	85.99	87.22*
CEs+WEs	93.77	90.49	92.10*	89.66	86.39	87.99*
CEs+DEs	93.63	90.33	91.95*	90.57	86.58	88.53*
WEs+DEs	94.07	90.74	92.37*	89.97	87.51	88.72*
All three	93.59	91.65	92.61*	89.11	87.10	88.09

Table 4 indicates the performance of the GRUs-based RNNs. Here, character embeddings were pivotal in the 2014 dataset but not on the 2016 dataset. The combination of GRUs with deep contextualized word representations achieved a good performance. The optimal F1-score in the 2014 dataset was received when all three embeddings were combined, while the combination of word embeddings and deep contextualized

word representations exhibited the second highest performance. The differences in F1-score are statistically significant at P=.008. Regarding the 2016 dataset, the optimal performance was produced when word embeddings and deep contextualized word representations were combined; combining character embeddings with these two embeddings deteriorated the performance of the system.

Comparing Tables 3 and 4, we see that GRUs demonstrated a comparable performance relative to the LSTM units and even outperformed the LSTM units in some combinations. Notably, the F1-score increased by 0.43 merely as a result of substituting the LSTM units with the GRUs for the 2014 dataset (which does not occur in the 2016 dataset). The differences in F1-score are statistically significant at P=0.002. However, the highest performance with the GRUs-based RNNs could not surpass that of the LSTM units-based RNNs for both datasets. The differences in F1-scores are statistically significant at P=0.004.

We also investigated the generalizability of both LSTM units-based and GRUs-based RNNs systems. To test the generalizability of each combinatio, we evaluated the performance of the model trained using the 2014 training set on the 2016 test set.

Table 5 – Generalizability of LSTM Units-Based RNNs

Embeddings	2014→2016			2016→2014		
	P	R	F1	P	R	F1
CEs	67.25	55.17	60.61*	57.67	62.76	60.11*
WEs	57.57	59.81	58.67*	57.08	42.65	48.82*
DEs	74.56	69.97	72.19*	70.95	69.51	70.22*
CEs+WEs	70.62	66.01	68.24	68.62	67.09	67.85
CEs+DEs	75.95	70.96	73.37*	69.37	70.84	70.10*
WEs+DEs	74.97	70.79	72.82*	70.92	70.42	70.67*
All three	77.51	68.30	72.61*	71.98	70.90	71.44*

Table 6 – Generalizability of GRUs-Based RNNs

Embeddings	2014→2016			2016→2014		
	P	R	F1	P	R	F1
CEs	72.85	54.02	62.04*	64.24	63.46	63.85*
WEs	61.08	58.87	59.95*	50.62	44.36	47.28*
DEs	71.78	70.17	70.97*	68.76	66.11	67.40*
CEs+WEs	76.68	66.82	71.41*	66.95	66.20	66.57*
CEs+DEs	74.55	71.04	72.75*	70.81	69.09	69.94*
WEs+DEs	76.82	69.19	72.81*	69.58	69.37	69.47*
All three	72.75	71.72	71.93*	73.18	69.12	71.09*

As Tables 5 and 6 demonstrate, combinations that performed effectively on training data also work effectively on new types of data. In addition, the introduction of deep contextualized word representations significantly increased the generalizability of both LSTM units and GRUs-based systems on both datasets. The differences in F1-score of the baseline and the models that include deep contextualized word representations are statistically significant at P<0.001. The gain in generalizability over the baseline is more significant than the gain in performance. The performance increase from the baseline to the highest F1-score of LSTM units-based system within the 2014 dataset was 1.15, while the increase in the performance while going across datasets was 5.13.

Discussion

Error Analysis

The distribution of PHI found by each RNN is shown in Table 7. Overall, both RNNs succeeded in finding 9,871 entities from the 2014 test set and 9,950 entities from the 2016 test set. In contrast, both units missed 1,302 entities in the 2014 test set and 3,339 entities in the 2016 test set. To investigate the strengths and weaknesses of each RNN, the number of entities found by only one of the units were counted. For example, "only GRUs found" indicates the number of PHI entities that all embeddings combinations with LSTM units failed to detect, but all embeddings combinations with the GRUs could find. These entities can be only found by GRUs regardless of embedding types, and they may reveal the empirical difference between LSTM units and GRUs.

Table 7 – Distribution of PHI by RNNs

	2014 Dataset	2016 Dataset
Total PHI in the test set	11,462	13,519
Both found	9,871	9,950
Both failed	1,302	3,339
Only GRUs found	243	104
Only LSTM units found	46	126

We extracted the sentences that include "only LSTM units found" and "only GRUs found" PHI entities and manually reviewed their characteristics. In many "only GRUs found" PHIs, the entities were directly adjacent to special characters such as slashes (/), colons (:), commas (,), and hyphens (-). For example, 102 DOCTOR entities, which are typically in the form of 'initial/name1/name2' or 'initial:name', could be found only by GRUs-based systems. It appears that GRUs are less sensitive to directly adjacent tokens in predicting the label of the current token. In contrast, PHI only found by LSTM units contain more contextually complex sentences. For example, in 41 cases, seasons and years (both are in DATE category) in sentences such as "healthy until spring of 0314 2126," "in the spring of 2128," and "in the summer of 2096" could not be found by GRUs, but they were detected by LSTM units. However, these findings need to be taken with a grain of salt given the small sample size.

Table 8 – Distribution of PHI by Embeddings

	2014 Dataset	2016 Dataset
Total PHI in the test set	11,462	13,519
Found only without CEs	347	175
Found only with CEs	131	249

Table 8 indicates the distribution of PHI for deep contextualized word representations with both RNNs. A total of 380 PHI entities were additionally found with these embeddings. These PHI entities were primarily centered around several entity types, such as PROFESSION (19%), HOSPITAL (15%), ORGANIZATION (15%), PATIENT (12%), and DOCTOR (11%). This distribution differs from the original PHI distribution of the test set: PROFESSION (5%), HOSPITAL (9%), ORGANIZATION (3%), PATIENT (7%), and DOCTOR (14%). From a practical perspective, 'PATIENT-DOCTOR' and 'HOSPITAL-ORGANIZATION' pairs are difficult to be differentiated from each other because of the semantic similarity between these PHI categories. Through manual review, we found these PHI entities were in complete sentences in most cases. This illustrates that a pre-trained context function returns better word representations when inputs are lengthy and sufficiently informative to make a machine understand context. Gains in the 2016 dataset are nearly twice as abundant as those in the 2014 dataset perhaps because EHRs in the 2016 dataset are more organized and structured than those in the 2014 dataset. As a result, the sentence detector worked more effectively on the 2016 dataset, providing cleaner input to the RNNs.

Miscellaneous

On average, GRUs show higher per-category F1-score on entity types CITY, COUNTRY, DOCTOR, HOSPITAL, IDNUM, PHONE, PROFESSION, and STATE whereas LSTM units perform better on AGE, DATE, MEDICALRECORD, ORGANIZATION, PATIENT, STREET, USERNAME, and ZIP. Interestingly, deep contextualized embeddings achieve higher per-category F1-score than both character and word embeddings regardless of entity types. Theoretically, the convergence of GRUs-based systems should be faster than LSTM units-based systems because GRUs have simpler structure and fewer parameters than LSTM units. This was confirmed by the experiments. For the 2014 dataset, convergence required an average of 37 epochs for the LSTM units-based systems, whereas for GRUs-based systems, convergence required an average of 28 epochs. Similarly, in the 2016 dataset, LSTM units required an average of 32 epochs, whereas GRUs required 30 epochs.

Conclusions

To automate de-identification tasks, de-identification must be highly accurate and generalizable. State-of-the-art de-identification systems are based on RNNs with bi-LSTM units. This study investigated avenues for improving the state-of-the-art by modifying types of recurrent units and embeddings. It also investigated how these modifications affect generalizability. We found that substituting LSTM units with the GRUs cannot significantly improve performance or generalizability. However, this does not imply the inferiority of GRUs relative to LSTM units on de-identification. GRUs outperformed LSTM units in some PHI types and could complement LSTM units in an ensemble. They could also converge to optimal parameters faster than LSTM units. The introduction of deep contextualized word representations on top of character embeddings and word embeddings was certainly helpful in increasing both the state-of-the-art performance and generalizability.

References

[1] D.A. Dorr, W.F. Phillips, S. Phansalkar, S.A. Sims, and J.F. Hurdle, Assessing the difficulty and time cost of de-identification in clinical narratives, *Methods Inf. Med.* **45** (2006) 246–252.

[2] M. Douglass, G.D. Clifford, a Reisner, G.B. Moody, and R.G. Mark, Computer-assisted De-identification of free text in the MIMIC II database, Comput. Cardiol. 31 (2004) 341–344.

[3] L. Sweeney, Replacing personally-identifying information in medical records, the Scrub system., *AMIA Annu Symp Proc.* (1996) 333–7.

[4] S.M. Thomas, B. Mamlin, G. Schadow, and C. McDonald, A successful technique for removing names in pathology reports using an augmented search and replace method., *AMIA Annu Symp*. (2002) 777–81.

[5] F.J. Friedlin, and C.J. McDonald, A Software Tool for Removing Patient Identifying Information from Clinical Documents, *J. Am. Med. Informatics Assoc.* **15** (2008) 601–610.

[6] O. Ferrández, B.R. South, S. Shen, F. Jeffrey Friedlin, M.H. Samore, and S.M. Meystre, BoB, a best-of-breed automated text de-identification system for VHA clinical documents, J. Am. Med. Informatics Assoc. **20** (2013) 77–83. doi:10.1136/amiajnl-2012-001020.

[7] H. Yang, and J.M. Garibaldi, Automatic detection of protected health information from clinic narratives, *J. Biomed. Inform.* **58** (2015) S30–S38.

[8] B. He, Y. Guan, J. Cheng, K. Cen, and W. Hua, CRFs based de-identification of medical records, *J. Biomed. Inform.* **58** (2015) S39–S46.

[9] Z. Liu, Y. Chen, B. Tang, X. Wang, Q. Chen, H. Li, J. Wang, Q. Deng, and S. Zhu, Automatic de-identification of electronic medical records using token-level and character-level conditional random fields, *J. Biomed. Inform.* **58** (2015) S47–S52.

[10] H.-J. Lee, Y. Wu, Y. Zhang, J. Xu, H. Xu, and K. Roberts, A hybrid approach to automatic de-identification of psychiatric notes, *J. Biomed. Inform.* (2017).

[11] Z. Liu, B. Tang, X. Wang, and Q. Chen, De-identification of clinical notes via recurrent neural network and conditional random field, *J. Biomed. Inform.* **75** (2017) S34–S42.

[12] A. Stubbs, C. Kotfila, and Ö. Uzuner, Automated systems for the de-identification of longitudinal clinical narratives: Overview of 2014 i2b2/UTHealth shared task Track 1, *J. Biomed. Inform.* **58** (2015) S11–S19.

[13] A. Stubbs, M. Filannino, and Ö. Uzuner, De-identification of psychiatric intake records: Overview of 2016 CEGS N-GRID shared tasks Track 1, *J. Biomed. Inform.* **75** (2017) S4–S18.

[14] F. Dernoncourt, J.Y. Lee, O. Uzuner, and P. Szolovits, De-identification of patient notes with recurrent neural networks, *J. Am. Med. Informatics Assoc.* **24** (2017) 596–606.

[15] J. Pennington, R. Socher, and C. Manning, Glove: Global Vectors for Word Representation, *Proc. 2014 Conf. Empir. Methods Nat. Lang. Process.* (2014) 1532–1543.

[16] Y. Bengio, N. Boulanger-Lewandowski, and R. Pascanu, Advances in optimizing recurrent networks, *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.* (2013) 8624–8628.

[17] S. Hochreiter, and J. Schmidhuber, Long Short-Term Memory, *Neural Comput.* **9** (1997) 1735–1780.

[18] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, (2014).

[19] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling, (2014) 1–9.

[20] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, Distributed Representations of Words and Phrases and their Compositionality, (2013) 1–9.

[21] M.E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, Deep contextualized word representations, (2018).

[22] M. Honnibal, and I. Montani, spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing, To Appear. (2017).

[23] D. Bahdanau, K. Cho, and Y. Bengio, Neural Machine Translation by Jointly Learning to Align and Translate, (2014) 1–15.

[24] R. Jozefowicz, W. Zaremba, and I. Sutskever, An Empirical Exploration of Recurrent Network Architectures, in Proc. 32nd Int. Conf. Mach. Learn., Lille, France, 2015.

[25] A. Karpathy, J. Johnson, and L. Fei-Fei, Visualizing and Understanding Recurrent Networks, (2015) 1–12.

[26] C. Chelba, T. Mikolov, M. Schuster, Q. Ge, T. Brants, T. Robinson, P. Koehn, P. Koehn, and T. Robinson, One Billion Word Benchmark for Measuring Progress in Statistical Language Modeling, INTERSPEECH. (2014).

[27] E.W. Noreen, Computer-intensive Methods for Testing Hypotheses: An Introduction, John Wiley & Sons, Inc, New York, 1989.

Address for correspondence

Kahyun Lee
klee70@gmu.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190216

Model Performance Metrics in Assessing the Value of Adding Intraoperative Data for Death Prediction: Applications to Noncardiac Surgery

Victor J. Lei[ab], Edward H. Kennedy[c], ThaiBinh Luong[d], Xinwei Chen[b], Daniel E. Polsky[b], Kevin G. Volpp[abe], Mark D. Neuman[bf], John H. Holmes[bg], Lee A. Fleisher[bf], Amol S. Navathe[abe]

[a] *Department of Medical Ethics and Health Policy, University of Pennsylvania, Philadelphia, Pennsylvania, USA,*
[b] *Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, Pennsylvania, USA,*
[c] *Department of Statistics & Data Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA*
[d] *Predictive Healthcare, University of Pennsylvania Health System, Philadelphia, Pennsylvania, USA,*
[e] *Veterans Health Administration, Department of Veterans Affairs, Philadelphia, Pennsylvania, USA,*
[f] *Department of Anesthesiology, University of Pennsylvania Health System, Philadelphia, Pennsylvania, USA,*
[g] *Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, Pennsylvania, USA*

Abstract

We tested the value of adding data from the operating room to models predicting in-hospital death. We assessed model performance using two metrics, the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC), to illustrate the differences in information they convey in the setting of class imbalance. Data was collected on 74,147 patients who underwent major noncardiac surgery and 112 unique features were extracted from electronic health records. Sets of features were incrementally added to models using logistic regression, naïve Bayes, random forest, and gradient boosted machine methods. AUROC increased as more features were added, but changes were small for some modeling approaches. In contrast, AUPRC, which reflects positive predicted value, exhibited improvements across all models. Using AUPRC highlighted the added value of intraoperative data, not seen consistently with AUROC, and that with class imbalance AUPRC may serve as the more clinically relevant criterion.

Keywords:
Medical Informatics; Hospital Mortality; Risk Assessment

Introduction

Clinicians routinely use data from various sources to make diagnostic assessments and treatment decisions. However, recent insights from behavioral economics and psychology highlight the limits of human decision-making, particularly with respect to synthesizing vast amounts of data with high variety and velocity [1-3]. The advent of big data science and predictive analytics holds the potential to improve key clinical tasks such as risk-stratifying patients at risk of poor outcomes. Doing so, however, will require careful alignment between methods and clinical use of predictions.

One key area of high clinical interest for predictive analytics is risk-stratification of patients for an outcome of death, particularly after surgery. There is expansive literature on this because of the frequency of noncardiac surgery. In 2010 alone, more than 25 million major noncardiac procedures were performed in the United States [4]. Though rates of postoperative mortality can be as low as 2%, postoperative mortality rates increase beyond 10% in high-risk subgroups [5-7]. Clinicians and researchers have thus developed various approaches to risk-stratify patients for postoperative mortality risk. The most widely used system is the American Society of Anesthesiologist (ASA) physical status classification score, which represents a clinician's assessment of a patient's risk of postoperative mortality based on data available before surgery starts [8]. However, the human decision-making element presents major inter-rater reliability issues with respect to accuracy and consistency of scores [9]. There are several other clinical risk scores that use data and patient characteristics before surgery, such as POSPOM, Risk Stratification Index (RSI) and Risk Quantification Index (RQI). [5; 10]. Such models obtain very high area under the receiver operating characteristic curve (AUROC) values ranging from 0.8 to 0.94. The Surgical Apgar Score takes a simpler approach and uses three intraoperative variables: estimated blood loss, lowest mean arterial pressure, and lowest heart rate measured during surgery to predict major postoperative complications [11]. The Surgical Apgar Score performs variably across studies with AUROC values from 0.6 to 0.86 [12-14]. One study added Surgical Apgar Score variables to RQI variables, but found little improvement in predicting postoperative death versus RQI itself [14]. This may be an indication that intraoperative information does not add value to postoperative mortality risk. However, there are two outstanding issues. First, it is still unclear whether and to what extent intraoperative data can improve the prediction of postoperative mortality. Second, assessing model performance with AUROC may be problematic because it is not aligned with the information clinicians seek, and thus other metrics need to be tested in assessing the value of intra-operative data.

Although AUROC serves as an important indicator for model performance, it is not necessarily aligned with how clinicians would use a prediction model in some settings. In the context of noncardiac surgery, death is a relatively rare outcome, so the importance of identifying high-risk patients who may die during their admission is more critical than identifying patients who do not die (the majority class). AUROC is immune to changes in class distribution because it is calculated only using sensitivity (i.e., recall, true positive rate) and specificity (i.e., false positive rate) [15] and therefore cannot distinguish between models that have equal sensitivity but different precision (i.e., positive predictive value (PPV)). In our clinical context, the precision--the probability of correctly identifying patient death out of all predicted deaths--is a more clinically meaningful measure than AUROC.

In this paper, we examine how adding intraoperative data to baseline (pre-hospitalization) and preoperative data can improve prediction of postoperative mortality using AUROC

and AUPRC to assess performance. We first show that AUROC metrics are very high even when using just baseline and preoperative data, which may lead to the conclusion that adding intraoperative data does not add any predictive information. We then use AUPRC, which reflects PPV as used in clinical decision-making, and demonstrate that adding intraoperative data results in improved predictions. These improvements are robust in multiple predictive modeling approaches. Lastly, we utilize clinician assessments of risk, specifically the ASA physical status score, to show that the AUPRC in models with intraoperative data contains substantial information that covers most of clinician assessment information.

Methods

Data Sources and Collection

Data was extracted from various databases across the University of Pennsylvania Health System (UPHS) and integrated into a comprehensive perioperative dataset (data containing administrative, preoperative, intraoperative, and postoperative information related surgeries). Administrative and clinical data were collected from a clinical data warehouse, Penn Data Store, and an electronic health record (EHR) (Epic Systems, USA). Intraoperative data was extracted from a perioperative data repository constructed from a set of transformation scripts run on EPIC's Clarity databases developed by the Multicenter Perioperative Outcomes Group who work to standardize perioperative information for academic research [16]. The study design and data access were approved by the University of Pennsylvania Institutional Review Board Protocol 824057.

Patient Inclusion and Exclusion Criteria

Patients 18 years or older who underwent major noncardiac surgery across four academic medical centers within the University of Pennsylvania Health System in Pennsylvania, United States between June 2012 to June 2017 were included in the sample. Patients who underwent multiple major surgeries during the same visit were excluded.

Major Noncardiac Surgery

Noncardiac surgery was identified using Current Procedural Terminology (CPT®) codes specification ranges for noncardiac surgeries 10021-32999 and 34001-69990 [17]. The Agency for Healthcare Research Quality Healthcare Cost Utilization Project (AHRQ HCUP) Surgery Flag Software was used to identify surgeries classified as major therapeutic procedures [18].

In-Hospital Mortality

The primary outcome for the predictive models was all-cause, in-hospital mortality, represented as a binary variable.

Model Performance

Cohort data was randomly split by the patient into training (60%), validation (20%), and test (20%) sets as defined by Hastie et al [4], with similar incidence of in-hospital mortality in the resulting sets. All models were assessed on test data. DeLongs and Wilcoxon-Mann Whitney tests were used to test for significant differences on 1000 Bootstrap samples for each model.

Imbalanced Outcome Class

Death is clearly a very important and clinically relevant postoperative complication and is quite infrequent for patients undergoing major noncardiac surgeries. This class imbalance may result in challenges in assessing model performance. We used two metrics to help assess overall model performance.

Area under the Receiver Operating Characteristic Curve (AUROC)

The AUROC metric represents model discrimination across various thresholds for a classification problem. The AUROC depends on sensitivity (recall, TPR) and specificity (FPR), and is the probability that the model will rank a randomly chosen positive case (e.g. death) higher than a randomly chosen negative case (e.g. non-death) [15]. AUROC values range from 0 to 1, with 1 indicating perfect model discrimination. When an AUROC is 0.5, it means the model has no class separation capacity and is no better than random classification.

Area under the Precision-Recall Curve (AUPRC)

Recent literature suggests other ways to examine model performance in skewed data. Davis and Goadrich introduced Precision-Recall curves for interpreting binary classification models in highly skewed datasets. In particular, optimizing on AUROC may not optimize the AUPRC [19]. Additionally, the AUPRC ignores the identification of true negative cases. In clinical practice, clinicians often care more about whether a patient is high risk for a poor outcome rather than low risk.

Defined mathematically, let Y and Z(t) be the true and predicted statuses at threshold t, respectively. As with AUROC, the AUPRC also considers sensitivity (i.e., recall, or the proportion high-risk classified as such)

$$\mathrm{Rec}(t) = \Pr\{Z(t)=1 \mid Y=1\}$$

but instead of specificity, it takes into account positive predictive value (i.e., precision, or the proportion classified as high-risk who actually are high-risk)

$$\mathrm{Prec}(t) = \Pr\{Y=1 \mid Z(t)=1\}$$

Then AUPRC is computed as the area under the curve traced across different thresholds t. As a simple example, suppose 5/10,000 patients die and we have two tests that correctly identify all 5 such patients. One test says 5 additional patients will die, while another says 20 additional patients will die. Clearly the first test is preferable; however, the false positive rates will both be 0.05% or less, whereas the PPVs are 50% and 20%, respectively. Thus, AUROC may not distinguish between the tests while AUPRC would.

Data Cleaning

EHR data frequently include missing data and erroneous values. We used a flexible approach to missing data by including dichotomous indicator variables for each feature, equaling one if the observation contained missing data or zero otherwise. This allowed the chance of death to differ across those who had missing values versus not, which is more flexible and less stringent than the common "missing completely at random" assumption required in multiple imputations. Additionally, to account for out-of-range values, we Winsorized outliers to the 1st or 99th percentile by feature.

Model Input Features

Each patient visit corresponded to a unique surgical record where features were calculated or extracted from the start of the admission until the end of the surgery. Features were split into three respective groups: Baseline, Preoperative, and Intraoperative. We also used ASA score as a separate feature to compare additive information from human judgement on top of the other objective data features. A sample of features included in models are shown in Table 2.

Baseline features included basic demographic information, like age, gender, race, and insurance type. Patient Elixhauser and Charlson comorbidity indices were also included, derived from the International Classification of Diseases (ICD9-CM/ICD10-CM) diagnostic codes [20-22].

The preoperative features contained clinical information related to the patient's admission, such as basic and extended lab measurements, medication use, and surgical procedure type. To categorize surgeries, we used AHRQ HCUP Clinical Classification Software to map each primary CPT code to 244 unique procedure groups [23]. Data for these variables was collected from the start of the admission up until the start of the surgery procedure.

Intraoperative features included descriptive intraoperative vital signs, such as minimum and maximum heart rate and blood pressure; fluid status, such as total fluid administration and estimated blood loss; and drug use, such as vasopressors and rescue medications. Data for this category was collected between the start and end of the surgery.

Modeling

We examined how adding intraoperative data to pre-hospitalization and preoperative data improved prediction across various statistical and machine learning classification models. We used four predictive modeling approaches: logistic regression, naïve Bayes, random forest, and gradient boosted machine (GBM). Logistic regression and naïve Bayes are well established statistical modeling techniques that are frequently used in predictive modeling. Random Forest and GBMs are robust machine learning ensemble classification methods. Large numbers of variables precluded the use of logistic models without feature reduction due to lack of convergence. All other models utilized all model inputs.

Feature Reduction for Logistic Regression

We used two approaches: (1) the Forward Selection (FORWARD) technique (SAS procedure HPGENSELECT), which sequentially adds variables based on model fit, terminating when no significant improvement is achieved by adding remaining features and (2) the Least Absolute Shrinkage and Selection Operator (LASSO) technique [4].

Random Forest and GBM Model Parameter Optimization

We executed a randomized grid search on our 60% training dataset for hyperparameter tuning, using three folds across 30 iterations and optimizing for PPV. We used the 20% validation set to evaluate, verify, and finalize our model parameters. Reported performance is based on the remaining 20% sequestered test dataset.

Software and Tools

Logistic regression with forward selection and LASSO models for this paper was generated using SAS software (Copyright © 2018 SAS Institute Inc). All other codes and predictive models were conducted in Python 3.6 with Pandas 0.23.3 and Scikit-learn 0.19.1 libraries.

Results

Patient Sample

Of the 71,147 adult single major noncardiac surgery encounters in our sample, 0.78% (n = 512) died in the hospital. The distribution of outcomes and preoperative clinical characteristics did not differ between training, validation, and holdout cohorts.

Patient characteristics differed significantly between those who did and did not die after surgery while in the hospital [Table 1]. Of note, the mean age of patients who died after surgery was significantly higher at 65 years (SD ± 16.1 years) vs 56 years (SD ± 16.8 years). Additionally, patients who died were proportionally more likely to be male, have Medicare, and a higher ASA score of 3 and 4 versus 2 and 3. There was also a significant difference in the mix of surgery types.

Table 1. Abbreviated Baseline Characteristics

Characteristic	In-hospital mortality cohort, n = 512 (%)	No in-hospital mortality cohort, n = 73,635 (%)
Age, mean (SD)	65 (16.1)	56 (16.8)
Gender		
Male	296 (57.81)	30,883 (41.94)
Race		
White	275 (53.71)	48,378 (65.7)
Insurance Class		
Commercial	143 (27.93)	35,341 (47.99)
Medicare	306 (59.77)	26,964 (36.62)
Surgery Type		
Breast/Skin	16 (3.12)	4,778 (6.49)
Endocrine	4 (0.78)	1,186 (1.61)
Ear/Nose/Throat	1 (0.2)	994 (1.35)
General	155 (30.27)	14,681 (19.94)
Gynecologic	2 (0.39)	4,425 (6.01)
Neurologic	89 (17.38)	9,347 (12.69)
Obstetrics	1 (0.2)	4,111 (5.58)
Ophthalmologic	2 (0.39)	137 (0.19)
Orthopedic	55 (10.74)	24,128 (32.77)
Thoracic	24 (4.69)	2,498 (3.39)
Tracheostomy	71 (13.87)	648 (0.88)
Transplant	3 (0.59)	948 (1.29)
Urologic	3 (0.59)	1,958 (2.66)
Vascular	86 (16.8)	3,566 (4.84)
ASA Rating		
ASA 2	12 (2.34)	33,590 (45.62)
ASA 3	232 (45.31)	33,891 (46.03)
ASA 4	227 (44.34)	2,693 (3.66)
ASA 5	39 (7.62)	42 (0.06)
ASA Unknown	2 (0.39)	170 (0.23)

Features

A total of 112 features were extracted for model development. 53 features were identified as baseline features. 34 features were identified as preoperative features. 25 features were identified as intraoperative or during surgery features. Finally, the ASA score was extracted.

Table 2. Model Features Examples

Baseline (n = 53)	Preoperative (n = 34)	Intraoperative (n = 25)
• Gender • Race • Insurance • Marital status • Elixhauser comorbidities	• CCS_Category • Serum creatinine level • Troponin-T level • White blood cell count • Insulin administration	• Lowest heart rate • Lowest mean arterial pressure • Max crystalloid administration • Total epinephrine dose • Total surgery duration

Model Performance

Table 3 compares the predictive performance of the various modeling approaches over sequential feature set additions. All models experienced the largest gain in predictive performance when adding preoperative features to baseline features, increasing AUROCs by 0.120 up to 0.272. Adding intraoperative features led to smaller increases (up to 0.059 for logistic regression with forward selection), with the LASSO model actually experiencing a 0.049 decrease in AUROC. AUROC increases were only significant for GBM and the logistic regression with forward selection models ($p = 0.036$ and $p < 0.001$, respectively). Further, adding the ASA score was not significant across models and did not lead to meaningful improvements for any modeling approach.

In contrast to AUROC, the values for AUPRC were much lower (as noted earlier, reflecting PPV rather than prioritizing low false positive rate). Also, in contrast to results observed for AUROC, the model performance on AUPRC demonstrated large increases for every model when adding intraoperative data to baseline and preoperative data. These gains were as large as 0.148 for logistic regression with forward selection and 0.075 for random forest. The AUPRC increase in each model was significant ($p < 0.001$), except when adding ASA score, which did not meaningfully increase further. These results were consistent across modeling approaches except for the LASSO model, which demonstrated worse performance with intraoperative data.

Discussion

Our study has two important findings. First, this study demonstrates that adding intraoperative data to models predicting postoperative death after noncardiac surgery yields meaningful improvements in performance. Second, the metric used to evaluate model performance was very important to inference, highlighting limitations of AUROC in some settings such as those with class imbalance where the minority class is more important clinically. Interestingly, further adding human clinician judgement after adding intraoperative data did not provide incremental improvements. Taken together, these results indicate that intraoperative data has important clinical information that can be used to better risk-stratify patients.

Importantly, the class imbalance that exists in many mortality prediction problems rendered the AUROC metric less useful in assessing the additional value of intraoperative data than an alternative metric, the AUPRC. In fact, without examining AUPRC, our study may have yielded the opposite main conclusion, erroneously inferring little value of intraoperative data in mortality prediction for noncardiac surgery. This point is further highlighted by the greater alignment between the clinical utility of the prediction and AUPRC, rather than AUROC, because in this setting of class imbalance (low death rate) PPV is clinically more important than false positive rate.

Lastly, it is noteworthy that while most modelling approaches yielded similar inferences on predictive performance improvements, there were substantial variations in performance and performance improvement across them. GBM, random forest, and logistic regression with forward selection seemed to perform the best and extract additional information from the intraoperative data.

Future work could examine the conditions under which methods perform better than others and should use mortality outcomes outside the hospital setting after 30 and 60 days to examine generalizability. This information may be important in understanding if predictive ability begins to stagger or decrease when the outcome of interest is farther away from the surgical exposure.

Conclusions

AUROC may not be the preferred metric for model performance in imbalanced datasets under certain conditions, such as when PPV is more important. Using AUPRC indicated that adding intraoperative data to baseline and preoperative data improved prediction of postoperative mortality. These improvements are robust to multiple machine learning model approaches. Further, combining intraoperative with preoperative data captures most of the information/predictive value within clinician assessments of risk as measured by the ASA physical status classification.

Acknowledgements

This project is funded, in part, under a grant with the Pennsylvania Department of Health. The Department specifically disclaims responsibility for any analyses, interpretations or conclusions. The corresponding author has nothing to disclose. Dr. Navathe reports receiving grants from Hawaii Medical Service Association, Anthem Public Policy Institute, and Oscar Health, personal fees from Navvis and Company, Navigant Inc., Lynx Medical, Indegene Inc., Sutherland Global Services, and Agathos, Inc.; personal fees and equity from NavaHealth; and an honorarium from Elsevier Press, none of which are related to this manuscript. Dr. Holmes receives funding from the Pennsylvania Department of Health, US Public Health Service, and the Cardiovascular Medicine Research and Education Foundation. Dr. Volpp reports grants from Humana during the conduct of the study; grants from Hawaii Medical Services Agency, Discovery (South Africa), Merck, Weight Watchers, and CVS outside of the submitted

Table 3. Model Performance

	Baseline		Baseline + Preoperative		Baseline + Preoperative + Intraoperative		Baseline + Preoperative + Intraoperative + ASA Score	
Model	**AUROC**	**AUPRC**	**AUROC**	**AUPRC**	**AUROC**	**AUPRC**	**AUROC**	**AUPRC**
Logistic Regression								
Forward Selection	0.750	0.033	0.870	0.151	0.929	0.297	0.936	0.295
LASSO	0.705	0.035	0.831	0.097	0.782	0.065	0.782	0.065
Naïve Bayes	0.728	0.019	0.859	0.045	0.888	0.065	0.896	0.066
Random Forest	0.716	0.018	0.907	0.201	0.921	0.276	0.936	0.260
GBM	0.645	0.014	0.917	0.235	0.938	0.282	0.945	0.282

* AUROC increases were only significant for GBM and the logistic regression with forward selection models ($p = 0.036$ and $p < 0.001$, respectively). AUPRC increases in all models were significant ($p < 0.001$), except for models adding ASA score.

work; he has received consulting income from CVS and VALHealth and is a principal in VALHealth, a behavioral economics consulting firm. All other authors declare no conflicts of interest.

References

[1] N.K. Arora, B.W. Hesse, B.K. Rimer, K. Viswanath, M.L. Clayman, and R.T. Croyle, Frustrated and confused: the American public rates its cancer-related information-seeking experiences, *J Gen Intern Med* **23** (2008), 223-228.

[2] I. Klerings, A.S. Weinhandl, and K.J. Thaler, Information overload in healthcare: too much of a good thing?, *Z Evid Fortbild Qual Gesundhwes* **109** (2015), 285-290.

[3] G. Miller, The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information, in, 1956, pp. 81-97.

[4] T. Hastie, R. Tibshirani, and J.H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2009.

[5] Y. Le Manach, G. Collins, R. Rodseth, C. Le Bihan-Benjamin, B. Biccard, B. Riou, P.J. Devereaux, and P. Landais, Preoperative Score to Predict Postoperative Mortality (POSPOM): Derivation and Validation, *Anesthesiology* **124** (2016), 570-579.

[6] R.M. Pearse, D.A. Harrison, P. James, D. Watson, C. Hinds, A. Rhodes, R.M. Grounds, and E.D. Bennett, Identification and characterisation of the high-risk surgical population in the United Kingdom, *Crit Care* **10** (2006), R81.

[7] R.M. Pearse, R.P. Moreno, P. Bauer, P. Pelosi, P. Metnitz, C. Spies, B. Vallet, J.L. Vincent, A. Hoeft, A. Rhodes, M. European Surgical Outcomes Study group for the Trials groups of the European Society of Intensive Care, and A. the European Society of, Mortality after surgery in Europe: a 7 day cohort study, *Lancet* **380** (2012), 1059-1065.

[8] A.S. Keats, The ASA classification of physical status--a recapitulation, *Anesthesiology* **49** (1978), 233-236.

[9] P. Cuvillon, E. Nouvellon, E. Marret, P. Albaladejo, L.P. Fortier, P. Fabbro-Perray, J.M. Malinovsky, and J. Ripart, American Society of Anesthesiologists' physical status system: a multicentre Francophone study to analyse reasons for classification disagreement, *Eur J Anaesthesiol* **28** (2011), 742-747.

[10] M.J. Sigakis, E.A. Bittner, and J.P. Wanderer, Validation of a risk stratification index and risk quantification index for predicting patient outcomes: in-hospital mortality, 30-day mortality, 1-year mortality, and length-of-stay, *Anesthesiology* **119** (2013), 525-540.

[11] A.A. Gawande, M.R. Kwaan, S.E. Regenbogen, S.A. Lipsitz, and M.J. Zinner, An Apgar score for surgery, *J Am Coll Surg* **204** (2007), 201-208.

[12] A.B. Haynes, S.E. Regenbogen, T.G. Weiser, S.R. Lipsitz, G. Dziekan, W.R. Berry, and A.A. Gawande, Surgical outcome measurement for a global patient population: validation of the Surgical Apgar Score in 8 countries, *Surgery* **149** (2011), 519-524.

[13] S.E. Regenbogen, J.M. Ehrenfeld, S.R. Lipsitz, C.C. Greenberg, M.M. Hutter, and A.A. Gawande, Utility of the surgical apgar score: validation in 4119 patients, *Arch Surg* **144** (2009), 30-36; discussion 37.

[14] M.A. Terekhov, J.M. Ehrenfeld, and J.P. Wanderer, Preoperative Surgical Risk Predictions Are Not Meaningfully Improved by Including the Surgical Apgar Score: An Analysis of the Risk Quantification Index and Present-On-Admission Risk Models, *Anesthesiology* **123** (2015), 1059-1066.

[15] T. Fawcett, An introduction to ROC analysis, *Pattern Recognition Letters* **27** (2006), 861-874.

[16] R.E. Freundlich and S. Kheterpal, Perioperative effectiveness research using large databases, *Best Pract Res Clin Anaesthesiol* **25** (2011), 489-498.

[17] CPT® (Current Procedural Terminology), in, American Medical Association, 2018.

[18] Surgery Flag Software in, AHRQ Healthcare Cost and Utilization Project, 2017.

[19] J. Davis and M. Goadrich, The relationship between Precision-Recall and ROC curves, in: *Proceedings of the 23rd international conference on Machine learning*, ACM, Pittsburgh, Pennsylvania, USA, 2006, pp. 233-240.

[20] M.E. Charlson, P. Pompei, K.L. Ales, and C.R. MacKenzie, A new method of classifying prognostic comorbidity in longitudinal studies: development and validation, *J Chronic Dis* **40** (1987), 373-383.

[21] A. Elixhauser, C. Steiner, D.R. Harris, and R.M. Coffey, Comorbidity measures for use with administrative data, *Med Care* **36** (1998), 8-27.

[22] H. Quan, V. Sundararajan, P. Halfon, A. Fong, B. Burnand, J.C. Luthi, L.D. Saunders, C.A. Beck, T.E. Feasby, and W.A. Ghali, Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data, *Med Care* **43** (2005), 1130-1139.

[23] Clinical Classifications Software (CCS) for ICD-9-CM, in, AHRQ Healthcare Cost and Utilization Project, 2017.

Address for Correspondence

Victor Lei, PharmD

Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania

423 Guardian Drive, Philadelphia, Pennsylvania, USA 19147

vlei@pennmedicine.upenn.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190217

Examining Reproducibility of Literature Search in Meta-Analysis

Fan Li[a], Pei-Yin Yang[b], Hong Kang[b], Xiaoqiu Chen[c], Chao Shang[d], Yang Gong[b]

[a] School of Medical Informatics, China Medical University, Shenyang, Liaoning Province, China,
[b] School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, Texas, United States,
[c] Department of ENT, Shengjing Hospital of China Medical University, Shenyang, Liaoning Province, China,
[d] Network Center of China Medical University, Shenyang, Liaoning Province, China

Abstract

Meta-analysis, a systematic retrieval from literature databases is an essential and prevailing method for combining data from multiple studies. Unfortunately, few studies have examined its rigor, which affects its reproducibility of results. We identified 22 meta-analyses on cervical cancer in PubMed for examining the parameters defined by PRISMA, relating to the rigor of literature retrieval. We found that 16 literature databases were used, and EMBASE was a leading resource, accounting for the highest frequency (81.82%). About half (45.45%) of the meta-analyses presented a complete, reproducible search strategy for at least one database. The ratio of included to retrieved articles after redundancy removal was only 6.58%, indicating low precision due to unclear or unreported processes. Our work serves as an initial step to examine the planning and execution of meta-analysis. Future efforts need to enhance reliability on literature retrieval in meta-analysis and compliance to PRISMA.

Keywords:

Meta-Analysis; Information Storage and Retrieval; Reproducibility of Results

Introduction

It has been an essential yet challenging task for clinicians to stay aprised on new knowledge from primary research papers while practicing in clinical settings [1]. Meta-analysis, the highest hierarchy of evidence regarding intervention questions [2], is defined as "the statistical synthesis of individual patient data from varying primary studies, leading to a quantitative summary of the pooled results" [3]. Meta-analyses assist clinical decision making, guide evidence-based medicine (EBM) and clinical practice, and serve as "the policy foundation for evidence-based practice guidelines, economic evaluations and future research agenda" [4].

The quality of the above-mentioned evidence is important. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) was developed for the reporting of systematic reviews and meta-analyses. PRISMA is a 27-item checklist and four-phase flow diagram that can examine whether the systematic reviews and meta-analyses are able to identify, appraise, and summarize research in an objective fashion so that clinicians would know whether the information is reliable for decision-making; however, most reported meta-analyses do not include such a quality assessment tool or do not report the method of assessment [5]. AMSTAR, an 11-item checklist, has been employed to assess the quality of meta-analysis and systematic review. Prior studies on quality assessments have identified that the Cochrane library was of a higher quality than others. Investigations of study selection bias and data extractions returned unclear results [6].

Systematic reviews require an unbiased and a reproducible search of data resources to identify as many relevant studies as possible. Reliability is the quality or state of being reliable, the extent to which an experiment, test, or measuring procedure yields the same results on repeated trials. In the field of academic research, reproducibility refers to that for the same research problem, independent research by other researchers can use the scheme provided by the authors to reproduce the experimental results [1]. PRISMA described Section Method "Search" as "Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated" and provided a guideline for researchers to examine the reproducibility of systematic reviews and meta-analyses [7]; however, researchers have conducted few meticulous studies on examining the reproducibility of meta-analyses.

This paper aims to examine the reproducibility of search strategy reported in meta-analyses. We chose cervical cancer, because providers can often detect it early, and sometimes prevent it entirely, by having regular Pap tests. When found early, cervical cancer is one of the most successfully treatable cancers. In addition, cervical cancer is the second most common cancer among females worldwide, with 80% of the cases occurring in sub-Saharan Africa, Central America, and South-Central Asia [8]. We aim to examine the reproducibility of search strategies reported by cervical-cancer related meta-analyses and reveal the opportunities and direction for clinical informaticians toward an enhanced rigor of meta-analyses.

Methods

First, we performed a systematic literature search in PubMed database to identify meta-analyses focusing on cervical cancer published from January 1st, 2013 to October 3rd, 2018. The search strategy utilized search terms as follows: (cervical cancer) AND (meta-analysis[PT] OR meta-analysis[tiab])

Titles and abstracts of the retrieved articles were screened manually and independently by two authors (FL and PY) using the eligibility criteria. The inclusion criteria were: 1) studies or systematic reviews with meta-analysis, data synthesis or quantitative overview; 2) studies focusing on cervical cancer. The exclusion criteria were: 1) comments or corrections for articles of meta-analysis; 2) narrative reviews or meta-analyses focusing on other cancers, which are different than cervical cancer, such as head and neck cancer. Figure 1 illustrates the overall strategy.

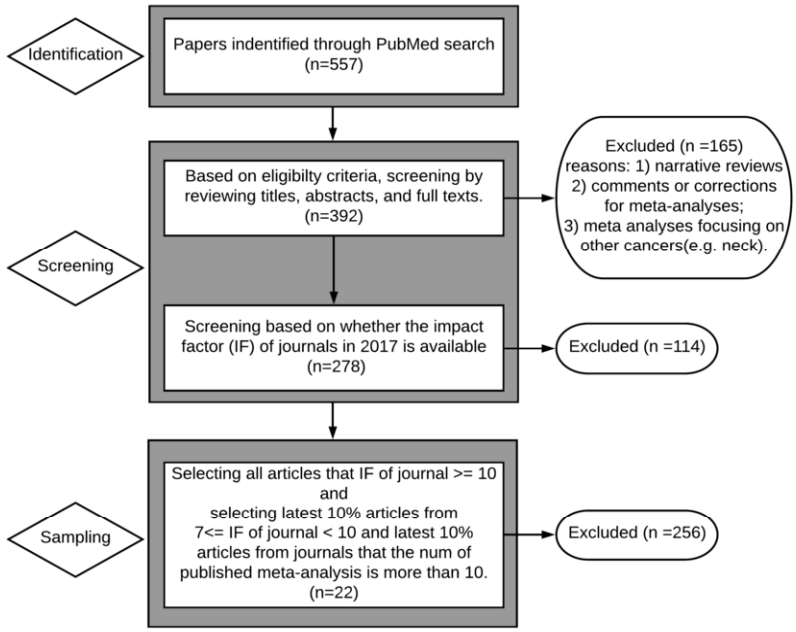


Figure 1- Flow Chart of Literature Search and Inclusion

Second, a Bibliographic Item Co-Occurrence Matrix Builder (BICOMB) analyzed the journal distribution of these articles [9]. The 2017 Journal Citation Report provided the impact factors (IF) of the journals. Next, we applied a stratified sampling strategy to the retrieved meta-analyses for further data extraction and evaluation: 1) all articles published in the journals with an IF higher than 10, 2) 10% of latest published articles in the journals with an IF from 7 to 10, and 3) 10% of latest published articles in the journals containing more than 10 articles, i.e. two gynecologic journals (*Gynecologic Oncology* and *Archives of Gynecology & Obstetrics*) and a comprehensive journal (*Plos One*).

Following the checklist of PRISMA, features of evaluation in this study were created, including journal name, publication year, electronic databases used, presence of full search strategy for all databases or at least one database, the number of retrieved articles before and after the removal of duplicates, and the number of included articles. Two authors (FL and PY) independently extracted the features from the full papers according to the checklist. Group discussions were held to resolve discrepancies involving additional authors when necessary. For each meta-analysis, we calculated the percentage of the number of included articles to the number of retrieved articles (before and after removal of duplicates).

Results

Search Results and Characteristics of Meta-Analyses

557 articles were retrieved from PubMed by applying the search strategy, among which 392 articles were included by applying the eligibility criteria. The 392 articles were published in a total of 157 journals, among which 123 (78.3%) journals, carrying 278 (70.9%) articles, possessed impact factors in the 2017 Journal Citation Report. A journal list with ranked IF in descending order and published in 2017 was created. Figure 2 illustrates the distribution of IF of journals and number of relevant meta-analyses on cervical cancer. The application of the stratified strategy resulted in 22 articles, shown in Table 1.

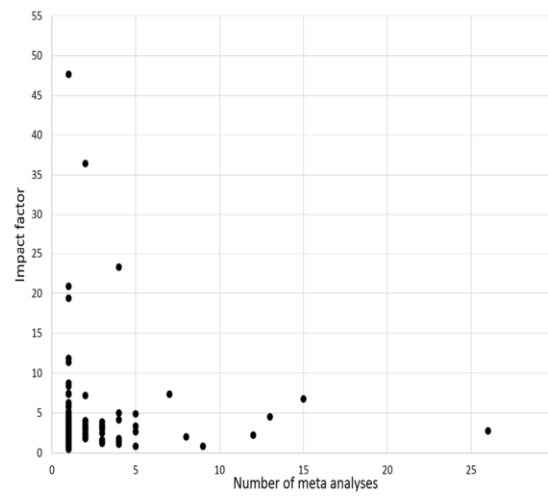


Figure 2- Distribution of Impact Factors of Journals and Number of Relevant Meta-Analyses on Cervical Cancer

Note: The journals (IF=2 to 7） contain more than 10 meta-analyses. Journals tend to publish meta-analyses on cervical cancer with IF < 10.

Table 1- Characteristics of Meta-Analyses Selected for Assessment

Author Year PMID	Journal	Search strategy§		# retrieved articles€		# included articles	% of included/retrieved articles£	
		1§	2§	1€	2€		1£	2£
Melnikow 2018 30140883	JAMA	Y*	Y	5232	2972	62	1.19	2.09
Arbyn 2017 29126708	Lancet. Oncology	N$	Y	N	N	93	N	N
Arbyn 2014 24433684	Lancet. Oncology	Y	Y	?л	?(884£)	97	N	N
Tainio 2018 29487049	BMJ	Y	Y	?(6275)	?	36	N	N
Fokom-Domgue 2015 26142020	BMJ	Y	Y	?	?(1049)	15	N	N
Kyrgiou 2014 25352501	BMJ	Y	Y	?	?(1697)	15	N	N
Kyrgiou 2016 27469988	BMJ	Y	Y	?	?(3021)	71	N	N
Arbyn 2017 27842420	Annals of Internal Medicine	N	Y	?	?(899)	24	N	N
Siristatidis 2013 23255514	Human Reproduction Update	Y	Y	7785	7785	9	0.12	0.12
Kelly 2018 29107561	Lancet HIV	Y	Y	605	407	16	2.64	3.93
Zard 2014 24657969	Autoimmunity Reviews	N	N	?	?(235)	7	2.98	N
Fisher 2013 23620381	International Journal of Epidemiology	N	N	1108	699	29	2.62	4.83
Hammer 2016 26661889	International Journal of Cancer	N	N	721	644	15	2.08	2.33
Li 2014 24308856	Alimentary Pharmacology & Therapeutics	N	N	565	393	63	11.15	16.03
Verdoodt 2015 26296294	European Journal of Cancer	N	N	376	252	16	4.26	6.35
de Lima 2018 29021084	Gynecologic Oncology	N	N	396	N	25	6.31	N
Charakorn 2018 29606483	Gynecologic Oncology	N	N	1797	1605	61	3.39	3.80
Zhang 2018 29641578	Plos One	N	N	1715	1342	20	1.17	1.49
Jin 2018 29554090	Plos One	N	N	2614	2588	N	N	N
Zhou 2017 29227998	Plos One	N	N	709	707	13	1.83	1.84
Ye 2018 29520664	Archives of Gynecology & Obstetrics	N	N	319	225	22	6.90	9.78
Feng 2018 29876746	Archives of Gynecology & Obstetrics	N	N	79	69	8	10.13	11.59

Notes: §: Search strategy 1 represents whether the full electronic search strategy for all databases is available. Search strategy 2 represents whether the full electronic search strategy for at least one database is available.

€: The # retrieved articles 1 is the number of retrieved articles before removal of duplicates, and 2 is that of those after removal of duplicates.

£: For % of included/retrieved articles, the first column represents the percentage of number of included articles to that of retrieved articles before removal of duplicates, and the second column represents the percentage of number of included articles to that of retrieved articles after removal of duplicates.

**: Y: available.; $: N: not available.; л: ?: not clear. ; £: The number in brackets is a rough estimate by the context.*

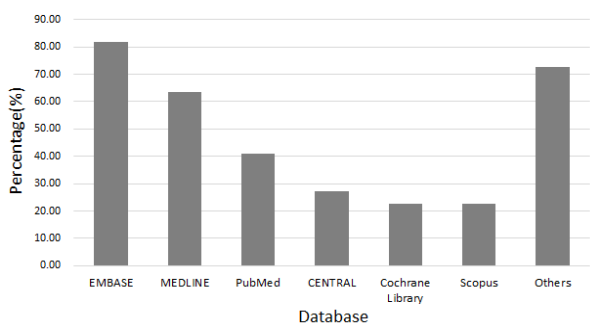


Figure 3- Literature Databases Searched in the Meta-Analyses

Notes: MEDLINE includes PubMed MEDLINE and OVID MEDLINE. Cochrane databases are part of Cochrane Library. Other resources include CINAHL, CNKI, ISI Web of Science, PsycINFO, Wanfang, CBM, Cochrane Database of Systematic Reviews, EBSCO, Google Scholar, ISI Proceedings, and PubMed/MEDLINE with a frequency ≤ 2.

Literature Databases Used in the Meta-Analyses

Seventeen literature databases (Involvement frequencies: Maximum = 9; Minimum = 2; Average = 4) were identified in the 22 meta-analyses. As illustrated in Figure 3, the top three databases involved in the meta-analyses were EMBASE (18 meta-analyses, 81.82%), MEDLINE (including PubMed MEDLINE and OVID MEDLINE, 14 meta-analyses, 63.64%), and PubMed (9 meta-analyses, 40.91%). Nine of 14 meta-analyses in which MEDLINE was searched (64.29%) did not specify the platform of MEDLINE .

Inclusion of One Full Search Strategy

Ten (45.45%) meta-analyses contained a full search strategy including any limits used for at least one database, among which eight (36.36%) meta-analyses reported the full search strategy for all databases. In the high impact journals (IF≥10), all meta-analyses presented a full search strategy for at least one database, while eight (80%) meta-analyses reported the full search strategy for all databases; however, among the low impact journals (IF<10), no meta-analyses presented a full search strategy even for at least one database.

Comparing the Number of Included and Retrieved Articles

The number of retrieved articles from each information resource was not always clearly included, shown in Table 1, resulting in the failures in determining whether or not the total amount of retrieved articles was subject to removal of duplicates. Twelve meta-analyses reported the number of retrieved articles before and after removal of duplicates and the number of included articles, among which the maximum number of retrieved articles after duplicate removal was 2,972, and its corresponding number of included articles was 62. Table 1 also presents the percentage of the number of included articles to the number of retrieved articles after duplicate removal. The percentage ranged from 1.49% to 16.03% (mean= 6.58%).

Discussion

Confusion on Literature Databases as Information Source

Based on the item regarding information sources in PRISMA 2009, it is necessary to describe all information sources (e.g., databases with dates of coverage and author contacts). Although we could provide information sources for the 22 meta-analyses we evaluated, choosing source information is frequently unclear due to the description of the names of databases and the relationship among databases. For example, PubMed and MEDLINE should not be used in the same meta-analysis, since MEDLINE is in fact a subset of PubMed. In addition, searching results may vary in different MEDLINE databases (e.g., PubMed, Ovid MEDLINE) [10]. More than one half of meta-analyses did not report the platform they used to search MEDLINE, which becomes a barrier in reproducing the results or repeating the search strategy.

The number of searched literature databases under the topic of cervical cancer varied between two and nine. The reasons of including multiple databases are extremely unclear and somewhat confusing. For example, databases like Google Scholar and ISI Proceedings are used in addition to specialized professional databases, such as PubMed and EMBASE. It is rarely reported how many additional hits were introduced by such an unclear inclusion of databases. These problems remain for future exploration, and further investigations should clarify the rules on use of databases in meta-analyses.

Poor Compliance on Inclusion of One Full Search Strategy

To establish a reasonable and detailed search strategy for each database and ensure the quality and reliability of meta-analysis, it is essential to enhance recall and precision of evidence. According to the PRISMA criteria, the search strategy should be repeatable for at least one database. Our findings indicate that only the strategies published in the high IF journals tend to be repeatable. It is necessary for meta-analysis to include at least one repeatable full search strategy for at least one database or for every database included in the search.

Lower Precision of Evidence Retrieval

Based on the data from 11 meta-analyses, we could obtain the ratio of the number of included articles to the number of retrieved articles after removal of duplicates, ranging from 1.49% to 16.03% (mean 6.58%). This percentage represents the precision of literature retrieval on meta-analyses. We found that authors usually developed a search strategy for higher recall so that the precision is much lower than general literature retrieval. The heavy workload of article screening may be subjective and prone to errors. This problem could be resolved by establishing a search strategy with clarified keyword limitations, which may reduce recall as a compromise. Therefore, systematic filters of potential eligible articles should be developed.

Lack of Reproducibility

The number of retrieved articles was not clearly presented for each information resource, which could be another barrier to repeat a meta-analysis. Table 1 shows that only three articles (13.64%) presented a full search strategy for all databases as well as the number of retrieved articles before and after removal of duplicates. The absence of full search strategy for all databases in large amounts of meta-analyses weakens the possibility of repeating them. Therefore, one full search strategy for at least one database in PRISMA is inadequate in meta-analysis. A full search strategy and number of article hits for each database should be a fundamental requirement of meta-analysis.

Recommendations

Based on the findings above, we suggest the following recommendations in order to make a meta-analysis repeatable at the literature retrieval stage. First, we propose to enhance the PRISMA criteria to present full electronic search strategies for all databases rather than for at least one. Researchers should strictly obey the item on presentation of detailed operable electronic search strategy according to PRISMA. Second, PRISMA should require that meta-analyses report the number of literature yielded from each database. Third, researchers who will participate in meta-analysis should be trained with the knowledge of information resource retrieval to control various problems that may arise from the literature search process and to ensure the repeatability of the results. In any case, reviewers and editors of journals should rigidly control the quality of literature retrieval for meta-analyses according to PRISMA.

Limitations and Future Work

Only 22 articles were included for the assessment on quality of literature search based upon a stratified sampling of 392 meta-analyses on cervical cancer. The limited sample size may have limited generalizability of the results. Nonetheless, the conclusion may be specific in cervical cancer and hold limited scalability to other gynecologic cancers. In the future, we will evaluate all the 392 meta-analyses by the AMSTAR checklist and perform a univariate and multivariate statistical analysis to investigate the factors influencing the quality of literature retrieval of meta-analyses.

Conclusions

The assessment of meta-analyses in regards to literature retrieval revealed poor performances in reporting quality of retrieval strategy and a low compliance of PRISMA. Overall, the reporting quality of literature retrieval in the meta-analysis on cervical cancer needs to be improved. Specifically, a clear justification is needed in describing the selected databases. PRISMA requires to include a complete search strategy toward reproducible results. The number of article hits retrieved from each and every database should be reported individually instead of a combined number which is a barrier for reproducibility of meta-analysis.

Acknowledgements

This work was supported by the Science and Technology Research Program of Department of Education of Liaoning Province, PR China (Grant No. LK201659). The study was also supported by UTHealth Innovation for Cancer Prevention Research Training Program Post-Doctoral Fellowship (RP160015).

References

[1] M.W. Russo, How to Review a Meta-analysis, *Gastroenterol Hepatol (N Y)* **3** (2007), 637-42.

[2] H. Szajewska, Evidence-Based Medicine and Clinical Research: Both Are Needed, Neither Is Perfect, *Ann Nutr Metab* **72 Suppl 3** (2018), 13-23.

[3] K. O'Rourke, An historical perspective on meta-analysis: Dealing quantitatively with varying study results, *J R Soc Med* **100** (2007), 579-82.

[4] D. Moher, D.J. Cook, S. Eastwood, I. Olkin, D. Rennie, D.F. Stroup, Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. Quality of Reporting of Meta-analyses, *Lancet* **354** (1999), 1896-900.

[5] S. Sanderson, I.D. Tatt, J.P. Higgins, Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: A systematic review and annotated bibliography, *Int J Epidemiol* **36** (2007), 666-76.

[6] J. Dong, G. Teng, T. Wei, W. Gao, H. Wang, Methodological Quality Assessment of Meta-Analyses and Systematic Reviews of Probiotics in Inflammatory Bowel Disease and Pouchitis, *PLoS One* **11** (2016), 1-11, doi:10.1371/journal.pone.0168785.

[7] A. Liberati, D.G. Altman, J. Tetzlaff, C. Mulrow, P.C. Gotzsche, J.P. Ioannidis, M. Clarke, P.J. Devereaux, J. Kleijnen, D. Moher, The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration, *PLoS Med* **6** (2009), doi: 10.1136/bmj.b2700.

[8] C.F.C.C. Collaboration, Reducing uncertainties about the effects of chemoradiotherapy for cervical cancer: A systematic review and meta-analysis of individual patient data from 18 randomized trials, *J Clin Oncol* **26** (2008), 5802-12.

[9] L. Cui, W. Liu, L. Yan, H. Zhang, Y.F. Hou, Y.N. Huang, H. Zhang, Development of a text mining system based on the co-occurrence of bibliographic items in literature, *New Technology of Library and Information Service* (2008), 70-75.

[10] A. Wanner, N. Baumann, Design and implementation of a tool for conversion of search strategies between PubMed and Ovid MEDLINE, *Res Synth Methods* (2018).

Address for correspondence

Yang Gong, MD, PhD
7000 Fannin St Suite 600, Houston, TX 77030, USA
Yang.Gong@uth.tmc.edu
Phone: +1 713 500 3547, Fax: +1 713 500 3907.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190218

Transforming Two Decades of ePR Data to OMOP CDM for Clinical Research

Daniel M. Lima[a], Jose F. Rodrigues-Jr[a], Agma J. M. Traina[a], Fabio A. Pires[a], Marco A. Gutierrez[b]

[a] *Institute of Mathematical and Computer Sciences (ICMC), University of São Paulo, São Carlos, São Paulo, Brazil,*
[b] *Heart Institute (InCor), Clinics Hospital, Faculty of Medicine, University of São Paulo, São Paulo, São Paulo, Brazil*

Abstract

This paper presents the extract-transform-and-load (ETL) process from the Electronic Patient Records (ePR) at the Heart Institute (InCor) to the OMOP Common Data Model (CDM) format. We describe the initial database characterization, relational source mappings, selection filters, data transformations and patient de-identification using the open-source OHDSI tools and SQL scripts. We evaluate the resulting InCor-CDM database by recreating the same patient cohort from a previous reference study (over the original data source) and comparing the cohorts' descriptive statistics and inclusion reports. The results exhibit that up to 91% of the reference patients were retrieved by our method from the ePR through InCor-CDM, with AUC=0.938. The results indicate that the method that we employed was able to produce a new database that was both consistent with the original data and in accordance to the OMOP CDM standard.

Keywords:

Medical Informatics; Health Information Exchange; Data Curation

Introduction

In the last decades, the field of Informatics has unveiled the, so-called, Big Data phenomenon; an intense increase of data gathering, exchanging and storage in several human activities. This phenomenon is outlined by the so called five Vs: velocity, volume, variety, veracity and value of data; posed as the big challenges for data analysis and processing [1]. Such characteristics are also common in Medical and Health Information Systems, whose databases have grown into huge amounts of patient information and health-related activities, in diverse formats, always online, and easily accessible from a mobile screen. If properly interconnected and treated, these systems offer interesting data sources for evidence-based research, such as Precision Healthcare, Population Health, Clinical Research, and more. [11]

However, this data exchange is often a significant challenge. Most of the Electronic Patient Record (ePR) systems were not explicitly designed for research; rather, they are organized by standards and structures which are local to the institutions they primarily serve, e.g. hospital facilities, clinics, pharmacies, health insurance companies, etc. Thus, interchange methods, protocols and architectures were designed to cope with this challenge, such as the HL7 standards [12] and the OpenEHR platform [13]. These approaches mediate the communication of near-line and online transaction processing (OLTP) systems, specifying standard data elements and transformations from their internal data to a common messaging format. Furthermore, data analysis follows an approach akin to online analytical processing (OLAP), using denormalized, coalesced and preprocessed data in a standard common database format.

OHDSI and the Common Data Model

In this context, the Observational Health Data Sciences and Informatics (OHDSI – www.ohdsi.org) initiative grew out of the Observational Medical Outcomes Partnership (OMOP) developing a mature data standardization model, the OMOP Common Data Model (CDM) [6]. Having a ready-to-use database in a standard common model such as the OMOP CDM simplifies the exchange and integration of standardized methods, applications, information and tools between clinical researchers; a critical feature for distributed research networks using patient-centric clinical databases. [7]

The CDM is a strong information model; its conceptual elements and their relationships are explicitly specified in a formal language, and every piece of information is connected to a standard term from SNOMED-CT. The CDM's Standard Clinical Tables include Person, Visits, Observations, Conditions, Death, Procedure occurrences, Drug exposures, Measurements and more detailed information such as Drug ingredients, and Condition modifiers. OHDSI also provides open-source CDM applications for visualization and statistical analysis of patient-exposure-outcome cohorts. [10]

The InCor data integration challenges

The Heart Institute (InCor) of São Paulo, Brazil, is one of the six institutes of the Clinics Hospital complex, University of São Paulo Medical School. In the last two decades, InCor has increased its commitment in integrating all the relevant information of its patients, successfully developing an ePR named SI³. The first version of SI was deployed in year 2000; currently, it stores the clinical history, examinations, procedures, surgeries, notes, laboratory tests, medication, bills, and more for 1.3 million patients. Since then, the system has continuously evolved, overcoming several challenges related to the exchange of information among different healthcare institutions and remote installations. Furuie et al., describe details of the system architecture [8], while a number of studies involved cohort selection based on information collected by the SI³ system [3-5].

However, the workload involved in extracting the relevant patient information from SI³ has motivated the adoption of new strategies. Recently, we started the mapping from the SI³ data model to a standard data model that can simplify the observational retrospective studies related to Clinical Research. In this paper, we present the steps related to the mapping between SI³ and the CDM data models to prepare a new standardized database, named InCor-CDM, that can be used with

the OHDSI toolset and a number of visual analytics tools. We measure the quality of the resulting InCor-CDM database using precision and recall statistics, when compared to the cohort generated by a previous study (gold standard).

Methods

Environment preparation

We prepared the InCor-CDM database environment by installing a PostgreSQL 10 DBMS, Java 10 JDK and Docker-compose in a Linux workstation. The database setup includes: (a) creating the required database users; (b) creating the OMOP CDM tables with the CommonDataModel/PostgreSQL scripts; and (c) importing the standard OMOP vocabularies from athena.ohdsi.org. All the OHDSI sources are available at github.org/OHDSI. Next, we installed the Achilles and Broadsea repositories required for the OHDSI web applications, configured the addresses and JDBC URLs and started their respective docker containers.

Database characterization

The InCor SI³ database is stored in an Oracle 12c instance, accessed with the Oracle JDBC connector and DBeaver SQL client. The first step to manage the database was sampling its tables and columns with the OHDSI's WhiteRabbit software. This application generates a spreadsheet with the most frequent values of each selected column, so we can inspect them and make decisions about which columns to ignore, especially those with irrelevant or missing values. Then, we use RabbitInAHat to parse the output of WhiteRabbit and draw relational data-flow diagrams for documentation.

Some namespaces in SI³ were ported from non-relational or older systems; such namespaces had no associated documentation, comments, constraints or foreign keys. For this reason, the database metadata had to be converted to a searchable JSON format with table and column names, types, comments, constraints and foreign keys. The column names were tokenized according to the naming scheme of the institution (e.g. abbreviations separated by underscore) and matched to similar columns in other tables; the goal was to find implicit relationships where the foreign keys were missing. Some attribute domains, such as internal record status codes, event sequence and timing diagrams, were documented from interviews with the support staff of InCor.

Patient de-identification

For de-identification purposes, personal information mapped to the CDM was limited to a minimum. Any key with a path to a patient primary key (and the PK itself) was **pseudonymized** [14], i.e., direct identification information such as citizenship document, phones numbers, addresses and names are not ported to the CDM, and the record primary key is exchanged to a *new id* (pseudonym), which is a random number drawn from an uniform distribution in the range 1×10^{10} and 9×10^{15} using Oracle's DBMS_RANDOM functions, addressing collisions with repeated sampling. This range was selected not to conflict with OMOP's standard concept ids (0 up to 2×10^{9} are reserved) and to be within the limits of JSON numbers (53-bit precision). InCor holds the mapping from the new ids to the original keys in a private table; the mapping is to be used for notifying the patient, or her/his physician, in case the result of a study can improve a patient's condition.

Numerical variables were truncated in order to satisfy a baseline level of **k-anonymity** [15], i.e., guaranteeing that any patient variable value have at least *k* patients with the same information, so no patient is uniquely identifiable. For example, event dates were truncated to yearly, monthly, daily or hourly precision where original precision were not needed. Records with spurious attributes (e.g., dates in the future, outside any visit, invalid range, null required field, missing keys) were discarded. Also, we only loaded data from patients born before 2010 (aged 18+), with at least one valid visit.

si3.pac_paciente

paci_id	tp_sexo	...
01721	M	...
01722	M	...
01723	F	...

keys

table	src_id	new_id
pac_paciente	01721	7369111123
pac_paciente	01722	1257321234
pac_paciente	01723	3038618654

gender_map

src	id
F	8532
M	8507

omop.person

person_id	gender_concept_id	...
7369111123	8507	...
1257321234	8507	...
3038618654	8532	...

Figure 1– Example: gender mapping from SI³ to the CDM.

Figure 1 shows sample data to illustrate this process. Each column from CDM person table (e.g., gender_concept_id) is extracted from the *source* table and column (e.g., pac_paciente. tp_sexo), and transformed with the appropriate domain map (e.g., value 'F' used in SI³ for female gender is mapped to CDM concept 8532). The *keys* and *gender_map* tables are populated beforehand. The standard CDM concept ids were searched in OHDSI's Athena – a web-based CDM vocabulary explorer. Then, observe the following query, which loads data from the SI³ PAC_PACIENTE table into the InCor-CDM omop.person table with remapped keys and concepts:

```
INSERT INTO omop.person
SELECT K.new_id          AS person_id,
       EXTRACT(YEAR FROM P.dt_nasc)
                         AS year_of_birth,
       COALESCE(G.id, 0) AS gender_concept_id
FROM si3.pac_paciente P
JOIN keys K ON  K.table='pac_paciente'
           AND K.src_id=P.paci_id
LEFT JOIN gender_map G ON P.tp_sexo=G.src;
```

This query operates on the table samples in Figure 1, where it is assumed that the *keys* table holds the patient's random new_id, and that the *gender_map* table is a domain map table defined as (src char(1); id integer) corresponding to values of (tp_sexo, gender_concept_id). Unmapped values receive code 0, meaning "unknown concept" in the standard CDM vocabulary. Related tables were joined to the patient PK as usual, with their PK also remapped by keys.new_id.

Coding translation

InCor SI³ uses the ICD 10 for diagnosis, a set of Brazilian vocabularies for coding clinical conditions, drugs, and procedures (TUSS is used by the Brazilian Health Care System for general terms and Brasindice for drugs – datasus.gov.br) and an internal coding system for generic billable items. Internal codes in use at InCor were inserted as new Concepts in the CDM (with ids mapped between 3×10^{9} and 9×10^{9}) under the "InCor" vocabulary, with Concept Relationships to standard concepts whenever this information was available in SI³. Initially, the records are inserted in the InCor-CDM with the original source codes, then the Concept Relationships from local to international codes are used to update the local InCor-CDM references to OMOP standardized terminologies, such as the SNOMED-CT,

RxNorm, and LOINC [6], while the preserving the original source code stored in the patient record.

Quality assessment

After loading all the CDM tables, we execute the Achilles analysis, which will report data quality issues as errors, warnings or notifications. It will also preprocess the demographic characterization of the database for the visualizations and reports; after that, we start the Atlas WebAPI and front-end servers. Atlas offers both a RESTful API and a graphical web interface to schedule the execution of OHDSI methods, automatically generating the queries on a properly built and configured CDM database.

The most relevant errors initially found were related to: (a) the referential integrity, e.g., events without associated visit or events without a valid person_id, which were discarded; (b) lack of condition_eras, drug_eras and observation_periods, which were imputed from the ePR records by collapsing all the events of a patient not apart for more than a year into a single era – in the CDM standard, an "event" era refers to a time period of interest where "events" are recorded in the ePR; (c) events with invalid date, e.g. a condition_occurrence with start date in a future date; (d) too many patients without a diagnostic or prescription; InCor-CDM keeps such records because they can be used in the condition or drug_exposure dashboards, regardless of previous diagnosis.

Statistical evaluation

After initial corrections, we evaluate the quality of InCor-CDM by using software OHDSI's Atlas to recreate a previous CVD patient cohort observed in a reference study that was executed over the InCor SI³ database by Abrahao et al. [9]. That reference study prepared a clean de-identified database named Pauá, based on a 2016 snapshot of SI³ records of patients, admissions, discharges, diagnoses, surgeries, PCI, medications, and laboratory tests. Within the Pauá database, Abrahao et al., verified the effect of statins on the survival rate of patients diagnosed with cardiovascular diseases. Our study will evaluate the InCor-CDM quality by defining a cohort with the same criteria and by verifying how many patients (with the same private InCor identifiers) were retrieved by each criteria given that reference study.

Results

In this section we evaluate InCor-CDM quality by selecting the same DCV cohort of a previous study in the Pauá database [9], by using Atlas over the InCor-CDM database. Our evaluation computes the Area Under the ROC Curve (AUC) based on the results of 12 executions of the DCV cohort at various settings. We compute predictive statistics for each cohort execution, using Pauá as a gold standard. The resulting AUC ranges from 0.5 (no different than random sampling) to 1.0 (reproduces exactly the same result as the gold standard).

Table 1– Databases cardinality (thousands of records).

Domain	SI³2016	Pauá	SI³2018	CDM
Person	1,116	323	1,346	946
Visit Occurrence	6,427	5,686	7,499	7,305
Condition Occur.	1,205	1,007	1,361	1,324
Procedure Occur.	45,024	144	53,945	51,479
Drug Exposure	83,283	2,775	100,052	38,962
Measurement	22,025	20,528	31,095	30,177
Death	17	21	18	18

In Table 1, we present the cardinality of the InCor-CDM in comparison to the SI³ and Pauá database. Note that the Procedure and Drug domains in the Pauá database have substantially less records; this is because they are restricted to surgeries and particular classes of drugs, and because Pauá only uses patients with at least one admission and diagnosis. Additionally, InCor-CDM is based on a more recent snapshot, named SI³2018 in the aforementioned table.

Cohort definition

To replicate the Pauá study, we created a cohort over the InCor-CDM database with the criteria below. Each list item directly corresponds to a HTML input field in the Atlas **cohort definition form**. These criteria are translated to SQL queries over our database in the CDM format (InCor-CDM).

- *Initial Event Cohort*: People having any of a **visit** occurrence of **Outpatient concept set** (with concept_id 9202, Outpatient Visit);
- *Additional Qualifying Inclusion Criteria*:
  - *Condition occurrence criteria*: with **at least 1** of **any condition**;
  - *Demographic criteria*: **age** greater than or equal **23** (in the censor window), matching the patients over 18 years old at the Pauá study start (1999);
  - *Demographic criteria*: with a **gender** of **MALE** (8507) or **FEMALE** (8532);
  - *Condition occurrence criteria*: of **CardioVascular Disease** (a concept set of concept ids from ICD-10 categories I20 to I25, I64 to I70 and G45, including descendants and mapped), with occurrence start between **2003-01-01** and **2013-12-31**, where event starts between **All** days before and **30** days after index end date (meaning the diagnosis was recorded around the time of the initial outpatient visit);
  - *Visit occurrence criteria*: with **at least 1** of **Outpatient concept set**, where event starts between **30** after and **All** days after index end date (a subsequent visit recording the outcome);
  - *Limit qualifying cohort to the* **earliest** *event*;
- *Era collapse gap size*: **1 day**;
- *Cohort censor window*: starting **2003-01-01** and ending **2013-12-31**.

After verifying the Concept Sets to have the correct concept ids, we generated the cohort on the InCor-CDM database and verified the cohort attrition report (the number of patients remaining in the selection after each filter) in Table 2:

Table 2– InCor-CDM CVD cohort attrition report in Atlas.

Criteria	n	%	Visualization
Initial	778,015	100.00	
a) Dx	351,205	45.14	
b) 18+	321,827	41.37	
c) M/F	303,847	39.05	
d) CVD	45,710	5.88	
e) 2nd-V	39,910	5.13	
f) People	39,498		

Table 2 summarizes the cohort definition criteria, where each abbreviation mean: (a) **Dx**: has a valid condition occurrence (diagnosis); (b) **18+**: over 18 years old; (c) **M/F**: Male or Female gender; (d) **CVD**: has any occurrence of ICD-10 I20 to I25, I64 to I70 or G45; (e) **2nd-V**: has a second outpatient visit occurrence more than 30 days after the index event (the initial visit); (f) **People**: lists how many people actually matched the events, because some patients have more than one episode. The baseline results achieved by the InCor-CDM CVD cohort indicate that 39,498 patients satisfy the criteria.

Table 3– Pauá reference CVD cohort attrition report.

Criteria	n	%	Visualization
Initial	313,894	100.00	
Dx	313,894	100.00	
18+	282,677	90.00	
M/F	263,339	83.87	
CVD	56,799	18.06	
2nd-V	27,698	8.80	
People	27,698		

In comparison, Table 3 displays an attrition report for the reference study (Pauá). The visualization was drawn with Python and GNU Gimp because Pauá's schema is not based on the CDM standard, and thus could not be used in Atlas. The reference cohort selected with the same criteria in the Pauá database has 27,698 patients (30% less than our first result), indicating the need of further refinement.

Initial Evaluation

All subject_ids in the cohort defined in Atlas were compared to the set of patient ids in the Pauá cohort. This comparison was executed by computing confusion matrices against the Pauá cohort result as a gold standard, then the derived scores: true and false positive ratios (TPR/FPR), positive and negative predictive power (PPV/NPV), accuracy (ACC) and F1-score. The confusion matrix the cohort #1 is given in Table 7, and derived scores make the first row of Table 8.

Table 7– Confusion matrix for cohort #1 (in Table 2).

InCor-CDM \ Pauá	P	N	Total
P	25,423	12,290	37,713
N	2,651	282,887	285,538
Total	28,074	295,177	323,251

Parameter refinement

Therefore, we noted cohort accuracy variations by creating additional cohorts with slightly adjusted parameters for the CVD criteria. Table 4 displays the attrition reports with increased periods for the qualifying criteria of CVD occurrence start date after index end date. Table 5 displays the attrition reports with increased periods for the qualifying criteria of Cardiopathy occurrence start date. We observe that increasing the collapse gap size to 7 and 14 days had no substantial effect on the results (Table 4).

Table 4– Varying condition start after index event (days).

Criteria \ days	7	14	21	30
Initial		778,015		
Dx, 18+, M/F		303,847		
CVD	44,967	45,255	45,484	45,710
People	39,055	39,203	39,342	39,498

Table 5– Varying condition start periods (years).

Criteria \ days	2003-2013	2000-2013	2000-2016
Initial		778,015	
Dx, 18+, M/F		303,847	
CVD	45,710	49,942	63,656
People	39,498	43,293	54,126

Table 6– Varying 2nd visit event start after index (days).

Criteria \ days	All	365	180	90
Initial		778,015		
Dx, 18+, M/F		303,847		
CVD	45,710	44,228	43,950	43,667
People	39,498	35,457	32,767	29,414

Then we evaluated all the patients selected in each cohort (from Tables 4, 5 and 6) by comparing them to the reference Pauá study, using the private Keys table to map the CDM person_ids to the SI³ patient ids (confidential to the institution), whose results are presented in Table 8. It should be noted that Pauá had an update in October 2016, after the reference study was published, and so we re-executed the query for its patient cohort, resulting in P=28,074 patients selected in the cohort (1.4% increase) and overall total P+N=323,251 patients included (3.0% increase).

Table 8– Predictive scores for each cohort.

#	TPR	FPR	PPV	NPV	ACC	F1
1	.905	.041	.674	**.990**	.953	.772
2	.901	.040	.678	**.990**	.954	.774
3	.903	.040	.677	**.990**	.954	.774
4	.904	.041	.676	**.990**	.954	.773
5	.905	.041	.674	**.990**	.953	.772
6	**.907**	.052	.623	**.990**	.944	.738
7	**.907**	.052	.622	**.990**	.944	.738
8	**.907**	.052	.620	**.990**	.943	.736
9	.889	.040	.680	**.990**	.954	.775
10	.877	.031	.727	.988	**.960**	**.795**
11	.829	.027	.743	.983	**.960**	.784
12	.754	**.023**	**.752**	.976	.957	.753

Evaluation of the ROC curve

We complete the evaluation by plotting the ROC curve from the predictive scores in Table 8. Only the patients existing in the Pauá database were used, i.e., only 37,713 patients of those retrieved in the cohort #1 also existed in the Pauá database. The highest scores of each column are highlighted in bold, e.g., cohort #10 exhibited the highest F1-score.

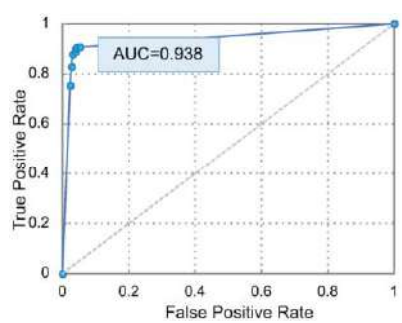


Figure 2– Empirical ROC curve for Table 8.

In this section we presented the results of the cohorts selected with Atlas from the InCor-CDM database. We also presented the effect of tweaking cohort parameters in the quality of the results (Table 8), which exhibited up to 80% F1-score, 75% Precision (PPV) and 91% Recall (TPR) at different settings. Figure 2 summarizes our results in an empirical ROC curve, exhibiting 0.938 of area under curve (AUC).

Discussion

The resulting InCor-CDM exhibits high agreement with the previous gold standard study [9]. This means that it is possible to estimate the same population-level effects (e.g. different medications) in both databases. InCor-CDM additionally benefits from the quality analyses implemented in the OHDSI Achilles tool, which warns about inconsistencies and errors found in the transformed data, and can be used for more advanced analysis in comparison with external CDM-based databases. For future work, we envision the comparison of patient cohorts between InCor-CDM and external CDM-based databases from other OHDSI work groups, further studying data quality, subpopulation characteristics between different institutes and evaluating risk scores for InCor patients.

Conclusions

We presented details of the migration process of a huge clinical database, the InCor's SI^3, to the OMOP CDM standard, an international format aimed at improving research in computer-aided medical systems. We presented the method used to extract, transform and load data between the databases commenting on the challenges regarding models, formats, terminology, and tools. We evaluated the quality of the resulting database, named InCor-CDM, by comparing cohorts obtained with the software OHDSI Atlas. We considered a previous cohort selection study used as ground truth; for a systematic comparison, we computed several information retrieval statistics and a ROC curve. The cohorts defined in Atlas exhibited from 62% to 75% precision, 75% to 91% of recall, 74% to 80% F1-score, and 0.938 of area under the ROC curve (AUC). The results indicate that the method that we employed was able to produce a new database that was both consistent with the original data and in accordance to the OMOP standard. The new database shall support a wide range of new research initiatives within the Heart Institute.

Acknowledgements

The authors thank FAPESP (grant #2018/11424-0), CNPq (grants #406550/2018-2 and #305580/2017-5) and InCor's Informatics Service: A. Longo, M. Rebelo, M. Yamaguti, R. Moreno, T. Vieira, V. Silva, V. Reis, W. Pereira and others.

References

[1] V.N. Gudivada, R. Baeza-Yates, and V.J. Raghavan, Big Data: Promises and Problems, *IEEE Computer* **48, 3** (2015), 20-23.

[2] A.K. Jha, D. Doolan, D. Grandt, T. Scott, and D.W. Bates, The use of health information technology in seven nations, *Int J Med Inform* **77, 12** (2008), 848-854.

[3] M.T. Abrahão, M.R.C. Nobre, and M.A. Gutierrez, Descriptive Statistics of 65000 Patients Treated for Myocardial Ischemia: data from routine electronic health records, *Evidence Live* **13** (2013).

[4] M.T. Abrahão, M.R.C. Nobre, and M.A. Gutierrez, The effectiveness of statins in the treatment of cardiovascular disease: Cross-sectional study with paired groups from electronic patient records, *Value Health* **16, 7** (2013), A518.

[5] M.T. Abrahão, T.J. Soares, F.A. Pires, M.A. Gutierrez, and M.R.C. Nobre, Data Warehouse com dados da Saúde Pública: Estudo de Caso sobre o Tratamento de Doença Cardiovascular Aterosclerótica no Estado de São Paulo, *CBIS* **12** (2010).

[6] G. Hripcsak, J.D. Duke N.H. Shah, C.G. Reich, V. Huser, M.J. Schuemie, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *MedInfo* **2015** (2015), 574-578.

[7] D. Madigan, P.B. Ryan, M. Schuemie, P.E. Stang, J.M. Overhage, A.G. Hartzema, et al. Evaluating the Impact of Database Heterogeneity on Observational Study Results. *Am J Epidemiol* **178, 4** (2013), 645-651.

[8] S.S. Furuie, M.F.S. Rebelo, R.A. Moreno, M. Santos, N. Bertozzo, G.H.M.B. Motta, et al. Managing Medical Images and Clinical Information: InCor's Experience, *IEEE Trans Inf Technol Biomed* **11, 1** (2007), 17-24.

[9] M.T. Abrahão, M.R.C. Nobre, and M.A. Gutierrez, A method for cohort selection of cardiovascular disease records from an electronic health record system, *Int J Med Inform* **102** (2017), 138-149.

[10] D. Yoon, E.K. Ahn, M.Y. Park, S.Y. Cho, P. Ryan, M.J. Schuemie, et al. Conversion and Data Quality Assessment of Electronic Health Record Data at a Korean Tertiary Teaching Hospital to a Common Data Model for Distributed Network Research, *Healthc Inform Res* **22, 1** (2016), 54-58.

[11] E.A. Voss, R. Makadia, A. Matcho, Q. Ma, C. Knoll, M.J. Schuemie, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *J Am Med Inform Assoc* **22, 3** (2015), 553-564.

[12] R.H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F.M. Behlen, P.V. Biron, and A. Shabo, HL7 clinical document architecture, release 2. *J Am Med Inform Assoc* **13, 1** (2006), 30-39.

[13] D. Kalra, T. Beale, and S. Heard, The openEHR foundation, *Studies in Health Technol & Inform* **115** (2005), 153-173.

[14] K. Pommerening, and M. Reng, Secondary use of the EHR via pseudonymisation. *Studies in Health Technol & Inform* **103** (2004), 441-446.

[15] R.J. Bayardo, and R. Agrawal, Data privacy through optimal k-anonymization, *ICDE* **21** (2005), 217-228.

Address for correspondence

Daniel Mário de Lima <danielm@usp.br> (+55 11 2661 5545)

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190219

Combining Structured and Unstructured Data for Predicting Risk of Readmission for Heart Failure Patients

Satish M. Mahajan[a], Rayid Ghani[b]

[a] Veterans Affairs Palo Alto Health Care System, Palo Alto, California, USA,
[b] Center for Data Science & Public Policy, University of Chicago, Chicago, Illinois, USA

Abstract

Researchers have studied many models for predicting the risk of readmission for heart failure over the last decade. Most models have used a parametric statistical approach while a few have ventured into using machine learning methods such as statistical natural language processing. We created three predictive models by combining these two techniques for the cohort of 1,629 patients from six hosptials using structured data along with their 136,963 clinical notes till their index admission, stored in the EMR system over five years. The AUCs for structured and combined models were very close (0.6494 and 0.6447) and that for the unstructured model was 0.5219. The clinical impact of the models using decision curve analysis showed that, at a threshold predicted probability of 0.20, the combined model offered 15%, 30%, and 70% net benefit over its individual counterparts, treat-all, and treat-none strategy respectively.

Keywords:

Heart Failure; Machine Learning; Electronic Health Records

Introduction

The Affordable Care Act of 2010 in the United States mandated penalties for health organizations with high readmission rates in an effort to improve quality of care and patient outcomes. The Centers for Medicare and Medicaid Services (CMS) implemented the law via the Readmissions Reduction Program (HRRP) [10]. This program has survived three repeal efforts in the past 18 months, and, in fact, has continued to improve its risk assessment strategies for readmissions. The risk formula now looks at readmission rates for six major diseases (heart failure, elective hip/knee arthroplasty, pneumonia, chronic obstructive pulmonary disease, myocardial infraction, and coronary artery bypass graft) for stratified hospital comparisons and uses it for prospective payments to hospitals for Medicare and Medicaid patients.

This change has launched the interest and development of risk prediction models for patient readmissions to hospitals. Availability of data in Electronic Health Records (EHR) systems has pushed this development further. Many variations of models for predicting readmissions have been developed, particularly for heart failure, in the last few years [8]. Heart failure (HF) epidemiology highlights this interest: HF affects 6.6 million adults in the United States alone and 550,000 new cases are reported annually [3]. It is the most frequent and expensive disease category for 30 day readmissions [2]. The chronicity of HF disease is further marked by repeated hospitalizations and consumes 70% of total cost related to the disease [4]. Moreover, as a part of the American Heart Association's Policy Statement on forecasting the future of cardiovascular diseases in the United States, Heidenreich et al. [5] provide projections for the 2010 to 2030 timeframe for HF: an increase of 215% in direct costs, 80% in indirect costs, and 25% in prevalence respectively.

A closer examination of HF predictive models for readmission reveals their specific characteristics: mortality and readmission appear to be influenced by different predictors and hence many models have preferred to treat them as independent outcomes. Many recent studies have used a 30-day timeframe for modeling readmission risk in line with the CMS lead, whereas older studies have used a somewhat variable timeframe ranging from seven days to one year [8]. The 30-day timeframe appears to be a strong representative of readmissions for many diseases, as a similar readmission pattern arises for different diseases based on time-to-event analyses [6]. Furthermore, most models have considered all-cause readmission to encourage hospitals to implement intervention programs to improve overall care—such as better discharge coordination, medication reconciliation, and post-discharge follow-up—instead of using a narrow set of condition-specific solutions. In terms of data sources used for predictor selection, recent models prefer to use EHR systems over registry or claims-based datasets [8].

Some predictive models [7] have used parametric algorithms derived mostly from generalized linear methods bolstered with modern strategies, such as variable class weighting, cross-validation, and regularization. These models emphasize the importance of interpretability; however, they require significant construction time especially with larger predictor sets and their interactions. Furthermore, they asymptotically become numerically unstable and even fail to build due to combinatorial explosion during variable selection process. Meanwhile, other predictive models [9] have deployed non-parametric methods mostly using machine learning techniques with large datasets comprising many predictors. This class of models focuses more on predictive accuracy at the cost of sensitivity to the configuration parameters of the underlying algorithms as well as to the features of datasets. Howerver, combining and comparing both the approaches with the same underlying patient population remains unexplored.

In this study, we built three predictive models: one based on the parametric approach using structured predictors readily available in the EHR system; the second based on a non-parametric approach using Natural Language Processing (NLP) on unstructurred data representing patient notes in the EHR system; and a third that combined predictors from both the approaches. We further compared discriminative power,

calibration, and usefulness of each model in clinical decision making.

Methods

Patient Population and Data Extraction

We used EHR data at six hospitals from the Veterans Health Administration system to derive and validate the predictive models for this observational retrospective cross-sectional study. We selected all the hospitals from one network of hospitals from the western region of the United States for this study. The datasets from all hospitals under the health system are extracted and loaded into a central repository on a regular basis and we leveraged this infrastructure for the data collection for our study.

Based on our review of past literature [8] describing the predictive models built for estimating risk of readmission for heart failure, we used variables from clinical, administrative, and psychosocial categories that are routinely collected during episodes of care. Application of the same data policy framework imposed by the EHR system helped us to develop common logic and programs for data extraction. Datasets for five consecutive calendar years from 2011 to 2015 were extracted using International Classification of Diseases version 9 – Clinical Modification (ICD-9-CM) codes for heart failure as a principal discharge diagnosis. We used 2011 – 2014 data for derivation cohort and 2015 data for validation cohort formations. This setup provided an external temporal validation for the models. Since the validation cohort consisted of patients from different regions, the setup also provided an external geographic validation for the models.

Primary Outcome

We adhered to the commonly adopted definition of any-cause 30-day readmission in the literature [6], which excludes hospitalizations with a length of stay less than one day and any elective hospitalizations. If the patient encountered multiple readmissions within the 30-day timeframe from the discharge date of the index admission, only the last readmission episode was considered so that the latest health status for the patient is used for predicting the risk of future readmission. This setup provided us with the sample of patients that was blinded for outcome by programming for the above rules. The sample was also statistically independent and mutually exclusive across the two classes of admission groups of non-readmitted and readmitted patients.

Predictors

Structured data predictors were broadly classified into clinical, administrative, and psychosocial categories based on prior literature survey and expert clinical consultation [8] and are shown in Table 1. All the variables were collected either on or before the discharge date but closest to it depending on the data availability.

Unstructured data for the patients in the cohort were obtained from clinical notes—such as history and physician note at admission, progress notes, social worker notes, and discharge summary—till the date of discharge for the index admission. Since the bulk extraction of the structured and unstructured data from the EHR system was carried for the study cohort, predictor assessment was blinded to the other predictor and outcome variables.

Table 1 – Structured Dataset

Clinical Set	Predictor Name
Labs & Vitals	Sodium
	Potassium
	Blood Urea Nitrogen
	Creatinine
	Hemoglobin
	Hematocrit
	Glucose
	Albumin
	B-Natriuretic Peptide
	Systolic/Diastolic Blood Pressure
	Pulse
	Respiratory Rate
Administrative & Psychosocial Sets	**Predictor Name**
Demographics	Age
	Sex (Male/Female)
	Race (Native American/Asian/ Black/Hispanic/White/Not specified)
	Marital Status (Married/Divorced/ Separated/Unknown)
	Insurance Type (Multiple Insurance/ Blue Cross-Shield/Special Medicare /Disability/Income Medicaid/None)
	Residential Area (Urban/Rural/Highly Rural/Unknown)
Pre-Index Admission Factors	Appointments in Past Year
	No Show to Appointment in Past Year
	ED Visits in Past Year
	Prior Diagnoses
	Admissions in Previous Year
	Telemetry Monitor During Ind. Adm. (Yes/No)
	Index Adm. via ED (Yes/No)
	Length of Stay
	Concurrent Procedures
Comorbidities and Concurrent Procedure	Alcohol Abuse (Yes/No)
	Cardiac Arrhythmia (Yes/No)
	Coronary Artery Disease (Yes/No)
	Cancer (Yes/No)
	Cardiomyopathy (Yes/No)
	Cerebrovascular Accident (Yes/No)
	Depression (Yes/No)
	Diabetes Mellitus (Yes/No)
	Drug Abuse (Yes/No)
	Functional Disability (Yes/No)
	Liver Disease (Yes/No)
	Lung Disease (Yes/No)
	Lung Disease (Yes/No)
	Protein Caloric Malnutrition (Yes/No)
	Psychiatric Disorder (Yes/No)
	Rheumatic Disease Group (Yes/No)
	Renal Disease Group (Yes/No)
	Vascular Disease Group (Yes/No)
	Aortic Valve Disorder (Yes/No)
	Cancer related (Yes/No)
	Cardiac Devices (Yes/No)
	Cardiac Surgery (Yes/No)
	Coronary Angioplasty (Yes/No)
	History of Mechanical Ventilation Devices (Yes/No)

Statistical Analysis

The extraction of data from the EHR system yielded a sample size of 1629 admissions with 114 variables with dummy variable formats for categorical variables. We did not have any missing values for the dependent variable; a few independent variables such as BNP and blood glucose had up to 5% missing values. We used multiple imputation by chained equations resampled over five imputed datasets for the missing values assuming missingness at random. We created separate dummy variables for missing values with higher rate of missingness for the categorical variables. This strategy allowed us to use all the available records in the analysis. Continuous variables were also examined for nonlinear effects and transformations were carried wherever necessary.

We extracted 54 different types of clinical notes totaling 136,963 (102,055 for derivation cohort and 34,908 for validaton cohort) for our study and could not find any notes for 10 patients for the timeframe of interest. We eliminated these patients from the structured dataset for fair comparison of models using structured and unstructured datasets. Our final structured dataset thus contained 1,619 total patients (1,279 for derivation cohort and 340 for validation cohort). The corpus datasets for derivation and validation cohorts were created separately and processed for tokenization and stop words. They were then treated with term frequency-inverse document frequency statistic for vectorization with bigram range and term frequency thresholds of 0.01 and 0.80 for minimum and maximum cutoff respectively. We linked the structured and unstructured datasets on patients and their admission episodes for creating a combined dataset.

We fitted all three models using logistic regression with their respective feature sets to predict the outcome of 30-day readmission and reported their results for the validation datasets. We did not apply any special variable selection methods in this study. All confidence intervals were calculated at 95% level using 2,000 stratified bootstrap replicates. We further calculated and plotted the discriminative power using c-statistic, precision, recall, and model calibration for each model along with the top 10 words most predictive of readmission in the unstructured data. We also carried out decision curve analysis using the three models to understand their relative utilities with respect to treat-none and treat-all strategies. We used Microsoft SQL Server Version 12 Release 2, R version 3.0.2 (The R Foundation for Statistical Computing Platform), and Python version 3.6.4 for data extraction and statistical analyses. This Pilot Project under Quality Improvement initiatives was exempted from the review of full committee Institutional Review Board.

Results

Our validation cohort's empirical readmission rate was 32.7% excluding repeated patient readmissions. We created and fitted three separate models with structured, unstructured, and the combined datasets. Figure 1(A) shows precision and recall plotted on the left and right Y-axis respectively against percentage of population on the X-axis for the models. The Area-Under-Curve (AUC) or c-statistic was calculated for the models and plotted as Receiver Operating Characteristic (ROC) curves as shown in Figure 2(A). The AUC values with 95% confidence interval were: structured model: 0.6494 [0.5885-0.7103]; unstructured model: 0.5219 [0.5157-0.5281]; combined model: 0.6447 [0.6386-0.6508].

Calibrations with ideal, apparent, and bias-corrected fits with 300 bootstrap repetitions for the models were carried out and plotted as actual probabilities on Y-axis against model-predicted probabilities on X-axis as shown in Figure 1(B). The word sets with the top 10 words that were most predictive of readmission versus non-readmission using notes are shown in Figure 2(B). The Feature Importance Plot for the combined dataset is shown in Figure 2(C).

Finally, we carried out decision curve analysis (DCA) [11] for the models to understand net benefit for each model at various thresholds of predicted probabilities. It is shown in Figure 2(D).

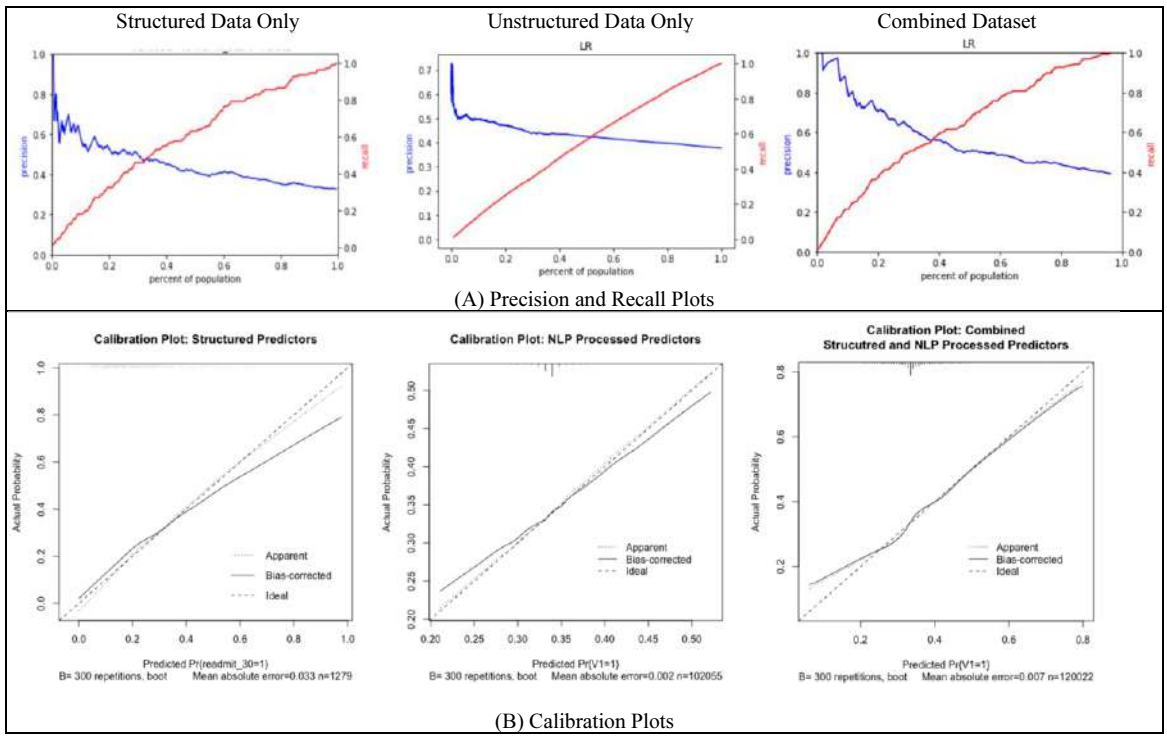
Discussion

As shown in Figure 1(A), precision and recall seem to have benefited from combining the structured and unstructured datasets especially when examining the 20% of the population with the highest precision: the combined dataset precision climbed to 70% from 45-50% precision of the individual datasets. Similarly, the combined dataset recall climbed to 40% from 20-30% recall of the individual datasets. Precision measures the accuracy of the model when it predicts an index admission as likely to have a readmission. Recall measures the ability of the model to find index admissions that have readmissions. As there is a tradeoff between these two measures, they can be varied and plotted for various thresholds of population as well.

As shown in Figure 2(A), the ROC curves represent the discriminative ability of the models by plotting the ratio of true positives to predicted positives on Y-axis against the ratio of true negatives to predicted negatives on X-axis. The curve that represents the maximum pull in the northwest direction of the graph signifies the best discrimination, while the diagonal axis represents random prediction. The unstructured data model had poor discrimination in our study, whereas the structured data model and combined model had similar discrimination with the combined model showing slightly better performance for non-readmitted patients. Our empirical datasets represented high class-imbalance (~30% readmitted versus 70% non-readmitted patients) and hence precision-recall curves might be better measures of model performance in the absence of tuning performed on the class-weight parameter.

Calibration plot scheme using bootstrapping method works as an efficient bias estimator since the difference between fits of the whole sample and a new sample is estimated by the difference between fits of the whole sample and a sample with replacement. The plots in Figure 1(B) show presence of over-fitting for predicted probabilities over 0.35 for all the models. The ideal fits with no over-fitting and no over-estimation are shown as the dotted diagonal axis at 45^{o} in the plots. All the models also show under-fitting below 0.35 of predicted probabilities. Moreover, all the models are under-estimated with positive Y-axis intercepts indicating need for better functional form for the models.

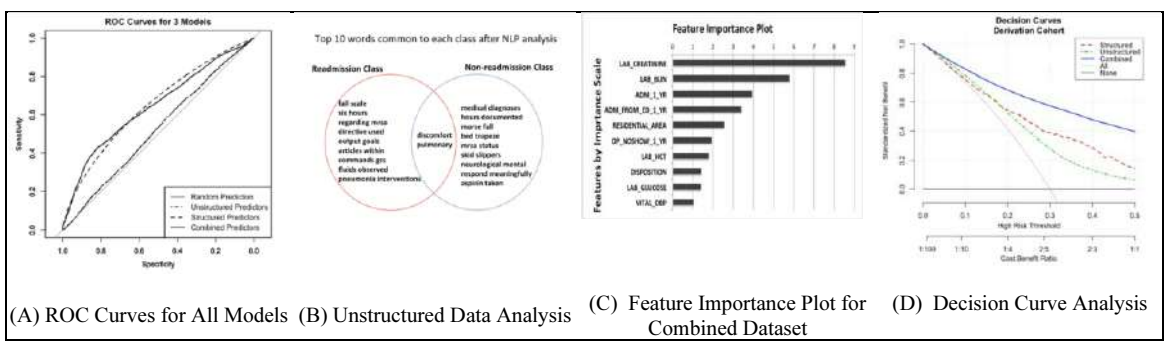
When we identified the top 10 words for both readmitted and non-readmitted patients as shown in Figure 2(B), there was only one dyad (*pulmonary discomfort*) common to both the classes. However, closer examination of the word sets reveals that some other words, even though written differently, are semantically equivalent: for example, word pairs such as *gcs commands* (Glasgow Command Scale) and *respond meaningfully*; *regarding mrsa* (Methicillin-Resistant Staphylococcus Aureus) and *mrsa status*. Such semantically



(A) Precision and Recall Plots

(B) Calibration Plots

Figure 1 – Plots with Validation Cohort for Models



(A) ROC Curves for All Models (B) Unstructured Data Analysis (C) Feature Importance Plot for Combined Dataset (D) Decision Curve Analysis

Figure 2 – Decision Analysis for Models

similar words need to be clustered together to further guide supervised learning. Given a large set of clinical documents for each patient episode, it would also be useful to further analyze specific documents with targeted word selections that have greater impact on readmission prediction.

The combined dataset has over 4900 predictors and hence we showed a plot of relative importance of the top 10 predictors in the model in Figure 2(C). Laboratory tests such as creatinine and hematocrit from the clinical set of structured predictors along with some administrative predictors such as patient hospital boarding characteristics and office visits status have strong influence on the readmission risk in our model. These predictors seem to be in line with the ones reported by the previous studies. Particularly, no predictor from the unstructured dataset made up to the top 10 list of the influencers of the patient readmission in the combined set.

Note that logistic regression is used as a baseline for the model development in all three. This is merely for convenience during methods development in this pilot study. The demonstrated methods could be used with any other parametric or machine learning models or the combination of both for structured and NLP datasets.

DCA is a method of evaluating usefulness of predictive models in clinical settings. It combines strengths of traditional biostatistical methods and decision-analytic methods while eliminating their limitations. For example, AUC for the model does not convey how high of an AUC is needed in order to deploy it in clinical practice. On the other hand, traditional decision analysis involves complex mathematics not suitable for continuous outcomes that is typical of predictive models. DCA is based on the relative harms caused by false positives and false negatives and is expressed as a threshold for

probabilities predicted by the models. As seen in Figure 2(D), the use of any predictive model provides positive standardized net benefit over both *treat none* (X-axis) and *treat all* (lowest curve cutting X-axis at risk threshold of 0.30) strategies. Either structured or unstructured data model appears to provide about the same net benefit of 55% over *treat none* and 15% over *treat all* strategies at risk threshold of 0.20. Moreover, the combined model (topmost curve in Figure 2(D)) shows consistently higher standardized net benefit over the other two models (15%) as well as *treat all* (30%) and *treat none* (70%) strategies.

Clinical Impact and Implications for Practice

We have combined and compared two seemingly different techniques for predicting 30-day readmissions for heart failure using most commonly available data in the EHR systems. Methods presented here pave the foundation for combining automated harvesting of predictors from unstructured data with carefully selected predictors from structured dataset. We have further demonstrated the ability to select the most useful model that has the greatest clinical impact. Decision analysis in this study has shown that the use of *some* predictive model is better than both using *no* model at all and treating *all* the admitted patients with the same readmission reduction program interventions. This appears to be valid despite having average descriminative capabilities (AUC = 0.6447) of the predictive model. Discharging physicians and case managers could make decisions for post-discharge care with their patients by identifying risks and benefits using such models. Even though the model development and associated activities of data processing pipeline might seem difficult to implement in practice, application of the selected risk threshold and outcome prediction could be captured in a simple mobile application for the clinicians in future implementations.

Limitations

This study represents a pilot project that brings together techniques for combining structured and unstructured data into a model for predicting readmissions for heart failure. It has used data from all the clinical notes for the patients till the index admission. Targeted use of notes with specific topic and word importance might help in better performing NLP system and remains unexplored in this study. Readmissions outside the index hosptials, even though deemed low, are not considered in this study. Finally, the study has applied external temporal validation but external geographical validation will further evaluate generalizability of this approach.

Conclusions

We built separate predictive models with the same EHR system's structured and unstructured data with two different modeling techniques. We compared the performance metrics for the two models and their combination. Our analysis showed that there was some benefit in combining these data although AUC for structured data model (0.6494) was not very different from the combined data model (0.6447). The combined model showed better results in terms of calibration plot and precision-recall curves. The decision curve analysis for assessing the clinical impact of practical usefulness of the models showed that the combined model offered 15% net benefit over its individual counterparts at a threshold predicted probability of 0.20. We have described the model development and validation efforts for this study using the Transparent Reporting of a multivariate prediction model for Individual Prognosis or Diagnosis (TRIPOD) guidelines [1].

Acknowledgements

Part of this work was carried under the collaboration between co-authors via 2018 Data Science Rotations for Advancing Discovery Trip program. This work was performed under operational and quality improvement initiatives at Veterans Affairs Palo Alto Health Care System and views expressed are those of the authors and not necessarily those of the system.

References

[1] G.S. Collins, J.B. Reitsma, D.G. Altman, and K.G. Moons, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement, *BMJ* **350** (2015), g7594.

[2] N. Fida and I.L. Pina, Trends in heart failure hospitalizations, *Curr Heart Fail Rep* **9** (2012), 346-353.

[3] A.A. Fitzgerald, L.A. Allen, and F.A. Masoudi, The evolving landscape of quality measurement for heart failure, *Ann N Y Acad Sci* **1254** (2012), 131-139.

[4] M.J. Hall and C.J. DeFrances, National Hospital Discharge Survey, in: *Advance data from vital and health statistics*, National Center for Health Statistics, Hyattsville, MD, 2001.

[5] P.A. Heidenreich, J.G. Trogdon, O.A. Khavjou, J. Butler, K. Dracup, M.D. Ezekowitz, et al., C. American Heart Association Advocacy Coordinating, C. Stroke, R. Council on Cardiovascular, Intervention, C. Council on Clinical, E. Council on, Prevention, A. Council on, Thrombosis, B. Vascular, C. Council on, C. Critical, Perioperative, Resuscitation, N. Council on Cardiovascular, D. Council on the Kidney in Cardiovascular, S. Council on Cardiovascular, Anesthesia, C. Interdisciplinary Council on Quality of, and R. Outcomes, Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association, *Circulation* **123** (2011), 933-944.

[6] H.M. Krumholz, L. Horwitz, C. Partovian, Z. Lin, J. Herrin, J. Grady, M. Conover, J. Montague, C. Dillaway, K. Bartczak, J. Ross, S. Bernheim, and E. Drye, Hospital-Wide (All-Condition) 30-Day Risk-Standardized Readmission Measure, in, Yale New Haven Health Services Corporation/Center, New Haven, CT, 2008.

[7] S.M. Mahajan, P. Burman, A. Newton, and P.A. Heidenreich, A Validated Risk Model for 30-Day Readmission for Heart Failure, *Studies in health technology and informatics* **245** (2017), 506-510.

[8] S.M. Mahajan, P.A. Heidenreich, B. Abbott, A. Newton, and D. Ward, Predictive models for identifying risk of readmission after index hospitalization for heart failure: A systematic review, *European Journal of Cardiovascular Nursing* (2018).

[9] S.M. Mahajan, A.S. Mahajan, R. King, and S. Negahban, Predicting Risk of 30-Day Readmissions Using Two Emerging Machine Learning Methods, *Studies in health technology and informatics* **250** (2018), 250-255.

[10] United States Center for Medicare and Medicaid Services, Readmissions Reduction Program (HRRP), 2018.

[11] A.J. Vickers and E.B. Elkin, Decision curve analysis: a novel method for evaluating prediction models, *Med Decis Making* **26** (2006), 565-574.

Address for correspondence

Satish M. Mahajan, VA Palo Alto Health Care System, Mailstop 118R, 3801 Miranda Ave, Palo Alto, CA 94304, USA.

Email: satish.mahajan@va.gov.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190220

Using Ensemble Machine Learning Methods for Predicting Risk of Readmission for Heart Failure

Satish M. Mahajan[a], Rayid Ghani[b]

[a] Veterans Affairs Palo Alto Health Care System, Palo Alto, California, USA,
[b] Center for Data Science & Public Policy, University of Chicago, Chicago, Illinois, USA

Abstract

Recently, researchers have been applying many new machine learning techniques for predicting the risk of readmission for heart failure. Combining such techniques through ensemble schemes holds a promise to further harness predictive performance of the resulting models. To that end, we examined two ensemble schemes and applied them to a real world dataset obtained from the EMR systems for 36,245 patients from 117 hospitals across the United States over five years. Both the ensemble schemes provided similar discriminative ability (AUC: 0.70, F1-score: 0.58) that was at least equal to or better than the base models that used a single machine learning method. The clinical impact of the models using decision curve analysis showed that at a threshold predicted probability of 0.40, the ensemble models offered 20% net benefit over the treat-all and treat-none strategies.

Keywords:
Patient Readmission; Heart Failure; Machine Learning

Introduction

Heart Failure (HF) is a chronic disease that affects 6.6 million adults in the United States alone and 550,000 new cases are reported annually [5]. The incidence of disease occurs mainly in older adults over 65 years who frequently have many other comorbidities. The main expense for the HF management is attributed to the hospitalizations (70% of total cost) [6] and the future financial burden forecasted until 2030 by the American Heart Association, estimates an increase of 215% in direct costs, 80% in indirect costs, and 25% in prevalence respectively [8]. It has thus become the most frequent and expensive disease for 30 day readmissions [4]. As a major insurance payer for older adults in the United States, the Centers for Medicare and Medicaid Services (CMS) has enacted penalties through its Readmissions Reduction Program (HRRP) for the healthcare organizations with high readmission rates [15]. The penalties are applied to the prospective payments to hospitals for medicare and medicaid patients. These developments have launched the interest and development of risk prediction models for patient readmissions to hospitals. Availability of data in Electronic Health Records (EHR) systems used in the hospitals has helped this development further.

Many variations of models for predicting readmissions have been developed, particularly for heart failure, over the last decade [12]. There are several interesting observations about the earlier models: many reserachers have modeled mortality and readmission as composite outcomes but that trend has declined recently since these outcomes appear to be influenced by different predictors. Along simliar lines, many older studies have used different timeframes for readmission ranging from seven to 365 days [12]. The more recent trend of using a 30-day timeframe appears to be a strong representative of readmissions for many diseases, as similar readmission patterns occur for different diseases based on time-to-event analyses [9]. Furthermore, most models have considered all-cause readmission to encourage hospitals to implement intervention programs to improve overall care; such as better discharge coordination, medication reconciliation, and post-discharge follow-up instead of using a narrow set of condition-specific solutions. In terms of data sources used for building predictive models, recent studies appear to use EHR systems more than registry or claims-based datasets due to the close data availability at the hospital level [12].

Focusing more on the methods used for building the models, and a recent systematic review of predictive models for HF [12], two main categories are noted: (a) interpretation oriented parametric and semi-parametric approaches such as logistic regression, elastic net, and LASSO; (b) predictive accuracy focused machine learning methods such as decision trees, random forests, neural nets, and gradient boosting. Recently, some studies [2; 11; 13] have also compared different methods for their effectiveness and discriminative abilities. The parametric methods suffer from long model building time and numerical instability with a large set of variables whereas machine learning methods are computation intensive with hyper-parameter tuning and show signs of over-fitting to training datasets. To this end, Laan van der et al. [10] have created a theoretical foundation to mitigate the shortcomings of individual methods and to harness the predictive power of different classes of models by combining them. In practice, humans have used such heuristic decision making in forms such as election voting and expert consultations. Perhaps the most well known example of such an approach over the last decade is from the winning team of Netflix's movie recommendation system contest. The team combined many models from the prior years' competitions to deliver the best possible performance for predicting and recommending movies for system users. In the healthcare domain, Shameer et al. [16] have recently experimented to predict readmission risk for HF patients by combining multiple models using correlation-based feature selection criteria for a large set of features extracted from the EHR system. Similarly, Turgeman et al. [17] have used two machine learning methods to overcome some of the limitations of individual methods while predicting risk of readmission for HF patients. However, use of newer and more advanced methods, called ensemble, for building risk prediction models for health outcomes remains largely unexlpored. In this study, we examined basic

properties and types of ensembles and applied them to a set of patient population to predict the risk of readmission for heart failure. We further explored their discriminative power, calibration, and usefulness in clinical decision making.

Methods

Patient Population and Data Extraction

We used EHR data at 117 hospitals from a large nationwide health care system in the United States to derive and validate the predictive models for this observational retrospective cross-sectional study. The datasets from all hospitals under the health system are extracted and loaded into a central repository on a regular basis and we leveraged this infrastructure for the data collection for our study.

Based on our review of the past literature [12] describing the predictive models built for estimating the risk of readmission for heart failure, we used variables from clinical, administrative, and psychosocial categories that are routinely collected during episodes of care. Application of the same data policy framework imposed by the EHR system helped us to develop common logic and programs for data extraction. Datasets for five consecutive calendar years from 2011 to 2015 were extracted using International Classification of Diseases version 9 – Clinical Modification (ICD-9-CM) codes for heart failure as a principal discharge diagnosis. We used 2011 – 2014 data for derivation cohort and 2015 data for validation cohort formations. This setup provided an external temporal validation for the models. As the validation cohort consisted of patients from all the hospitals, it also provided external geographic validation for the models within the United States.

Primary Outcome

We adhered to the commonly adopted definition of any-cause 30-day readmission in the literature [9], which excludes hospitalizations with a length of stay less than one day and any elective hospitalizations. If the patient encountered multiple readmissions within the 30-day timeframe from the discharge date of the index admission, only the last readmission episode was considered so that the latest health status for the patient is used for predicting the risk of future readmission. This setup provided us with the sample of patients that was blinded for outcome by programming for the above rules. The sample was also statistically independent and mutually exclusive across the two classes of admission groups of non-readmitted and readmitted patients, even though machine learning based methods do not impose this constraint.

Predictors

Structured data predictors were broadly classified into clinical, administrative, and psychosocial categories based on prior literature survey and expert clinical consultation [12] and are shown in Table 1. All the variables were collected either on or before the discharge date but closest to it depending on the data availability. Since the bulk extraction from the EHR system was carried for the study cohort, predictor assessment was blinded to the other predictor and outcome variables.

Statistical Analysis

The extraction of data from the EHR system yielded a sample size of 36,245 admissions with 72 variables including dummy variable formats for categorical variables. We counted 27,714 admissions for the derivation cohort and 8,531 admissions for the validation cohort with the overall observed readmission rate of 35.7%. We did not have any missing values for the dependent variable; a few independent variables such as BNP

Table 1 – Structured Dataset

Clinical Set	Predictor Name
Labs & Vitals	Sodium
	Potassium
	Blood Urea Nitrogen
	Creatinine
	Hemoglobin
	Hematocrit
	Glucose
	Albumin
	B-Natriuretic Peptide
	Systolic/Diastolic Blood Pressure
	Pulse
	Respiratory Rate
Administrative & Psychosocial Sets	**Predictor Name**
Demographics	Age
	Sex (Male/Female)
	Race (Native American/Asian/ Black/Hispanic/White/Not specified)
	Marital Status (Married/Divorced/ Separated/Unknown)
	Insurance Type (Multiple Insurance/ Blue Cross-Shield/Special Medicare /Disability/Income Medicaid/None)
	Residential Area (Urban/Rural/Highly Rural/Unknown)
Pre-Index Admission Factors	Appointments in Past Year
	No Show to Appointment in Past Year
	ED Visits in Past Year
	Prior Diagnoses
	Admissions in Previous Year
	Telemetry Monitor During Ind. Adm. (Yes/No)
	Index Adm. via ED (Yes/No)
	Length of Stay
	Concurrent Procedures
Comorbidities and Concurrent Procedure	Alcohol Abuse (Yes/No)
	Cardiac Arrhythmia (Yes/No)
	Coronary Artery Disease (Yes/No)
	Cancer (Yes/No)
	Cardiomyopathy (Yes/No)
	Cerebrovascular Accident (Yes/No)
	Depression (Yes/No)
	Diabetes Mellitus (Yes/No)
	Drug Abuse (Yes/No)
	Functional Disability (Yes/No)
	Liver Disease (Yes/No)
	Lung Disease (Yes/No)
	Lung Disease (Yes/No)
	Protein Caloric Malnutrition (Yes/No)
	Psychiatric Disorder (Yes/No)
	Rheumatic Disease Group (Yes/No)
	Renal Disease Group (Yes/No)
	Vascular Disease Group (Yes/No)
	Aortic Valve Disorder (Yes/No)
	Cancer related (Yes/No)
	Cardiac Devices (Yes/No)
	Cardiac Surgery (Yes/No)
	Coronary Angioplasty (Yes/No)
	History of Mechanical Ventilation Devices (Yes/No)

and blood glucose had up to 5% missing values. Some methods such as decision trees and its variants accept predictors with mising values, whereas some other methods such as logistic regression need all predictor values to be present. Hence, for consistency we used multiple imputation by chained equations resampled over five imputed datasets for the missing values assuming missingness at random. We created separate dummy variables for missing values with higher rate of missingness for the categorical variables [1; 7]. This strategy allowed us to use all the available records in the analysis. Continuous variables were also examined for nonlinear effects and transformations were carried wherever necessary.

We investigated 10 different base learning models in this study. The base learners were chosen to represent various method families such as linear models, tree based and bagging models, boosting-based models, and neural networks. The base learners were run individually for the optimal AUC (Area Under Curve) performances by tuning their respective hyper-parameters. Figure 1 shows general setup of an ensemble system. The given dataset is applied to each *base learner* that implements a predictive model to predict the outcome of interest. A set of these predicted outcomes is then applied to a final model called a *meta learner* that uses some method to produce the final predictive outcome that is used to score the risk on the data that is not seen previously. We used Super learner and Subsemble schemes to combine base learner outputs using varying number of folds, partitions, and meta learner models to understand the effect of variations in these parameters on the ensemble performance. More details about the ensemble structure are provided in the Discussion section below.

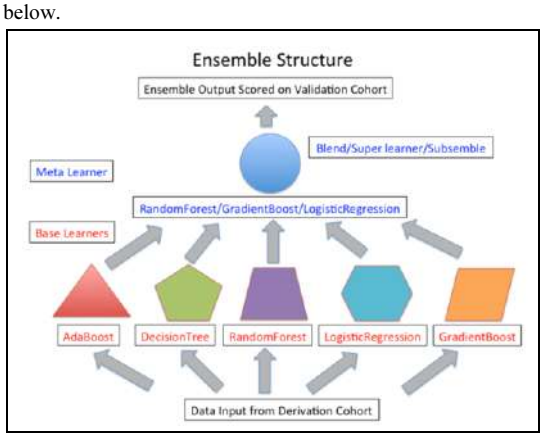


Figure 1 – Ensemble Structure

All confidence intervals were calculated at 95% level using 2000 stratified bootstrap replicates. We further calculated an AUC and a F1-score for each base learner and ensemble and also plotted a calibration plot and ROC (Receiver Operating Characteristics) curves for showing discriminative power. We also carried out decision curve analysis for the base learner and two ensemble schemes to understand their relative utilities with respect to treat-none and treat-all strategies.

We used Microsoft SQL Server Version 12 Release 2, R version 3.0.2 (The R Foundation for Statistical Computing Platform), and python version 3.6.4 for data extraction and statistical analyses. This Pilot Project under Quality Improvement initiatives was exempted from the review of full committee Institutional Review Board.

Results

We fitted 10 different base models with the derivation cohort by tuning with their respective hyper-parameters to optimize for the AUC using the validation cohort. The models and their individual performances are shown in Table 2.

The ensemble methods were tried with LogisticRegression, RandomForest, GradientBoost, and NeuralNetwork as meta learners. There was no performance difference in terms of root mean square errors and the AUCs between the Super learner and Subsemble schemes for the same meta learner. Furthermore, ensembles performed at the same level when the number of folds and partitions of the cross-validation scheme were changed. The final ensemble schemes used 16-fold, 16-partition combination on the validation cohort with GradientBoost and RandomForest meta learners for the Super learner and Subsemble respectively. The F1-score in Table 2 represented harmonic mean between precision and recall for the models.

Table 2 – Base Learners and Ensembles

Base Learner	AUC [95% CI]	F1-Score
RandomForest	0.6922 [0.6791-0.7058]	0.5780
AdaBoost	0.6612 [0.6496-0.6747]	0.5726
LogisticRegression	0.6914 [0.6801-0.7030]	0.5797
GradientBoost	0.6725 [0.6611-0.6840]	0.5676
DecisionTrees	0.6192 [0.6081-0.6326]	0.5437
ExtraTrees	**0.6993 [0.6885-0.7107]**	**0.5841**
KNearestNeighbor	0.6498 [0.6385-0.6611]	0.5592
NeuralNetwork	0.6962 [0.6854-0.7076]	0.5869
NaiveBayes	0.0095 [0.0041-0.0679]	0.5388
SupportVectorMachine	0.6869 [0.6748-0.6980]	0.5797
Ensemble		
Super Learner (using GradientBoost)	**0.6987 [0.6872-0.7101]**	**0.5843**
Subsemble (using RandomForest)	0.6914 [0.6798-0.7030]	0.5747

The KNearestNeighbor as a base learner was chosen merely to highlight the differentiation in the curves in Figure 2(A) and Figure 2(C). The AUC or C-statistic was calculated for the base learner and both the ensembles and was plotted as ROC curves for comparison as shown in Figure 2(A).

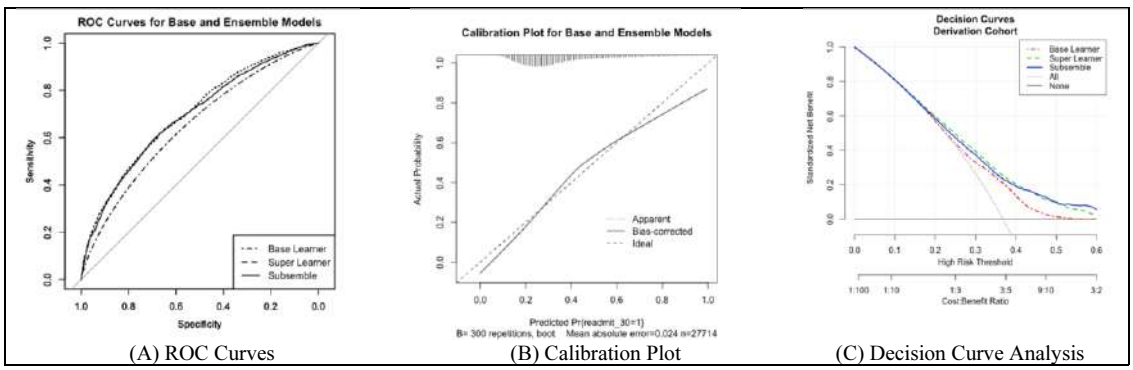
Calibration with ideal, apparent, and bias-corrected fits with 300 bootstrap repetitions for the derivation cohort was carried out and plotted as actual probabilities on Y-axis against model-predicted probabilities on X-axis as shown in Figure 2(B).

Finally, we carried out Decision Curve Analysis (DCA) [18] for the base learner and both the ensembles to understand net benefit for each model at various thresholds of predicted probabilities and is shown in Figure 2(C).

Discussion

Characteristics of Ensemble System

The seemingly simple ensemble structure shown in Figure 1 is based on a few important principles: (a) A meta learner in the ensemble has to learn optimal combination of the base learners' predictions. In its simplest form, this could mean to implement simple or weighted majority-voting algorithm if the



(A) ROC Curves (B) Calibration Plot (C) Decision Curve Analysis

Figure 2 – Performance Plots for Base and Ensemble Models

outcomes of the base learners are discrete numbers or a class such as that resulting from the support vector machine algorithm. For continuous outcome predictions from models such as logistic regression and boosting, simple arithmetic combinations (mean, sum, or product) could be applied.

In practice, most ensembles apply complex combination of algorithms to arrive at the final predictive outcome from the meta learner. (b) The ensembles improve their performance by correcting the base learners' prediction errors. Moreover, the prediction errors need to be relatively uncorrelated to improve the overall ensemble performance. This principle partly relies on the choice of the base learners and their hyper-parameter tuning. A use of a set of base learners is embraced from the *no free lunch theorem*, which suggests that no single algorithm is optimal to solve every problem. Another important point highlighted by this principle is that combining the base learner outputs in the ensemble does not necessarily lead to the performance that is guaranteed to be better than the best base learner; rather, it reduces the likelihood of choosing a suboptimal model with poor performance. (c) A part of the improvement in the ensemble performance comes from the ability of the base learners to see diverse data during their training phase. This objective is achieved by implementing various strategies discussed below that give rise to different types of ensembles.

Blend ensemble divides the input dataset into two parts, each of which is used by the base learner layer and the meta learner layer. Each layer in turn divides its data into a proportion of a train set and a test set. As in any supervised learning scheme, the learners use the train set to fit their models and the test set to assess the predictive performance. This ensemble is useful for training large datasets in a short time but has a shortcoming that the input dataset is used only once by the system.

Super learner ensemble uses cross-validation methods by formulating the minimization of the cross-validated risk over the base learners as a new least squares regression problem which itself is carried out by the meta learner using cross-validation itself, thereby preventing over-fitting of the cross-validated risk. The training time for this ensemble could be long, especially with large datasets using a large number of cross-validation folds, but the performance is much better than blend ensemble, especially with a high variance noisy data [10].

Subsemble ensemble partitions the full dataset into subsets of observations, fits a specified base learner's underlying algorithm on each subset, and uses a novel form of K-fold cross-validation to output a prediction function that combines the subset-specific fits. At the meta learner layer, this ensemble evaluates the performance relative to the underlying algorithm fit just once on the full dataset and hence is useful for small, medium, and large datasets [14]. Since our dataset comprised of one data record for one patient admission/readmission, it did not empirically fall into high volume, high velocity dataset. Moreover, each layer of the Blend ensemble uses only half the dataset. Hence, in order to effectively use our data, we did not use Blend ensemble structure. We applied our dataset to Super learner and Subsemble by systematically varying the parameters such as folds, partitions, and underlying meta learner. We did not notice significant difference in the performance of these two ensembles for the same set of parameters except that the Subsemble was able to complete the computations in 1/3rd amount of time taken by the Super learner. This might be attributable to its strategy to further partition data in each fold and work on all the partitions in parallel. We observed some but not significant performance differential between the two schemes if the underlying meta learners are different. This could be confirmed by looking at mostly overlapping ROC curves for the ensembles in Figure 2(A); the ensembles, however, show significantly better discriminative ability than the base learner.

Calibration plot in Figure 2(B) shows presence of over-fitting for predicted probabilities in the lower and upper tertiles, whereas under-fitting in the middle tertile. The ideal fits with no over-fitting and no over-estimation are shown as dotted line diagonal axis at 45^{o} in the plot. The dataset further demonstrates some over-estimation with negative Y-axis intercept indicating need for better functional form for the model.

DCA is based on the relative harms caused by false positives and false negatives and is expressed as a threshold for probabilities predicted by the models. As seen in Figure 2(C), the use of *any* predictive model provides positive standardized net benefit over both *treat none* (X-axis) and *treat all* (lowest curve cutting X-axis at risk threshold of 0.37) strategies. The model based on either ensemble scheme appears to provide about the same net benefit that is consistently better than the base learner model as well as the *treat all* and *treat none* strategies across all the predicted probabilities above 0.20. For example, at threshold probability of 0.40, the ensemble based models show 20% net benefit over the *treat all* and *treat none* strategies.

Clinical Impact and Implications for Practice

Researchers are developing many new methods for effective predictive modeling for outcomes important for patients and healthcare organizations alike. We have examined, applied, and compared various individual as well as ensembled

machine learning techniques for predicting 30-day readmissions for heart failure using most commonly available data in the EHR systems. Decision analysis in this study has shown that the use of *some* predictive model is better than both using *no* model at all and treating *all* the admitted patients with the same readmission reduction program interventions. Discharge physicians and coordinators as well as case managers could make post-discharge decisions with their patients by taking into account risks and benefits by applying such models in practice. Even though the model development and associated activities of data processing pipeline might seem daunting to implement in practice, the predicted risk threshold and outcome could be easily displayed in a simple mobile application for the clinicians.

Limitations

This study represents a pilot project that explored techniques for combining different machine learning models for predicting readmissions for heart failure using a limited dataset. It also used data from different hospitals but within the same health system. Many machine learning algorithms are known to work better with a large dataset and this study could be further explored in that setup.

Conclusions

In this study, we built predictive models with the EHR system data using various machine learning techniques independently, and in combination by deploying two different ensemble schemes. We compared the perfmance metrics for the base models and their ensembles. Our analysis showed that the ensemble techniques deliver the performance that is at least as good as the best performing base model. The decision curve analysis for assessing the clinical impact of practical usefulness of the models showed that the ensemble based models offered 20% net benefit over not using any model at a threshold predicted probability of 0.40. We have described the model development and validation efforts for this study using the Transparent Reporting of a multivariate prediction model for Individual Prognosis or Diagnosis (TRIPOD) guidelines [3].

Acknowledgements

Part of this work was carried under the collaboration between co-authors via 2018 Data Science Rotations for Advancing Discovery Trip program. This work was performed under operational and quality improvement initiatives at Veterans Affairs Palo Alto Health Care System and views expressed are those of the authors and not necessarily those of the Department of Veterans Affairs or other affiliated organizations.

References

[1] R. Amarasingham, B.J. Moore, Y.P. Tabak, M.H. Drazner, C.A. Clark, S. Zhang, W.G. Reed, T.S. Swanson, Y. Ma, and E.A. Halm, An Automated Model to Identify Heart Failure Patients at Risk for 30-Day Readmission or Death Using Electronic Medical Record Data, *Med Care* **48** (2010), 981–988.
[2] A. Artetxe, A. Beristain, and M. Graña, Predictive models for hospital readmission risk: A systematic review of methods, *Computer Methods and Programs in Biomedicine* **164** (2018), 49-64.
[3] G.S. Collins, J.B. Reitsma, D.G. Altman, and K.G. Moons, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement, *BMJ* **350** (2015), g7594.
[4] N. Fida and I.L. Pina, Trends in heart failure hospitalizations, *Curr Heart Fail Rep* **9** (2012), 346-353.
[5] A.A. Fitzgerald, L.A. Allen, and F.A. Masoudi, The evolving landscape of quality measurement for heart failure, *Ann N Y Acad Sci* **1254** (2012), 131-139.
[6] M.J. Hall and C.J. DeFrances, National Hospital Discharge Survey, in: *Advance data from vital and health statistics*, National Center for Health Statistics, Hyattsville, MD, 2001.
[7] B.G. Hammill, L.H. Curtis, G.C. Fonarow, P.A. Heidenreich, C.W. Yancy, E.D. Peterson, and A.F. Hernandez, Incremental value of clinical data beyond claims data in predicting 30-day outcomes after heart failure hospitalization, *Circ Cardiovasc Qual Outcomes* **4** (2011), 60-67.
[8] P.A. Heidenreich, J.G. Trogdon, O.A. Khavjou, J. Butler, et al, Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association, *Circulation* **123** (2011), 933-944.
[9] H.M. Krumholz, L. Horwitz, C. Partovian, Z. Lin, J. Herrin, J. Grady, M. Conover, J. Montague, C. Dillaway, K. Bartczak, J. Ross, S. Bernheim, and E. Drye, Hospital-Wide (All-Condition) 30-Day Risk-Standardized Readmission Measure, in, Yale New Haven Health Services Corporation/Center, New Haven, CT, 2008.
[10] M.J. Laan van der, E.C. Polley, and A.E. Hubbard, Super Learner, *U.C. Berkeley Division of Biostatistics Working Paper Series* **Working Paper 222** (2007).
[11] S.M. Mahajan, P. Burman, and M. Hogarth, Analyzing 30-Day Readmission Rate for Heart Failure Using Different Predictive Models, *Studies in health technology and informatics* **225** (2016), 143-147.
[12] S.M. Mahajan, P.A. Heidenreich, B. Abbott, A. Newton, and D. Ward, Predictive models for identifying risk of readmission after index hospitalization for heart failure: A systematic review, *European Journal of Cardiovascular Nursing* (2018).
[13] B.J. Mortazavi, N.S. Downing, E.M. Bucholz, K. Dharmarajan, A. Manhapra, S.X. Li, S.N. Negahban, and H.M. Krumholz, Analysis of Machine Learning Techniques for Heart Failure Readmissions, *Circ Cardiovasc Qual Outcomes* **9** (2016), 629-640.
[14] S. Sapp, M.J. van der Laan, and J. Canny, Subsemble: an ensemble method for combining subset-specific algorithm fits, *J Appl Stat* **41** (2014), 1247-1259.
[15] United States Center for Medicare and Medicaid Services, Readmissions Reduction Program (HRRP), 2018.
[16] K. Shameer, K.W. Johnson, A. Yahi, R. Miotto, L. Li, D. Ricks, J. Jebakaran, P. Kovatch, P.P. Sengupta, S. Gelijns, A. Moskovitz, B. Darrow, D.L. David, A. Kasarskis, N.P. Tatonetti, S. Pinney, and J.T. Dudley, Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: a case-study using Mount Sinai heart failure cohort, in: *Biocomputing 2017*, pp. 276-287.
[17] L. Turgeman and J.H. May, A mixed-ensemble model for hospital readmission, *Artif Intell Med* **72** (2016), 72-82.
[18] A.J. Vickers and E.B. Elkin, Decision curve analysis: a novel method for evaluating prediction models, *Med Decis Making* **26** (2006), 565-574.

Address for correspondence

Satish M. Mahajan, VA Palo Alto Health Care System,
3801 Miranda Ave, Palo Alto, CA 94304, USA.
Email: satish.mahajan@va.gov.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190221

Application of Machine Learning and Grocery Transaction Data to Forecast Effectiveness of Beverage Taxation

Xing Han Lu[a,b*], Hiroshi Mamiya[a*], Joseph Vybihal[b], Yu Ma[c], David L. Buckeridge[a]

[a] *Surveillance Lab, McGill Clinical and Health Informatics, McGill University, Montreal, Quebec, Canada*
[b] *School of Computer Science, McGill University, Montreal, Quebec, Canada*
[c] *Desautels Faculty of Management, McGill University, Montreal, Quebec, Canada*

Abstract

Sugar Sweetened Beverages (SSB) are the primary source of artificially added sugar and have a casual association with chronic diseases. Taxation of SSB has been proposed, but limited evidence exists to guide this public health policy. Grocery transaction data, with price, discounting and other information for beverage products, present an opportunity to evaluate the likely effects of taxation policy. Sales are often non-linearly associated with price and are affected by the prices of multiple competing brands. We evaluated the predictive performance of Boosted Decision Tree Regression (B-DTR) and Deep Neural Networks (DNN) that account for the non-linearity and competition across brands, and compared their performance to a benchmark regression, the Least Absolute Shrinkage and Selection Operator (LASSO). B-DTR and DNN showed a lower Mean Squared Error (MSE) of prediction in the sales of most major SSB brands in comparison to LASSO, indicating a superior accuracy in predicting the effectiveness of SSB taxation. We demonstrated the application of machine learning methods and large transactional data from grocery stores to forecast the effectiveness food taxation.

Keywords:

Beverages; Machine learning; Public policy

Introduction

Unhealthy diet is the leading preventable cause of global death and disability, claiming 11 million lives and 241 million disability adjusted lost life years in 2012 [1]. Diet-related chronic diseases, such as obesity, cardiovascular diseases, cancers, and type-2 diabetes mellitus impose a considerable burden on society and individuals. Taxation has been proposed as a public health policy to discourage the purchasing of unhealthy foods [2], most notably Sugar Sweetened Beverages (SSB). The primary source of artificially added sugar with a recognized association with obesity and major chronic diseases [3,4], SSB consists of beverages such as soda (carbonated soft drinks), fruits drinks, sports and energy drinks each containing many product brands (e.g. Coca-Cola™ and Pepsi™ in the category of soda). The expected effectiveness of taxation is determined by the magnitude of reduction in SSB purchasing that is likely to occur with an increase in the price of SSB. Formally, this value is called the *price elasticity of demand* and is quantified as the percent reduction in product purchased in response to a one percent increase in price.

Grocery transaction data are generated by scanning products at the time of purchasing from a large number of retail food outlets and can be used to estimate changes in SSB purchasing from price fluctuations. The data include the quantity of food products sold and ourchase details, such as the price and promotions (e.g. discounting, in-store display, and flyer). Although rarely used for public health research, grocery transaction data can be used to predict SSB sales conditional on pricing, promotions, consumer demographic and economic attributes of the store neighborhood (e.g. income and family size). Because sales of a product are influenced by its features (*focal features*), but also by the features of competing products in the same store (*competing features*), the prediction of beverage purchasing must take into account the influence of numerous competing brands.

Due to correlations in price and promotion across many food products, feature selection is critical. Researchers previously performed ad-hoc dimensionality reduction, such as aggregating product sales and features into broader SSB categories or modeling only a small number of brands [5]. These approaches masks the complex patterns of competition among individual food products, potentially resulting in biased estimate of price elasticity due to aggregation bias or omitted confounders. Therefore, prediction at the level of individual food items or brands is critical.

More importantly, associations between product features and sales are non-linear (i.e. deal-effect curve), and multiple product features can jointly affect sales through statistical interactions, or non-additive effects due to competitive interference and synergistic effect of promotions. The shape of the relationship between sales and product features is determined by a mix of factors including: 1) a threshold effect, where the quantity of sales remains constant until the price or promotion reaches a threshold value; 2) a saturation effect, where the growth of sales plateaus as consumers are de-sensitized by price reduction or promotions above certain levels; 3) a cross-item deal effect, where the sales of brands are affected by the price of a large competing brands; 4) an interaction between features of multiple brands e.g., price reduction of a brand fails to increase sales due to promotional activities of a larger competing brand; and 5) an interaction between features of the same brand e.g. the effect of discounting can be enhanced by promotions of the same brand [6].

While parametric estimators (e.g. linear regression) are traditionally used to model product demand, manual specification of non-linear functions and interactions is not feasible with dozens or hundreds of competing product features. In contrast, non-parametric algorithms, such as decision trees and artificial neural networks, naturally incorporate non-linear associations and interactions. Although typically used for classification, these algorithms can be used in a regression context where the output variable is numeric (i.e. sales).

Table 1 - Description of Brand-level, Temporal, and Store Neighborhood Predictive Features of Sugar Sweetened Beverage (SSB) sales. Each transaction record consists of sales (target variable) of specific SSB brands and these features at a given week and store. The third column describes either the number of categories (if categorical), or the range of values (if numerical).

Feature description	Type	Category count or numeric range
Brand-level features		
Chain code where product was sold	Categorical	3
Percent price discount (%)	Numeric	0 - 98.225
Prices in Canadian cents	Numeric	0.001 - 1399
Display advertisement frequency	Numeric	0 – 1
Flyer advertisement frequency	Numeric	0 – 1
Brand name	Categorical	154
Store code where product was sold	Categorical	74
Temporal features		
Month of Sale	Categorical	12
Week of Sale	Categorical	53
Store neighborhood features		
Proportion of post-secondary certification	Numerical	0.435 - 0.891
Average family size	Numerical	2 – 3
Proportion of family with child	Numerical	0.207 - 0.530
Proportion of single parent family	Numerical	0.103 - 0.274
Median family income ($/family)	Numerical	38046 - 108735
Proportion of immigrants	Numerical	0.005 - 0.606
Number of dwellings (families)	Numerical	2885 - 58710
Total population (inhabitants)	Numerical	7460 - 133570
Dwelling density (families/km2)	Numerical	1.145 - 6289.960
Total		
Log of Weekly Sales of brand	Numerical	-1.403 - 11.761

To date, SSB taxation is rarely implemented in developed nations, and the price elasciticy of demand is for the most part estimated by observational data using parametric demand estimation models that tend to suffer the limitations mentioned above [7]. Traditional econometric approaches (e.g. linear regression) to estimate the effect of price on SSB (price elasticity of SSB demand) is infeasible when modling the interaction of a large number of beverage products. We thus aim to evaluate the accuracy of non-parametric learning algorithms for predicting the sales of SSB from scanner grocery transaction data.

Data

We obtained weekly transaction records of food products purchased from 44 stores sampled to be geographically representative of three large retail grocery chains in the province of Quebec, Canada between 2008 and 2013. The data were indexed by time (week), store identification code, product name, price, and three promotional activities: discounting, in-store display (placement of a product in a prominent location) and flyer advertising. Price indicates the dollar amount paid by the consumer at the time of purchase (i.e. net price), while discounting is the percent reduction of the purchase price from the regular price calculated as the maximum purchase price in a three month trailing window [8].

Among many types of beverages, we were interested in predicting the sales of SSB, or artificially sweetened beverages which included soda, energy and sports drink, carbonated fruit beverage, frozen fruit juice, fruit drink, and carbonated soda water. There were 2,608 distinct SSB products defined by brand, flavor, and package type. Because products in the same brand tended to exhibit similar pricing and promotional patterns, we aggregated the value of sales, pricing and promotion into a smaller set of 154 distinct SSB brands, such as Coca-Cola™ and Pepsi™. Brand-level predictive features (i.e. price, discounting, display, and flyer advertisement) were calculated as the mean (price and discounting) and proportion promotion (display and flyer) across the products belonging to the brand.

Let $t := week, i := brand, j := store$. There were 1,509,280 weekly transaction records for the 154 SSB brands across all stores, with each record representing the brand-specific sales denoted as Y_{ijt}, which was the target variable and defined as the natural-log of the sales of brand i in store j at week t. The sales quantity was standardized to the U.S Food and Drug Administration serving size of 240 milliliters. Although the log transformation was relevant to parametric regression modeling only [9], we applied this transformation in accordance with existing practice in demand modeling.

The vector of brand-level focal features was denoted as X_{ijt} (Table 1, Brand-level features). We let S_j be the features of the store where the products were sold (categorical indicator of chain and store identification code) and store neighborhood socio-economic and demographic features that may influence food purchasing were obtained from the 2011 Canadian census (Table 1, Store neighborhood features). We let M_t and W_t represent categorical features indicating the month and week for each record to account for temporal fluctuations in purchasing (e.g. increases on holidays). As noted above, sales of a brand depend on the pricing and promotion of that brand (*focal brand features*) and on the features of popular competing brands (*competing brand features*). Because a few brands accounted for most of the market share in each SSB category (e.g. Coca Cola™ and Pepsi™ have nearly 70% of share in the soda category), their brand features have a strong influence on the sales of other brands. Thus, we extracted price and promotions of twenty brands with the highest market

Table 2 - Mean Squared Error of most popular brands of Sugar Sweetened Beverages

	Pepsi	Coca Cola	Seven Up	Crush	Sprite	Canada Dry	Nestle
LASSO	0.51	0.44	0.46	0.35	0.45	0.38	0.41
B-DTR	**0.17**	**0.16**	0.22	0.28	**0.22**	**0.24**	0.41
DNN	0.19	0.23	**0.21**	**0.23**	0.23	0.27	**0.31**

share among SSB that are denoted as C_{kjt}. The dimension of each feature vector was: (X_{ijt}, 245), (C_{kjt}, 80), (S_j, 9), (M_t, 12), and (W_t, 53).

Methods

We used two non-parametric methods: an ensemble of Decision Trees with Adaptive Boosting (B-DTR) and a fully-connected Deep Neural Network (DNN). The baseline model was a regularized linear parametric model (LASSO, or Least Absolute Shrinkage Selection Operator). The DNN was implemented in Keras [10], and the other models were implemented in Scikit-Learn [11]. Unless otherwise noted, normalization was done using standard mean shifting and variance scaling.

LASSO regression identifies a sparse set of features through shrinkage via L_1 regularization [12] and has been previously used for demand forecasting in high-dimensional feature space [13], even though explicit specification of non-linear features (e.g. spline) becomes unrealistic when modeling the sales of a large number of brands. We selected the regularization parameter λ by iterating over a range of values and selecting the one with lowest average mean squared error (MSE) through three-fold Cross Validation.

Decision Tree Regression is a rule-based learning algorithm that identifies a binary segmentation of predictive features, where the cut-point for each feature represents a decision boundary that minimizes the prediction loss (e.g sum of squared errors) for a target vector Y_{ijt}. The partitioning ends when pre-specified criteria, such as a maximum number of branches or a minimum number of observations at each terminal node, are met. The decision trees implement the Classification and Regression Trees algorithm due to its ability to predict numerical values [14].We used Drucker's improved Adaptive Boosting [15] meta-estimator to form an ensemble of 100 weak learners. The weight of each learner was determined by a linear loss. Each learner was a Decision Tree with varying depths, set to a maximum depth of 30 nodes. The value of each node was determined by the partition that best minimized the MSE.

The DNN model with the best results had four fully connected layers. Adam optimization was used to enable convergence with large data and noisy gradients [16]. The optimum values of exponential decay rates and fuzzy factors were selected based on training stability and the ability to converge. The network weight parameters were initialized using Normalized Initialization [17] to accelerate convergence. We trained the model using mini-batches of 128 samples to leverage the richness of the data and to provide inherent regularization [18], while maintaining a stable training process. We chose the activation function to be a Rectified Linear Unit due to its biological properties and strong experimental results on high-dimension datasets [19], due in part to its non-linearity, which allows the DNN to learn complex relationships between features.

The DNN had an input layer dimension of 389, and fed a 400-dimension vector to the first hidden layer. The first hidden layer output a 100-dimension vector to the next layer with a L_1 regularization and ReLU activation. The last hidden layer output was a 25-dimension vector to the output layer. The final layer outputs a single numerical value corresponding to the predicted log of sales, using a linear activation function to take into account negative target values (brands with extremely low sales has negative log values).

We extracted the first five years (2008-2012) of the transaction data for training and validation. We randomly sampled 90% of these data as the training set for learning algorithm parameters, leaving the remaining 10% as the validation set for evaluating the prediction accuracy of the algorithms. The final year (2013) of data was reserved to estimate prediction accuracy, measured as Mean Squared Error (MSE). Data were managed using Numpy, Pandas and PostgreSQL.

Results

The MSE for the prediction of all SSB brands in the 2013 transaction data was 0.67, 0.72, and 0.91 for DNN, B-DTR, and LASSO, respectively.

At the individual brand level, DNN, B-DTR, and LASSO showed best predictive performance for 80, 31, and 21 brands, respectively. Prediction error of seven most popular SSB brands driving overall sales of SSB is presented in Table 2. The DNN and B-DTR had comparable prediction accuracy for these brands, while LASSO showed the lowest accuracy except for the Nestle brand. Category-level prediction is of public health interest, since each beverage category can have different health effects. We thus calculated a category-level MSE (Table 3). The DNN had the highest prediction accuracy for the soda category, the most important source of added sugar consumption. LASSO showed the lowest prediction accuracy for all the categories.

Table 3 – Mean Squared Error for each Beverage Category, Across all Stores

	Soda	Soda Water	Energy Drinks	Frozen Juice	Frozen Drinks
B-DTR	0.650	0.743	0.956	**0.600**	**0.631**
DNN	**0.609**	**0.647**	**0.691**	0.660	0.640
LASSO	0.756	0.852	0.819	1.228	0.887

Using the most accurate predictive algorithm (DNN), we generated price elasticity, which is the predicted percent reduction in SSB sales conditional on increased beverage prices in reference to the predicted SSB sales due to observed SSB price. The elasticity across all stores was -1.74 (95% Confidence Interval [CI]: -0.89, -2.69), implying that the increase of SSB price by 1% results in the expected reduction of SSB purchasing by 1.74%.

Discussion

The superior prediction accuracy demonstrated by B-DTR and DNN over LASSO is likely due to their ability to model non-linear relationships and interactions across predictive features of the 154 brands. Conventional approach to model low-dimentional data, such as Ordinary Least Square and its extension adapted for a high-dimentional varaible space (LASSO) may be a suboptimal approach in predicting the sale of SSB in a competitive retail environment due to its linear constraint Although it is theoretically possible to manually specify appropriate non-linear functional forms guided by model-fit criteria (e.g. Akaike's Information Criterion) in LASSO, this approach is not feasible when the number of competing brands grows large.

The estimated own-price elasticity was slightly higher than the mean estimate from a systematic review of previously reported observatioinal studies in the United States (-1.21, range: -0.71 to -3.87) [7] but was lower than the elasticity (-3.1) estimated by the exsiting natural experiment in Berkley, California [20].

Non-parametric learning algorithms, in particular DNN as demonstrated by the superior prediction accuracy in our study, are free from the linearity assumptions. They are a promising alternative to predict the sales of a large number of brands, thus providing a better quality of public health evidence to guide policy on SSB taxation.

Future work includes in-depth investigation of store-level difference in the estimated effectiveness of taxation, or price elasticity. Identification of store-level features (e.g. promotion and the number of competing items) and neighborhood features driving differential store-level elasticity is a critical public health interest, since the analysis allows the characterization of communities that are less likely to benefit from taxation and consequently in need of community-specific interventions addressing local obstacles of healthy eating. As well, variation of prediction error (MSE) across stores and beverage categories warrants thorough investigation to improve the performance of the models.

A limitation of our study, as in the majority of research estimating price elasticity using observational transaction data, is the lack of the sales data generated by the simultaneous increase of SSB price across all beverages due to taxation, which may induce purchasing patterns different from the observed price fluctuation without the taxation. Thus, our ability to formerly validate the predictive performance of these algorithms under SSB taxation is limited. Future research therefore includes the evaluation of these algorithms in a (rare) setting where the taxation is applied. Nevertheless, our study adds an important contribution to the existing methodological research in estimating the price elasticity of demand using the transaction of a large number of food products.

From a public health perspective, unique aspects of our study include the evaluation of the effectiveness of health policy using a large amount of transactional data, which have become recently available to public health researchers. More importantly, analytical strategies for learning food demand from high-dimensional data were also lacking. Our novel approach to predicting SSB sales using machine learning methods should improve the accuracy of estimating the effectiveness of SSB taxation, while demonstrating the effective use of scanner transaction data for public health research. In addition, our analytical framework offers a scalability in terms of geographic coverage (e.g. expansion to national-level analysis) and food categories that are not modelled in this study but are often proposed as part of SSB including 100% fruit juice and sweetened milk, albeit having much lower market share than soda.

Conclusions

Overall, our study demonstrated the utility of non-parametric machine learning algorithms in predicting unhealthy beverage sales in the presence of complex interactions across a large number of SSB brands and non-linear effect of sales and product attributes. The higher accuracy of the non-parametric learning algorithms, in particular DNN, over parametric algorithm highlights its utility in predicting the potential effect of SSB taxation.

References

[1] M.H. Forouzanfar, L. Alexander, H.R. Anderson, V.F. et al., Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: A systematic analysis for the global burden of disease study 2013, *The Lancet* **386** (2015), 2287–2323. doi:10.1016/S0140-6736(15)00128-2.

[2] A.M. Thow, S. Downs, and S. Jan, A systematic review of the effectiveness of food taxes and subsidies to improve diets: Understanding the recent evidence, *Nutr Rev* **72** (2014), 551–565. doi:10.1111/nure.12123.

[3] M.A.C. Escobar, J.L. Veerman, S.M. Tollman, M.Y. Bertram, and K.J. Hofman, Evidence that a tax on sugar sweetened beverages reduces the obesity rate: a meta-analysis, *BMC Public Health* **13** (2013), 1072.

[4] F.B. Hu, Resolved: there is sufficient scientific evidence that decreasing sugar-sweetened beverage consumption will reduce the prevalence of obesity and obesity-related diseases, *Obes Rev Off J Int Assoc Study Obes* **14** (2013), 606–619. doi:10.1111/obr.12040.

[5] P. Bajari, D. Nekipelov, S.P. Ryan, and M. Yang, Demand estimation with machine learning and model combination, National Bureau of Economic Research, 2015. http://www.nber.org/papers/w20955 [accessed April 22, 2016].

[6] H.J. Van Heerde, P.S. Leeflang, and D.R. Wittink, Semiparametric analysis to estimate the deal effect curve, *J Mark Res* **38** (2001), 197–215.

[7] L.M. Powell, J.F. Chriqui, T. Khan, R. Wada, and F.J. Chaloupka, Assessing the potential effectiveness of food and beverage taxes and subsidies for improving public health: a systematic review of prices, demand and body weight outcomes, *Obes Rev* **14** (2013), 110–128. doi:10.1111/obr.12002.

[8] J.S. Raju, The Effect of Price Promotions on Variability in Product Category Sales, *Mark Sci* **11** (1992) 207–220.

[9] P. Leeflang, J.E. Wieringa, T.H.A. Bijmolt, and K.H. Pauwels, Modeling Markets: Analyzing Marketing Phenomena and Improving Marketing Decision Making, Springer-Verlag, New York, 2015. //www.springer.com/gp/book/9781493920853 (accessed April 26, 2018).

[10] F. Chollet, and others, Keras, 2015.

[11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and others, Scikit-learn: Machine learning in Python, *J Mach Learn Res* **12** (2011), 2825–2830.

[12] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Stat Soc Ser B Methodol* (1996) 267–288.

[13] S. Ma, and R. Fildes, A retail store SKU promotions optimization model for category multi-period profit maximization, *Eur J Oper Res* **260** (2017), 680–692. doi:10.1016/j.ejor.2016.12.032.

[14] L. Breiman, Classification and regression trees, Routledge, 2017.

[15] H. Drucker, Improving Regressors Using boosting techniques, *ICML* **97** (1997), 107–115..

[16] D.P. Kingma and J. Ba, Adam: A Method for Stochastic Optimization, *CoRR* **abs/1412.6980** (2014). http://arxiv.org/abs/1412.6980 (accessed April 1, 2019).

[17] X. Glorot, and Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, *Proc Thirteen Int Conf Artif Intell Stat* (2010),249–256.

[18] Y. Bengio, Practical recommendations for gradient-based training of deep architectures, in: Montavon, G, Orr, G.B, Muller, K-R. (eds.) *Neural Netw. Tricks Trade* (2nd ed.) (pp. 437–478), Springer, Berlin,2012..

[19] X. Glorot, A. Bordes, and Y. Bengio, Deep sparse rectifier neural networks, in: Proc. Fourteenth Int. Conf. Artif. Intell. Stat., 2011: pp. 315–323.

[20] J. Falbe, H.R. Thompson, C.M. Becker, N. Rojas, C.E. McCulloch, and K.A. Madsen, Impact of the Berkeley excise tax on sugar-wweetened beverage consumption, *Am J Public Health* **106** (2016) 1865–1871. doi:10.2105/AJPH.2016.303362.

Address for correspondence

* These authors contributed equally to this work.

Corresponding author name: Hiroshi Mamiya

Address: 1140 Ave Pine, Montreal, Quebec, Canada, H1A 1A3

E-mail: hiroshi.mamiya[at]mail.mcgill.ca

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190222

A Mathematical Morphology-Based Filter for Noise Reduction and Detail Preservation in Low-Dose Dental CT Images

Rômulo Marconato Stringhini[a], Daniel Welfer[a], Marcos Cordeiro d'Ornellas[a], Daniel Fernando Tello Gamarra[b]

[a] *Department of Applied Computing (DCOM), Federal University of Santa Maria, Santa Maria, Rio Grande do Sul, Brazil*
[b] *Department of Electrical Energy Processing (DPEE), Federal University of Santa Maria, Santa Maria, Rio Grande do Sul, Brazil*

Abstract

During the acquisition on a low-dose radiation computed tomography (CT) scan, images are usually marked by heavy noise and undesired artifacts, which dramatically reduce its applicability in the image processing workflow. A noise reduction and detail preservation filter based on mathematical morphology is presented in this paper. The filter is geared to allow control of an opening operator followed by a systematic contrast limited adaptive histogram equalization (CLAHE) in conjunction with a reconstruction by dilation in last stage. A quantitative metric built on peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and mean-squared error (MSE) were applied to check noise reduction, detail preservation, and performance. The results obtained by the proposed filter were compared with those obtained in the literature, showing very good results: compared with the best-tested filter, the filter had a gain of 7.91% on PSNR, 7.57% on SSIM and 37.8% on MSE.

Keywords:

Imaging, Radiography, Dental, Digital, Tomography, X-Ray Computed, Image Processing, Computer-Assisted

Introduction

Computed tomography, commonly known as CT, is a painless, non-invasive diagnostic tool that uses a specialized form of X-ray coupled with computer technology to produce cross-sectional images (scans) of soft tissue, organs, bone and blood vessels. Although scans have been applied to improve diagnostic imaging capabilities, the exposure to sufficiently high levels of radiation increases cancer risk. Radiation doses during CT procedures are higher than conventional radiography and are linked to an increased risk of cancer [1].

Research [1] suggests that low-dose CT scans provide effective radiation risk reduction when compared to traditional levels. However, the radiation dose will generate an image corrupted by noisy and undesired artifacts, interfering the diagnosis to a certain extent. Many algorithms have been proposed in the literature to reduce the noise, usually classified in both spatial and frequency domain. Well-known spatial domain filters include Gaussians, Median, Wiener filters. Fourier transforms, wavelet transforms, and anisotropic diffusion are some filters related to frequency domain techniques. Even though these filters produce a significant noise reduction and enhances image quality, they are not designed for edge preservation and noise reduction of images [2].

Median and Gaussian low-pass filters [3-4] are applied over the entire image by following the same procedure. A sliding neighborhood operation (sliding window) is an operation performed one pixel at a time, determining the value of any given pixel in the output image by applying an algorithm to the values of the neighborhood of the corresponding input pixel. For the median filter, the centered pixel receives the median value of its neighborhood pixels while the Gaussian filter allows only low-pass frequencies through, which is the center of the frequency transformed image. The drawbacks of such filters are that they tend to break up image edges, produce false noise edges in the presence of small signal-to-noise ratios, and cannot suppress noise distributions.

The Wiener filter [5] performs an optimal interaction between reverse filtering and noise smoothing. It generates an estimate of the original image and minimizes the square error overall mean. Tomasi and Manduchi [6] have proposed a non-interactive edge-preserving smoothing method called bilateral filtering. Average similar and nearby pixel values replace the value of the processing pixel. The bilateral filter averages the small and weakly correlated differences between noise-caused intensities.

He and Tang [7] introduced Guided Filtering, using a guidance image, an input image, and an output image. The pixels from the guidance image are processed by a sliding window *MxN* and applied to the input image. The value of the pixels is calculated according to the color and similarity of the space. The output image will be processed once this calculation is done.

Greece [8] suggested a bitonic filter where the signal is deemed bitonic. A bitonic sequence is one that monotonically increases to a peak and monotonically decreases. That is, if it has only one local maximum or minimum value, or no maximum or no minimum, a signal is considered bitonic.

Dabov et al [9] proposed the block matching 3D (BM3D) filter based on sparse representation in the transform domain. The filter gathers fragments of similar 2D images into 3D arrays of data, called groups. The BM3D filter is split into two steps. A block-wise estimate is performed in the first step, called basic estimate, where similar fragments are grouped in a 3D group. After that, a hard-thresholding collaborative is done, bringing the fragments back to their original position for aggregation. The final estimate is called the second step. This step uses the baseline estimate output to perform an improved grouping followed by Wiener collaborative filtering. The fragments are then aggregated to result in the final image of the output.

Operators of mathematical morphology are also used to reduce noise. In addition, morphological operators are able to detect boundaries, edges, and important details. These operators use a certain shape and size structuring element [2]. The two basic operators are dilation and erosion. These operators describe the interactions of an image *f* with a structuring element *B*.

The dilation operator δ of an image *f* by a structuring element *B* in a certain pixel *x* is the maximum value of *f* in the window defined by *B* centered in *x.* This operator is described by (1) below.

1. $\left(\delta^B(f)\right)(x)=\max_{b\in B} f(x+b)$

Contrarily, the erosion ε is the minimum value of *f* in the window defined by *B* centered in *x* and is described below by (2).

2. $\left(\varepsilon^B(f)\right)(x)=\min_{b\in B} f(x+b)$

While the dilation operator increases sets and eventually connects them if the size of the structuring element is greater than the space between them, the erosion operator decreases sets and disconnects them. If combined, these two operators can represent two other useful morphological operators: opening and closing operators.

The opening operator γ consists in dilating an image *f* by a structuring element *B* after eroding *f* with the same structuring element. The γ operator can be defined as:

3. $\gamma^B=\delta^B\left(\varepsilon^B(f)\right)$

On the other hand, the closing operator φ consists in eroding an image *f* by a structuring element *B* after dilating *f* with the same structuring element and can be expressed as:

4. $\varphi^B=\varepsilon^B\left(\delta^B(f)\right)$

The opening operator removes clear areas in the image where the structuring element does not fit [10], breaks narrow joints and eliminates narrow reliefs. Similarly, the closing operator removes dark areas in which the structuring element does not fit [10], widens narrow gaps and eliminates small holes.

This paper presents an image filter to reduce noise and preserve details in low-dose dental CT images. The proposed filter is based on the morphological opening operator, followed by the contrast limited adaptive histogram equalization (CLAHE) algorithm and morphological image reconstruction by dilation. Experimental results demonstrated that the proposed filter reduced the noise and preserved edges, borders and details better than the compared filters. In order to validate our approach, PSNR, SSIM and MSE were used as quantitative metrics.

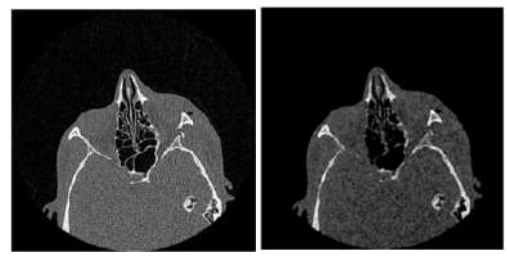
This paper is organized as follows: The Proposed Method section presents in details the proposed filter. The method validation is displayed in the Results section, comparing it to different filters. A discussion on the previous section is presented in the Discussion section. The work is summarized in the last section with the findings and future work.

Proposed Method

The first step is to preprocess the input noisy image, denoted as *F* to get its gray scale level, denominated as *G.* A low-dose dental CT input noisy image is illustrated by Figure 1 (a).

After that, in order to reduce noisy artifacts, the opening operator is applied to *G* with a diamond shaped structuring element with size 2 called *B*. The output image of this step, denominated *K* is illustrated by Figure 1 (b) and described as:

5. $K=\delta^B\left(\varepsilon^B(G)\right)$



(a) *(b)*

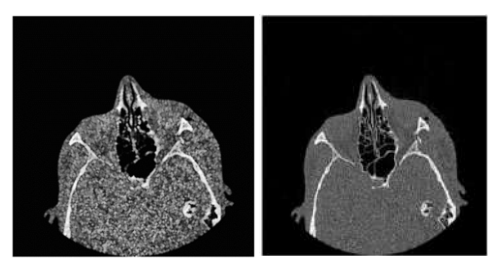
Figure 1 – (a) Low-dose dental CT noisy image. (b) Resulting image K after the opening operator.

Once the opening operator analyzes the visual results in *K*, it is possible to see that some edges, borders and small structures have been degraded. We used the CLAHE enhancement algorithm before restoring these fragments to enhance all the preserved details. CLAHE is an improved version of the adaptive histogram equalization (AHE) [11]. The algorithm basically divides an image into regions of nearly the same sizes. The number of regions is usually 64 for a 512x512 image. Then, with these regions, three different groups are formed. The first group, called corner regions, is made up of four corner regions of the image. The second group, called the Border Region (BR), includes the 24 border regions. Lastly, the third group is called the internal regions (IR) and it has all the other regions. The histogram of each region is calculated and enhanced after this grouping separation. These processed groups are combined back to use bilateral interpolation to perform the output image. The output image *C* is shown by Figure 2 (a).

The filter's last step is to restore the degraded edges, boundaries and small structures. To do this, we used dilation to reconstruct a morphological image. In the marker image C, dilations are done until the contour fits under the mask image G. The final output image, called Q, is the reconstruction of the morphological image between these images by dilation and is described by (6):

6. $Q=R_G^{\delta}(C)=\delta_G^i(C)$

where *i* is such that $\delta_G^i(C)=\delta_G^{(i+1)}(C)$.The result image of the proposed method following these steps is shown by Figure 2 (b).



(a) *(b)*

Figure 2 – (a) Image C. (b) Final output image Q.

Experimental Results and Validation

A dataset of 991 low-dose dental CT images corrupted by natural noise was taken to test the performance of the proposed method. Median filter, Wiener filter, Gaussian filter, guided filter, bilateral filter, BM3D filter, and bitonic filter performance were compared. The parameters were set to their optimal values for each filter tested.

The proposed algorithm was evaluated using quantitative quality metrics of peak signal-to-noise natio (PSNR), mean-square error (MSE) and structural similarity (SSIM). The mean-square error (MSE) between the original image f and the processed image g is computed as

7. $$\text{MSE}=\frac{1}{\text{MxN}}\sum_{i=1}^{\text{MxN}}\left(g(x,y)-f(x,y)\right)^2$$

where M x N specifies the size of the image, $g(x,y)$ is the processed image and $f(x,y)$ is the original image. The PSNR describes the quality of the reconstructed image after applying any technique on it and is described as:

8. $$\text{PSNR}=10\log_{10}\left(\frac{255^2}{\text{MSE(f, g)}}\right)$$

A high PSNR value means a good quality of the processed image. The structural similarity (SSIM) measures the similarity between the enhanced picture and the original picture [12]. A near-1 SSIM value means that the images are identical in structure. The structural similarity of the reconstructed image g between the original image f and g can be formulated as:

9. $$\text{SSIM(f, g)}=\text{l(f, g)}*c(\text{f,g})*s(\text{f,g})$$

where

$$l(\text{f,g})=\frac{2\mu_f\mu_g+\text{C}_1}{\mu_f^2+\mu_g^2+\text{C}_1},c(\text{f,g})=\frac{2\sigma_f\sigma_g+\text{C}_2}{\sigma_f^2+\sigma_g^2+\text{C}_2},s(\text{f,g})=\frac{\sigma_{fg}+\text{C}_3}{\sigma_f\sigma_g+\text{C}_3}$$

with $C_1=(k_1L)^2$, $C_2=(k_2L)^2$, $C_3=\text{C}_2/2$, $k_1=0.01$and $k_2=0.02$ by default and L is a dynamic range of the pixel values. The terms $l(f,g)$ refers to the luminance comparison where μ is the average of f and g; $c(f,g)$ is the contrast function where σ is the variance of f and g and $s(f,g)$ is the structure comparison function. When the SSIM value is close to 1, it means that the processed image had the structures preserved and is similar with the original one. The performance evaluation in terms of PSNR for each tested filter is shown by the graphic on Figure 3.

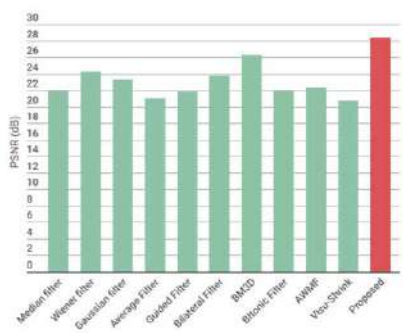


Figure 3 - Average PSNR values for each tested filter.

According to the results, it can be seen that our proposed method outperformed all the tested filters with an average PSNR of 28.35dB, 7.91% more efficient than the BM3D technique.

A low MSE value means better results, that is, the resulting image did not have much loss of information. That being said, it is possible to confirm, according to the graphic shown by Figure 4, that the proposed method had a better average MSE value for all processed images between the tested filters. The average value was 116.41, 37.8% better than the BM3D technique with 187.38.

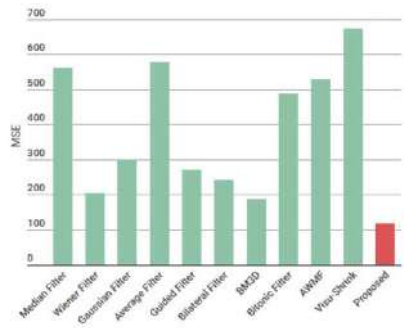


Figure 4 - Average MSE values for each tested filter.

In terms of structural similarity (SSIM), the proposed method also performed better than the tested filters. The average value obtained by the proposed filter on this quality metric was 0.71, 7.57% better than BM3D, with 0.66. The results for the SSIM metric are shown by Figure 5.

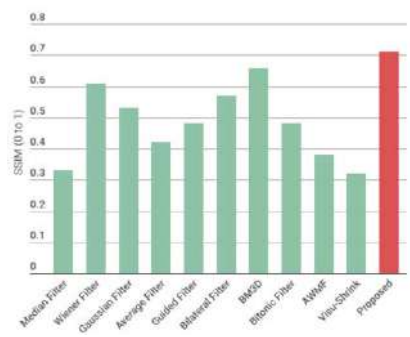
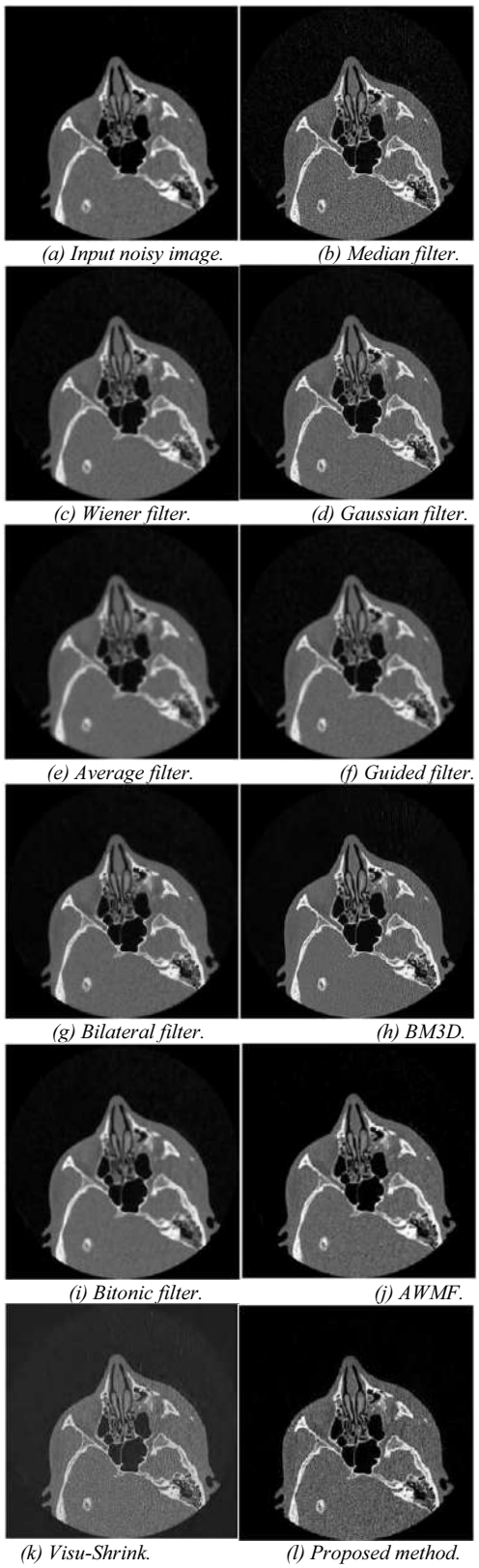


Figure 5 - Average SSIM values for each tested filter.

Table 1 – PSNR, SSIM and MSE average values for all the evaluated filters.

Filters	PSNR	SSIM	MSE
Median Filter	21.92	0.33	561.95
Wiener Filter	24.21	0.61	204.31
Gaussian Filter	23.32	0.53	299.64
Average Filter	21.08	0.42	579.03
Guided Filter	21.83	0.48	271.27
Bilateral Filter	23.72	0.57	242.76
BM3D	26.27	0.66	187.38
Bitonic filter	21.91	0.48	488.18
AWMF	22.33	0.38	529.16
Visu-Shrink	20.74	0.32	673.65
Proposed	**28.35**	**0.71**	**116.41**



(a) Input noisy image. *(b) Median filter.*
(c) Wiener filter. *(d) Gaussian filter.*
(e) Average filter. *(f) Guided filter.*
(g) Bilateral filter. *(h) BM3D.*
(i) Bitonic filter. *(j) AWMF.*
(k) Visu-Shrink. *(l) Proposed method.*

Figure 6 – Visual performance of each tested filter.

Discussions

The performance results of each filter evaluated above are discussed in this section. In addition to quantitative comparisons, subjective visual comparisons are also made between images processed by all filters evaluated. Figure 6 shows the visual results and interpretation of low-dose dental CT images obtained from each tested filter.

Figure 6 (a) is a noisy low-dose dental computed tomography taken as input for each evaluated filter. Figure 6 (b) is the denoised image processed by the median filter, which it can be noticed that the noise was well removed but it blurred some small structures and edges, which can also be proved by its low SSIM value, 0.33, and its high MSE average value of 561.95.

On Figure 6 (c), the image was processed by the Wiener filter. Visually speaking, it is still possible to see some noise and that the image was moderately blurred with a slightly almost transparent line across the image rounded border. On the other hand, the average values for the Wiener filter proves that it has a reasonable filtering performance and it does not degrade small structures and edges.

The image processed by the Gaussian filter, illustrated in Figure 6 (d), was blurred just like the Wiener and the median filter.

The average filter left the image in Figure 6 (e) with the same visual results as in Figure 6 (d), although the average filter had a lower PSNR value compared to the Gaussian filter. The high average MSE value shows a significant information loss for the average filter and the low average SSIM value shows that the filter degrades structures.

For the PSNR and SSIM metrics, the guided filter had nearly the same average values as the average filter, which can be seen in Figure 6 (f) because the noise was still present, and the image was blurred. Figure 6 (g) is the image processed by the bilateral filter. The filter reduced the noise quite well and had a slightly better result in visual terms than the Wiener filter.

The image processed by the BM3D, Figure 6 (h), showed no damage to its structures and details, proving to be the best among the filters evaluated by its average SSIM value of 0.66. The filter had a good performance in terms of noise reduction under the image's circular structure, which encompasses all important details. The filter also had the best average value for PSNR and MSE between the filters tested. Some minor noisy pixels can still be noticed in the dark area of the image, but they do not affect the image's main structures.

The bitonic filter preserved edge and details in Figure 6 (i) better than the median filter, demonstrated by its higher average SSIM value, and had nearly the same visual results and noise reduction performance as the Gaussian filter. The average weighted median filter (AWMF) had nearly the same visual results as the median filter as shown in Table 1 by the mean values. The difference is that smaller structures were slightly less blurred by the AWMF than the median filter. Visu-Shrink was the worst filter among the comparison. We can see the poor noise reduction and detail preservation performance according to Figure 6 (k). Also, in Table 1, it describes its low performance in quantitative values.

The resulting image of the proposed method can be analyzed through Figure 6 (l). Our method exceeded all the above-mentioned filters in both quantitative and visual ways. The average PSNR of 28.35, 7.91 percent higher than the BM3D, shows that our method reduced the noise that affected the quality of the image better. Our method also retained important details and smaller structures 7.57 percent higher than the

BM3D and had almost no information loss, proven by its low average MSE of 116.41.

Conclusion

In this paper we presented an effective noise reduction and detail preservation filter based on mathematical morphology or low-dose dental computed tomography images. The filter works with the opening morphological operator, CLAHE enhancement algorithm and a morphological image reconstruction by dilation. Extensive experiments have been conducted on almost 1000 low-dose dental CT images to demonstrate the superior noise reduction and detail preservation performance of the proposed method. Experimental results indicated that the proposed filter outperformed the evaluated filters by providing higher quantitative values on PSNR, SSIM and MSE metrics with 28.35dB, 0.71 and 116.41, respectively. Compared to the best evalutated filter, the BM3D, our filter had a gain of 7.91% on PSNR, 7.57% on SSIM and 37.8% on MSE. For future work, we intend to develop an additional technique to this proposed filter in order to obtain better noise reduction performance, and, consequently, better visual results.

Acknowledgements

This study was financed in part by the Coordination of Improvement of Higher Level Personnel – Brazil (CAPES) – Finance Code 001.

References

[1] Committee to Assess Health Risks From Exposure to Low Levels of Ionizing Radiation and National Research Council, *Health Risks From Exposure to Low Levels of Ionizing Radiation BEIR VII Phase 2,* Washington, DC: The National Academies Press, 2006.

[2] S. Yang and J. Li, The design of composite adaptive morphological filter and applications to Rician noise reduction in MR images, *International Journal of Imaging Systems and Technology* **25** (2015), 15-23.

[3] J. W. Tukey, Exploratory Data Analysis, Reading, MA, 1971.

[4] F. Bergholm, Edge focusing, *IEEE Transactions on Pattern Analysis Machine Intelligence* (1987), 726-741.

[5] N. Wiener, Extrapolation, Interpolation and Smoothing of Stationary Time Series, *The MIT Press,* 1964.

[6] C. Tomasi and R. Manduchi, Bilateral Filtering for Grayscale and Color Images, *Sixth International Conference on Computer Vision,* 1998, 839-846.

[7] K. He and X. Tang, Guided image filtering, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **6** (2013), 1397-1409.

[8] G. Treece, The bitonic filter: linear filtering in an edge-preserving morphological framework, *IEEE Transactions in Image Processing,* **25** (2016), 5199-5211.

[9] K. Dabov, A. Foi, and K. Egiazarian, Video denoising by sparse 3D transform-domain collaborative filtering, in: *2007 15th European Signal Processing Conference, IEEE*, 2007, pp. 145-149.

[10] V. Naranjo, R. Lloréns, M. Alcañiz, and F. López-Mir, Metal artifact reduction in dental CT images using polar mathematical morphology, *Computer Methods and Programs in Biomedicine* **102** (2011), 64-74.

[11] M. Abdoli, Evaluation strategies for metal artifact reduction approaches in CT: A literature survey, *Frontiers in Biomedical Technologies* **1** (2014).

[12] S. Serikawa and H. Lu, Underwater image dehazing using joint trilateral filter, *Computers & Electrical Engineering* **40** (2014), 41-50.

Address for correspondence

Rômulo Marconato Stringhini. Preferable method for contact is through e-mail. Rua Tuiuti 2272, Santa Maria, Rio Grande do Sul, Brazil. E-mail: rstringhini@inf.ufsm.br. Phone number: +5555984033667.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190223

Knowledge Learning Symbiosis for Developing Risk Prediction Models from Regional EHR Repositories

Jing Mei[a], Eryu Xia[a]

[a] *IBM Research, Beijing, China*

Abstract

Secondary use of regional EHR data suffers several problems, including data selection bias and limited data size caused by data incompleteness. Here, we propose knowledge learning symbiosis (KLS) as a framework to incorporate domain knowledge to address the problems and make better secondary use of EHR data. Under the framework, we introduce three main categories of methods: knowledge injection to input features, objective functions, and output labels, where knowledge-enhanced neural network (KENN) was first introduced to inject knowledge into objective functions. A case study was conducted to build a cardiovascular disease risk prediction model on the type 2 diabetes patient cohort using regional EHR repositories. By incorporating a well-established knowledge risk model as domain knowledge under our KLS framework, we increased risk prediction performance both on small and biased data, where KENN showed the best performance among all methods.

Keywords:

Electronic Health Records, Machine Learning, Risk Assessment

Introduction

Adoption of electronic health record (EHR) systems has increased dramatically in recent years. In 2008, 9.4% of US hospitals adopted a basic EHR system, while the number climbed to 83.8% in 2015 [1]. Along with the increasing use comes the accumulation of EHR data. Efforts of secondary use of EHR data have been made for clinical research purposes such as risk prediction [2,3], patient similarity [4] and treatment recommendation [5].

Compared to traditional risk prediction research, studies based on EHR have several different features [6]. Traditionally, risk prediction models are developed from large cohort studies [7,8], which have predefined inclusion criterion, long-term and regular follow-up, specified metrics to collect, and protocols for adjudicating outcomes. EHR-based risk prediction studies, however, are based on data collected irregularly (only when the patient uses medical services at a certain place), and collect only metrics that are necessary for each specific encounter. These lead to data bias, data inaccuracy and data incompleteness. One resulting feature of EHR-based risk prediction studies is that missing values are frequent in such studies and they are always missing not at random. Traditionally, missing values can be dealt with by either data imputation or removal of incomplete cases. Data imputation hardly work on EHR-based studies due to the extremely high missing rate, and for instance, the missing rate of regular HbA1c tests (every 3 months for diabtic patients) can be higher than 90% from our experience. Another way to deal with missing value is the removal of incomplete cases, which not only decreases the cohort size significantly, but also creates bias on the selected cohort compared to the original cohort and affects the results and conclusions [9]. Another feature of EHR-based risk prediction studies is that since there are no predefined inclusion criterion and data is generated de-facto, there may be severe bias between the population the prediction model is built upon, and the population the model is to be used upon. As an example, patients who visit community hospitals usually have less severe conditions than those who visit emergency rooms though they may have the same diagnosis code. This thus renders a model built on one population hard to be generalized to other populations. Generalization has always been an important concern of risk prediction models. In a review of cardiovascular disease risk prediction models [10], authors pointed out that though risk prediction models may have good discrimination performance in the original cohort, the performance was lower in external validation cohorts, and results in different cohorts were largely different. This shows the problem of overfitting and lack of generalizability, which gets exacerbated when using EHR data. As a summary, EHR-based risk prediction studies suffer from small data size and data bias, and require carefully designed techniques to increase both model performance and model generalization based on the available dataset.

We thus propose knowledge learning symbiosis (KLS) as a framework to incorporate domain knowledge to increase model performance and generalization, and to improve secondary use of EHR data. Here we use domain knowledge to refer to the clinical insights accumulated with time, validated extensively, and accepted widely by the medical community, primarily written into clinical guidelines as logical rules, entity relationships, or mathematical formulas, etc. While domain knowledge is easy to interpret and well-accepted by the community, they suffer from the drawbacks of prone to outdating, too general for precision medicine, and limited availability. By data mining of EHR data, we gather real-world evidence [11,12] using data documented from medical practices among heterogeneous sets of patients in actual clinical settings. Real-world evidence, on the contrary to domain knowledge, can easily and automatedly be updated and are specialized for individual cases, but suffer from the problems of dependency on large amount of high-quality data, are prone to overfitting, and difficult to interpret. KLS thus represents an attempt which uses domain knowledge to complement real-world evidence to leverage the advantage of both.

Previous studies for KLS have shown its benefits in increasing treatment recommendation accuracy [13]. In this study, we explore the capacity of KLS in overcoming the two major

flaws of EHR data for secondary use. We have two experiment scenarios: (1) learning from small data; and (2) learning form biased data, with the main aim of building accurate and generalizable risk prediction models.

Methods

The study was driven by the need to build a localized risk prediction model for 4-year atherosclerotic cardiovascular disease (ASCVD) on a type 2 diabetes patient cohort using 5-year de-identified EHR data of a city in China. ASCVD is the leading cause of morbidity and mortality for individuals with diabetes and is the largest contributor to the direct and indirect costs of diabetes [14].

Cohort Construction

The study cohort was constructed as follows. We included patients with at least one diabetes-related diagnosis and one diabetes-related prescription. Date of the first diabetes-related diagnosis or diabetes-related prescription, whichever is earlier, was defined as the index date. The observation window was defined as one year before the index date, while prediction window was defined as four years after the index date. Patients were excluded if they: (1) had at least one diagnosis of gestational diabetes or type 1 diabetes (to focus on a cohort of type 2 diabetes mellitus [T2DM] patients); (2) were under 18 years old; (3) did not have gender recorded; (4) had observation of ASCVD before index date; (5) had no medical records within the observation window; or (6) had no medical records after the prediction window (right-censored). The EHR repository has medical records of over 3.6 million patients in total. This constructed cohort includes 54,482 patients. Outcome was defined as having at least one ASCVD event observed in the prediction window.

Feature Construction

Features were constructed based on: (1) previously reported risk factors; (2) medical events in the EHR data. Nine risk factors previously reported in literature were identified: gender, age, race, total cholesterol (TC), high-density lipoproteins cholesterol (HDL-C), systolic blood pressure (SBP), antihypertensive treatment, diabetes, and smoking. As our study was focused on a Chinese T2DM cohort, race and diabetes were not included. For TC, HDL-C, SBP, antihypertensive treatment, and smoking, the latest report in the observation window was used if more than one record was available. We also constructed features to reflect medical events in the observation window. By grouping all diagnosis into 22 categories based on the 22 chapters of ICD-10 codes, we constructed another 22 bool features. For such a feature denoting a disease category, patients with at least one diagnosis falling into this category received a value of 'true'.

Extensive missing values were observed, especially in laboratory tests. Only 4,722 (8.7%) of the patients had at least one HDL-C value, and 6,986 (12.8%) patients had at least one TC value, which would render imputation strategies difficult. We thus included 4,143 (7.6%) cases without missing values in seven previously reported risk factors. Of the 4,143 cases, 1,535 (37%) had a positive outcome.

Domain Knowledge

Here we used the Pooled Cohort Equations (PCE) as our knowledge model. The PCE model is a widely validated risk prediction model to estimate the 10-year ASCVD risk and has been recommended in ACC/AHA guidelines [14]. Its difference from our study is that: (1) Cohort. While PCE is built on a multi-race general cohort, we attempt to build our model on a Chinese type 2 diabetes cohort; and (2) Period. While PCE is a 10-year risk prediction model, we attempt to build a 4-year prediction model. Its similarity to our study is that: (1) Outcome. Both are predicting ASCVD risk; and (2) though it is a 10-year prediction model, it is a Cox model, where the predictive value for 10-year risk is positively correlated with its predictive value for 4-year risk. PCE has the 9 risk factors as listed in our feature construction, and a previous study [22] elaborated different experiments on extending PCE risk factors with ICD-10 codes. For each patient in the studied cohort, we calculated its risk score using the PCE model as the knowledge risk score. Actually, we did not use any cutoff of the PCE score, considering the cutoff itself might increase the imprecision in different cohorts.

KLS Methodologies

Methods for KLS can be divided into three main categories based on where knowledge is injected: (1) into input features; (2) into objective functions; and (3) into output labels. When taking KLS as a fusion of knowledge and data, these three categories correspond to early fusion at the feature level, intermediate fusion in the learning process, and late fusion (or output ensemble). Schematic representations of KLS methodologies are shown in Figure 1.

Formally, we denote $P = \{P_1 \ldots, P_n\}$ as the set of patients. Given a patient $P_i (1 \le i \le n)$, we define:

- $\boldsymbol{x}_i = (f_i^1, \ldots, f_i^m)$ as the feature vector;
- y_i as the patient's ASCVD outcome (1 for positive cases and 0 for negative cases);
- z_i as the knowledge risk score calculated using the PCE model (continuous value between 0 and 1).

We implemented methods in each of the three categories as elaborated below, where l denotes the loss function (cross entropy as used here), θ represents the learning parameters.

[KLS#1] Knowledge Injection to Input features

The knowledge risk score z_i is added as an additional input feature, thus making the learning objective function to be: $argmin_\theta \frac{1}{n}\sum_{i=1}^{n} l\big(y_i, \sigma_\theta(\boldsymbol{x}_i, z_i)\big)$. Here we employ two base learning methods: logistic regression (LR, denoted as LR-K) and multi-layer perceptron (denoted as NN-K).

[KLS#2] Knowledge Injection to Objective Functions

We regard the knowledge risk score as another learning target to be leveraged in the learning process. We implemented two methods, described below and illustrated in Figure 1 (bottom).

1. Teacher-student network (TSNN) adapted from [15].

TSNN is composed of two independent networks: the teacher network ρ and the student network σ, which learn from each other and are optimized iteratively. The objective function of teacher network is: $argmin_\theta \frac{1}{n}\sum_{i=1}^{n} l(\sigma_\theta(\boldsymbol{x_i}), \rho_\theta(\boldsymbol{x_i})) + \pi_T l(z_i, \rho_\theta(\boldsymbol{x_i}))$ and the objective function of student network is: $argmin_\theta \frac{1}{n}\sum_{i=1}^{n} l(\rho_\theta(\boldsymbol{x_i}), \sigma_\theta(\boldsymbol{x_i})) + \pi_S l(y_i, \sigma_\theta(\boldsymbol{x_i}))$, where π_T and π_S are weighting parameters. After convergence, both networks can be used for prediction.

2. Knowledge-enhanced neural network (KENN).

KENN is a modular neural network composed of three modules: knowledge representation network, data representation network, and decision network. The objective

function is defined as: $argmin_\theta \frac{1}{n}\sum_{i=1}^{n}(1-\pi)l(y_i,\sigma_\theta(\boldsymbol{x_i})) + \pi l(z_i,\sigma_\theta(\boldsymbol{x_i}))$, where π is the weighting parameter to adjust the relative knowledge importance.

[KLS#3] Knowledge Injection to Output Labels

Let $s_i = \sigma_\theta(\overline{x_i})$ and denote the predictive score of a pure data-driven model (LR used in this study). We propose knowledge injection into output labels with a meta-learning function φ, where the objective function is $argmin_\theta \frac{1}{n}\sum_{i=1}^{n} l(y_i,\varphi(z_i,s_i))$. In Figure 1 (top right), we illustrate decision fusion using LR as the meta model, which is trained on data learning predictive score and the knowledge risk score.

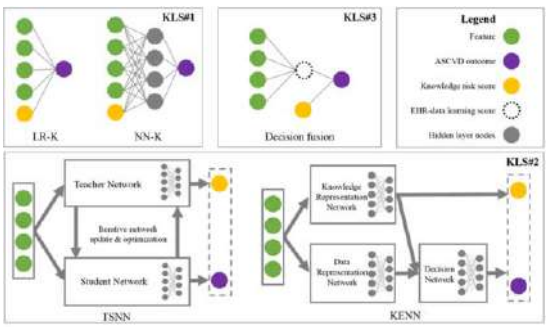


Figure 1. Schematic Representations of KLS Methodologies

Experimental Design

We designed two experiments to explore KLS's capability in dealing with small data and biased data to build accurate and generalizable risk prediction models (Figure 2, EX-1 for small data, EX-2 for biased data). For all studies, we evaluated eight methods: one knowledge-based method (PCE as applying PCE model without modification), two pure data-driven models (LR, and NN implemented as multi-layer perceptron), and five KLS models (LR-K, NN-K, TSNN, KENN, and Decision fusion).

LR, LR-K and Decision fusion were implemented using the Python library 'sklearn'. NN, NN-K, TSNN, and KENN were implemented using the Python library 'keras' with 'theano' backend. We trained each neural network for 50 epochs, with the activation function of softmax and a dropout (of 0.5) layer to avoid overfitting before the last Dense layer. All hidden layers are set up with 64 cells, and the batch size is 32. Finally, we evaluated the prediction performance in terms of AUC (area under curve) based on ROC (receiver operating characteristic), as ROC tries all of the cut-points and plots the sensitivity and specificity.

[EX-1] Learning from Small Data

In the extreme case where we have extremely large amount of data that fully captures information in the knowledge model, KLS would have no advantage over data-driven methods. To the other extreme, where we have extremely small amount of data, knowledge-based methods would be the best as data provides little information. In this experiment, we explored KLS methods with data size between the two extremes. Briefly, for each round, we first randomly selected 30% of instances as the testing data. We then randomly selected a certain proportion of the remaining instances as the training data. All eight models were trained on the training data and evaluated on the testing data by AUC (Area Under the ROC Curve) for risk prediction. We included eight proportion settings: 1%, 2%, 5%, 10%, 20%, 30%, 50% and 70%. For each proportion, 50 rounds of experiments were repeated.

[EX-2] Learning from Biased Data

Bias is unavoidable when using EHR data for building risk prediction models, such as in age, race, and disease severity distribution. Such bias can render the model hard to generalize and apply to clinical practice. In this experiment, we explored the performance of each method in learning from biased data. Since bias is hard to quantify, we intentionally generated data bias to make it easy to evaluate. Briefly, models were trained on a selected sub-cohort (patients with hypertension) and tested on the general cohort. In practice, in each run, we first randomly selected 30% of instances as the testing data. From the remaining 70% of instances, all patients with hypertension were included in the training data. Each of the eight models was trained with the training data and evaluated using the testing data in terms of prediction AUC. Fifty rounds were run.

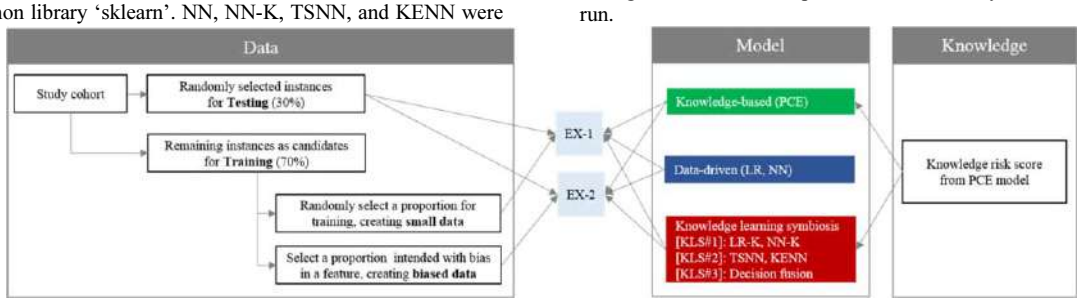


Figure 2. Schematic Representation of Experimental Design

Results

[EX-1]

A knowledge-based model and two purely data-driven models were used for comparison. Five KLS models were included and compared to assess the performance of different knowledge injection strategies. Risk prediction performance was evaluated by calculating average AUC on the testing data from 50 rounds. For TSNN, which was composed of two networks, results from both networks were reported (TSNN-S for results from the student network, and TSNN-T for results from the teacher network). Experiment results for EX-1 are summarized in Figure 3. In all plots, PCE had a stable AUC of around 0.65. Data-driven models (LR and NN) had increasing AUC with the increase of sample size. When the sample size was small (proportion less than 10%), NN had better performance compared to LR, but as the sample size increased, the performance of LR was better than NN. In KLS#1, the line of LR-K overlapped with LR, while the line of NN-K overlapped with NN, suggesting that in our study, knowledge injection to input features cannot improve model performance. Similarly, in KLS#3, the line of Decision fusion

overlapped when the proportion was larger than 1%. These could possibly be explained by the fact that the knowledge risk score was not a strong signal, since it had only an AUC of around 0.65 on the study cohort. Using such a weak signal as feature or label may not be able to increase the performance to a discernible scale.

In KLS#2, however, the performance was different. While TSNN, regardless of TSNN-T or TSNN-N, had unsatisfactory performance, KENN was shown to increase the performance. When the proportion was between 2% and 20%, KENN had the best performance compared to all other methods. When the proportion increased beyond 20%, different methods tended to have similar performance. From these experiments, we concluded that when using small data, KLS had the effect of increasing risk prediction performance. However, this increase was not prevalent for all KLS methods. Among all methods evaluated, only KENN can bring increase to the performance, and this increase was consistent for a number of sample sizes.

[EX-2]

In EX-2 of learning from biased data, we created biased data for training and evaluated the performance on the general data. Here we trained models using patients with hypertension and tested models on the general study cohort. In all runs, the training data (hypertension patients in the cohort after setting aside 30% samples for testing) constituted around 19% of the study cohort, and thus included for comparison the risk prediction performance in EX-1 when the training proportion was 20% as the prediction performance without bias. The results are summarized in Figure 4. All other methods outperformed PCE, suggesting the power of data in training risk prediction models. When trained with unbiased data, LR, NN, KENN, LR-K, decision fusion, and NN-K showed similar performance, suggesting that 20% of unbiased data is sufficient to capture features from knowledge models and that all methods, if properly trained, can have comparable performance. This, however, changed when the models were trained with biased data (bars with dense patterns). Two KLS methods: KENN and LR-K, showed better prediction AUC compared to both data-driven methods, suggesting that KLS can benefit learning from biased data, and increase model generalization. Again, KENN had the best performance among all models, proving its success as a KLS model for learning from biased data.

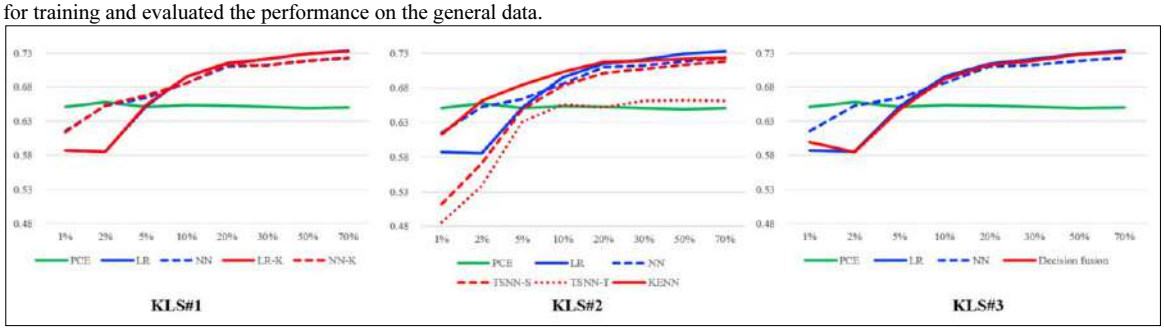


Figure 3. Performance comparison of learning from small data in EX-1 using mean AUC. The y-axis shows the mean AUC. Three plots were generated to show the performance of different KLS categories. In all plots, the knowledge-based model was colored green, data-driven models were colored blue, and KLS models were colored red.

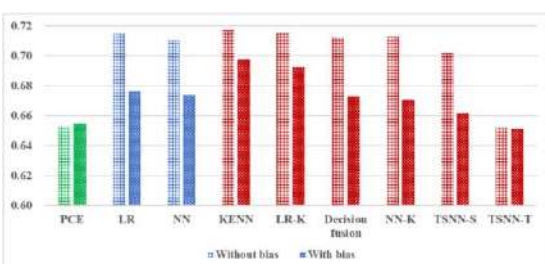


Figure 4. Performance comparison of learning from biased data in EX-2 using mean AUC. The y-axis shows the mean AUC. Knowledge-based model, data-driven models, and KLS models were plotted in green, blue, and red, respectively. Performances while trained with biased data was shown as bars with dense pattern, while performances trained with unbiased data was shown as bars with sparse pattern.

Discussion

Deep Learning in EHR Analytics

Recently, healthcare analytics have been moving its focus from well-designed prospective registry studies or randomized clinical trials toward gathering real-world evidence from EHR data. EHR data is accumulating at an amazing speed, and has some properties compared to traditional healthcare data: (1) it is collected unintentionally and includes information from all aspects of clinical practices; and (2) it is highly heterogeneous. These properties, data complexity and heterogeneity, attract researchers' attention to deep learning techniques [16]. Despite the promising results obtained from deep EHR analysis, there remain several unsolved challenges, and as pointed out in [17], data volume and data quality are the two key issues. After looking into the data issues of our local EHR repositories, we summarized the major problems to overcome in EHR analytics as: (1) learning from small data; and (2) learning from biased data. Therefore, we proposed the KLS framework to address the two problems. Experimental results demonstrated that KLS can address the problems by increasing model generalization and robustness.

Transfer Learning for KLS

Our proposal of knowledge-learning symbiosis is from an application perspective, as a way to leverage knowledge-based and data-driven analytics to improve clinical decision support. Technically, similar topics include transfer learning [18] and domain adaptation [19]. Transfer learning is a technical domain that focuses on using knowledge gained while solving one problem for a different, but related problem. It calls for less sample and less learning time to solve the new, yet related problem and is thus related to our target of learning from small and biased data. Domain adaptation is a research field associated with transfer learning. It is useful when we have a

well-performing model, trained from a source data distribution, and aim to build a model on a different, but related target data distribution. This is similar to our target of learning from biased data. Both transfer learning and domain adaptation are broad research domains with a few common used methodologies. An example practice of transfer learning is in medical imaging [20], which utilizes transfer learning by using medical images to fine-tune the convolutional neural network, trained with general-purposed ImageNet images. Another example is that logic rules are transferred into neural networks, followed by refining the neural networks using a few examples [21]. Many domain adaptation methods include space projection, and not necessarily re-training the model. In our study, we followed the idea of transfer learning and proved its usefulness in healthcare analytics. However, since our knowledge model was relatively simple (a Cox regression model) with a limited number of features, we did not use complex methodologies as designs. In the future, when we have more expressive knowledge like logic rules, we would consider more complex ways of encoding knowledge, and leverage commonly-used, transfer learning methods for KLS purposes.

Small, High-Quality Data or Large Noisy Data?

There are a few questions we can ask in the big data era. Is big data always better than small data? Can the problem of noise be solved when we have sufficient data? Which would be better: small high-quality data or large noisy data? Our results provide us a glimpse of the comparison between small data and big data. In Figure 3, all models have better performance when the training set size increases, but the increase gradually becomes smaller. The results illustrate that for these machine learning methods, the importance of training set size decreases when the size is already large. Another insight from our result comes from a comparison between EX-1 and EX-2. In EX-2, we intentionally created biased data, which can be regarded as a kind of data noise. We can refer to the models' performance trained on 19% of biased data and locate the proportion of unbiased data used in EX-1 that has achieved comparable performance. For KENN, 7% of unbiased data can achieve equivalent performance. For LR, NN, LR-K, NN-K, and Decision fusion, the numbers are 7%, 7%, 10%, 7%, and 7%, respectively. This suggests that compared with using noisy data, using high-quality data can build a model of comparable performance using less data. From different understanding of the questions, two approaches can be taken for future study: learning from small data and learning from big, but noisy data.

Conclusions

In this study, we proposed knowledge learning symbiosis as a framework to address the challenge of learning from small data and biased data. For experiments, we built localized prediction models using regional EHR repositories with the help of a well-established, knowledge risk model. The performance results were promising, which demonstrated that knowledge-learning symbiosis methods can increase risk prediction performance, both on small and biased data, compared to purely data-driven methods, especially for KENN which was first introduced in this study and showed the best performance among all methods.

References

[1] J. Henry, Y. Pylypchuk, T. Searcy, V. Patel, Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2015, ONC Data Brief No. 35, 2016, 1-11.

[2] Y. Cheng, F. Wang, P. Zhang, J. Hu, Risk Prediction with Electronic Health Records: A Deep Learning Approach. SIAM Int Conf Data Min. 2016, 432–40.

[3] A. Rajkomar, E. Oren, et al., Scalable and Accurate Deep Learning for Electronic Health Records. *Nature Partner Journals, Digital Medicine*, 2018.

[4] J. Sun, F. Wang, J. Hu, S. Edabollahi, Supervised patient similarity measure of heterogeneous patient records. *ACM SIGKDD Explor Newsl.* **14(1)** 2012, 16.

[5] E. Xia, J. Mei, G. Xie, X. Li, Z. Li, M. Xu, Learning Doctors' Medicine Prescription Pattern for Chronic Disease Treatment by Mining Electronic Health Records: A Multi-Task Learning Approach, *AMIA* 2017.

[6] B.A. Goldstein, A.M. Navar, M.J. Pencina, J.P.A. Ioannidis, Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review, *J Am Med Inform Assoc.* **24(1)** 2017, 198–208.

[7] P.W. Wilson, R.B. D'Agostino, D. Levy, A.M. Belanger, H. Silbershatz, W.B. Kannel WB, Prediction of coronary heart disease using risk factor categories, Circulation, **97(18)** 1998, 1837–47.

[8] A. Patel, S. MacMahon, J. Chalmers, B. Neal, L. Billot, M. Woodward, et al., Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes, *N Engl J Med.* **358(24)** 2008, 2560–72.

[9] J.G. Ibrahim, H. Chu, M. Chen, Missing Data in Clinical Studies: Issues and Methods. *J Clin Oncol.* **30(26)** 2012, 297–303.

[10] S. van Dieren, J.W.J. Beulens, A.P. Kengne, L.M. Peelen, GEHM. Rutten, M. Woodward, et al., Prediction models for the risk of cardiovascular disease in patients with type 2 diabetes: a systematic review, *Heart* **98(5)** 2012, 360–9.

[11] Real World Evidence: A New Era for Health Care Innovation, 2015.

[12] E. Xia, H. Liu, J. Li, J. Mei, X. Li, E. Xu, X. Li, G. Hu, G. Xie, Gathering Real World Evidence with Cluster Analysis for Clinical Decision Support, *Stud Health Technol Inform.* **245** 2017, 1185–9.

[13] J. Mei, H. Liu, X. Li, G. Xie, Y. Yu, A Decision Fusion Framework for Treatment Recommendation Systems. *Studies in Health Technology and Informatics*, 2015, 300–4.

[14] R.H. Eckel, J.M. Jakicic, J.D. Ard, J.M. De Jesus, N. Houston Miller, V.S. Hubbard, et al., 2013 AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: A report of the American college of cardiology/American heart association task force on practice guidelines. Vol. 63, *Journal of the American College of Cardiology*, 2014, 2960–84.

[15] Z. Hu, X. Ma, Z. Liu, E. Hovy, E. Xing E, Harnessing Deep Neural Networks with Logic Rules. *ACL*, 2016, 1–18.

[16] B. Shickel, P.J. Tighe, A. Bihorac, P. Rashidi, Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis, *IEEE Journal of Biomedical and Health Informatics*, 2017;

[17] R. Miotto, F. Wang, S. Wang, X. Jiang, J.T. Dudley, Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform*, 2017;

[18] L.Y. Pratt, Discriminability-Based Transfer between Neural Networks. *Adv Neural Inf Process Syst.* 1993, 204–11.

[19] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, J.W. Vaughan. A theory of learning from different domains. *Mach Learn.* **79(1–2)** 2010, 151–75.

[20] D.S. Kermany, M. Goldbaum, W. Cai, CCS. Valentim, H. Liang, S.L. Baxter, et al., Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell.* **172(5)** 2018, 1122–1131.e9.

[21] G.G. Towell, J.W. Shavlik, Knowledge-based artificial neural networks. *Artif Intell.* **70(1–2)** 1994, 119–65.

[22] J. Mei, E. Xia, X. Li, G. Xie, Developing Knowledge-enhanced Chronic Disease Risk Prediction Models from Regional EHR Repostitories, Arxiv:1707.09706, 2017.

Address for correspondence

Jing Mei, PhD. IBM Research China

Email: meijing@cn.ibm.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190224

What Is a Chronic Disease? A Contribution Based on the Secondary Use of 161 Million Discharge Records

Emeric Mellot[a], Thibaut Balcaen[b], Matthieu Calafiore[c], Guillaume Bouzillé[d], Jean-Baptiste Beuscart[e], Grégoire Ficheur[e], Emmanuel Chazard[e]

[a] Univ. Lille, CERIM EA2694, General practitioner, F-59000 Lille, France
[b] Univ. Lille, CERIM EA2694, F-59000 Lille, CH Saint Quentin, F-02100 Saint Quentin, France
[c] Univ. Lille, CERIM EA2694, F-59000 Lille, General practitioner, F-59150 Wattrelos, France
[d] Univ Rennes, CHU Rennes, Inserm, LTSI – UMR 1099, F-35000 Rennes, France
[e] Univ. Lille, CHU Lille, CERIM EA2694, F-59000 Lille, France

Abstract

Several definitions of chronic diseases exist. The objective is to reuse a nationwide medical-administrative database (PMSI) to estimate the lifespan of diagnostic codes, hence the chronicity of the corresponding diseases. We analyzed 162 million inpatient stays from 2008 to 2014, and estimate the lifespan of every ICD-10 code for every patient, identified by a unique imprint. We calculated 200 indicators for different time and survival values, and selected the ones that maximized the area under the ROC curve (AUC) drawn by comparison against 4 chronic disease classifications: CCI, ALD, result from the analysis of ICD-10 labels, and a handmade list. The best indicator was the time to reach a survival of 4.5%. It enables to get the following AUC: 78.9% compared with CCI, 90.3% compared with ALD, 75.1% compared with labels analysis, and 91.5% compared with the handmade list. This indicator enables to classify 23,349 ICD-10 codes from "most chronic" to "most acute". The 100 most chronic codes are listed.

Keywords:
Chronic disease, Patient discharge, Big data.

Introduction

Chronic conditions are the world epidemic of the 21st century, accounting for about two-third of all deaths [1–4]. For the WHO, a chronic condition is of long duration and generally slow progression [5]. Other definitions propose a chronic/non-chronic classification according to time thresholds which vary from 3 to 12 months [2,6]. In France, the compulsory health insurance provides with a list of 30 categories of chronic conditions, called "ALD List" [7]. In the USA, the American Agency for Healthcare Research and Quality publishes the Chronic Condition Indicator (CCI) which provides a chronic/acute status for all ICD-10-CM codes [8].

The French nationwide hospital discharge database (PMSI) comprehends all the impatient stays from both profit and non-profit hospitals. This exhaustive nationwide database comprehends about 25 to 30 million inpatients stays per year [9]. For each inpatient stay, the data notably comprehend the ICD-10 principal and secondary diagnostic codes, the anonymized date of the stay, and a unique cryptographic patient identifier. This identifier enables to follow the readmissions of any patient without disclosing his identity, even in different hospitals. The PMSI data are initially collected for billing purposes.

Our objective is to reuse the big data [10,11] from the PMSI to estimate the temporal persistence of the ICD-10 codes of French patients, in order to propose a ranking of the ICD-10 codes according to their chronicity.

Methods

Population of interest

We processed the anonymous data from the French PMSI from 2008 to 2014, which comprehends 162 million stays after exclusion of sequential treatments (e.g. hemodialysis, radiotherapy, etc.). An authorization was issued from the French agency for data protection (CNIL) [9].

Computation step

The stays were sorted by inpatients identifiers thanks to their unique imprints. Patients who wished to remain anonymous (e.g. for an abortion) accounted for 0.4% of all stays and were excluded. The sample then comprehended 46 million patients, 161 million inpatient stays, and 503 million ICD-10 codes.

For each inpatient, we listed all the codes found during the observation period (2008-2014), then we looked when they appeared and disappeared, to calculate their lifespans. We assumed the date of every diagnostic code was the first day of the inpatient stay. Then, if a code was first found in a stay at date *t1* and last found in a stay at date *t2*, its lifespan was *(t2 – t1 + 1)*. In order to simplify computations, we did not implement right-censoring. A code "disappeared" when it was not followed by itself (or any code having the same 3 first digits) within 720 days (this method was determined after several tests, using a 10% training sample). For instance, "J18.1" and "J18.9" both relate to "J18". For a given patient, the presence of "J18.9" will extend the lifespan of "J18.1". From our experience, those codes variations are more often due to inaccurate coding than patient's disease evolution.

This way, for every ICD-10 code and every patient, we could compute a duration. Those durations were plotted through survival curves which, in the absence of censoring, correspond to empirical distribution functions of the durations [12]. From those curves, we could compute the survival at a given time

(we tested 100 times, from 1 to 100 days), or the time to get a certain survival (we tested 100 survival rates, from 50% to 0,5%). For each ICD-10 code, we could then store 200 values.

We used the R programming language with the additional packages data.table, survival, and proc [13].

Validation step

Overview

The external validation was performed against 4 lists of diseases, presented hereafter:

- "**CCI**": list of chronic ICD-10-CM codes provided by the American Agency for Healthcare Research and Quality published in 2017 [4,8], and mapped to French ICD10 codes.
- "**ALD**": list of 30 "long duration diseases" provided by the French compulsory insurance in 2015, and mapped to 122 3-digit ICD-10 codes [7]
- "**ICDlabel**": the result from text analysis of the wordings of all French ICD-10 codes [14,15]
- "**Handmade100**": a list of 100 chronic diseases identified from the ICD-10 from our knowledge.

Preparation of the "CCI" list

The 2017 American CCI provides the chronicity of 64,955 ICD-10-CM codes [4,8] while the 2017 French ICD-10-FR has 40,519 codes. A mapping between both classifications was realized as follows. For 6,003 codes, a direct matching was possible. Then, when a French code did not have any match, the ICD-10-CM codes starting with the same characters were considered. For a given ICD-10-FR code, if all the corresponding codes were chronic or non-chronic in the CCI classification, the French code could be classified as chronic or non-chronic. If the corresponding codes were of mixed type, the French code couldn't be classified. We retrieved 11,034 codes, 10,487 of which were found in our population. We kept the 4,323 chronic codes and selected the 4,323 most frequent non-chronic codes to obtain a total list of 8,626 codes.

Preparation of the "ALD" list

Among the 23,349 codes of our database, 1,733 were matching the 122 3-digits codes from the official 2015 ALD list [7]. We made the simple assertion that all the conditions which were not in the ALD list were non-chronic.

Preparation of the "ICDlabel" list

We searched for the keywords "acute" and "chronic" in the free-text labels of the 2017 French version of ICD-10 [14,15]. We removed all the results which contained terms like "sub-acute", "not specified as acute or chronic" or "antecedent of acute/chronic", etc. We then obtained 283 "acute" codes and 236 "chronic" codes. Among acute codes, we only kept the 236 most frequent ones.

Preparation of the "Handmade100" list

For each ICD-10 code of the database, we calculated how many patients had this code at least once, in order to get the cumulated prevalence of those codes during the period. We sorted the codes by decreasing prevalence, reviewed them, and selected the codes for which we fully agreed with the CCI classification, until we got 100 chronic and 100 acute codes.

Indicators selection

For this step, the statistical individual was the patient. The aim was to discriminate "chronic" codes from "acute" codes (other codes were ignored). For each of the 200 available indicators, we draw a ROC curve [16] against each of the 4 validation lists and calculated the AUC. For each of the 4 validation lists, we then selected the indicator that obtained the highest AUC. With the best indicator, we then selected the threshold giving the closest values of sensibility and specificity (respectively *Se* and *Sp*, which are only provided for descriptive purposes). In some cases, a sensibility analysis was also performed, as described hereafter.

Results

Example of survival curve

For a typical chronic condition, such as end-stage renal disease (Figure 1), the survival curve begins with a moderate drop, then decreases slowly. The drop corresponds to the patients in which either the code is present once, or our algorithm failed to detect any recurrence of the code. The survival at the end of our study is always zero as we did not implement right-censoring.

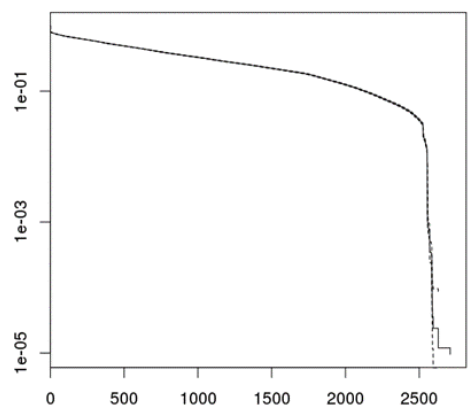


Figure 1. Survival curve of a chronic code, N180 (end-stage renal disease; time in days; Y axis in logarithmic scale)

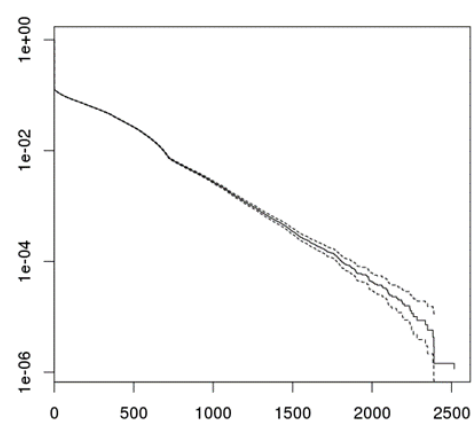


Figure 2. Survival curve of an acute code, E559 (vitamin D deficiency; time in days; Y axis in logarithmic scale)

For a typical non-chronic condition, such as vitamin D deficiency (Figure 2), the initial drop is important, and the survival rate rapidly approaches zero.

It is worth noting that these curves do not show the actual duration of the diseases, but only enable to estimate their chronicity.

Indicator selection

Overview

In this step, for each validation list, and for each of the 200 indicators, we draw a ROC curve and selected the indicator associated with the highest AUC value.

Indicator selection for "CCI"

Compared with the "CCI" list (4,323 chronic codes, and 4,323 acute codes), the best indicator is the time required to reach a survival of 4.5%. It obtains an AUC of 0.789 [0.780; 0.799] (Figure 3). The best point, with a threshold of 182 days, gets a specificity of 72.4% and a sensitivity of 72.7%.

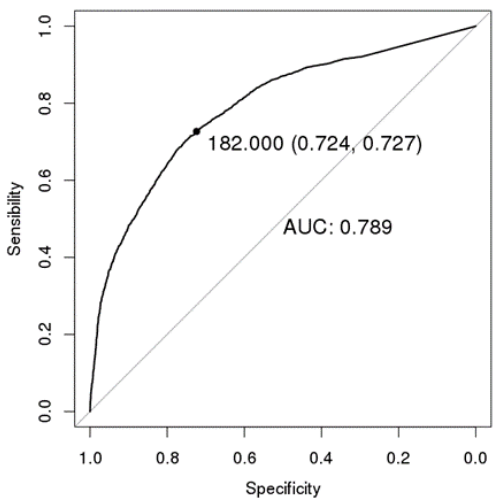


Figure 3. ROC curve of the indicator "time until survival=4.5%" against the "CCI" validation list

A first sensibility analysis, restricting the evaluation to the 6,003 codes (over 8,646) having an exact match between ICD-10-FR and ICD-10-CM, enables to select the same indicator too, with AUC=0.8, Sp=73.2% and Se=73.3% for a threshold of 218 days.

In a second sensibility analysis, we only included the codes that were used in only *n* patients, and reran the analysis. With the threshold n≥10, the AUC rises up to 0.817 (with the same indicator). For a threshold n≥1000, the AUC rises up to 0.83 (with another indicator, the survival rate at 85 days).

The third sensibility analysis consisted of including all the 10,487 codes from the CCI data. We then obtain an AUC of 0.796 with the same indicator, and Sp=73.2% and Se=73.4% for a threshold of 174 days.

Indicator selection for "ALD"

Compared with the "ALD" list (1,733 chronic codes, and 21,616 non-chronic codes), the best indicator is the time required to reach a survival of 7%. It enables to get an AUC of 0.903 [0.897; 0.909] (Figure 4). The best point, with a threshold of 168 days, gives Sp=Se=82.7%.

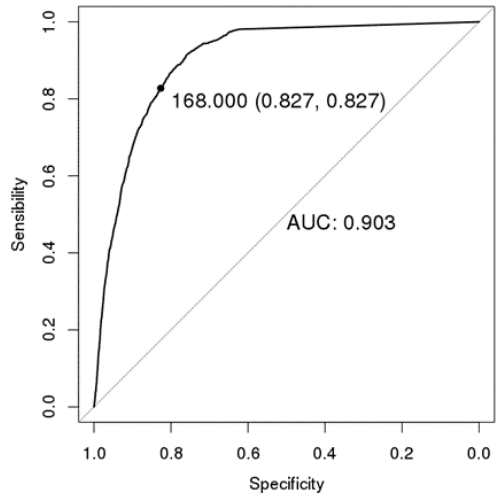


Figure 4. ROC curve of the indicator "time until survival=7%" against the "ALD" validation list

The sensibility analysis consisted of choosing manually the same indicator as previously, namely the time required to reach a survival of 4.5% (instead of 7%). This indicator obtains an AUC of 0.899, and the best threshold (322 days) enables to reach Sp=Se=82.2%.

Indicator selection for "ICDlabel"

Compared with the "ICDlabel" list, the best indicator is the time required to each a survival of 4.5%. It obtains an AUC of 0.751 [0.707; 0.794]. The best point, with a threshold of 210 days, gives Sp=Se=61.9%.

The sensibility analysis consists of comparing this list of codes to the "CCI" list. The comparison between both lists leads to a specificity of 81.1% and a sensibility of 85.5%. By removing the codes from the "ICDlabel" list which are not in the "CCI" list, we obtain an AUC of 0.774 for the same indicator, and the best threshold is 200 days, with Se=70.6% and Sp=70.1%.

Indicator selection for "Handmade100"

Compared with the "Handmade100" list (100 chronic codes, and 100 acute codes), the best indicator is the time required to reach a survival of 4.5%. It obtains an AUC of 0.915 [0.876; 0.954]. The best point, with a threshold of 258 days, enables to reach Sp=Se=84%.

The first sensibility analysis consisted of refactoring this list by exact matching with the CCI list, and not by expert-operated construction, for both the chronic codes list, and the acute codes list. We then obtain an AUC of 0.81 with the same indicator, and the threshold of 317 days enables to reach Sp=Se=74%.

The second sensibility analysis did the same but with the 100 less frequent chronic and acute codes. We then obtain AUC=0.911 for the same indicator, and the threshold of 102 days leads to Sp=85% and Se=86%.

List of "chronic" ICD-10 codes

The four validation steps enabled to identify that, among the 200 indicators, the indicator "time until survival=4.5%" was the most accurate, as it was nearly always selected regarding the AUC criteria. The ICD-10 codes were sorted by decreasing value of this indicator, which enabled to select the 100 "most chronic" conditions, which are presented in Table 1.

Table 1. List of top-100 chronic ICD-10 codes

#	Code	Wording
1	Q861	Fetal hydantoin syndrome
2	E761	Mucopolysaccharidosis, type II
3	D595	Paroxysmal nocturnal hemoglobinuria
4	M834	Aluminum bone disease
5	E763	Mucopolysaccharidosis, unspecified
6	D800	Hereditary hypogammaglobulinemia
7	E760	Mucopolysaccharidosis, type I
8	M0511	Rheumatoid lung disease with rheumatoid arthritis of shoulder
9	E849	Cystic fibrosis, unspecified
10	Z941	Heart transplant status
11	Z942	Lung transplant status
12	E840	Cystic fibrosis with pulmonary manifestations
13	E848	Cystic fibrosis with other manifestations
14	N180	End stage renal disease
15	G243	Spasmodic torticollis
16	Z943	Heart and lungs transplant status
17	G245	Blepharospasm
18	G513	Clonic hemifacial spasm
19	G241	Genetic torsion dystonia
20	G242	Idiopathic nonfamilial dystonia
21	E841	Cystic fibrosis with intestinal manifestations
22	Z948	Other transplanted organ and tissue status
23	D570	Sickle-cell anemia with crisis
24	Z9481	Bone marrow transplant status
25	E710	Maple-syrup-urine disease
26	E711	Other disorders of branched-chain amino-acid metabolism
27	G244	Idiopathic orofacial dystonia
28	Q812	Epidermolysis bullosa dystrophica
29	C83	Non-follicular lymphoma
30	M8336	Adult osteomalacia due to malnutrition, leg
31	N16	Renal tubulo-interstitial disorder in diseases classified elsewhere
32	D86	Sarcoidosis
33	Z944	Liver transplant status
34	E762	Other mucopolysaccharidosis
35	B676	Echinococcus multilocularis infection, other and multiple sites
36	Z940	Kidney transplant status
37	E752	Other sphingolipidosis
38	D562	Delta-beta thalassemia
39	D571	Sickle-cell disease without crisis
40	M0508	Felty's syndrome, other joints
41	G120	Infantile spinal muscular atrophy, type I [Werdnig-Hoffman]
42	M0720	Psoriasic spondylopathy, multiple sites
43	E740	Glycogen storage disease
44	M05	Rheumatoid arthritis with rheumatoid factor
45	E21	Hyperparathyroidism and other disorders of parathyroid gland
46	M349	Systemic sclerosis, unspecified
47	M340	Progressive systemic sclerosis
48	Z491	Renal dialysis
49	N18	Chronic kidney disease (CKD)
50	M0758	Enteropathic arthropathies, other joint
51	W443	Foreign body, eye
52	E661	Drug-induced obesity
53	M0503	Felty's syndrome, wrist
54	D839	Common variable immunodeficiency, unspec.
55	M058	Other rheumatoid arthritis with rheumatoid factor
56	M341	CR(E)ST syndrome
57	E662	Morbid (severe) obesity with alveolar hypoventilation
58	M0801	Unspecified juvenile rheumatoid arthritis, shoulder
59	B24+0	Pre-AIDS
60	D812	Severe combined immunodeficiency with low or normal B-cell numbers
61	M9411	Relapsing polychondritis, scapular region
62	B675	Echinococcus multilocularis infection of liver
63	M0833	Juvenile rheumatoid arthritis (seroneg.), wrist
64	Q819	Epidermolysis bullosa, unspecified
65	E220	Acromegaly and pituitary gigantism
66	Q442	Atresia of bile ducts
67	D817	Major histocompatibility complex class II deficiency
68	M0505	Felty's syndrome, hip
69	K51	Ulcerative colitis
70	G710	Muscular dystrophy
71	G121	Other inherited spinal muscular atrophy
72	L103	Brazilian pemphigus [fogo selvagem]
73	M348	Other forms of systemic sclerosis
74	E723	Disorders of lysine and hydroxylysine metabolism
75	M339	Dermatopolymyositis, unspecified
76	E771	Defects in glycoprotein degradation
77	E241	Nelson's syndrome
78	K501	Crohn's disease of large intestine
79	M0581	Other rheumatoid arthritis with rheumatoid factor of shoulder
80	E84	Cystic fibrosis
81	M0580	Other rheumatoid arthritis with rheumatoid factor of unspecified site
82	K508	Crohn's disease of both small and large intestine
83	D838	Other common variable immunodeficiencies
84	M059	Rheumatoid arthritis with rheumatoid factor, unspecified
85	K509	Crohn's disease, unspecified
86	Z992	Dependence on renal dialysis
87	D572	Sickle-cell/Hb-C disease
88	M072	Psoriasic spondylopathy
89	M0590	Rheumatoid arthritis with rheumatoid factor, unspecified, multiple site
90	B677	Echinococcus multilocularis infection, unspec.
91	G248	Other dystonia
92	Z949	Transplanted organ and tissue status, unspec.
93	Z9488	Other transplanted organ and tissue status
94	D830	Common variable immunodeficiency with predominant abnormalities of B-cell numbers and function
95	E724	Disorders of ornithine metabolism
96	M330	Juvenile dermatopolymyositis
97	E753	Sphingolipidosis, unspecified
98	M6110	Myositis ossificans progressiva, unspec. site
99	N188	Other chronic kidney disease
100	M0753	Enteropathic arthropathies, wrist

Discussion

In this paper, we proposed an automated method to identify chronic conditions from ICD-10 codes. We validated the results against four different classifications, and obtained good and stable results. We finally proposed a list of 100 "most chronic" ICD-10 codes.

French rules specify that a diagnostic code should be used when the disease has had a significant impact on the medical management of the patient, which means not only the disease was still active, but also the disease had to be actively treated during the inpatient stay. According to that definition, our work should not enable to identify chronic conditions, but conditions which induce a chronic need for care.

We provide a ranking of the ICD-10 codes according to their chronicity instead of a binary classification. It is possible to calculate and compare the indicator for different periods or different inpatient populations. Results can be used for many applications without the need for a threshold.

Our work is based on years of nationwide data validated by physicians. This database does not comprehend ICD-9 codes. It does not rely on patient self-report, is not limited in time like a cross-sectional study, is not limited to local morbidity registries and does not depend on the judgment of a group of experts.

As our data relate to inpatient stays, codes for non-severe diagnostics are scarce. There is poor tracking, and stays can be several years apart. The evolution of a disease can be different between outpatients and inpatients, e.g. the reason for the admission can be a surgery that cures the chronic condition. The data may be biased due to per-service pricing, as coders pay more attention to codes that bring money to their hospital.

The CCI classification did not perfectly fit our task. It has additional conditions to define whether a code is a disease or not, thus a code can be chronic without being a chronic condition. Our population consisted of inpatients, while the CCI is determined for all patients. As the CCI is made for the American population, some discrepancies might result from differences between French and American morbidities.

The list of the French ALD was not a very good control list either, as it only relates to few chronic conditions, which are far better coded in the PMSI than other conditions. It gives an estimate of how our indicator would perform if the coding was perfect.

Conclusion

We designed a quantitative indicator of the chronicity of the ICD-10 diagnostic codes by reusing 161 million stays of the French hospital nationwide discharge database from 2008 to 2014. This indicator has positive and negative likelihood ratios of 2.63 and 0.38 respectively when compared to the CCI from the AHRQ, and of 4.78 and 0.21 respectively when compared to the ALD from the French public health insurance. This indicator enabled to rank 23,349 ICD-10 diagnostic codes according to their chronicity.

References

[1] World Health Organization, Noncommunicable diseases now biggest killers, *WHO*. (2008). https://www.who.int/mediacentre/news/releases/2008/pr14/en/ (accessed November 25, 2018).

[2] G. Anderson, and J. Horvath, The growing burden of chronic disease in America, *Public Health Rep*. **119** (2004) 263–270. doi:10.1016/j.phr.2004.04.005.

[3] P. Piot, A. Caldwell, P. Lamptey, M. Nyrirenda, S. Mehra, K. Cahill, and A. Aerts, Addressing the growing burden of non-communicable disease by leveraging lessons from infectious disease management, *J Glob Health*. **6** (2016) 010304. doi:10.7189/jogh.06.010304.

[4] W. Hwang, W. Weller, H. Ireys, and G. Anderson, Out-of-pocket medical spending for care of chronic conditions, *Health Aff (Millwood)*. **20** (2001) 267–278. doi:10.1377/hlthaff.20.6.267.

[5] World Health Organization, Major Noncommunicable diseases and their risk factors, *WHO*. (2018). http://www.who.int/ncds/introduction/en/ (accessed November 25, 2018).

[6] A. Clarke, What is a chronic disease? The effects of a re-definition in HIV and AIDS, *Soc Sci Med*. **39** (1994) 591–597.

[7] Ameli (French compulsory health insurance), Affection de longue durée (ALD), (2018). https://www.ameli.fr/assure/droits-demarches/maladie-accident-hospitalisation/affection-longue-duree-ald (accessed November 25, 2018).

[8] Healthcare Cost and Utilization Project (HCUP), Beta Chronic Condition Indicator (CCI) for ICD-10-CM, (2018). https://www.hcup-us.ahrq.gov/toolssoftware/chronic_icd10/chronic_icd10.jsp (accessed November 25, 2018).

[9] French Technical Agency for Hospital Information (ATIH), Nationwide data access for healthcare facilities, researchers and public institutions, (2018). https://www.atih.sante.fr/acces-aux-donnees-pour-les-etablissements-de-sante-les-chercheurs-et-les-institutionnels (accessed November 25, 2018).

[10] E. Baro, S. Degoul, R. Beuscart, and E. Chazard, Toward a Literature-Driven Definition of Big Data in Healthcare, *Biomed Res Int*. **2015** (2015). doi:10.1155/2015/639021.

[11] E. Chazard, G. Ficheur, A. Caron, A. Lamer, J. Labreuche, M. Cuggia, M. Genin, G. Bouzille, and A. Duhamel, Secondary Use of Healthcare Structured Data: The Challenge of Domain-Knowledge Based Extraction of Features, *Stud Health Technol Inform*. **255** (2018) 15–19.

[12] E.L. Kaplan, and P. Meier, Nonparametric Estimation from Incomplete Observations, *Journal of the American Statistical Association*. **53** (1958) 457–481. doi:10.1080/01621459.1958.10501452.

[13] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2011. http://www.R-project.org/.

[14] French Technical Agency for Hospital Information (ATIH), ICD-10 FR 2017 for PMSI usage, (2017). https://www.atih.sante.fr/cim-10-fr-2017-usage-pmsi (accessed November 25, 2018).

[15] World Health Organization, International Statistical Classification of Diseases and Related Health Problems, 2018. http://www.who.int/classifications/icd/en/ (accessed November 25, 2018).

[16] J. Fan, S. Upadhye, and A. Worster, Understanding receiver operating characteristic (ROC) curves, *CJEM*. **8** (2006) 19–20.

Address for correspondence
Emmanuel Chazard emmanuel.chazard@univ-lille.fr CERIM, medicine school, 1 av Oscar Lambret, 59035 Lille cedex, France.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190225

No Structural Differences Are Revealed by VBM in 'De Novo' Parkinsonian Patients

Verónica Muñoz Ramírez[a,b], Florence Forbes[b], Pierrick Coupé[c], Michel Dojat[a]

[a] *Univ. Grenoble Alpes, Inserm U1216, CHU Grenoble Alpes, Grenoble Institut Neurosciences, 38000 Grenoble, France*
[b] *Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LJK Laboratoire Jean Kuntzmann, 38000 Grenoble, France*
[c] *Laboratoire Bordelais de Recherche en Informatique, UMR 5800, Univ. Bordeaux, Talence, France*

Abstract

The identification of brain morphological alterations in newly diagnosed PD patients (i.e. 'de novo') could potentially serve as a biomarker and accelerate diagnosis. However, presently no consensus exists in the literature possibly due to several factors: small size cohorts, differences in segmentation techniques or bad control of false positive rates. In this study, we use the CAT12 pipeline, to seek for morphological brain differences in gray and white matter of 66 controls and 144 de novo PD patients from the PPMI database. Moreover, we search for subcortical structure differences using the VolBrain pipeline. We found no structural brain differences in this de novo Parkinsonian population, neither in tissues using a whole brain analysis nor in any of nine subcortical structures analyzed separately. We conclude that some results published in the literature may appear as false positives and we contest their reproductibility.

Keywords:

Biomarkers, Magnetic Resonance Imaging, Brain

Introduction

Parkinson's Disease (PD) is a complex neurodegenerative disorder that affects more than 10 million people worldwide [1]. It is mainly characterized by the depletion of dopaminergic neurons situated in the substantia nigra that consequentially disturbs the functions of subcortical nuclei and triggers cortical neuropathological changes causing a plethora of heavily disabling motor and non-motor symptoms [2].

In general, the diagnosis of PD takes place after the manifestation of motor symptoms, which have been found to occur once 50 % of nigrostriatal neurons are lost and dopamine levels are dropped by 80 % [2], [3], creating an urgent need to detect PD biomarkers at the earliest pre-clinical stages of illness possible [4].

The study of morphological brain differences between pathological and healthy groups could potentially identify key regions affected during the PD prodromal phase to better understand PD pathophysiology and its treatment. Magnetic Resonance Imaging (MRI) has positioned itself as a valuable tool for the non-invasive study of the brain's structure. Many automated non operator-dependent techniques have been developed for the analysis of structural MRI data. Voxel-based morphometry (VBM) is the most popular, it allows the detection of subtle morphometric group differences at voxel level [5].

In order to elucidate the nature of morphological differences in de novo PD patients, we investigated 210 subjects from the PPMI (Parkinson Progressive Markers Initiative) through both 1) the well-established Computational Anatomy Toolbox (CAT12) (University of Jena) via the current version of the Statistical Parametric Mapping (SPM12) software and 2) via a new online platform: volBrain [6]. Both pipelines have complementary strengths that are exploited in this study: volBrain performs state of the art quality segmentation of subcortical nuclei [7] and CAT12 facilitates group analysis. Furthermore, we looked for quantitative differences between the tissue classification performed by the two approaches, both including partial volume estimation.

Methods

It is well-known that gathering large cohorts of subjects is a time and resource-consuming task. This is why several efforts have been made by the community to generate databases that benefit for more than one research group. The PPMI (Parkinson Progression Markers Initiative) project is a longitudinal study that gathers data from 35 centers that follows PD patients for five years. The database is openly available for researchers and contains, among other clinical test results, structural MRI images at baseline for 412 patients and 182 healthy subjects.

The scans being heterogeneous, we chose to pool data acquired with the same acquisition parameters, notably magnetic field and scanner manufacturer, to eliminate any additional sources of bias. As a result, our study included 144 de novo PD patients (age: 61.30 ± 9.06; sex: 53 F, 91 M) and 66 healthy controls (age: 60.12 ± 11.39; sex: 23 F, 43 M) from the PPMI database. The structural T1-weighted MRI images extracted were acquired with a 3T Siemens Trio Tim scanner with repetition time (TR) = 2300 ms; echo time (TE) = 2.98 ms; flip angle = 9 degrees; field of view (FOV) = 240 x 256 mm; matrix size : 240 x 256; thickness = 1mm. We note that although T2-weighted images are generally prefered to the delineation of brain structures in neurodegenerative diseases, the available scans on PPMI are provided with low-resolution and thus barely suitable for VBM studies.

Using the CAT12 pipeline

Imaging data were first analyzed using the CAT12 toolbox included in SPM12. All 3D T1-w MRI scans follow a pre-processing protocol including intensity normalization, bias and noise-correction with the Spatially Adaptive Non-Local Means (SANLM) filter introduced in [8] that removes spatially varying noise while maintaining edges. Then the images were spatially normalized using an affine and non-linear (DARTEL and Geodesic Shooting) registration to a reference template brain. Tissue segmentation served to classify the MRI scans into gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) components. CAT12 integrates a classical Markov Random Field and the Adaptive Maximum Posterior (AMAP) technique that reduces the dependency on Tissue Probability Maps (TPM). In addition, the segmentation approach uses a Partial Volume Estimation (PVE), taking the three pure tissue classes

as input and estimating two additional mixed classes: GM-WM and GM-CSF. This allows for more precise segmentation as single voxels are likely to contain more than one tissue type. Next, the total intracranial volumes (TIV) were estimated for each subject and the segmented images were modulated by scaling with the amount of volume changes due to non-linear spatial registration, so that the total amount of grey matter in the modulated image remains the same as it would be in the original image.

The resulting images, appearing in Figure 1, were smoothed with an isotropic Gaussian kernel (8mm), and ready for statistical analysis.

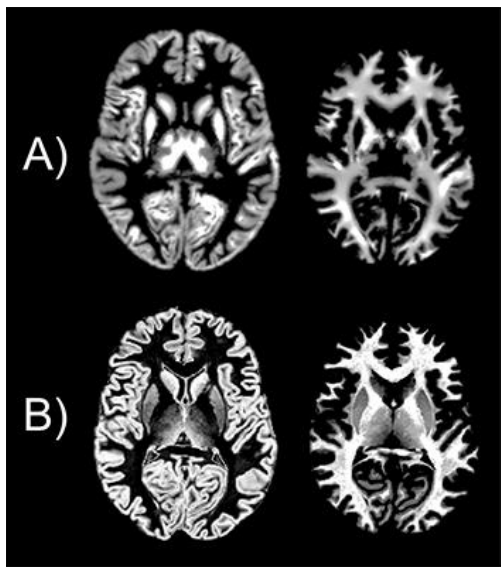


Figure 1– A) CAT12 GM and WM segmentations (modulated) B) volBrain GM and WM segmentations (raw).

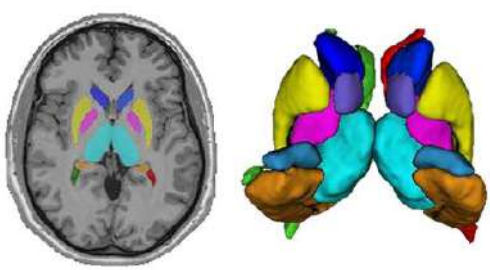


Figure 2– Segmented structures using volBrain.

Using the volBrain pipeline

In parallel, imaging data were analyzed via the volBrain online platform. This system not only provides a state-of-the-art segmentation of the brain tissues (WM, GM, CSF and TIV) (Figure 1), but also segments brain regions like the cerebrum, cerebellum and brainstem; the ventricules; and GM structures such as the putamen, the caudate, the globus pallidus, the thalamus, the hipocampus, the amygdala, and the accumbens [9], as shown in Figure 2.

The multi-template method employed to segment the above mentioned structures considers non-local label fusion schemes using a library built from the manual segmentation of 50 subjects.

The segmentation performed by volBrain provides results in the native and MNI space along with a report containing normality bounds corresponding to the age and sex of the considered subject. These bounds were estimated from the IXI dataset containing 600 normal subjects covering most of adult lifespan.

The pipeline starts by some pre-processing steps. The image is denoised using a SANLM filter, goes through a rough inhomogeneity correction using the N4 method, is registered to the MNI space with a linear affine transformation, goes though a fine SPM based inhomogeneity correction and intensity normalization. Then, segmentation takes place. Tissue classification is obtained by the TMS method that robustly estimates the mean values of the different tissues by excluding partial volume voxels from the estimation jointly with the use of an unbiased robust mean estimator. Partial Volume Coefficients (PVC) are computed from the mean values and completely leave aside tissue probability maps. Next, GM and WM are divided into cerebrum, cerebellum and brainsteam, discriminating between the two hemispheres; and last, subcortical structure segmentation is performed.

VolBrain results analysis by CAT12

Since some subcortical structures of the brain are impacted by PD, we decided to do VBM analysis for the regions provided by volBrain. To do this, we brought volBrain output images to the template space of CAT12 by applying the forward deformation DARTEL field. Once in the same space, the segmented images were used as input for the subsequent statistical analysis. For tissue segmentation analysis (GM & WM), corresponding volBrain's PVC maps were, similarly to CAT12's PVE maps, spatially smoothed with a 8mm kernel.

Statistical analysis

We chose to employ a two-sample T-test to compare the CAT12 modulated tissue maps (GM and WM) of patients versus controls with a general linear model (GLM) where age, sex, and TIV were entered as covariates. The same test was effectuated on volBrain's PVC maps.

A recent study investigating the high rate of false positive present in VBM studies recommends the use of the same group size to detect morphological differences between two groups [10]. Following this recommendation, we repeated our analysis five times to compare the tissue maps of 66 controls versus 66 randomly selected patients using sampling with replacement technique. Their age and sex characteristics are summarized on Table 1.

Table 1– Characteristics of the original study population and the 5 sub-samples of patients equal in size to the control group

	Age	Sex
Controls	60.1 ± 11.4	43 M, 23F
Patients	61.3 ± 9.1	91 M, 53F
PD sample 1	61.0 ± 8.7	40 M, 26F
PD sample 2	60.6 ± 9.7	41 M, 25F
PD sample 3	61.7 ± 9.5	44 M, 22F
PD sample 4	61.9 ± 8.7	38 M, 28F
PD sample 5	59.6 ± 8.6	38 M, 28F

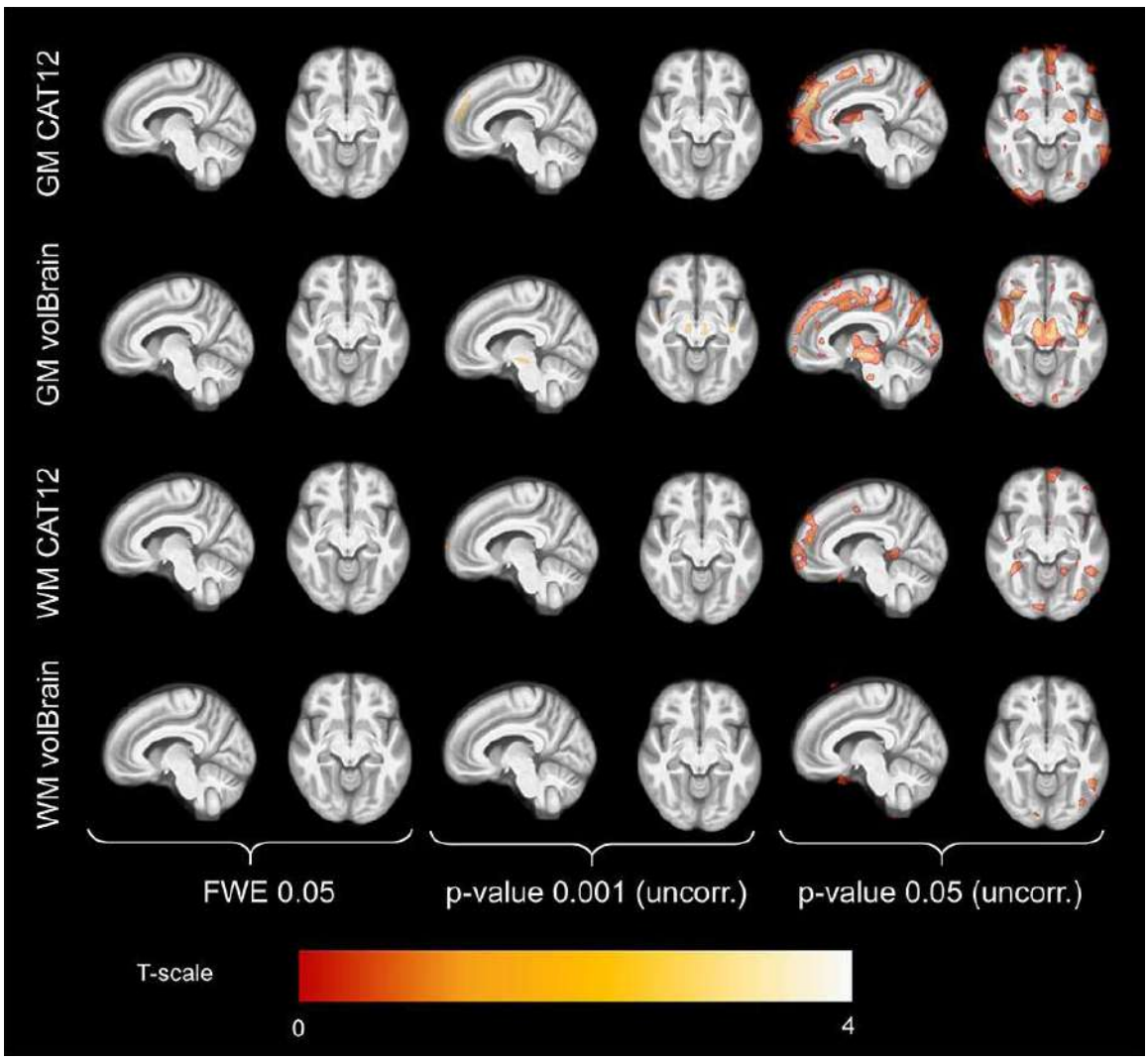


Figure 3– Comparison of PD patients vs controls. Clusters detected for GM an WM diminution in patients using CAT12 and volBrain for different statistical thresholds. The selected slices in the template's MNI space are x=99 and z=64.

Results

When choosing a p-value of 0.05 with Family Wise Error (FWE) correction for multiple comparisons, no voxels survive the difference analysis between PD patient and control groups with tissue map computed with CAT12 or volBrain. In order to replicate some literature results (exploratory study), we decreased the statistical threshold to $p<0.001$ and $p<0.05$ and refrained from any type of correction. Several clusters were then found in PD patients showing volume decrease both in GM and WM as seen in Figure 3.

Also, two-sample T-test comparisons of each independent subcortical structure (computed by volBrain) failed to detect any differences in GM and WM contents $p<0.05$ while FWE corrected. Differences were found in the caudate nucleus, the hipocampus and the putamen for an uncorrected p-value of 0.001.

"Small volume" analysis in SPM12 was used as well to study possible morphometric changes in the substantia nigra, key structure in PD research, using volBrain maps. We observed that differences were only present in gray matter for an uncorrected p-value of 0.001 and did not survived multiple comparison correction.

For all of the 5 new equal size sub-populations (see Table 1), no differences were found in GM or WM for $p< 0.05$ FWE corrected, whilst several significant clusters appeared for an uncorrected p-value of 0.001, esspecially in the frontal cortex for gray matter.

Discussion

Using two recent approaches for accurate segmentation of tissues (CAT12 and volBrain) and subcortical structures (volBrain) we failed to detect robust structural differences in de novo PD patients and healthy controls. We took special care to consider a relatively large cohort of subjects, consider the

effects of an unbalanced number of patients and controls and correct for multiple comparisons. We controlled for multiple comparison using FWE approach, which is known nevertheless to produce some false positives [11]. Following these precautions, no morphological differences were found in PD patients, neither on whole brain GM and WM group analysis or on the analysis of several subcortical structures separately.

In the literature, several studies have reported structural brain differences in PD patients compared to controls. However, these findings tend to be contradictory. In studying a different PD population than in our study, Summerfield and colleagues detected gray matter loss on the right hipocampus, the left anterior cingulate region and the left superior temporal gyrus (p=0.001 uncorrected) in PD patients (n= 13) compared to controls (n=13) [12]. Nyberg and colleagues found an augmentation in the volume of the hipocampus (p=0.03 uncorrected) of PD patients (n=21) and shape deformations of the right accumbens nucleus (p=0.005 uncorrected) compared to controls (n=20) [13]. Radziunas and colleagues observed that PD patients (n=28) with sleep disturbances had bigger ventricules and smaller hipocampus (p-FDR<0.05) than healthy controls (n=28) [14].

Similarly to our study, some VBM studies used the PPMI database and reported structural differences in PD patients. Jia and colleagues noted gray matter losses (p-FWE<0.001) in the fronto-parietal areas and the caudate nucleus, as well as an increase in the size of the limbic and paralimbic areas, the globus pallidus and the putamen of PD patients (n=89) [15] versus controls (n=55) using SPM8.

This lack of consensus on the morphological differences present in de novo PD patients may be due to a variety of factors.

Some studies were carried out on small cohorts, no more than 60 subjects in total, so one may argue that the inconsistencies could be resolved with a larger cohort more representative of the population.

Although, in [10], it was brought to light that sample size does not appear to influence false positive rate, a small sample may incorrectly represent a pathological population, hindering the reproductibility of results.

We note that there is a wide variety of softwares for pre-processing MRI images (i.e. SPM, Freesurfer, FSL), all using different techniques that will inevitably influence the final statistical results as proven by [16] on the study of Multiple Sclerosis. By combining the latest improvements on VBM analysis present in CAT12 (notably denoising and partial volume estimation) with the state of the art segmentations of volBrain [7] we sought to reduce estimation bias considerably.

Finally, correction for multiple comparison is vital to reduce the false positive rate, even if it is not perfect to hope providing robust and reproducible results [11]. Exploratory studies, which use lenient statistical thresholds, could be interesting to indicate some trends in the observed population, that should be confirmed by more robust studies. Then, in our exploratory action (p<0.001 uncorrected) we were able to reproduce some GM results reported in [17]. In the case of [15], the tests were FWE corrected, but the the study was effectuated on VBM8 while, according to [18], the CAT12 toolbox can contribute to more robust detection compared to VBM8.

Regarding the differences we observed between the tissue classification with CAT12 versus volBrain, raw volBrain's PVC maps seem to better distinguish the presence of gray and white matter in the subcortical nuclei. However, as in this study, no morphological robust differences were found between PD patients and controls, a more in depth investigation would be necessary to pertinently test the performances of both methods of partial volume estimation.

In order to further this research, other morphometric methods should be explored, notably Surface Based Morphometry (SBM) and Deformation Based Morphometry (DBM) [17].

Conclusions

In sight of the lack of morphological differences, we suspect that early PD biomarkers may lie on the physiological properties of the Parkinsonian brain and could be investigated through quantitative MRI techniques.

Finally, we reinforce the message that VBM is a delicate technique involving many parameters that should be handled with care to avoid false positive influencing the final results.

Acknowledgements

VMR is supported by a grant from NeuroCoG IDEX UGA in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02).

Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org.

PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including Abbvie, Allergan, Avid, Biogen, BioLegend, Bristol Meyers Squibb, Denali, GE Healthcare, Genentech, GSK, Lilly, Merck, MSD, Pfizer, Piramal, Roche, Sanofi, Servier, Takeda, Teva, UCB.

References

[1] Parkinson's Foundation, "Parkinson's Prevalence Project," 2018. [Online]. Available: http://parkinson.org/Understanding-Parkinsons/Causes-and-Statistics/Statistics.

[2] M. Politis, "Neuroimaging in Parkinson disease: from research setting to clinical practice," *Nat. Publ. Gr.*, vol. 10, pp. 708–722, 2014.

[3] T. R. Barber, J. C. Klein, C. E. Mackay, and M. T. M. Hu, "Neuroimaging in pre-motor Parkinson's disease," *NeuroImage Clin.*, vol. 15, pp. 215–227, Jan. 2017.

[4] W. Poewe, "The natural history of Parkinson's disease," *J. Neurol.*, vol. 253, no. SUPPL. 7, pp. 2–6, 2006.

[5] J. Ashburner and K. J. Friston, "Unified segmentation," *Neuroimage*, vol. 26, no. 3, pp. 839–851, Jul. 2005.

[6] J. V. Manjón and P. Coupé, "volBrain: An Online MRI Brain Volumetry System," *Front. Neuroinform.*, vol. 10, p. 30, Jul. 2016.

[7] J. V. Manjón *et al.*, "Automatic thalamus and hippocampus segmentation from MP2RAGE: comparison of publicly available methods and implications for DTI quantification," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 11, no. 11, pp. 1979–1991, 2016.

[8] J. V. Manjón, P. Coupé, L. Martí-Bonmatí, D. L. Collins, and M. Robles, "Adaptive non-local means denoising of MR images with spatially varying noise

levels," *J. Magn. Reson. Imaging*, vol. 31, no. 1, pp. 192–203, Jan. 2010.

[9] P. Coupé, J. V. Manjón, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins, "Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation," *Neuroimage*, vol. 54, no. 2, pp. 940–954, Jan. 2011.

[10] C. Scarpazza, S. Tognin, S. Frisciata, G. Sartori, and A. Mechelli, "False positive rates in Voxel-based Morphometry studies of the human brain: Should we be worried?," *Neurosci. Biobehav. Rev.*, vol. 52, pp. 49–55, May 2015.

[11] A. Eklund, T. E. Nichols, and H. Knutsson, "Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 28, pp. 7900–5, Jul. 2016.

[12] C. Summerfield *et al.*, "Structural brain changes in parkinson disease with dementia: A voxel-based morphometry study," *Arch. Neurol.*, vol. 62, no. 2, pp. 281–285, Feb. 2005.

[13] E. M. Nyberg *et al.*, "Morphologic changes in the mesolimbic pathway in Parkinson's disease motor subtypes.," *Parkinsonism Relat. Disord.*, vol. 21, no. 5, pp. 536–40, May 2015.

[14] A. Radziunas *et al.*, "Brain MRI morphometric analysis in Parkinson's disease patients with sleep disturbances," *BMC Neurol.*, vol. 18, no. 1, p. 88, Dec. 2018.

[15] X. Jia, P. Liang, Y. Li, L. Shi, D. Wang, and K. Li, "Longitudinal Study of Gray Matter Changes in Parkinson Disease.," *AJNR. Am. J. Neuroradiol.*, vol. 36, no. 12, pp. 2219–26, Dec. 2015.

[16] A. Versteeg *et al.*, "Grey Matter Atrophy in Multiple Sclerosis: Clinical Interpretation Depends on Choice of Analysis Method," *PLoS One*, vol. 11, no. 1, p. e0143942, 2016.

[17] Y. Zeighami *et al.*, "Network structure of brain atrophy in de novo parkinson's disease," *Elife*, vol. 4, no. September 2015, 2015.

[18] F. Farokhian, I. Beheshti, D. Sone, and H. Matsuda, "Comparing CAT12 and VBM8 for Detecting Brain Morphological Abnormalities in Temporal Lobe Epilepsy," *Front. Neurol.*, vol. 8, p. 428, Aug. 2017.

Address for correspondence

Grenoble Institut des Neurosciences
Bâtiment Edmond J. Safra
Chemin Fortuné Ferrini, Site Santé
38706 La Tronche Cedex
Ph. 04 56 52 05 92
e-mail: veronica.munoz-ramirez@univ-grenoble-alpes.fr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190226

Enhancing Prediction Models for One-Year Mortality in Patients with Acute Myocardial Infarction and Post Myocardial Infarction Syndrome

Seyedeh Neelufar Payrovnaziri[a], Laura A. Barrett[a], Daniel Bis[b], Jiang Bian[c], Zhe He[a]

[a] *School of Information, Florida State University, Tallahassee, Florida, USA,*
[b] *Department of Computer Science, Florida State University, Tallahassee, Florida, USA*
[c] *Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, Florida, USA*

Abstract

Predicting the risk of mortality for patients with acute myocardial infarction (AMI) using electronic health records (EHRs) data can help identify risky patients who might need more tailored care. In our previous work, we built computational models to predict one-year mortality of patients admitted to an intensive care unit (ICU) with AMI or post myocardial infarction syndrome. Our prior work only used the structured clinical data from MIMIC-III, a publicly available ICU clinical database. In this study, we enhanced our work by adding the word embedding features from free-text discharge summaries. Using a richer set of features resulted in significant improvement in the performance of our deep learning models. The average accuracy of our deep learning models was 92.89% and the average F-measure was 0.928. We further reported the impact of different combinations of features extracted from structured and/or unstructured data on the performance of the deep learning models.

Keywords:

Electronic Health Records, Machine Learning, Deep Learning

Introduction

In 2016, the top two death causes were heart disease and cancer, accounting for 44.9% of all deaths in that year [1]. Based on a recent report from the American Heart Association, cardiovascular disease and stroke are accounted for tremendous economic and health-related burdens in the United States and worldwide [2]. Acute myocardial infarction (AMI) is an event of myocardial necrosis caused by the unstable ischemic syndrome. It is the leading cause of mortality worldwide [3]. Appropriate management of AMI and timely interventions play a key role in reducing mortality from cardiovascular diseases. Nevertheless, this requires us to understand the past trends and patterns of AMI-related mortality and subsequently to inform the design of future tailored interventions based on the available data and models [4][5].

Prediction models have been increasingly used in hospital settings to assist with risk prediction, prognosis, diagnosis, and treatment planning, ultimately leading to better health outcomes for patients. For example, predictive modeling can inform personalized care based upon health conditions of each individual patient [6]. Specifically, mortality prediction models estimate the probability of death for a group of patients based on their characteristics including the severity of their illness and many other associated risk factors for death [7]. They are important complementary tools to assist in clinical decision-making [8][9]. In current clinical practice, score-based mortality prediction systems, such as the series of the acute physiology and chronic health evaluation (APACHE) scoring system, are widely used to help determine the treatment or medicine should be given to patients admitted into intensive care units (ICUs) [10]. Nevertheless, these scoring systems have significant limitations, e.g., 1) they are often restricted to only few predictors; 2) they have poor generalizability and may be less precise when applied to specific subpopulations other than the original population used for the initial development; and 3) they need to be periodically recalibrated to reflect changes in clinical practice and patient demographics [6]. The wide adoption of electronic health record (EHR) systems in healthcare organizations allows the collection of rich clinical data from a huge number of patients [11]. Large EHR data enables one to 1) build more precise prediction models considering a wider range of patient characteristics; 2) be able to refresh these prediction models more frequently with less engineering efforts; and 3) improve the quality of these prediction models with fewer issues such as the common generalization problem [12].

One contemporary approach to build these prediction models is to use Machine Learning (ML) methods. ML is a field of computer science closely related to artificial intelligence that has drawn significant attention in the last few years. ML methods can be used to extract patterns and to predict different outcome variables (e.g., mortality) based on a training dataset. They have been shown to improve the predictive power in many real-world prediction tasks; and especially on biomedical problems, ML methods can lead to a better prognosis with richer predictors compared to traditional statistical approaches [6][13]. Most ML methods require significant feature engineering efforts, which rely on a deep understanding of the data and their underlying relationships with the outcome variable. Traditional artificial neural networks, even though relaxed the requirements of feature engineering, have a limited number of layers, connections and learning capacity because of the constraints of their computational power. In recent years, with the fast growing evolutions in both computer hardware (e.g., graphics processing unit, GPU) and training algorithm developments (e.g., the backpropagation algorithm that fine-tunes the whole network toward optimized representations [14]), deep learning systems now have the ability to use multi-layer architecture to learn patterns based on raw input data in every layer, in which features are not engineered by human but are learned from data automatically.

In recent years, a number of studies have deployed different deep learning architectures to predict mortality using EHR data. For example, Du et al. used a deep belief network (DBN) to

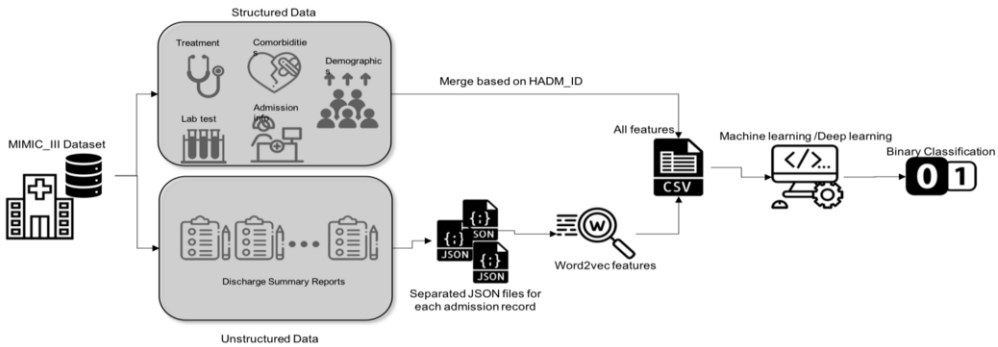


Figure 1– The workflow of the study (Icons made by https://www.flaticon.com)

predict critical care patient's 28-days mortality [15]. Zahid et al. used self-normalizing neural networks to predict 30-day mortality and hospital mortality in ICU patients [16]. Rajkomar et al. proposed a new representation of raw medical data and used deep learning to predict multiple medical events including in-hospital mortality 24 hours after admission [17]. However, these studies either did not consider free-text data in their feature sets or were only concentrated on short-term mortality prediction such as 24-hour mortality, for which any interventions might be too late.

In a previous study [18], we built a number of machine learning models using structured EHR data including admission information, demographics, diagnoses, treatments, laboratory tests, and chart values. The aim of the study was to predict one-year mortality in patients diagnosed with AMI or PMI. We compared the prediction results of these different machine learning models (i.e., shallow learners such as random forest and adaboost); and then compared the prediction performance of the best performing shallow learners to a deep learning model—a fully connected neural network. The results showed that the deep learning model enhanced recall and F-measure metrics (i.e., from a recall of 0.744 to 0.820; and a F-measure of 0.715 to 0.813) while preserving a good prediction accuracy of 82.02%.

In this study, we advance our previous work by adding unstructured data to the previous models. Word embedding features are extracted from free-text discharge summaries and added to the structured features. This study aims to improve the deep learning model performance using the mixture of both structured and unstructured data, which will be called mixed data throughout this paper. Also, the best performing shallow learners from the previous study are compared once again with the deep learning model using the same mixed data. Further, we examine the performance of the deep learning model using the unstructured set of data only, as well as five different combinations of the structured and unstructured data. We aim to determine which set of features contributes the most in enhancing the performance of deep learning models.

Methods

In this section, we first briefly introduce our preparation of the structured data as well as the free-text data. Our goal was to build and compare deep learning and traditional machine learning (i.e., shallow learner) models to predict one-year mortality in ICU patients with AMI and PMI. Many tasks in natural language processing (NLP) have benefited from neural word representations. These representations do not treat words as symbols; but rather can capture the semantics of the words and reflect their semantic similarities. These methods that represent words as dense vectors are referred to as "neural embeddings" or "word embeddings". Word embeddings have been proven to benefit a variety of NLP tasks [19]. Then, we briefly introduce the best performing shallow learners from the previous study, which was used to build new models based on the new mixed dataset. Then, we explain the architecture of the deep learning model. A workflow of this study is depicted in Figure 1.

Dataset Processing

Data Source and Patient Cohort

We used the data from the Medical Information Mart for Intensive Care III (MIMIC-III). MIMIC-III is a freely accessible, de-identified critical care patient database developed by the MIT Lab for Computational Physiology [20]. The latest version of the MIMIC-III dataset includes information about 58,000 admissions to the Beth Israel Deaconess Medical Center in Boston, Massachusetts from 2001 to 2012. Using the International Classification of Diseases, Ninth Revision (ICD-9) codes of 410.0-411.0 (Acute myocardial infarction, Postmyocardial infarction syndrome), we identified 5,436 admissions into our experiment dataset.

Structured Data Processing

The structured data in MIMIC-III include admission information (e.g., total days of admission, initial emergency room diagnoses, etc.), demographics (e.g., age at admission, gender, etc.), treatment information (e.g., cardiac catheterization, cardiac defibrillator, and heart assist anomaly, etc.), comorbidity information (e.g., cancer, endocrinology, etc.) and lab and chart values (e.g., cholesterol ratio, alanine transaminase, etc.). We selected these features based on the features used in similar studies. For details, see [18]. They were further refined and limited by their availabilities in MIMIC-III. To ensure that there was only one admission per instance, duplicates were removed. If duplicates existed because of multiple treatments or comorbidities for the same admission, all of them were counted. Regarding the demographics, since age and death age for people over 89 years old were masked in MIMIC-III by adding 211 to the actual age, we changed them back by subtracting 211 from their value. Some lab values were entered with a '0' and associated with a note of 'see comment'. Thus, 0 values were removed from the lab. Also, the lab or chart values that were biologically invalid were removed. We replaced removed values with the mean value of each feature column. The data was imbalanced with 30% positive and 70%

negative cases. The outliers were removed based on the interquartile range rule [18]. Data values were normalized between 0 and 1.

Unstructured Data Processing

The unstructured data were retrieved based on the corresponding admission IDs in the structured dataset using NOTEEVENTS table of MIMIC-III, from discharge summaries associated with each admission. Discharge summaries are the main method to communicate a patient's plan of care to the next provider [21]. Thus they include rich information about a patient's condition and treatments. Skip-gram model is a neural embedding method to learn an efficient vector representations of words from unstructured text data. These representations of words encode many linguistic regularities and patterns. The Skip-gram model finds the word representations that can predict the surrounding words [22]. The resulting dense vectors are called word embeddings. In this work, we opted to use document embeddings which is the average of word embeddings vectors for the words in the discharge summary of an admission, because the order of the words is not associated with the outcome (i.e., mortality). We used the Word2Vec algorithm in the Gensim library, a free Python library for processing plain text [23]. We used the embeddings pre-trained with Gensim using scientific articles (i.e., PubMed abstracts and PubMed Central full texts [24]).

Case Labeling

The goal of this study was to predict one-year mortality, i.e., whether a patient will die within a year after admission or not. Thus, the admission records of the patients who died within a year were labelled as positive instances, and of those who did not die within a year were labelled as negative instances.

Predictive Modeling

Machine Learning Models

Waikato Environment for Knowledge Analysis (WEKA) is a freely available Java-based software developed at the University of Waikato, New Zealand. Based on the results from our previous study [18], simple logistic and logistic model trees (LMT) classifiers in WEKA produced the best results using the dataset of structured features including admission information, demographics, treatment information, comorbidity information, lab values, and chart values. The simple logistic classifier in WEKA, builds linear logistic regression models. LMT in WEKA builds classification trees with logistic regression functions at the leaves [25].

Deep Learning Model

The deep learning model we used in this work consists of four layers (i.e., the input layer, two hidden layers, the classification layer). Figure 2 shows the deep neural network architecture used in this study. We used the Keras library [26] running on top of the Tensorflow framework [24], as well as a number of other Python packages including SciPy [27], Scikit-learn [28], NumPy [29], and Pandas [30].

The deep neural network we used for this study had 2 hidden layers fully connected with 400 neurons in each layer. The input dimension was 279. We used hyperbolic tangent activation function in hidden layers, and softmax activation function in the classification layer. We used the stochastic gradient descent method for optimization and categorical cross entropy as the loss function. To avoid over-fitting, we used L2 regularization in each hidden layer as well as dropout with a rate of 0.3. Batch size was 100 and epoch size was 60. In each hidden layer we applied batch normalization. All the deep learning architecture settings were chosen based on an extensive examination of different values and their impact on the overal performance. Since the data size was limited, we considered 10-fold cross-validation technique for model validation. We shuffled the data before each run.

Model Evaluation

We ran each algorithm 10 times. In each run, the data was shuffled randomly and 10-fold-cross-validation was employed to evaluate the performance (90% for training and 10% for testing). The performance metrics (i.e., accuracy, precision, recall and F-measure) were averaged after 10 folds.

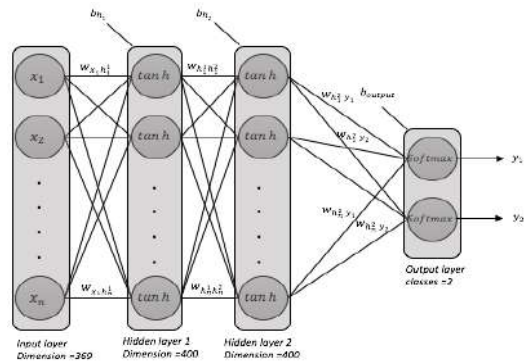


Figure 2 – a fully connected deep neural network architecture: two hidden layers, each with 400 neurons, initial weights=random uniform, initial bias=zeros, learning rate =0.001

The accuracy metric reports the model overall performance on the test set; however, recall and precision metrics of these models are more important in our task. If the actual outcome for a patient is mortality within a year, recall metric evaluates how many times the model was able to predict this correctly that a patient died within a year (true positive) out of all the patients who actually died within a year (true positive + false negative). Precision, on the other hand, evaluates how many times a correct prediction (true positive) happened out of all positive predictions made by the model regardless of their correctness (true positive + false positive). False negative in this study means that a patient who is predicted to live within a year actually died. False positive in this study means a patient who is predicted to have died within a year did not die. F-measure evaluates the balance between these two metrics. Although the receiver operator characteristics (ROC) curve is another popular evaluation metric, its interpretation requires caution when used with imbalanced datasets [31]. Since our dataset is imbalanced, we used precision-recall plot for the visual evaluation of the binary classifier.

Results

In our previous study [18], we first compared the performance of various machine learning models on each set of structured features separately and then compared them to the performance of machine learning models on the combined dataset (admission + treatment + lab and chart values + demographics + comorbidities). We observed that LMT and simple logistic models achieved the best accuracy of 85.12% on the combined dataset. The recall values were low (from 0.499 to 0.660). Only the J48 classifier yielded a precision of 0.993 using the admission dataset alone, while other performance metrics decreased notably comparing to using the combined dataset.

Then we showed that a deep learning model can enhance the performance. Our deep learning model achieved 82.02% accuracy, while boosted recall and F-measure metrics to 0.820 and 0.813, respectively. All features used in the previous study were derived from structured data.

Table 1– Comparing machine learning models to deep learning model based on the mixed dataset

Model	Accuracy	Precision	Recall	F-measure
LMT	85.78%	0.856	0.621	0.724
Simple Logistic	85.71%	0.863	0.623	0.723
Deep Learning	**92.89%**	**0.931**	**0.929**	**0.928**

In this work, we first compared the performance of machine learning and deep learning models on the mixed dataset (i.e., features from both structured and unstructured data). Then, we created different combinations of structured data with unstructured data to examine which set of features has more predictive power for our classification task. Table 1 shows the performance of the two top performing traditional machine learning models (as obtained from our previous study) and a deep learning model on the mixed dataset. The deep learning model outperformed the best shallow learners considerably.

In Table 1, we can see that the precision values of shallow learners are higher than their recall values, which means they are exact but not complete. A low recall value indicates a large number of false negatives (i.e., incorrectly classified as not dying within a year), which is suboptimal in this classification task. The dimension of data in our previous study was 79 considering only features from structured data. Adding features derived from unstructured data increased the total number of features and increased the input data dimension up to 279. Table 2 illustrates the comparison between the previous work and current study. We can see from the results that shallow learners did not benefit from more features (and higher data dimensionality). Accuracy slightly improved, while precision slightly dropped. Recall improved less than 0.03. Unlike the shallow learners, our deep learning model showed considerable improvements with more than 10% increase in accuracy and ~10% improvement in both precision and recall. Further, we were interested in comparing the performance of deep learning models using only free-text features vs. using different combinations of structured and free-text features. Results are summarized in Table 3.

Table 2– Comparing machine learning and deep learning models based on structured dataset vs. mixed dataset (structured + unstructured),-p means previous study, -c means current study

Model	Accuracy	Precision	Recall	F-measure
LMT-p	85.12%	0.867	0.594	0.705
LMT-c	85.78%	0.865	0.621	0.724
Simple Logistic-p	85.12%	0.867	0.549	0.705
Simple Logistic-c	85.71%	0.863	0.623	0.723
Deep Learning-p	82.02%	0.831	0.820	0.813
Deep Leanirng-c	92.89%	0.931	0.929	0.928

From the results we observed, demographic and admission information are two key groups of structured features in enhancing the deep learning model. Demographic information in this dataset includes age at admission, gender, religion, ethnicity and marital status. Admission information includes total days of admission, discharge location and initial ER diagnosis as AMI or rule out AMI.

Table 3– Comparing using unstructured data only vs. different combinations with structured data in the deep learning model

Data	Accuracy	Precision	Recall	F-measure
Free text	81.83%	0.836	0.818	0.816
Free text + lab results	83.61%	0.853	0.836	0.833
Free text + treatment	84.15%	0.850	0.841	0.840
Free text + comorbidity	84.69%	0.856	0.846	0.842
Free text + demographics	87.37%	0.881	0.874	0.872
Free text + admissions	88.25%	0.885	0.882	0.881

The combination of admission information with free-text features produced an accuracy of 88.25% in the deep learning model; while, the accuracy of the deep learning model based on the combination of demographics data with free-text features was 87.37%. We compared these two models to the accuracy of another deep learning model based on the complete mixed dataset, which produced an accuracy of 92.89%. Figure 3 illustrates the precision-recall curve resulted after 10 rounds of deep learning algorithm run. Table 4 shows a comparison of other recent works in mortality prediction using deep learning methods on EHR data.

Table 4– Comparing recent work in mortality prediction using deep learning methods on EHR data

Paper	Mortality Prediction Task	AUC	ACC
Payrovnaziri et al. *(this paper)*	1-year	0.916	92.89%
Du et al.[15]	28-days	Not reported	86%
Zahid et al.[16]	30-days/hospital	0.8445/0.86	Not reported
Rajkomar et al.[17]	24 h after admission	0.92-0.94	Not reported

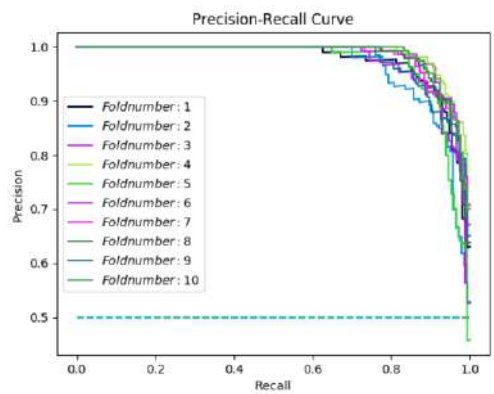


Figure 3– Precision-Recall Curve, after 10 runs average precision = 0.931, average recall = 0.929

Discussion and Conclusions

In this work, we enhanced our previous deep learning model by combining unstructured and structured data to predict one-year

mortality in ICU patients with AMI and PMI. For unstructured data, we extracted word embedding features from discharge summaries of each patient admission. While these word embedding features had no impact on the shallow learners, the performance of our deep learning model increased and achieved an accuracy of 92.89%, precision of 0.931, recall of 0.929 and F-measure of 0.928.

Our findings suggest that a richer data dimension through adding features from unstructured data will enhance deep learning model performance. We also confirmed our previous findings that initial emergency room diagnosis, gender, age, and ethnicity are important factors for the prediction of one-year mortality. One limitation worth noting is that using ICD-9 CM codes for cohort identification may introduce some noise. But this noise should not impact the findings of this study. In future work, we are interested in: 1) designing deep neural network ensembles that have the potential to further improve the model performance; 2) exploring the unstructured sequential data through other state-of-the-art models such as recurrent neural networks and long short-term memory (LSTM) techniques; and 3) the potential to enrich the textual features by extracting Unified Medical Language System (UMLS) concepts from the free-text data.

References

[1] M. Heron, Deaths: Leading Causes for 2016, (n.d.) 77.

[2] Heart Disease and Stroke Statistics—2018 Update: A Report From the American Heart Association, (n.d.) 426.

[3] G.W. Reed, J.E. Rossi, and C.P. Cannon, Acute myocardial infarction, *The Lancet*. **389** (2017) 197–210.

[4] R.L. McNamara, K.F. Kennedy, et al., Predicting In-Hospital Mortality in Patients With Acute Myocardial Infarction, *Journal of the American College of Cardiology*. **68** (2016) 626–635.

[5] E.S. Ford, U.A. Ajani, et al., Explaining the Decrease in U.S. Deaths from Coronary Disease, 1980–2000, *New England Journal of Medicine*. **356** (2007) 2388–2398.

[6] K. Ng, A. Ghoting, et al., PARAMO: A PARAllel predictive MOdeling platform for healthcare analytic research using electronic health records, *Journal of Biomedical Informatics*. **48** (2014) 160–170.

[7] M. Poukkanen, S.T. Vaara, et al., Predicting one-year mortality of critically ill patients with early acute kidney injury: data from the prospective multicenter FINNAKI study, *Critical Care*. **19** (2015) 125.

[8] J.A. Golden, Deep Learning Algorithms for Detection of Lymph Node Metastases From Breast Cancer: Helping Artificial Intelligence Be Seen, *JAMA*. **318** (2017) 2184–2186.

[9] D.S.W. Ting, C.Y.-L. Cheung, et al., Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes, *JAMA*. **318** (2017) 2211–2223.

[10] D.C. Bouch, and J.P. Thompson, Severity scoring systems in the critically ill, *Contin Educ Anaesth Crit Care Pain*. **8** (2008) 181–185.

[11] C. Xiao, E. Choi, and J. Sun, Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review, *J Am Med Inform Assoc*. **25** (2018) 1419–1428.

[12] K. Jung, and N.H. Shah, Implications of non-stationarity on predictive modeling using EHRs, *Journal of Biomedical Informatics*. **58** (2015) 168–174.

[13] M. Motwani, D. Dey, D.S. Berman, et al., Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis, *Eur Heart J*. **38** (2017) 500–507.

[14] R. Miotto, F. Wang, et al., Deep learning for healthcare: review, opportunities and challenges, *Brief Bioinform*. 19(6):1236-1246.

[15] H. Du, M.M. Ghassemi, and M. Feng, The effects of deep network topology on mortality prediction, EMBC 2016: pp. 2602–2605.

[16] M.A.H. Zahid, and J. Lee, Mortality prediction with self normalizing neural networks in intensive care unit patients, 2018 IEEE EMBS BHI 2018, 2018: pp. 226–229.

[17] A. Rajkomar, E. Oren, et al., Scalable and accurate deep learning with electronic health records, *Npj Digital Medicine*. **1** (2018) 18.

[18] L.A. Barrett, S.N. Payrovnaziri, et al., Building Computational Models to Predict One-Year Mortality in ICU Patients with Acute Myocardial Infarction and Post Myocardial Infarction Syndrome, in: AMIA Summits on Translational Science Proceedings, 2019.

[19] O. Levy, and Y. Goldberg, Neural Word Embedding as Implicit Matrix Factorization, in: Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, and K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 27, 2014: pp. 2177–2185.

[20] A.E.W. Johnson, T.J. Pollard, et al., MIMIC-III, a freely accessible critical care database, *Scientific Data*. **3** (2016) 160035.

[21] S.M. Dean, A. Gilmore-Bykovskyi, et al., Design and Hospitalwide Implementation of a Standardized Discharge Summary in an Electronic Health Record, *Jt Comm J Qual Patient Saf*. **42** (2016) 555-AP11.

[22] T. Mikolov, I. Sutskever, et al., Distributed Representations of Words and Phrases and their Compositionality, in: C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems, 2013: pp. 3111–3119.

[23] gensim: topic modelling for humans, (n.d.). https://radimrehurek.com/gensim/models/word2vec.html (accessed November 2, 2018).

[24] M. Abadi, A. Agarwal, et al., TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, (n.d.) 19.

[25] M. Hall, E. Frank, et al., The WEKA data mining software: an update, *ACM SIGKDD Explorations Newsletter*. **11** (2009) 10.

[26] Deep Learning for humans. Contribute to keras-team/keras development by creating an account on GitHub, Keras, 2018. https://github.com/keras-team/keras (accessed November 6, 2018).

[27] SciPy.org — SciPy.org, (n.d.). https://www.scipy.org/ (accessed November 6, 2018).

[28] F. Pedregosa, G. Varoquaux, et al. Scikit-learn: Machine Learning in Python, *MACHINE LEARNING IN PYTHON*. (n.d.) 6.

[29] NumPy — NumPy, (n.d.). http://www.numpy.org/ (accessed November 24, 2018).

[30] Python Data Analysis Library — pandas: Python Data Analysis Library, (n.d.). https://pandas.pydata.org/ (accessed November 24, 2018).

[31] T. Saito, and M. Rehmsmeier, The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets, *PLoS One*. **10** (2015).

Address for correspondence

Zhe He, PhD. School of Information, Florida State University, 142 Collegiate Loop, Tallahassee, Florida, USA. Email: zhe@fsu.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190227

Identifying Phytochemicals from Biomedical Literature Utilizing Semantic Knowledge Sources

Indra Neil Sarkar, PhD, MLIS[a], Wayne Law, PhD[b], Michael J. Balick, PhD[b]

[a]*Center for Biomedical Informatics, Brown University, Providence, RI, USA*
[b]*Institute for Economic Botany, New York Botanical Garden, Bronx, NY, USA*

Abstract

Chemicals derived from plants (phytochemicals) are major concepts of interest in the study of medicinal plants. To date, efforts to catalogue and organize phytochemical knowledge have resorted to manual approaches. This study explored the potential to leverage publicly accessible semantic knowledge sources for identifying possible phytochemicals. Within the context of this feasibility study, putative phytochemicals were identified for more than 4,000 plants from the Medical Subject Headings Supplementary Concept Records and the Semantic MEDLINE Database. An examination of phytochemicals identified for five selected plant species using the method developed here reveals that there is a disparity in electronically catalogued phytochemical knowledge compared to information from Dr. Duke's Phytochemical and Ethnobotanical Databases maintained by the United States Department of Agriculture. The results therefore suggest that semantic knowledge sources for biomedicine can be utilized as a source for identifying potential phytochemicals and thus contribute to the overall curation of plant phytochemical knowledge.

Keywords:
Plants, Medicinal; Knowledge Bases; Phytochemicals

Introduction

A major facet of ethnobotany, the study of human uses of plant species, is the identification of chemicals that may have an active role in potential medicinal effects [1]. Such chemicals, referred to as "phytochemicals," are identified through a range of extraction and analysis techniques [2]. Reports of phytochemicals associated with a given plant species are then catalogued in monographs or articles that provide description of their actions and constituency. A major foundational step in evaluating the potential medicinal utility of a given plant species therefore requires a listing of associated phytochemicals. The process for identifying and recording phytochemical information is mostly manual, labor intensive, and costly.

A limited number of electronic databases exist, including Natural Products Alert (NAPRALERT) [3] and Dr. Duke's Phytochemical and Ethnobotanical Databases (Dr. Duke's) [4], and are artifacts of manually curated resources (e.g., more than 200,000 articles for NAPRALERT and a limited number of monographs for Dr. Duke's). Maintenance and updating information within such databases can be difficult due to challenges in available resources that are not able to keep up with the growing volume of knowledge.

Within biomedicine, there have been significant advances in developing computational approaches for identifying relevant entities from electronically accessible literature resources. Such approaches commonly utilize publicly accessible biomedical knowledge sources, which are enriched with semantic information that facilitates the inference of putative relationships. Those resources of note include the Medical Subject Headings (MeSH) Supplemental Concept Records (SCR) [5], the Unified Medical Language System [6] Metathesaurus (UMLS Meta), the UMLS Semantic Network (UMLS SN), and the Semantic MEDLINE Database (SemMedDB) [7].

MeSH SCR provides an index of chemicals, drugs, and other concepts of interest to MeSH descriptors (which are, in turn, used for cataloguing biomedical artifacts such as publications in MEDLINE). UMLS Meta is a collection of more than two million biomedical concepts collected from more than 200 classifications, codings, thesauri, and controlled vocabularies organized based on synonymy and discernable relationships between concepts. UMLS SN is a set of broad subject categories that organize concepts from UMLS Meta and relationships among them [8]. Finally, SemMedDB is a database of more than 80 million semantic predications (subject-predicate-object triples), which have been extracted by the SemRep [9] natural language processing tool and underpin the Semantic MEDLINE system [10].

Biomedical knowledge resources are thus designed with at least two purposes, to:

1. Facilitate information retrieval tasks; and
2. Support identification of putative relationships between biomedical concepts.

With regards to the latter, a number of studies have demonstrated how the aforementioned resources can be leveraged for the identification of disease risk factors [11], clinical adverse events [12], disease relationships based on genetic knowledge [13], gene-disease relationships [14], as well as many others. To date, there have been no studies exploring the potential to leverage biomedical knowledge sources for the identification of phytochemicals.

The purpose of this feasibility study was to develop an approach to identify phytochemicals from MeSH SCR and SemMedDB using biomedical concepts indexed in UMLS Meta. In addition to identifying putative phytochemicals, a detailed manual comparison of predicted phytochemicals was done for five selected plant species.

Methods

Recent versions of MeSH, UMLS Meta, UMLS SN, and SemMedDB were accessed from a local MySQL database. The list of plant species that were analyzed for this study originated from NCBI Taxonomy. Processing and analysis of data were done through programs written in Julia [15]. A graphical overview of the process for identifying chemicals associated with plants is shown in Figure 1.

Identification of Chemicals from MeSH SCR

MeSH was queried for each plant species name using an exact match of all plant names from NCBI Taxonomy, resulting in a set of entry terms. For each entry term, the associated MeSH descriptor was identified. When available, additional relevant MeSH descriptors and associated entry terms were identified through entries in the "See Also" (FX) field. Only FX field entries of UMLS semantic type "Organic Chemical" (T109; 'orch') or "Pharmacological Substance" (T121; 'phsu') were included. For the final set of MeSH descriptors, chemical names and corresponding UMLS Meta concept unique identifiers (CUIs) were retrieved by querying the MeSH SCR. The overall process used to identify chemicals from MeSH SCR is graphically depicted in Figure 1 as A1-A7 (green arrows; top half of figure).

Identification of Chemicals from SemMedDB

For each plant species name from NCBI Taxonomy, three queries were made of strings indexed in UMLS Meta to identify a query set of UMLS Meta CUIs:

1. The plant species name itself;
2. The plant species name plus the word "extract"; and
3. The set of MeSH entry terms determined as an artifact of the previous identification of chemicals from the MeSH SCR chemical identification process.

All UMLS Meta CUIs for the resulting query set were required to be of semantic type "Plant" (T002; 'plnt'). SemMedDB was then queried with the query CUIs for the given plant for predications that included one of the following predicate types: "ISA", "LOCATION_OF", or "CONVERTS_TO."

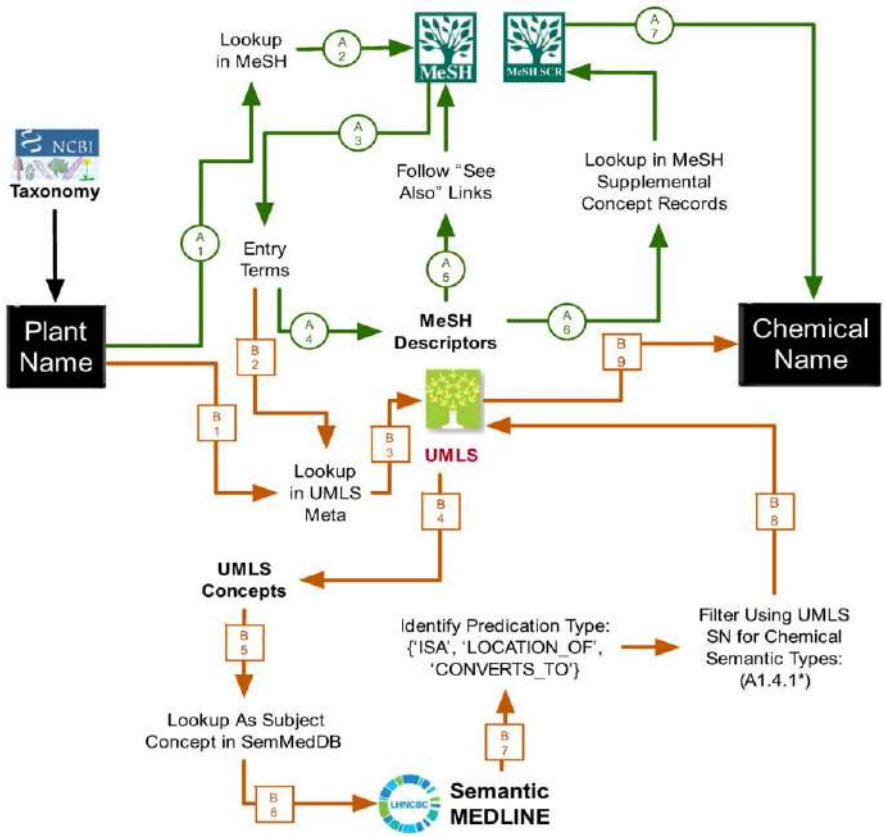


Figure 1. Overview of Approach for Identifying Chemicals Associated with Plants. Two types of knowledge sources were used to identify chemicals associated with plant species as listed in NCBI Taxonomy: (A) MeSH Supplemental Concept Records (SCR); and (B) Semantic MEDLINE. For MeSH SCR, first the full set of MeSH descriptors and entry terms were determined (A1-A5) and then the MeSH SCR was queried to identify associated chemicals (A6-A7). In addition to the NCBI Taxonomy plant species name (B1), entry terms from the MeSH SCR search (B2) were used to identify relevant concepts in the UMLS Meta (B3-B4). These concepts were then searched in Semantic MEDLINE as subject concepts (B6) and restricted to specific predicate types of interest (B7) and filtered for chemical related semantic types (B8-B9).

Table 1: Summary of agreement between chemicals associated with selected plants based on information from Dr. Duke's versus the developed approach (phyotkb_chem). For each of the five plant species examined in detail, the number of chemicals identified by the developed approach from Medical Subject Headings Supplemental Concept Records (SCR) or SemMedDB (SM), as unique to phytokb_chem (u_p), shared by both phytokb_chem and Dr. Duke's (c), unique to Dr. Duke's (u_d), or chemicals that were catalogued in both phytokb_chem and Dr. Duke's but were not identified as either as being associated with a given plant species ($u_{\neg p \neg d}$). For each plant species, respective values for the F-measure (fm), Matthews Correlation Coefficient (mcc), and Cohen's Kappa (κ) are shown.

Plant species (common name)	SCR	SM	u_p	c	u_d	$u_{\neg p \neg d}$	*fm*	*mcc*	κ
Calendula officinalis L. (Marigold)	2	24	24	1	61	5481	0.023	0.103	0.017
Cannabis sativa L. (Marijuana)	42	46	77	9	63	5476	0.011	0.204	0.101
Papaver somniferum L. (Poppy)	5	64	54	14	156	5375	0.118	0.279	0.102
Senna alexandrina Mill. (Senna)	19	10	25	2	29	5518	0.069	0.179	0.064
Solanum lycopersicum L. (Tomato)	0	896	810	85	125	5177	0.154	0.154	0.105

The UMLS Meta CUIs for the objects were retrieved. The candidate objects were filtered for concepts contained within the "Chemical" (T003; 'chem') hierarchy of the UMLS SN (A1.4.1*). The process used to identify chemicals from SemMedDB is graphically depicted in Figure 1 as B1-B11 (orange arrows; bottom half of figure).

Evaluation for Selected Plant Species

For a chosen set of five plant species with known phytochemical properties (*Calendula officinalis* L. [Marigold], *Cannabis sativa* L. [Marijuana], *Papaver somniferum* L. [Poppy], *Senna alexandrina* Mill. [Senna], *Solanum lycopersicum* L. [Tomato]), associated chemicals were retrieved from Dr. Duke's. A complete list of chemicals listed in all plants catalogued in Dr. Duke's was also retrieved.

The full set of Dr. Duke's chemicals were mapped to UMLS Meta CUIs by direct lookup; those chemicals that could not be mapped to a UMLS Meta CUI were not included in the evaluation. The set of chemicals retrieved through the process developed in this study (phytokb_chem) were compared to the set of chemicals from Dr. Duke's for each plant species of interest. In addition to a proportional analysis of the chemicals suggested to be associated with each plant species, three agreement statistics were calculated with Dr. Duke's and phytokb_chem, serving as reference standards to each other:

1. F-Measure (*fm*), which is the harmonic mean of precision and recall for a set of classifications:

$$fm = \frac{2 \cdot c}{(2 \cdot c) + u_p + u_d}$$

2. Matthews Correlation Coefficient (*mcc*), which quantifies the correlation between two classification systems:

$$mcc = \frac{(c \cdot u_{\neg p \neg d}) - (u_p \cdot u_d)}{\sqrt{(2 \cdot c) \cdot (c + u_p) \cdot (u_{\neg p \neg d} + u_p) \cdot (u_{\neg p \neg d} + u_d)}}$$

3. Cohen's Kappa (κ), which ascertains the quality of the relative accuracy of a given classification system as a function of true accuracy compared to random accuracy:

$$\kappa = \frac{\left(\frac{c + u_{\neg p \neg d}}{c + u_{\neg p \neg d} + u_p + u_d}\right) - \left(\frac{\left((u_{\neg p \neg d} + u_p) \cdot (u_{\neg p \neg d} + u_d)\right) + \left((u_d + c) \cdot (u_p + c)\right)}{(c + u_p + u_d + u_{\neg p \neg d})^2}\right)}{1 - \left(\frac{\left((u_{\neg p \neg d} + u_p) \cdot (u_{\neg p \neg d} + u_d)\right) + \left((u_d + c) \cdot (u_p + c)\right)}{(c + u_p + u_d + u_{\neg p \neg d})^2}\right)}$$

Where c represents the number of chemicals in common between Dr. Duke's and phytokb_chem; u_p represents the number of chemicals unique to phytokb_chem; u_d represents the number of chemicals unique to Dr. Duke's; and $u_{\neg p \neg d}$ represents the number of chemicals that are catalogued in either phytokb_chem or Dr. Duke's but not accounted for by u_p or u_d.

Results

A total of 127,597 plant species were searched in MeSH SCR and SemMedDB, resulting in 4,361 plants having at least one chemical. The mean number of chemicals per plant species was 9±100 [95%CI 6-12], with an inclusive range of 1 to 5,589 chemicals per plant species. For the five plant species manually examined relative to Dr. Duke's, there was little overall agreement (summary of agreement statistics shown in Table 1), suggesting that the approach developed here offers potentially synergistic information about chemicals associated with plant species. Detailed results are available as supplementary information at:

https://sites.google.com/a/brown.edu/phytokb/medinfo2019

Discussion

Identification of phytochemicals is an essential aspect in the study of medicinal plants. In advance of the laborious process of developing and implementing screening programs for ascertaining the phytochemicals for a given plant species, it is paramount to have an understanding of what is known about the plant species. To date, the process of cataloguing known phytochemicals relies on a manual curation process. Screening programs, such as those previously led by the National Cancer Institute [16], have led to the identification of active chemicals that have led to significant drugs (e.g., taxol [17]). A major challenge with such screening programs is the difficulty associated with developing appropriate protocols for reliable extraction of phytochemicals [18].

With recent interest in exploring plants as a potential source for new or complementary therapies [19,20], there is a great need to develop robust methodologies for identifying knowledge about plants lest effort be wasted in identifying already known (but perhaps lost) phytochemical knowledge. As with other sectors of biomedicine that have leveraged biomedical knowledge sources, there is a significant opportunity to utilize automated approaches to determine the latest recorded information. This feasibility study demonstrates that there is a strong discordance between existing resources (e.g., Dr. Duke's) and phytochemicals that can be identified using the MeSH Supplemental Concept Records and SemMedDB.

In considering biomedical knowledge sources that may be of utility for identifying phytochemicals, this study focused on using MeSH SCR and SemMedDB. The results of this study

suggest that MeSH SCR and SemMedDB offer complementary knowledge about chemicals. MeSH SCR data are based on drug and chemical information that appear in biomedical literature and can be mapped to MeSH descriptors. As with the MeSH thesaurus, this well-curated list of associations can generally be trusted as *bona fide* associations. However, like the MeSH thesaurus, it is limited in the scope and range of chemicals listed based principles for indexing MEDLINE. Resources like the MeSH thesaurus and MeSH SCR are primarily designed to facilitate information retrieval tasks (i.e., to identify relevant literature for a given query). Nonetheless, the results of this study demonstrate that there is utility in leveraging SCR information to identify potential phytochemicals.

In contrast to MeSH SCR, SemMedDB consists of a detailed set of predications that have been determined using a natural language processing (NLP) system (SemRep). The predications underpin the Semantic MEDLINE system, facilitating the process of retrieving relevant biomedical literature based on semantic relationships between concepts. SemRep based data have been used in a number of text mining studies (e.g., to identify interactions between drugs [21] or proteins [22] from MEDLINE). There have been previous studies in leveraging NLP approaches for studying medicinal plants [23-25]). The present study is, to the best of our knowledge, the first to leverage a predicate extraction system like SemRep for the identification of plant-chemical associations.

A major limitation in the evaluation of the quality of extracted relationships is the limited availability of suitable reference standards. In this study, the phytochemical candidates identified by the developed system were compared to a popular electronic catalogue of phytotherapy knowledge (Dr. Duke's). However, it is important to note that such resources are not "true" reference standards, the information catalogued is incomplete and also may contain inconsistent errors (e.g., chemicals like alcohol are often listed as chemical components, but these are likely artifacts of the phytochemical extraction process [26]). There is some information about putative medicinal uses, but these are supported by limited primary literature. Other resources, like NAPRALERT, have references to primary literature but also may suffer from incorrectness or incompleteness. In this study, the comparison was done with Dr. Duke's because it is a freely accessible resource. Future work will expand to include comparisons to NAPRALERT (which requires a fee for complete access to data on par with Dr. Duke's).

Ultimately, it is envisioned that the work presented in this study will be seen as complementary to resources like Dr. Duke's and NAPRALERT. However, there are still some areas of potential improvement that we consider essential before the results of the developed system could be seen as a reliable catalogue. The first is addressing the issue of identifying chemicals mentioned in literature that are associated with plants only because of the extraction process (e.g., the aforementioned alcohol). While this only occurs in a small percentage of the extractions, it may be easily addressed through a combination of stop words (which would consist of chemicals commonly used for extractions) and information theoretic approaches to identify chemicals that occur commonly and may not be of interest (e.g., water).

Another area of improvement that is planned includes developing a confidence score for the chemicals identified for each plant species. For example, there may be utility in weighting certain predicates in combination with their relative frequency in order to identify chemicals that are of interest for a particular plant species. A more detailed evaluation is needed for each of the data sources for more than five plant species; five plants was chosen to support a manual assessment to verify potential for species that are known to have phytochemicals. Nonetheless, the results of this initial study suggest that there is promise in leveraging biomedical knowledge sources for identifying phytochemicals that are not currently available in electronic phytochemical resources such as Dr. Duke's or NAPRALERT.

Acknowledging that a given plant species has at least one thousand phytochemicals, another important contribution of automated systems such as the one developed here is a tabulation of reported phytochemicals that may supplement data which are in resources like Dr. Duke's or NAPRALERT. In doing so, one can identify plant species that have been heavily studied (e.g., those with more than 1,000 phytochemicals, such as *Solanum lycopersicum* L. [Tomato] as identified either by the process developed here or in Dr. Duke's) versus those that have not had extensive phytochemical analyses shared in accessible resources (e.g., *Calendula officinalis* L. [Marigold] which had only 86 phytochemicals identified either by the process developed here or in Dr. Duke's amidst its long history of medicinal use [27]). The results of this study may thus be used to continually curate available electronic literature sources and, in combination with electronic resources like Dr. Duke's or NAPRALERT, provide a metric for identifying plant species that have been understudied. This would be especially valuable in the context of studying plants that may have therapeutic indications (e.g., based on ethnobotanical survey knowledge).

Biomedical knowledge sources are principally aimed at cataloguing and retrieving information mostly relative to disease knowledge. This can make it challenging to identify non-traditional biomedical concepts, such as plant species. In this study, a combination of MeSH entry terms, MeSH "see also" entries, and UMLS string lookups were used to identify the array of relevant concepts for a given plant species. The source used for taxonomic plant species name in this study was NCBI Taxonomy, which is limited to organisms that have associated data in other databases within NCBI (e.g., GenBank), and does not necessarily reflect the full set of plant species that are catalogued in more comprehensive resources like the International Plant Name Index [28]. Similarly, the set of chemical names indexed within UMLS Meta is not necessarily the full set of known chemicals (i.e., the 5405±138 catalogued chemicals that could not be mapped to UMLS; these were excluded from the evaluation of this study). Future work will therefore need to utilize a more robust approach for identifying plant species names and chemical names that can be added as source concepts for SemRep. Finally, the three target predicate types ("ISA", "LOCATION_OF", or "CONVERTS_TO") used for this study were determined through manual examination of predications associated with plant concepts within SemMedDB; there may additional predications that could be considered (including those that may be added to SemRep) in future work.

This study focused on analyzing biomedical literature indexed in MEDLINE. It is our expectation that the techniques described here may be used to develop approaches to identify phytochemical knowledge from other sources of electronic knowledge (e.g., PubMedCentral or the Biodiversity Heritage Library).

Conclusions

Knowledge about phytochemicals is embedded across a number of resources, including biomedical literature. This

study demonstrates how semantic biomedical knowledge sources can be leveraged to identify potential phytochemicals from literature based resources, focusing on the MeSH Supplement Concept Records and the Semantic MEDLINE Database. The results suggest that automated approaches, such as developed here, can identify a largely non-overlapping, complementary set of potential phytochemicals compared to an existing manually curated resource (Dr. Duke's Phytochemical and Ethnobotanical Databases).

Acknowledgements

This study was supported in part by the the National Institutes of Health grants R01LM011963 and U54GM115677. The content is solely the responsibility of the authors and does not reflect the official views of the National Institutes of Health.

References

[1] D.P. Briskin, Medicinal plants and phytomedicines. Linking plant biochemistry and physiology to human health, *Plant Physiol* **124** (2000), 507-514.

[2] J.B. Harborne, *Phytochemical Methods: A Guide to Modern Techniques of Plant Analysis*, Chapman and Hall, London ; New York, 1980.

[3] W.D. Loub, N.R. Farnsworth, D.D. Soejarto, and M.L. Quinn, NAPRALERT: Computer Handling of Natural Product Research Data, *J Chem Inf Comput Sci* **25** (1985), 99-103.

[4] U.S. Department of Agriculture Agricultural Research Service, Dr. Duke's Phytochemical and Ethnobotanical Databases, 1992-2016.

[5] U.S. National Library of Medicine, MeSH Record Types, 2014.

[6] O. Bodenreider, The Unified Medical Language System (UMLS): Integrating biomedical terminology, *Nucleic Acids Res* **32** (2004), D267-270.

[7] H. Kilicoglu, D. Shin, M. Fiszman, G. Rosemblat, and T.C. Rindflesch, SemMedDB: A PubMed-scale repository of biomedical semantic predications, *Bioinformatics* **28** (2012), 3158-3160.

[8] A.T. McCray, An upper-level ontology for the biomedical domain, *Comp Funct Genomics* **4** (2003), 80-84.

[9] T.C. Rindflesch and M. Fiszman, The Interaction of Domain Knowledge and Linguistic Structure in natural Language Processing: Interpreting Hypernymic Propositions in Biomedical Text, *J Biomed Inform* **36** (2003), 462-477.

[10] T.C. Rindflesch, H. Kilicoglu, M. Fiszman, G. Rosemblat, and D. Shin, Semantic MEDLINE: An Advanced Information Management Application for Biomedicine, *Information Services & Use* **2011** (2011), 15-21.

[11] M. Fiszman, G. Rosemblat, C.B. Ahlers, and T.C. Rindflesch, Identifying Risk Factors for Metabolic Syndrome in Biomedical Text, *AMIA Annu Symp Proc* (2007), 249-253.

[12] G.B. Melton and G. Hripcsak, Automated detection of adverse events using natural language processing of discharge summaries, *J Am Med Inform Assoc* **12** (2005), 448-457.

[13] I.N. Sarkar, A Vector Space Model Approach to Identify Genetically Related Diseases, *J Am Med Inform Assoc* **19** (2012), 249-254.

[14] C.B. Ahlers, M. Fiszman, D. Demner-Fushman, F.M. Lang, and T.C. Rindflesch, Extracting semantic predications from Medline citations for pharmacogenomics, *Pac Symp Biocomput* (2007), 209-220.

[15] The Julia Programming Language (2019). https://julialang.org (Accessed April 1, 2019).

[16] M. Suffness and J. Douros, Current Status of the NCI Plant and Animal Product Program, *J Nat Prod* **45** (1982), 1-14.

[17] M.E. Wall and M.C. Wani, Camptothecin and Taxol: from Discovery to Clinic, *J Ethnopharmacol* **51** (1996), 239-253.

[18] N.R. Farnsworth, Biological and phytochemical screening of plants, *J Pharm Sci* **55** (1966), 225-276.

[19] A.S. Bouldin, M.C. Smith, D.D. Garner, S.L. Szeinbach, D.A. Frate, and E.M. Croom, Pharmacy and herbal medicine in the US, *Social Science & Medicine* **49** (1999), 279-289.

[20] J.P. Kelly, D.W. Kaufman, K. Kelley, L. Rosenberg, T.E. Anderson, and A.A. Mitchell, Recent trends in use of herbal and other natural products, *Arch Intern Med* **165** (2005), 281-286.

[21] R. Zhang, M.J. Cairelli, M. Fiszman, G. Rosemblat, H. Kilicoglu, T.C. Rindflesch, S.V. Pakhomov, and G.B. Melton, Using semantic predications to uncover drug-drug interactions in clinical data, *J Biomed Inform* **49** (2014), 134-147.

[22] I.N. Sarkar and T.C. Rindflesch, Discovering protein similarity using natural language processing, *Proc AMIA Symp* (2002), 677-681.

[23] V. Sharma and I.N. Sarkar, Leveraging Biodiversity Knowledge for Potential Phyto-Therapeutic Applications, *J Am Med Inform Assoc* **20** (2013), 668-679.

[24] V. Sharma and I.N. Sarkar, Leveraging concept-based approaches to identify potential phyto-therapies, *J Biomed Inform* **46** (2013), 602-614.

[25] P. Srinivasan and B. Libbus, Mining MEDLINE for implicit links between dietary substances and diseases, *Bioinformatics* **20 Suppl 1** (2004), i290-296.

[26] S. Sasidharan, Y. Chen, D. Saravanan, K.M. Sundram, and L. Yoga Latha, Extraction, Isolation and Characterization of Bioactive Compounds from Plants' Extracts, *Afr J Tradit Complement Altern Med* **8** (2011), 1-10.

[27] E. Basch, S. Bent, I. Foppa, S. Haskmi, D. Kroll, M. Mele, P. Szapary, C. Ulbricht, M. Vora, and S. Yong, Marigold (Calendula officinalis L.): An evidence-based systematic review by the Natural Standard Research Collaboration, *J Herb Pharmacother* **6** (2006), 135-159.

[28] J. Croft, N. Cross, S. Hinchcliffe, E.N. Lughadha, P.F. Stevens, J.G. West, and G. Whitbread, Plant Names for the 21st Century: The International Plant Names Index, a Distributed Data Source of General Accessibility, *Taxon* **48** (1999), 317.

Address for correspondence

Indra Neil Sarkar, PhD, MLIS, FACMI

Brown University

Box G-R

Providence, RI 02912 USA

neil_sarkar@brown.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190228

Impact of De-Identification on Clinical Text Classification Using Traditional and Deep Learning Classifiers

Jihad S. Obeid[a], Paul M. Heider[a], Erin R. Weeda[b], Andrew J. Matuskowitz[c], Christine M. Carr[c,a], Kevin Gagnon[d], Tami Crawford[a], Stephane M. Meystre[a]

[a]Biomedical Informatics Center, Medical University of South Carolina, Charleston, SC, USA,
[b]Department of Clinical Pharmacy and Outcome Sciences, Medical University of South Carolina, Charleston, SC, USA,
[c]Department of Emergency Medicine, Medical University of South Carolina, Charleston, SC, USA,
[d]Department of Computer Science, University of South Carolina, Columbia, SC, USA

Abstract

Clinical text de-identification enables collaborative research while protecting patient privacy and confidentiality; however, concerns persist about the reduction in the utility of the de-identified text for information extraction and machine learning tasks. In the context of a deep learning experiment to detect altered mental status in emergency department provider notes, we tested several classifiers on clinical notes in their original form and on their automatically de-identified counterpart. We tested both traditional bag-of-words based machine learning models as well as word-embedding based deep learning models. We evaluated the models on 1,113 history of present illness notes. A total of 1,795 protected health information tokens were replaced in the de-identification process across all notes. The deep learning models had the best performance with accuracies of 95% on both original and de-identified notes. However, there was no significant difference in the performance of any of the models on the original vs. the de-identified notes.

Keywords:

Machine Learning, Data Anonymization, Natural Language Processing

Introduction

The use of electronic health records (EHR) data to support research is growing in the translation research community [1]. Significant amounts of those data are trapped in free-text throughout a variety of clinical notes [2,3]. Provider notes often contain patient names, dates of services, and other types of Protected Health Information (PHI). In the United States, the Health Insurance Portability and Accountability Act (HIPAA) protects the confidentiality of patients [4], and based on the Common Rule, researchers are required to obtain either an informed consent from patients or a waiver of informed consent from an Internal Review Board (IRB) in order to use data for research [5]. As a result, automated de-identification in order to make text data more accessible for research is being investigated [6]. The impact of de-identification on the fidelity of information content as well as the utility of the data for research purposes, and information retrieval are also being evaluated [7,8]. In this paper, we examine the impact of de-identification on machine learning-based text classifiers. This experiment was conducted in the context of an automated approach for the detection of altered mental status (AMS) in emergency department (ED) physician notes for the purpose of decision support in the evaluation and management of pulmonary embolism [9–11].

Methods

This study was approved by the IRB at the Medical University of South Carolina (MUSC) under protocol # Pro00080055. We extracted ED physician notes from the MUSC Research Data Warehouse, which contains data extracted from the EHR system. The text notes span a period of 6 years, exported from one commercial EHR system. The notes were enriched with records from adult patients with concurrent International Classification of Diseases (ICD)-10 codes indicating AMS (e.g. codes under the R41 ICD-10 code hierarchy, which includes symptoms and signs involving cognitive functions and awareness) and an equal number of records from patients without AMS ICD codes as controls or negative cases. Using regular expressions to identify the section headers, a total of 8,194 clinical notes were segmented into the different components of a clinical record including the history of present illness (HPI), physical exam, assessment, etc.

Labeling

The parsed notes were imported into REDCap [12] and made available to clinical members of our research team (two ED physicians, a Pediatrician and a Doctor of Pharmacy) for review and labeling as either AMS or not AMS. The team was instructed to look for any signs or symptoms of AMS in the context of pulmonary embolism as described in the literature [9,10] within the HPI, e.g. disorientation, lethargy, stupor, somnolence, confusion, coma, loss of consciousness, or syncope. They were also asked to drop repetitive notes for patients with frequent ED visits in order to minimize bias in testing the models. Cases that were deemed not clear cut AMS by a reviewer, were labeled by consensus after consultation with other team members. The team completed the labeling on 1,113 HPI text notes from 849 patients, with 487 notes labeled as AMS, and 626 labeled as non-AMS. A sub-sample of 100 notes was labeled independently by two labelers in order to estimate the interrater reliability.

De-identification

For automated de-identification, we used a system based on 'BoB' (best-of-breed clinical text de-identification application) [13,14] with a previously demonstrated precision of 93% and

recall of 76% on MUSC data [15]. This system combines high precision algorithms (e.g., regular expression matching) and high recall algorithms, e.g., Conditional Random Fields (CRF) and Support Vector Machines (SVM). After all models are run on a document, their individual PHI determinations are consolidated into a single coherent output using an ensemble method. Tokens tagged as PHI can be replaced with tags, (e.g. a name replaced with a single generic token "[***PHI***]" or a corresponding PHI-type or class name e.g. "[*** Patient ***]") or "resynthesized" (i.e., PHI is replaced with realistic surrogates so a name will be replaced with a different, randomly sampled name). For the purposes of this study, we selected to use the multi-class token replacements in the de-identified output.

Text Processing

We used R version 3.5.1 [16] for constructing the machine learning pipelines, and Keras [17] and TensorFlow [18] for constructing and training the deep learning models. We ran both the original text (i.e., not de-identified) and de-identified text through several traditional bag-of-words (BOW) text classifiers, and word-embedding (WE)-based deep learning models. Text processing for the BOW models was done using the quanteda package [19] and included lower casing, punctuation removal, stop-word removal, stemming, and tokenization. The BOW word frequencies were normalized using term-frequency, inverse document frequency (tf-idf) a weighted approach for term discrimination [20]. For the WE models, we used Keras for text processing, which included lower casing, sentence segmentation, punctuation removal, and tokenization, followed by sequence padding to ensure that all sequences have the same length.

Baseline Approaches

As a baseline, we used regular expressions (regex) on lowercased notes to identify AMS key words as described in the literature in the context of pulmonary embolism [9,10] (e.g. altered, disorientation, disoriented, lethargy, stupor, confusion, syncope, etc.). The regex algorithm was refined after several iterations of testing against the labeled data to include other patterns based on the root words and abbreviations. We also examined the accuracy of ICD codes against the labeled data.

BOW-based Machine Learning Models

The traditional models included: Naïve Bayes Classifier (NBC) with uniform priors, smoothing of 1; Lasso (LASS) with a default alpha=1 and lambda=NULL [21]; Single Decision Tree (SDT) [22] with a maximum depth of 20; Random Forrest (RF) [23] with 201 trees and the number of variables randomly sampled as candidates at each split (mtry)=150; SVM Type 1 with a radial basis kernel [24]; and a Simple Multilayer Perceptron (SMP) artificial neural network with 64-node input and hidden layers.

Word Embedding - Deep Learning Models

The deep learning model was based on the architecture described by Kim [25]. However, instead of using parallel channels for the embedding layer, we used either a pre-trained layer using word2vec (W2V) [26] or word embedding without pre-training (vectors randomly initialized using a uniform distribution) with either of 50 (D50) or 200 (D200) dimensions per word. The W2V weights were derived by pre-training the W2V skip-gram model on all 8,194 HPI (original, not de-identified) notes using a window of 5 words in each direction and 200 dimensions per word. The next layer was a convolutional layer or convolutional neural network (CNN) with multi-filter sizes (3, 4 and 5), 200 nodes each, with global maxpooling, followed by a merge tensor, a fully connected 200 node layer, then a single sigmoid activation output node. A dropout rate of 0.2 was used after both the embedding layer and the last dense layer. Other deep neural network architectures were tested including Kim's and CNN's with larger window sizes; however, we chose the above architecture due to its superior performance in our hands for the purpose of demonstrating the impact of de-identification.

Training and Evaluation

Due to the relatively small number of labeled clinical notes, all the models were trained and evaluated using 5-fold train/test cycles, i.e. the test set in each of those runs was used as an unseen holdout set. Moreover, in order to ensure further consistency in results, the experiment was repeated 5 times by bootstrap sampling using different random seeds. Therefore, all models were run a total of 25 times on different train/test sets. In order to allow comparison between models, the same train/test sets were used during each cycle for all the models. The same random seeds were used for both the original and de-identified sets, thus ensuring identical partitioning of train/test sets and that the models trained on original notes were trained on the de-identified version of the same original notes. The performance metrics, including area under the receiver operating characteristic curve (AUC), were calculated based on the pooled predictions of the test data from the 25 runs. The caret package was used for k-fold, bootstrap sampling and calculations of the metrics [27].

Results

Dataset Statistics

The BOW matrices generated after lower casing, stop-word removal and stemming had token dimensions of 5,200 and 4,925 for the original and de-identified sets respectively. The WE sequences, which did not undergo stop-word removal and stemming, had vocabulary sizes of 7,260 and 6,923 for the original and de-identified sets respectively. The HPI note sizes ranged from 21 to 716 words with a median of 174 words for both original and identified sets. The corpus of the labeled HPI notes included a total of 207,475 tokens.

De-identification Results

The automated de-identification resulted in the replacement of 1,795 PHI tokens from a variety of types or classes of PHI (Table 1), which is less than 1% of all tokens in the corpus. The most prevalent replacements within our HPI data were related to health care unit names (such as "MUSC" or "Gastrointestinal" unit) and ages. The different PHI classes and their prevalences are listed in Table 1.

Table 1– Breakdown of the Numbers of PHI Tokens Replaced by the De-identification System.

PHI Token Class	n	%
Health care unit names	558	31.1%
Ages greater than 89	512	28.5%
Dates	360	20.1%
Provider names	128	7.1%
Patient names	122	6.8%
Street or City	73	4.1%
State or Country	16	0.9%
Phone numbers	15	0.8%
Other organization names	10	0.6%
Other IDs	1	0.1%
Total	**1795**	**100.0%**

Baseline Analyses

We had a fairly high interrater reliability between labelers (Cohen's Kappa = 0.94). Using the ICD codes listed in table 2, the accuracy of concurrent ICD codes assignments associated with the labeled clinical notes was 81.1% (precision 72.2%, recall 92.4%). The presence of any of these codes was considered as positive for AMS otherwise the note was considered negative for AMS by ICD.

Table 2– List of ICD-9 and ICD-10 Codes Considered to be AMS in the Context of Pulmonary Embolism.

Code Set	ICD Code	Diagnosis Name
ICD9	780.0x	Alteration of consciousness
ICD9	780.2	Syncope and collapse
ICD9	780.97	Altered mental status
ICD9	799.5x	Signs and symptoms involving cognition
ICD10	R40.x	Somnolence, stupor and coma
ICD10	R41.0	Disorientation, unspecified
ICD10	R41.8x	Other symptoms and signs involving cognitive functions and awareness
ICD10	R41.9	Unspecified symptoms and signs involving cognitive functions and awareness
ICD10	R55	Syncope and collapse

The accuracy of classification using the refined regex patterns against the labeled notes was 88.1% (precision 81.3%, recall 94.7%).

Machine Learning Performance

There was no discernable difference in performance of any of the models between original and de-identified text, with significant overlap in the 95% confidence intervals of the AUC's for all the models across original vs. de-identified (Figure 1). Table 3 shows the performance of the models, along with the differences in performance between original and de-identified text (Δ's).

The RF model was the best performer in the BOW models with AUC's of 0.978, in both original and de-identified texts. Both CNN's (D50 and D200) with the randomly initialized word embeddings had the best overall performance with AUC's near 99% and average accuracies near 95% across both original and de-identified text, exceeding those of the W2V model.

Discussion

The de-identification resulted in the replacement of 1,795 PHI tokens out of a total of over 200 thousand tokens in our sample of 1,113 HPI notes. Table 1 demonstrates the extent of de-identification that the notes were subjected to. The results show negligible difference in performance of text classifiers on original vs. de-identified HPI notes, across all types of machine learning models. The deep learning models in particular seem to perform exceedingly well in both environments.

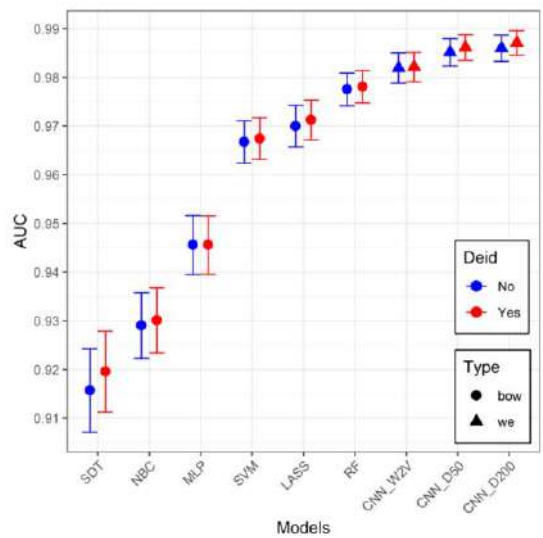


Figure 1. AUC values and 95% confidence intervals for all the models for both original and de-identified (Deid) data.

We hypothesized that the replacement of a number of different named entities with uniform classes of tokens, which slightly reduces the vocabulary size in our corpus and therefore the number of features that the BOW and WE models have to deal with, could potentially improve machine learning performance due to reduced dimensionality and noise in the data. If such an advantage exists, it could not be definitively demonstrated in our results. Several of the models seem to have a slight advantage when applied to the de-identified set, but given the 95% confidence intervals' overlap between original and de-identified, this difference is not significant. It is worth noting that such a result would also be expected if a PHI 'resynthesis' process were used (i.e., replacement of PHI with realistic

*Table 3– Performance of models. (Acc=accuracy, Prec=precision; other abbreviations are in the text, **Δ** = De-identified – Original, 95% CI=95% confidence interval)*

	Original				De-identified				Δ	
Model	**AUC (95% CI)**	**Acc**	**Prec**	**Recall**	**AUC (95% CI)**	**Acc**	**Prec**	**Reecall**	**AUC**	**Acc**
RF	0.978 (0.974-0.981)	0.924	0.938	0.885	0.978 (0.975-0.981)	0.928	0.943	0.890	0.001	0.004
LASS	0.970 (0.966-0.974)	0.906	0.955	0.824	0.971 (0.967-0.975)	0.907	0.958	0.823	0.001	0.001
SVM	0.967 (0.962-0.971)	0.907	0.905	0.879	0.967 (0.963-0.972)	0.908	0.907	0.880	0.001	0.001
MLP	0.946 (0.940-0.952)	0.885	0.875	0.860	0.946 (0.940-0.952)	0.883	0.869	0.863	0.000	-0.002
NBC	0.929 (0.922-0.936)	0.842	0.776	0.898	0.930 (0.923-0.937)	0.848	0.782	0.903	0.001	0.006
SDT	0.916 (0.907-0.924)	0.911	0.921	0.870	0.920 (0.911-0.928)	0.913	0.927	0.870	0.004	0.002
CNN D200	**0.986 (0.983-0.989)**	0.946	**0.946**	0.929	**0.987 (0.985-0.990)**	**0.949**	**0.949**	**0.934**	0.001	0.004
CNN D50	0.985 (0.982-0.988)	**0.948**	0.945	**0.934**	0.986 (0.984-0.989)	0.948	0.947	**0.934**	0.001	0.001
CNN W2V	0.982 (0.979-0.985)	0.939	0.941	0.918	0.982 (0.979-0.985)	0.937	0.936	0.919	0.000	-0.001

surrogates), which is an option in the automatic de-identification system described above, and is based on a large and diverse database of possible surrogates (e.g., all last names found in the U.S. national census). The resynthesis might allow for conservation of diversity of the original PHI.

The fact that automated de-identification did not reduce the accuracy of machine learning performance should be of interest to the translational research community. Clinical text corpora are critical for biomedical informatics research in domains such as machine learning, ontology annotations, predictive modeling and precision medicine, to name a few. Automated de-identification should make such corpora more accessible at scale for such research. While automated de-identification technologies have matured significantly in recent years, regulatory guidelines still lag behind. Looking ahead, we envision several governance models in which de-identified text corpora could be made available with appropriate data use agreements with minimal oversight and review by ethics boards. Institutional research leaders at academic medical centers should work closely with their offices of research oversight and local informatics experts to make such resources more accessible.

Regarding the performance of deep learning vs. the traditional classifiers, the deep learning classifiers seem to significantly outperform the BOW-based classifiers with the exception of RF, which had a performance approaching that of the CNN's. Not surprisingly, all machine learning models outperformed ICD codes on accuracy, which is consistent with reports in the literature on coder errors, such as misattribution, unbundling, and upcoding [28,29]. However, only the better performing models (MLP and above) had better accuracies than the optimized regex classifier. The accuracies of the deep learning models were particularly far superior to the regex accuracy. However, none of the models outperformed the high recall demonstrated by our regex classifier. However, the regex approach required significant fine-tuning specific to the detection of AMS keywords in order to yield this performance. As such, the regex approach is more difficult to generalize to other classification problems. Finally, it should be noted that the W2V initialized WE models, did not outperform the randomly initialized WE models. In fact they seem to have consistently lower AUC's, but with overlapping 95% confidence intervals. The lower performance is possibly due to the relatively small amount of pre-training data, for example, compared to models trained on all of PubMed.

Limitations

Our text corpus represents data from one health system, on a single EHR system making it difficult to draw generalizations about performance in other environments. We also examined the performance of machine learning through the narrow prism of a simple text classifier to identify AMS in one type of EHR clinical text, namely the HPI. Future work should include expanding the study to other institutions and examining other types of notes and machine learning tasks, such as predictions of outcomes.

Conclusions

Despite the limitations outlined above, this simple experiment demonstrates the power of an automated de-identification pipeline, and the preservation of text features that are necessary for the performance of both deep learning models as well as the more traditional machine learning models used in text classification tasks.

Acknowledgements

This project was supported in part by the South Carolina SmartState Program, the South Carolina Clinical & Translational Research Institute, with an academic home at the Medical University of South Carolina, through the National Institutes of Health - National Center for Advancing Translational Sciences grant number UL1-TR001450, and the Delaware-CTR Accel program through the National Institute of General Medical Sciences grant number U54-GM104941.

References

[1] J.S. Obeid, L.M. Beskow, M. Rape, R. Gouripeddi, R.A. Black, J.J. Cimino, P.J. Embi, C. Weng, R. Marnocha, and J.B. Buse, A survey of practices for the use of electronic health records to support research recruitment, *Journal of Clinical and Translational Science* **1** (2017), 246–252.

[2] S.M. Meystre, G.K. Savova, K.C. Kipper-Schuler, and J.F. Hurdle, Extracting information from textual documents in the electronic health record: a review of recent research, *Yearbook of medical informatics* **17** (2008), 128–144.

[3] C. Shivade, P. Raghavan, E. Fosler-Lussier, P.J. Embi, N. Elhadad, S.B. Johnson, and A.M. Lai, A review of approaches to identifying patient phenotype cohorts using electronic health records, *Journal of the American Medical Informatics Association* **21** (2014), 221–230.

[4] HIPAA Privacy Rule, 45 CFR Part 160, Part 164(A,E)., U.S. Department of Health and Humans Services, 2002.

[5] Federal Policy for the Protection of Human Subjects ('Common Rule, *HHS.Gov*. (2009). https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html (accessed November 20, 2018).

[6] S.M. Meystre, F.J. Friedlin, B.R. South, S. Shen, and M.H. Samore, Automatic de-identification of textual documents in the electronic health record: a review of recent research, *BMC Medical Research Methodology* **10** (2010), 70.

[7] S.M. Meystre, Ó. Ferrández, F.J. Friedlin, B.R. South, S. Shen, and M.H. Samore, Text de-identification for privacy protection: a study of its impact on clinical text information content, *Journal of Biomedical Informatics* **50** (2014), 142–150.

[8] L. Deleger, K. Molnar, G. Savova, F. Xia, T. Lingren, Q. Li, K. Marsolo, A. Jegga, M. Kaiser, L. Stoutenborough, and I. Solti, Large-scale evaluation of automated clinical note de-identification and its impact on information extraction, *Journal of the American Medical Informatics Association* **20** (2013) 84–94.

[9] J. Donzé, G. Le Gal, M.J. Fine, P.-M. Roy, O. Sanchez, F. Verschuren, J. Cornuz, G. Meyer, A. Perrier, M. Righini, and D. Aujesky, Prospective validation of the Pulmonary Embolism Severity Index, *Thrombosis and haemostasis* **100** (2008), 943–948.

[10] P. Prandoni, A.W.A. Lensing, M.H. Prins, M. Ciammaichella, M. Perlati, N. Mumoli, E. Bucherini, A. Visonà, C. Bova, D. Imberti, S. Campostrini, S. Barbar, and PESIT Investigators, Prevalence of Pulmonary Embolism among Patients Hospitalized for Syncope, *New England Journal of Medicine,* **375** (2016) 1524–1531.

[11] G. Costantino, M.H. Ruwald, J. Quinn, C.A. Camargo, F. Dalgaard, G. Gislason, T. Goto, K. Hasegawa, P. Kaul, N. Montano, A.-K. Numé, A. Russo, R. Sheldon, M. Solbiati, B. Sun, and G. Casazza, Prevalence of

Pulmonary Embolism in Patients With Syncope, *JAMA Internal Medicine,* **178** (2018),
[12] P.A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, and J.G. Conde, Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support, *Journal of Biomedical Informatics* **42** (2009), 377–381.
[13] O. Ferrández, B.R. South, S. Shen, F.J. Friedlin, M.H. Samore, and S.M. Meystre, BoB, a best-of-breed automated text de-identification system for VHA clinical documents, *Journal of the American Medical Informatics Association* **20** (2013), 77–83.
[14] Y. Kim, P. Heider, and S.M. Meystre, Ensemble-based methods to improve de-identification of electronic health record narratives., in: AMIA Annual Symposium Proceedings 2018, San Francisco, CA, n.d.: pp. 663–672.
[15] S.M. Meystre, P. Heider, Y. Kim, A. Trice, and G. Underwood, Clinical text automatic de-identification to support large scale data reuse and sharing: pilot results., in: AMIA Annual Symposium Proceedings 2018, San Francisco, CA, n.d.: p. 2069.
[16] R Core Team, R: A Language and Environment for Statistical Computing., (2018).
[17] F. Chollet, Keras, (2018). https://keras.io/ (accessed November 20, 2018).
[18] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Y. Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, TensorFlow: large-scale machine learning on heterogeneous systems, (2018). https://www.tensorflow.org/ (accessed November 20, 2018).
[19] K. Benoit, K. Watanabe, H. Wang, P. Nulty, A. Obeng, S. Müller, and A. Matsuo, quanteda: An R package for the quantitative analysis of textual data, *Journal of Open Source Software*. **3** (2018), 774.
[20] G. Salton, C.S. Yang, and C.T. Yu, A theory of term importance in automatic text analysis, *Journal of the American Society for Information Science* **26** (1975), 33–44.
[21] J. Friedman, T. Hastie, and R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, *Journal of Statistical Software* **33** (2010), 1–22.
[22] L. Breiman, Classification and regression trees, Chapman & Hall/CRC, New York, N.Y., 1984. http://lib.myilibrary.com?id=1043565 (accessed December 6, 2018).
[23] L. Breiman, Random Forests, *Machine Learning* **45** (2001), 5–32.
[24] J. Weston, and C. Watkins, Multi-class support vector machines, Citeseer, 1998.
[25] Y. Kim, Convolutional neural networks for sentence classification, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Doha, Qatar, 2014: pp. 1746–1751.
[26] T. Mikolov, K. Chen, G. Corrado, and J. Dean, Efficient Estimation of Word Representations in Vector Space, (2013). https://arxiv.org/abs/1301.3781v3 (accessed November 20, 2018).
[27] M. Kuhn, The caret Package, n.d. http://topepo.github.io/caret/index.html (accessed December 6, 2018).
[28] K.J. O'Malley, K.F. Cook, M.D. Price, K.R. Wildes, J.F. Hurdle, and C.M. Ashton, Measuring diagnoses: ICD code accuracy, *Health Services Research* **40** (2005), 1620–1639.
[29] W.-Q. Wei, C.L. Leibson, J.E. Ransom, A.N. Kho, P.J. Caraballo, H.S. Chai, B.P. Yawn, J.A. Pacheco, and C.G. Chute, Impact of data fragmentation across healthcare centers on the accuracy of a high-throughput clinical phenotyping algorithm for specifying subjects with type 2 diabetes mellitus, *Journal of the American Medical Informatics Association* **19** (2012), 219–224.

Address for correspondence

Jihad S. Obeid, M.D.
Medical University of South Carolina
135 Cannon St. Suite 405
MSC200
Charleston, SC, 29425, USA
Email: jobeid@musc.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190229

Evaluating the Impact of Data Representation on EHR-Based Analytic Tasks

Wonsuk Oh[a], Michael S. Steinbach[b], M. Regina Castro[c], Kevin A. Peterson[d], Vipin Kumar[b], Pedro J. Caraballo[e], Gyorgy J. Simona[f]

[a] Institute for Health Informatics, University of Minnesota, Minneapolis, MN, USA,
[b] Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA,
[c] Division of Endocrinology, Diabetes, Metabolism and Nutrition, Mayo Clinic, Rochester, MN, USA,
[d] Department of Family Medicine and Community Health, University of Minnesota, Minneapolis, MN, USA,
[e] Department of General Internal Medicine, Mayo Clinic, Rochester, MN, USA,
[f] Department of Medicine, University of Minnesota, Minneapolis, MN, USA

Abstract

Different analytic techniques operate optimally with different types of data. As the use of EHR-based analytics expands to newer tasks, data will have to be transformed into different representations, so the tasks can be optimally solved. We classified representations into broad categories based on their characteristics, and proposed a new knowledge-driven representation for clinical data mining as well as trajectory mining, called Severity Encoding Variables (SEVs). Additionally, we studied which characteristics make representations most suitable for particular clinical analytics tasks including trajectory mining. Our evaluation shows that, for regression, most data representations performed similarly, with SEV achieving a slight (albeit statistically significant) advantage. For patients at high risk of diabetes, it outperformed the competing representation by (relative) 20%. For association mining, SEV achieved the highest performance. Its ability to constrain the search space of patterns through clinical knowledge was key to its success.

Keywords:

Data Science, Electronic Health Records, Data Mining

Introduction

The widespread adoption of electronic health records (EHRs)[1] enables new kinds of analytics such as explicitly modeling population heterogeneity or identifying benefit groups for an intervention[2,3]. It is well understood that different analytics tasks and techniques operate optimally on different types of data[4]. For example, association pattern mining requires binary or categorical data[5] and most regression models assume that the predictor variables have an additive effect[6]. Data, as it exists in the EHR, is not ideal for many analytics tasks.

A data representation is a transformation of data into a format amenable to a particular analytic technique. Data transformations are not new, e.g., log or rank transformations of non-normally distributed variables[7] have been a mainstay for decades. The recent success of deep learning in some applications[8] has put data representation into the spotlight and is, at least in part, attributed to the underlying data representation. In this work, we propose a data representation, which is specific to the clinical domain and represents data at a high level and enriches it with clinical knowledge.

Specifically, SEV augments the original data with a set of ordered or partially ordered binary variables, combining information about patients' state from multiple perspectives: therapies, diagnoses, and whether or not the laboratory results or vital signs are normal and/or achieve a typical therapeutic target. These variables are (at least partially) ordered: the variable 'patient is under control with first-line oral therapy', represents a lower severity than the variable 'patient is not under control despite last-line therapy'. These variables are highly interpretable, as they follow clinical reasoning and incorporate clinical knowledge.

To make the discussion concrete, we carry out our study in the context of type 2 diabetes (T2D). Diabetes is a common disease with severe complications[9], affecting 29.1 million Americans. T2D can be prevented or delayed through lifestyle modifications and/or pharmacological treatment[10], hence identifying patients at high risk is of high importance. From a technical perspective, T2D is an ideal evaluation platform, as it exhibits common challenges: T2D is heterogeneous; risk factors are correlated and not necessarily additive; and the time frame between the risk factors and the onset of diabetes can be as long as 20 years, which makes missing data inevitable[2,11].

We encode diabetes risk factors, hyperlipidemia, hypertension, and obesity as SEVs (a set of SEVs for each disease) and perform two clinical tasks related to type-II diabetes. The first task is to predict the onset of diabetes using a Cox model and the second one is to model population heterogeneity in terms of the risk of T2D incidence using association pattern mining. We will compare SEV to five other data presentations, including the original data. The main objective is to study the characteristics of the data representations.

Related Works

Data representations transform the data into a new data set. For a *dimensionality-expanding* representation, the new data has more features than the original data, while for a *dimensionality-reducing* representations, it has fewer. The key concern in dimensionality reduction is information loss. Representations can also be *outcome-specific* or *outcome-independent*. Outcome-specific data representations are specific to a

particular study end point (outcome) and can potentially limit the information loss that is relevant to the outcome, while *outcome-independent* representations do not consider an outcome.

Outcome-specific Representation

A severity score (SS)[12] quantifies disease burden with respect to some outcome of interest. For example, the Framingham diabetes score[13] associates disease burden, defined by a handful of risk factors, with the risk of developing diabetes (an outcome). Severity score is a dimensionality-reducing representation, as it summarizes numerous original risk factors into a single number, which is proportional to the burden conferred by those risk factors on some outcome.

Outcome-independent Representations

Outcome-independent representations transform the original data into a new set of features, typically with a different dimensionality. Many currently existing representations, such as principal component analysis (PCA)[14] and nonnegative matrix factorization (NMF) [15] have the specific aim of reducing the problem dimensionality. PCA is a statistical procedure that transforms a set of features into a new (ordered) set of orthogonal features (called principal components) in a manner such that each subsequent component explains maximal amount of residual variance, and NMF factorizes the original matrix into two matrices having only non-negative values, in a way that each subsequent component captures maximal amount of the residual information. Dimensionality reduction is achieved by using only the first few components.

Deep neural networks (DNNs)[8] are computational models that are inspired by neural networks in animal brains and have recently achieved considerable success. Much of this success is attributed to the data representation of these techniques, which is known as de-noising autoencoders (DAE). DAEs consist of successive layers of transformations, where the outcome of each layer is the input to the next. Each layer is thought to extract features that are higher-level than those of the previous layer. The criterion for the goodness of the transformation is the reconstruction error, which is a measure of how well an autoencoder can reconstruct the original data from its output.

Severity Encoding Variables

Severity Encoding Variables (SEV) is our proposed outcome-independent representation. The purpose of SEV is to summarize the numerous facets of a disease into a single hierarchical variable. Nodes at the same level in the hierarchy are fully or partially ordered.

The construction of the hierarchy replicates the clinical reasoning steps of determining the severity of a certain disease. Reasoning involves a sequence of questions: (i) are lab results and vital signs present and normal, (ii) has an intervention been initiated, and if it has, how aggressive is it (first-line treatment, combination therapy, etc.), and (iii) has a diagnosis been recorded. Accordingly, the first split (at the root) produces three nodes: patient with missing, normal, and abnormal lab results. Next, we reason about medications. Each of the three nodes can be split indicating whether treatment has been initiated and how aggressive those treatments are. The final question splits the nodes based on the presence of diagnoses.

Figure 1 illustrates the SEV for hyperlipidemia. At the root of the hierarchy, we ask whether lab results (LDL, HDL and TG) are normal (if they are not missing) and which (if any) are

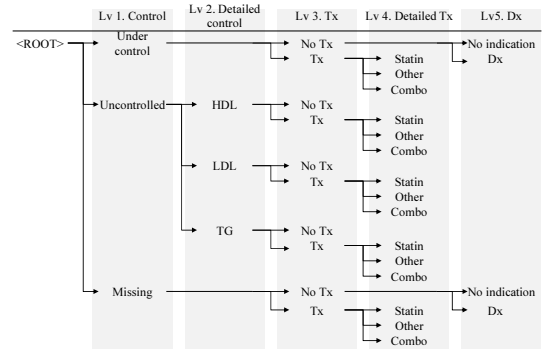


Figure 1: Sample Severity Encoding Variable Hierarchy for Hyperlipidemia. Abbreviations used: Treatment (Tx), Diagnosis (Dx), High-density lipoprotein (HDL), Low-density lipoprotein (LDL), Triglycerides (TG).

Table 1. Categorization of the Data Representations.

	Outcome-Specific	Outcome Independent	
		Data-Driven	Knowledge-Driven
Dimensionality-Reducing	SS	PCA DAE-9	SEV
Dimensionality-Expanding		DAE-34	SEV

abnormal. At the next level, we reason using medications. For example, does a patient under control use medications? If medications are used, are they first-line medications (statins in case of HL), other drugs, or combinations of drugs? On the last level, we reason using diagnoses. Naturally, diagnoses are most helpful if no other indication of disease exists.

For analysis, the hierarchy can be cut at any level and the nodes at that level are taken as binary variables. For example, cutting the hierarchy at the top-most level results in a set of three binary variables: 'patient is under control', 'patient is not under control', and 'laboratory results are missing'. These variables are partially ordered: being under control could (but does not have to) indicate lower severity than not being under control, but 'lab results are missing' is not comparable to the other two in terms of severity. Cutting the hierarchy at the (say) third level yields 10 leaves and incorporates information about medication use. One of these leaves would be 'patient has abnormal LDL despite medication'. By changing the level at which the hierarchy is cut, we can increase the number of leaves (and information content).

SEV is a framework for representing diseases as hierarchies induced by a sequence of clinical decisions; it is not a set algorithm for modeling all diseases. Recall that SEV is outcome independent; once a SEV is constructed, it can be used for multiple study end-points. The diseases that we build SEVs for are predictors of the outcome and the construction of the SEV can (and possibly should) depend on the disease that we build the SEV for. Depending on the disease in question, a different ordering of the same clinical questions could yield a more clinically meaningful hierarchy, and other diseases may incorporate altogether different questions (for example, stage and grade of cancer). We have not observed substantial changes in predictive performance in terms of the ordering of the questions.

Materials and Methods

Data, Cohort Construction and Study Design

Mayo Clinic, located in Rochester, MN, provides primary care to a large population. Resources available at Mayo Clinic are described elsewhere[16]. After IRB approval, a cohort of 75,317 patients aged 18 or older on 01/01/2005 with research consent was constructed. The cohort was followed from the baseline of 01/01/2005 until the end of 2015. To determine patients' baseline status, we retrospectively collected diagnoses of obesity, hyperlipidemia, hypertension, and prediabetes; laboratory test results for lipid panels and fasting plasma glucose (FPG); vital signs (blood pressure, and body mass index [BMI]); demographic information (age, gender); and medications for hypertension and hyperlipidemia. From the cohort, we excluded patients with preexisting diabetes at or before baseline (11,897 patients) and suspicion of diabetes (3 patients with fasting plasma glucose > 125 ml/dL and 2 patients taking anti-diabetic drugs), resulting in a final cohort of 63,415 patients.

Comparative Representation

Severity Scores (SS): A severity score is computed for each diabetes risk factor (obesity, hypertension, hyperlipidemia, pre-diabetes) quantifying the risk factor's contribution to diabetes. While all features could be combined into a single severity score (analogously to the Framingham score), we compute a severity score for each risk factor, combining only the features that are related to the specific risk factor. Modeling the risk factors separately allows us to retain the relationships among them.

For each risk factor, the corresponding SS is the linear prediction from a Cox model, whose independent variables are the data elements that describe the risk factor in question and the dependent variable is diabetes outcome. Missing blood pressure measurements were imputed using mean imputation and a bias-correcting indicator variable signaling whether imputation was performed for each patient was added.

Principal Component Analysis (PCA): In this study, logistic principal component analysis (PCA)[14] is applied to the risk factors, resulting in a single set of principal components. We kept the first 9 principal components because additional components are unable to explain significant amounts of variation. PCA is thus a dimensionality-reducing, outcome-independent representation.

Deep autoencoder (DAE): For this study, we used two configurations, tuned via cross-validation. Both used the hyperbolic tangent activation function, had two hidden layers with 20 nodes on the first layer and had 9 and 34 nodes on the second layer, respectively. The first configuration (DAE-9) has the lowest reconstruction error among configurations that *reduce* the dimensionality of the problem, while the 34-node configuration (DAE-34) has the lowest reconstruction error among all configurations. DAE-9 is a dimensionality-reducing representation, while DAE-34 is a dimensionality-expanding representation.

Severity Encoding Variables (SEV): A severity encoding was constructed for each of the four risk factors of diabetes independently. The hierarchy was cut at the leaf level, making it dimensionality-expanding (there are more nodes in the hierarchy than original features).

The Two Tasks

Regression Analysis: The objective is to measure the impact of the data representations on the predictive performance of estimating patients' 8-year risk of T2D. Risk factors (lab results, vital signs, diagnoses (ICD-9 billing code rolled up into categories), and prescriptions rolled up into NDF-RT pharmaceutical subclasses) are determined at baseline. We use this information transforming into the five new representations. The sixth representation is RAW, the original (untransformed) data. Six Cox proportional hazard models are constructed using age, gender and each of the six data representations as independent variables. Backwards elimination is applied.

Association Pattern Analysis: The central concept in association pattern mining is an *item*, which is a binary variable such as 'presence of hyperlipidemia diagnosis' or 'LDL ≥ 130 mg/dL'. Items are combined into conjunctive sets, called *itemsets* (e.g. 'LDL ≥ 130 mg/dL AND diagnosis of hyperlipidemia'). The association of an itemset with the outcome is measured through *confidence*, which is the fraction of patients presenting with the outcome among patients who present with all conditions in the itemset (fraction of patients who developed diabetes among those with LDL ≥ 130 mg/dL and diagnosis of hyperlipidemia in our example). Association pattern mining systematically enumerates all itemsets and computes their confidence. In the Classification Based on the Association (CBA) framework[17], the risk of diabetes for a patient is the confidence of the highest-confidence rule that applies to that patient.

Continuous variables (age, severity scores, scores from PCA and DAE) are categorized into deciles (with backwards elimination discarding superfluous categories) and laboratory results and vital signs are dichotomized using the American Diabetes Association[18] cutoffs. Of interest are the number of patterns and their predictive performance. A data representation that can achieve higher predictive performance with a lower number of rules is preferable.

Results

Regression Analysis

Figure 2 (a) shows the concordance of the various data representations as box plots. The top, middle, and bottom line in each box correspond to the upper quartile, median, and the lower quartiles of the concordances estimated from the 1,000 bootstrap replications, respectively. The representations are ordered left to right by the number of features they produce.

While all performance differences are statistically significant, some are not substantial. Our population consists of relatively healthy patients, hence all methods achieved high discrimination. A more clinically meaningful question is to accurately estimate diabetes in risk patients who are at relatively high risk and may actually benefit from an intervention. To this end, we consider patients with Framingham score of at least 20 and in Figure 2 (b), we present the predictive performance of the Cox model on the 6 data representations on these 2,493 patients.

Association Analysis

Association rule mining can discover an exponentially large number of patterns, many of which can be coincidental. The parameter that controls the number of patterns is *Minimum Support in Cases (minsupC)*, the number of cases (patients who developed diabetes) to whom the pattern applies. Figure 3

displays the concordance and number of patterns discovered as a function of minsupC.

Discussion

As the paradigm for clinical studies continues to shift toward precision medicine, the range of tasks that clinical data analysis is used for will broaden. Since these newer tasks may operate optimally with different data representations, understanding existing and developing new data representations will become increasingly important. In this manuscript, we proposed a new data representation, Severity Encoding Variables, which represents diseases at a high level and is enriched with clinical knowledge. We compared SEV to five other existing data representations using two clinical tasks.

Assessing the Risk of Incident Diabetes through Regression

The key concern in regression is information loss. The two dimensionality expanding methods, SEV and DAE-34, achieved the highest performance, as they can extract more information (e.g. SEV encodes some interactions and Deep Autoencoders can encode non-linearities). While the performance difference between these two methods in the entire population was minimal (although statistically significant), when we focused on the subpopulation with very high Framingham score (20 or higher), the performance gap widened substantially and SEV outperformed DAE by 20%. Given their high risk of developing diabetes, this is precisely the group of patients for which we need to estimate the risk accurately so that we can effectively target preventive measures to the patients most in need.

Mechanistically, SEV's performance advantage stems primarily from interactions. It can distinguish between patients who have similar lab results at baseline but are in very different states of severity: e.g. patients who are not yet pharmaceutically treated are very different from those who are already undergoing combination therapy at baseline. Despite having similar (abnormal) lab results, the latter patients are at a disproportionately higher risk and interaction among the various facets of the disease are required to model this correctly. Second, SEV can handle missing data without imputation, identifying that the presence of the diagnosis code is more important in patients who have no available lab results than in patients where the lab results already suggest the presence of the disease.

While interactions among various facets of a disease partly explain how SEV achieves high performance, selecting the right interactions is important. Some classification methods, such as decision trees or association rules, are capable of automatically discovering interactions, however, as our experiment with association rules demonstrates, finding the right combination of interactions is non-trivial.

Dimensionality-reducing data representations did not perform well. Dimensionality reduction can reduce noise and can also lead to information loss. Given that our problem is "tall", the number of patients far exceeds the number of variables, dimensionality reduction led to information loss. Among the dimensionality-reducing methods, SS takes the diabetes outcome into account, and hence managed to preserve most of the outcome-related information, achieving a reasonable performance with the smallest number of features. PCA and DAE-9 are outcome-independent, and have suffered greater outcome-related information loss than SS despite having more features.

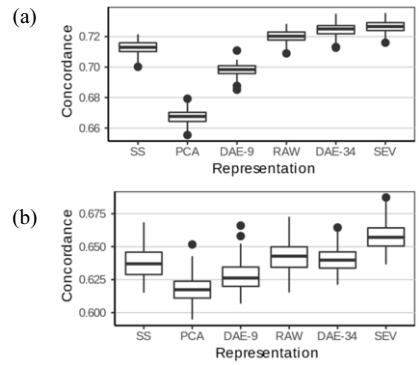


Figure 2: (a) Performance comparison of data representations for the regression task. (b) Comparison of concordance on subpopulation with Framingham score ≥ 20.

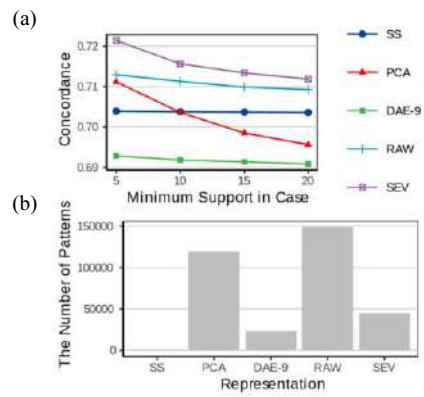


Figure 3: (a) Comparison of the predictive performance of the association patterns discovered using the various data representations as a function of the minimum support in cases (minsupC) (b) The number of association patterns discovered using the various data representations. (minsupC=5)

Modeling Patient Population Heterogeneity through Association Pattern Mining

On this task, SEV performed substantially (and statistically significantly) better than others. The association mining algorithm itself performs dimensionality expansion by forming combinations of the features the data representation provides. To find high-risk patients, we typically focus on patterns that occur in small patient groups, which can yield less reliable risk estimates and higher predisposition to overfitting (finding patterns that happen to randomly *coincide* with diabetes). Different data representations offer different mechanisms to reduce overfitting. The severity scores reduce the number of items an itemset can have. For example, for SS, there are only 5 axes (demographics, obesity, hypertension, hyperlipidemia and prediabetes), each of which is categorized into multiple bins. Since a patient cannot fall into two different bins along the same axis, the maximal number of conditions in a pattern is 5, which seriously limits the number of patterns. Some patterns have as many as 11 conditions in the RAW representation.

SEV, the data representation that achieved the highest performance on association pattern mining, applies a different mechanism. SEV uses the same dichotomization as RAW, but SEV combined these dichotomized variables into predefined "sub-patterns". For instance, the SEV item 'lipids under

control' is a combination of three RAW items: LDL is normal AND HDL is normal AND TG is normal. These higher-level items constrain the space of possible patterns (based on clinical knowledge) and thus reduce the tendency for overfitting.

Generalizability

We tested the data representations with a regression model and association pattern mining to highlight certain characteristics of the SEV representation. We believe that these results generalize to other classification methods, as well. First, the SEV representation offers a high-level clinical description of the diseases enhancing clinical interpretability of the models. Second, SEV can improve predictive performance by automatically handling missing lab results and by incorporating clinically meaningful high-order interactions. Third, as we have mentioned earlier, some methods have the ability to discover interactions, and discovering high-order interaction is non-trivial. Currently, there are no classification methods that can do all three well.

Limitations

Unlike the data-driven representations, the construction of the SEV requires clinical expertise. Most of the effort is spent on classifying diagnoses into categories and determining pharmaceutical subclasses for drugs. This effort is not specific to SEVs; even the RAW representation had access to these higher-level categorizations. The effort that is specific to SEV is determining whether lab results and vital signs are normal and whether a drug is first-line or last-line medication. This information is often readily available from practice guidelines, such as the American Diabetes Association guidelines for diabetes. The effort to include this information is small, but non-negligible. However, SEV is outcome-independent, thus once a hierarchy for a risk factor or disease is defined, it can be used for numerous outcomes without the need to change it.

Conclusions

For both regression and association pattern mining, SEV provides the highest performance, substantially higher than the other data representations in a high-risk subpopulation, where accurate risk assessment is particularly important to appropriately target preventive measures. Besides having the highest performance, SEV produces clinically interpretable models and can also handle missing values.

Acknowledgements

This work was supported by NIH award LM011972, NSF awards IIS 1602394 and IIS 1602198. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies.

References

[1] P.B. Jensen, L.J. Jensen, and S. Brunak, Mining electronic health records: towards better research applications and clinical care, *Nature Reviews Genetics* **13** (2012), 395.

[2] W. Oh, E. Kim, M.R. Castro, P.J. Caraballo, V. Kumar, M.S. Steinbach, and G.J. Simon, Type 2 Diabetes Mellitus Trajectories and Associated Risks, *Big data* **4** (2016), 25-30.

[3] E. Kim, W. Oh, D.S. Pieczkiewicz, M.R. Castro, P.J. Caraballo, and G.J. Simon, Divisive hierarchical clustering towards identifying clinically significant pre-diabetes subpopulations, in: *AMIA Annual Symposium Proceedings*, American Medical Informatics Association, 2014, p. 1815.

[4] Y. Bengio, A. Courville, and P. Vincent, Representation learning: a review and new perspectives, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35** (2013), 1798-1828.

[5] R. Agrawal, T. Imieliński, and A. Swami, Mining association rules between sets of items in large databases, in: *Acm Sigmod Record*, ACM, 1993, pp. 207-216.

[6] J. Tolles and W.J. Meurer, Logistic regression: relating patient characteristics to outcomes, *JAMA* **316** (2016), 533-534.

[7] D.S. Moore, G.P. McCabe, and B.A. Craig, Introduction to the Practice of Statistics, 6th ed., W. H. Freeman, (2007).

[8] B. Shickel, P.J. Tighe, A. Bihorac, and P. Rashidi, Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis, *IEEE Journal of Biomedical and Health Informatics* **22** (2017), 1589-1604.

[9] Centers for Disease Control and Prevention, National Diabetes Statistics Report, 2017, Atlanta, GA, 2017.

[10] Diabetes Prevention Program Research Group, Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin, *New England journal of medicine* **346** (2002), 393-403.

[11] G. Hulsegge, A. Spijkerman, Y. Van Der Schouw, S.J. Bakker, R. Gansevoort, H. Smit, and W. Verschuren, Trajectories of metabolic risk factors and biochemical markers prior to the onset of type 2 diabetes: the population-based longitudinal Doetinchem study, *Nutrition & Diabetes* **7** (2017), e270.

[12] J.O. Pelz, A. Stojadinovic, A. Nissan, W. Hohenberger, and J. Esquivel, Evaluation of a peritoneal surface disease severity score in patients with colon cancer with peritoneal carcinomatosis, *Journal of Surgical Oncology* **99** (2009), 9-15.

[13] P.W. Wilson, J.B. Meigs, L. Sullivan, C.S. Fox, D.M. Nathan, and R.B. D'Agostino, Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study, *Archives of internal medicine* **167** (2007), 1068-1074.

[14] A.J. Landgraf and Y. Lee, Dimensionality reduction for binary data through the projection of natural parameters, *arXiv preprint arXiv:1510.06112* (2015).

[15] J.C. Ho, J. Ghosh, S.R. Steinhubl, W.F. Stewart, J.C. Denny, B.A. Malin, and J. Sun, Limestone: High-throughput candidate phenotype generation via tensor factorization, *Journal of Biomedical Informatics* **52** (2014), 199-211.

[16] J.L. St Sauver, B.R. Grossardt, B.P. Yawn, L.J. Melton III, J.J. Pankratz, S.M. Brue, and W.A. Rocca, Data resource profile: the Rochester Epidemiology Project (REP) medical records-linkage system, *International Journal of Epidemiology* **41** (2012), 1614-1624.

[17] B. Liu, W. Hsu, and Y. Ma, Integrating classification and association rule mining, in: *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, 1998, pp. 24-25.

[18] American Diabetes Association, 2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2018, *Diabetes care* **41** (2018), S13-S27.

Address for correspondence

Gyorgy J. Simon

Institute for Health Informatics, 8-100 Phillips Wangensteen Building, 516 Delaware St. SE, Minneapolis, MN 55455

Email: simo0342@umn.edu. Phone: 1 (612) 626-3364

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190230

Visual Analytics for Congestive Heart Failure Mortality Prediction

Rema Padman[a], Ofir Ben-Assuli [b], Tsipi Heart[b], Nir Shlomo [c], Robert Klempfner[c]

[a] The H. John Heinz III College of Information Systems and Public Policy, Carnegie Mellon University, Pittsburgh, USA
[b] Faculty of Business Administration, Ono Academic College, Kiryat Ono, Israel
[c] Heart Center, Sheba Medical Center, Ramat Gan, Israel

Abstract

Several indices exist to classify Congestive Heart Failure (CHF) patients' propensity for early mortality; however, they are primarily based on limited data and are not intuitive to use at the point of care. We investigate a novel, data-driven, risk assessment and visualization approach to investigate mortality prediction of CHF patients using data retrieved from an intensively digitized hospital's data repository. Combining well-known, computationally efficient, dimensionality reduction (DR) methods with 2-d information visualization, the method classifies and visualizes CHF patients into high and low risk groups, contextualized by the factors driving their classification. The DR method performed similar to logistic regression (LR), but visualized the classification and its significant factors at the population level, individual level and the potential impact of interventions for an individual patient. These are encouraging results in favor of the proposed visualization approach, and contributes to the current focus on advancing patient care via large-scale visual analytics.

Keywords:

risk assessment, computer graphics, heart failure

Introduction

The widespread adoption of Health Information Technology (HIT) by healthcare organizations in many parts of the world has generated vast repositories of detailed patient-level data as a result of routine care delivery [1]. Efficient and effective use of multi-dimensional, multi-sourced data for clinical decision support at the point of care is further exacerbated by the time-constrained and stressed environments where most of health care is provided – in the general practitioner (GP)'s office, hospitals and emergency care institutions [2]. Visually intuitive methods and tools that support clinicians for early identification of high-risk patients with Congestive Heart Failure (CHF), as well as patients with better adherence to treatment recommendations and shared decision-making, can improve both clinical decision-making and patient health outcomes. The objective of this study is to demonstrate a multi-level visualization of CHF patients' disease risk classification that combines statistical dimensionality reduction (DR) methods with innovative information visualization to enhance efficient and effective clinical decision support at the point of care. The visual presentation provides a simple, intuitive tool for clinicians to quickly identify high-risk patients and understand key factors affecting their condition at a snapshot in time.

Congestive Heart Failure (CHF)

Based on recent data, nearly 5 million US adults are currently living with CHF as are more than 25 million adults worldwide [3]. CHF affects people of all ages, from children and young adults to the middle-aged and elderly. Almost 1.4 million individuals with CHF in the U.S are under 60 years of age. CHF is responsible for 11 million physician visits each year in the US, and more hospitalizations than all forms of cancer combined (https://www.emoryhealthcare.org/heart-vascular/wellness/heart-failure-statistics.html). In fact, CHF is the leading cause of death for both men and women, and the most common diagnosis in hospital patients age 65 years and older. More than half of those who develop CHF die within 5 years of diagnosis, and sudden death is common in patients with CHF, occurring at a rate of six to nine times that of the general population [4, https://www.cdc.gov/heartdisease/facts.htm].

Several indices exist to classify CHF patients' propensity for early mortality. Among the most popular are the Enhanced Feedback for Effective Cardiac Treatment (EFFECT) [5]; the Emergency department-based models of Society of Chest Pain Centers, and the Ottawa Heart Failure Risk Scale [6].

Current prediction tools are largely based on limited data, rather than on patient-level big data derived from various data sources [7, 8]. Furthermore, most prior studies either used explanatory rather than predictive statistics, or utilized traditional, albeit sometimes complex, statistical models [9, 10].

Visual Analytics (VA)

Data or information visualization refers to graphical *presentation* of complex computerized results in formats that enable rapid comprehension of complex situations by the viewers, but does not necessarily involve *analysis* tasks or *interactivity* [11]. In contrast, VA requires collaborative interaction between the user and the computer. It is defined as the science of analytical reasoning facilitated by interactive visual interfaces [12]. VA combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision-making on the basis of very large and complex data sets. Using VA, the results are presented via visual interactive platforms for decision makers to explore and gain insights and new knowledge. The approach is to marry the big data processing capabilities of analytics with the human intuitive capabilities of interactive visualization.

Benefits of using visualization tools

Innovative algorithms and tools that combine statistical machine learning and optimization with information visualization and electronic health data may reduce clinicians' information processing load and improve their ability to assess risk of disease onset and related complications at the point of care [2, 13]. Information visualization utilizes the high bandwidth processing capabilities of the human visual system to more efficiently perform interactive data exploration and glean important insights [14, 15]. A critical element in visualization is the incorporation of an expert user, such as the clinician decision-maker, in the interpretation of the data. This may make visualization methods particularly useful for cardiology care where clinicians desire flexibility for customizing assessments to the needs of their unique patient

populations. To the best of our knowledge, this is the first study on computationally driven visualization techniques for improving CHF risk assessment.

We applied the visual analytics method to two subsets of the uniquely comprehensive cardiology dataset from a major academic medical center in Israel. We compare the risk visualization approach with Logistic Regression (LR) results to highlight the unique value of visual analytics in clinical decision making for complex health conditions.

Methods

Study setting and data

The Sheba Medical Center is the largest Israeli hospital, located in the center of the country. It is a leading public academic medical center, annually handling over a 1.5 million patient visits and ~200K emergency visits, and conducts more than two million medical tests of all types. Sheba is an extensively computerized hospital, where an integrated EMR system is used in the ED and all inpatient departments, fully replacing paper-based medical records. In recent years, Sheba has established a research-focused data warehouse which collects data describing imaging results, manual monitored parameters keyed into the clinical information system in the intensive care unit (patient temperature, blood pressure, pulse, blood oxygen saturation, weight, height, etc.), medical devices data (ECG, Echocardiography examinations, Cardiac Catheterization, Nuclear Imaging), the National Population Registry, the National Cancer Registry and the Sheba Executive Survey data.

Each patient admitted to the hospital, via the ED or otherwise, who is diagnosed with a cardiovascular condition as a primary diagnosis, has data uploaded to the data warehouse. All patient information is then retrieved from the other systems using the unique patient ID as it appears in the Israeli Population Registry. This unique identifier allows accurate location of patient data. The records have the potential to contain above thousand features for each patient, albeit sparsely populated. Such a health data repository is both rich and unique for large-scale data analysis, but there are many data challenges. While return of investment of such data integration projects is uncertain [16], one of the use cases proposed for health data is to improve clinicians' decisions at the point of care.

In this preliminary study, we use a subset of the uniquely comprehensive Sheba data warehouse for our analysis, which provides many relevant variables affecting readmissions and mortality and covers all possible patients treated at Sheba over the recent five years. Consequently, this represents the whole population of interest instead of a limited, skewed sample.

Data sample

We employed k-means cluster analysis on a dataset of 7,168 patients hospitalized at the Sheba Medical Center for heart disease between 2010 and 2017. For the demonstration of the visual analytics approach, we selected two specific clusters: younger (N= 204, average age 63 (±15.5)) and older patients (N=367, average age 79.12 (±10.57). Table 1 shows the descriptive statistics of the younger cohort. The other cohort is omitted due to page limit constraints.

Logistic regression to predict patients' propensity for early mortality

Among the many analytical methods currently available, we chose logistic regression (LR) as a baseline method to compare with the visual analytics approach. LR is a widely used and accepted statistical method in clinical medicine [17].

Visual analytics approach

We apply prior work on dimensionality reduction and information visualization for disease risk assessment [2, 13] to CHF data. Specifically, the visual analytics approach incorporates Principal Component Analysis (PCA) and Fisher's Linear Discriminant Analysis (LDA) to develop informative two-dimensional (2-D) projections of multi-dimensional patient data and classify them for CHF-related risk assessment [2, 13]. Included in this step are the identification of appropriate data normalization procedures for the dimensionality reduction methods and the disparate measurement of the data attributes. A set of feasible methods for pre-processing and projecting high-dimensional patient data to 2-D plots are incorporated into the framework so that multiple visual enhancements that may augment a user's analysis can be obtained. Prior results show that the framework may generate models which visually classify a large patient population with accuracy comparable to that produced by common statistical methods [2, 13].

Starting with a d-dimensional space (the d features of interest in the current study), PCA uses an eigenvalue decomposition to find d orthogonal linear combinations that explain the most variance in the data [18]. The principal components are ordered based on the amount of variance explained. LDA, on the other hand, is explicitly concerned with classification and takes into account data labels [19]. In a dataset labeled with k classification levels, LDA finds a k-1 dimensional projection of the d-dimensional data that maximizes the ratio of between-class variance to within-class variance. LDA was chosen because it stratifies patients into risk groups. PCA ensures that patient observations are maximally scattered in the reduced space. The data are projected into 2-D space as scatter plots using the optimal linear combinations found by PCA and LDA, with LDA finding a decision boundary for classifying patients into high and low risk categories. Contextualization of the scatter plot is critical for ease of use and interpretability, and is provided by anchoring the risk factors around the scatter plot, effectively circumscribing them in a circle. Each anchor's direction of attraction (θi) is defined using the ratio of that feature's weights on the vertical and horizontal components of the two-dimensional space:

$$\theta_i = \begin{cases} \tan^{-1}\ (d_{i1} / d_{i2}) & d_{i2} \geq 0 \\ \pi + \tan^{-1}\ (d_{i1} / d_{i2}) & d_{i2} < 0 \end{cases} \qquad (1)$$

where d_{i1} is the weight of feature i on the vertical axis component, and d_{i2} is the weight of feature i on the horizontal axis component. Finally, the significance of each anchor can be represented by the size of the anchor (S_i), computed, for simplicity, as proportional to either d_{i1} or d_{i2}. Further details of the method and its evaluation with diabetes data are reported in [2, 13].

These methods have been implemented in an RShiny-based platform (https://shiny.rstudio.com/) that allows loading, viewing, descriptive analyses, predictive analyses and visual analytics of any dataset. We employ these methods and tools to a subset of the Sheba data sample with a set of selected features to investigate the population distribution between low- and high-risk patients, and examine factors emerging as most critical for high-risk patients. The risk prediction results from this phase are compared against the results from LR to generate additional insights, particularly to identify patients who are noted to be at the boundary of high and low risk levels by the two approaches.

Results

The two data subsets represent older group (OG) and younger group (YG) of CHF patients, on average, that are randomly retrieved from the large dataset. In the future, we intend to use the complete dataset, possibly even over a longer duration, and include additional data on demographics, diagnoses and current and past health indicators, medications, and lab and imaging results.

Descriptive statistics

The mean and proportion of indicators that differed significantly between the younger group (YG- alive and deceased) are highlighted in Table 1 (The older group (OG) is also discussed but the table is not included due to page limits). As evident in Table 1, among factors affecting significantly likelihood of mortality are age (older more likely), prior CHF, dyslipidemia, hypertension and diabetes mellitus. Among lab tests, patients with lower level of hemoglobin in the first test (indication of anemia), and lower GFR MDRD (measure of renal function) were more likely to die. Additional factors were lower systolic and diastolic blood pressure among patients who died. Among the older patients (not shown as a table), age was a significant factor (older more likely to die), lower GFR MDRD and lower systolic and diastolic blood pressure. For this cohort, however, two unique factors affected the likelihood of older patients' mortality: higher sodium level in the blood and lower level of hematocrit. Interestingly, factors significantly affecting the likelihood of mortality in younger patients were derived from medical history and risk factors, whereas for older patients they were derived mainly from lab tests and physical examination.

Table 1: Data description of the younger population

Indicator and Description	Total	Alive (0)	Deceased (1)	P value
N:Total admissions included	204	160	44	
Patient age (mean ± sd)	63 (15.6)	61 (15.5)	70 (13.4)	<0.001
Male (%)	144 (71)	114 (71)	30 (68)	0.835
Ejection fraction (mean ± sd)	38.08 (17)	37.38 (17)	40.66 (16)	0.266
Past Percutaneous Coronary Infusion (%)	48 (23)	36 (22)	12 (27)	0.645
Past Myocardial Infarction (%)	55 (27)	44 (27)	11 (25)	0.889
Congestive Heart Failure (%)	117 (57.4)	90 (56.2)	27 (61.4)	0.663
Past Stroke (%)	15 (7.4)	13 (8.1)	2 (4.5)	0.632
Peripheral arterial disease (%)	7 (3.4)	5 (3.1)	2 (4.5)	1.000
Dyslipidemia (%)	82 (40.2)	58 (36.2)	24 (54.5)	0.044
Hypertension (%)	101 (49.5)	73 (45.6)	28 (63.6)	0.052
Diabetes Mellitus (%)	64 (31.4)	43 (26.9)	21 (47.7)	0.014
Chronic Obstructive Pulmonary Disease (%)	10 (4.9)	6 (3.8)	4 (9.1)	0.290
FIRST Hemoglobin (mean ± sd)	12.49 (1.91)	12.76 (1.81)	11.52 (1.96)	<0.001
Left Ventricular Function (%)	133 (65.2)	107 (66.9)	26 (59.1)	0.435
SPAP (mean ± sd)	47.33 (13.00)	47.09 (13.46)	48.23 (11.25)	0.606
Glomerular Filtration Rate MDRD (mean ± sd)	60.39 (30.35)	67.40 (29.10)	34.91 (19.24)	<0.001
SODIUM (mean ± sd)	138.01 (4.23)	137.91 (3.59)	138.39 (6.05)	0.506
Hematocrit (mean ± sd)	37.04 (6.33)	37.20 (5.65)	36.43 (8.39)	0.478
ACE/ARBs (%)	176 (86.3)	141 (88.1)	35 (79.5)	0.223
Chronic Aldactone (%)	127 (62.3)	105 (65.6)	22 (50.0)	0.086
Length of Stay (mean ± sd)	8.60 (7.11)	8.27 (6.72)	9.81 (8.36)	0.206
Atrial Fibrillation (%)	52 (25.5)	37 (23.1)	15 (34.1)	0.200
Systolic Blood Pressure (mean ± sd)	120.27 (24.00)	124.82 (21.44)	103.73 (25.74)	<0.001
Diastolic Blood Pressure (mean ± sd)	67.82 (14.21)	70.86 (11.86)	56.80 (16.56)	<0.001
First Glucose (mean ± sd)	123.00 (54.19)	120.15 (52.06)	133.36 (60.83)	0.153
First SGOT (mean ± sd)	37.78 (45.01)	39.48 (49.53)	31.61 (21.03)	0.306
First INR (mean ± sd)	1.45 (0.81)	1.42 (0.80)	1.54 (0.81)	0.385
Echo Fractional Shortening (mean ± sd)	26.05 (12.33)	25.28 (12.22)	28.85 (12.49)	0.089
Body Mass Index (mean ± sd)	27.73 (4.51)	27.73 (4.65)	27.70 (3.98)	0.966
Echo LV Mass Index (mean ± sd)	117.14 (29.05)	118.21 (29.72)	113.24 (26.38)	0.316

Table 2: Comparing DR and LR on statistically significant factors driving risk

DR		LR	
Mortality Younger	Mortality Older	Mortality Younger	Mortality Older
Pre-MI	Any-RE-CHF	DM	HCT
Pre-CVA	First-HGB	GFR MDRD (lower)	GFR MDRD (lower)
Pre-COPD	SPAP, Sodium level	Sodium level	Sodium level
First HGB	DBP, Glucose	SBP (lower)	SBP (lower)
HCT	Any-ARBs		
Any-ACE	HOS-Duration		
Pre-A.Fib	Pre-AFib		

Table 3: Comparison across Dimensionality Reduction (DR) vs. Logistic Regression (LR)

Mortality: Young vs. Old Patients - Comparison across methods				
	Young		Old	
Performance measures	DR	LR	DR	LR
Accuracy	0.853	0.863	0.668	0.807
Recall	0.864	0.795	0.867	0.722
Precision	0.613	0.648	0.415	0.586
F1 Score	0.717	0.714	0.561	0.647
Confusion matrix				
Actual / Predicted	DR	LR	DR	LR
Died / High risk patients (TP)	20	35	78	65
Died / Low risk patients (FN)	0	9	12	25
Not died / High risk patients (FP)	60	19	110	46
Not died / Low risk patients (TN)	124	141	167	231

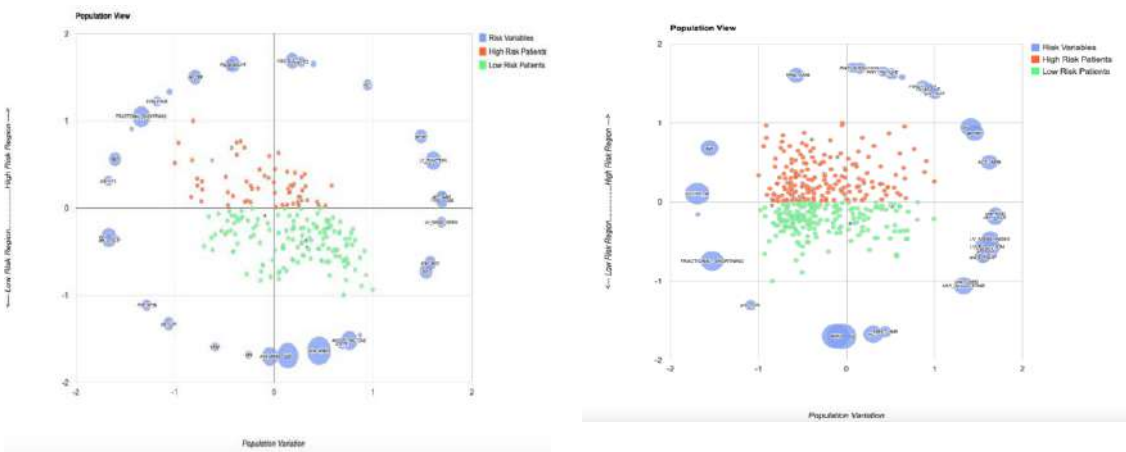


Figure 1: Population-level mortality risk prediction for younger (top) and older (bottom) patient clusters

Dimensionality reduction and population risk visualization

Figure 1 displays the population-level mortality risk prediction for younger and older patients, respectively, using the statistical dimensionality reduction approach described above. It is interesting to note that the distribution of mortality risk in the respective patient cohorts (indicated by red points for high risk and green for low risk patients), the factors driving risk in each group (indicated by the blue circles at the top of the displays and listed in Table 2) and protective factors (indicated by the blue circles at the bottom of the displays – compliance with medications) are distinct and different for each subset. These visualizations provide new insights about customized, cohort-driven population risk management. Individual patient-level and intervention-level analysis can be performed analogously, as described in [2, 13].

The DR technique classified almost 70% of the younger group to have low-risk of mortality, while 30% were classified as high-risk (Figure 1, top). In this cohort, there are seven factors in the high risk region (top of the graph) that are driving risk for this population while eight factors at the bottom of the graph are protective.

Comparison of DR and LR

Table 2 shows factors that have been determined to be statistically significant in driving the classification as high or low risk. It is interesting to note that that DR identifies a different set of factors for the younger vs. the older populations whereas the factors are quite similar in the LR solution. DR also identifies a larger group of modifiable factors that clinicians can intervene on compared to LR. These findings need to be evaluated with larger data sets and validated extensively with practicing clinicians and the clinical literature.

We compared the two methods using three measures (Table 3): Accuracy (TP + TN divided by the number of participants), Recall (TP divided by TP + FN) and Precision (TP divided by TP + FP). DR outperformed LR on the Recall measure for both cohorts, an important indicator for mortality.

Discussion

The DR analysis provides a visual classification of the patient cohort into low- and high-risk regions. Factors affecting the specific classification can be derived and visualized as anchors around the scatter plot (circles), as can the magnitude of the effect of individual anchors (size of the circle). It is important to demonstrate the validity of this classification method in the context of multiple disease conditions. This has been done by comparing the visual analytic results to those of logistic regression, a common classification method. As presented in Tables 1, 2 and 3, DR outperformed LR in discriminating high risk patients (TP), while LR performed better in identifying low risk patients (TN). Although results of other comparison measures (accuracy, precision) were inconclusive, the capacity of DR to more precisely identify high-risk patients, without likewise increasing the falsely identified patients (especially FN) is encouraging (Table 3). Minimizing the FN is the major criterion in actual practice. Moreover, DR identified a higher number of significant factors affecting high-risk patients specifically for older, readmitted patients (Table 2).

While expert physicians are generally able to classify regular patients to low- or high-risk categories, patients with multiple comorbidities often are more difficult to classify correctly. We found that the DR tool was able to classify patients as either low- or high-risk, thus aiding the clinician to refine their assessment. Both options are important: classifying low-risk patients as high-risk is costly and causes unnecessary worry and inconvenience for the patients, and classifying high-risk patients as low-risk may be dangerously life threatening. This approach offers a unique opportunity to deliver cognitively-guided capabilities that have the potential to move statistical risk models closer to the frontlines of clinical practice using visualization techniques, and facilitate the goal of meeting clinician and patient information needs to improve care quality and health outcomes. The demonstrated visualization techniques may benefit multiple stakeholders, including patients, clinical practitioners, researchers and policy-makers, with generalizability to many risk assessments in clinical practice.

Conclusions

We have demonstrated the promise of an innovative statistical model-based approach to create a visualization tool (dimensionality reduction and 2-D visualization). The advantages are evident in the lower number of FNs and identification of modifiable significant factors that differ between younger and older populations. In terms of model characteristics, such as accuracy and recall, results are not conclusive. Additional analyses with larger patient cohorts need to be carried out.

Future research includes many avenues such as extending the methodology to allow time-varying risk factors, optimizing feature selection and decision boundary alignment and evaluation against several other statistical and machine learning methods.

References

[1] E. Scheurwegs, K. Luyckx, L. Luyten, W. Daelemans, and T. Van den Bulcke, Data integration of structured and unstructured sources for assigning clinical codes to patient stays, *JAMIA* 2016, doi: 10.1093/jamia/ocv115.

[2] C.A. Harle, D.B. Neill, and R. Padman, Information visualization for chronic disease risk assessment, *IEEE Intelligent Systems* **27** (2012), 81-85.

[3] G. Savarese and L.H. Lund, Global Public Health Burden of Heart Failure, *Cardiac failure review* **3** (2017), 7-11.

[4] American Heart Association, Cardiovascular Disease: A Costly Burden for America, Projections Through 2035, in, www.healthmetrics.heart.org, 2017.

[5] F.J. Martín-Sánchez, V. Gil, P. Llorens, P. Herrero, J. Jacob, C. Fernández, Ò. Miró, and A.H.F.W.G.o.t.S.S.o. E.M.I. Group, Barthel Index—Enhanced Feedback for Effective Cardiac Treatment (BI—EFFECT) Study: Contribution of the Barthel Index to the Heart Failure Risk Scoring System Model in Elderly Adults with Acute Heart Failure in the Emergency Department, *Journal of the American Geriatrics Society* **60** (2012), 493-498.

[6] I.G. Stiell, J.J. Perry, C.M. Clement, R.J. Brison, B.H. Rowe, S.D. Aaron, A.D. McRae, B. Borgundvaag, L.A. Calder, and A.J. Forster, Prospective and Explicit Clinical Validation of the Ottawa Heart Failure Risk Scale, With and Without Use of Quantitative NT-pro BNP, *Academic Emergency Medicine* **24** (2017), 316-327.

[7] S.E. Mitchell, J. Martin, S. Holmes, C.van D. Lukas, R. Cancino, M. Paasche-Orlow, C. Brach, and B. Jack, How Hospitals Reengineer Their Discharge Processes to Reduce Readmissions, *J Health Qual.* **38** (2016), 116-126.

[8] A.J. Sommerfeld, A.D. Althouse, J. Prince, and G.W. Hickey, Obstructive Sleep Apnea is Associated with Increased Readmissions in CHF Patients, *Journal of Cardiac Failure* **22** (2016), S88.

[9] R. Amarasingham, F. Velasco, B. Xie, C. Clark, Y. Ma, S. Zhang, D. Bhat, B. Lucena, M. Huesch, and E.A. Halm, Electronic medical record-based multicondition models to predict the risk of 30 day readmission or death among adult medicine patients: validation and comparison to existing models, *BMC Medical Informatics and Decision Making* **15** (2015), 39.

[10] J. Futoma, J. Morris, and J. Lucas, A comparison of models for predicting early hospital readmissions, *Journal of Biomedical Informatics* **56** (2015), 229-238.

[11] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, Visual analytics: Definition, process, and challenges, In: Kerren A., Stasko J.T., Fekete J.D., and North C., (eds.), *Information visualization, Lecture Notes in Computer Science, vol 4950.* Springer, Berlin, Heidelberg, 2008, pp. 154-175.

[12] K.A. Cook and J.J. Thomas, Illuminating the path: The research and development agenda for visual analytics, in, Pacific Northwest National Lab., Richland, WA, 2005.

[13] C.A. Harle, D.B. Neill, and R. Padman, An information visualization approach to classification and assessment of diabetes risk in primary care, in: *Proceedings of the 3rd INFORMS Workshop on Data Mining and Health Informatics (DM-HI 2008)*, 2008.

[14] J.J. Caban and D. Gotz, Visual analytics in healthcare–opportunities and research challenges, Oxford University Press, 2015.

[15] B. Shneiderman, The eyes have it: A task by data type taxonomy for information visualizations, In: *Proceedings, IEEE Symposium on Visual Languages*, 1996, pp. 336-343.

[16] L.E. Price, K. Shea, and S. Gephart, The Veterans Affairs's Corporate Data Warehouse: Uses and Implications for Nursing Research and Practice, *Nursing Administration Quarterly* **39** (2015), 311-318.

[17] A. Dag, A. Oztekin, A. Yucel, S. Bulur, and F.M. Megahed, Predicting heart transplantation outcomes through data analytics, *Decision Support Systems* **94** (2017), 42-52

[18] I. Joliffe, Principal Component Analysis. Springer, 2002.

[19] T. Hastie, J. Friedman, and R. Tibshirani, Linear methods for classification, *The Elements of Statistical Learning*, Springer, 2001, pp. 79-113.

Address for correspondence

Prof. Rema Padman, The H. John Heinz III College of Information Systems and Public Policy, Carnegie Mellon University, Pittsburgh, USA, rpadman@cmu.edu.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190231

Provenance Solutions for Medical Research in Heterogeneous IT-Infrastructure: An Implementation Roadmap

Marcel Parciak[a], Christian Bauer[a], Theresa Bender[a], Robert Lodahl[a], Björn Schreiweis[b], Erik Tute[c], Ulrich Sax[a]

[a] *Department of Medical Informatics, University Medical Center, Göttingen, Germany,*
[b] *Institute for Medical Informatics and Statistics, University Hospital Schleswig-Holstein and Kiel University, Kiel, Germany,*
[c] *Peter L. Reichertz Institute for Medical Informatics, Hannover Medical School, Hannover, Germany*

Abstract

Research data generated in large projects raise challenges about not only data analytics but also data quality assessments and data governance. The provenance of a data set – that is the history of data sets – holds information relevant to technicians and non-technicians and is able to answer questions regarding data quality, transparency, and more. We propose an implementation roadmap to extract, store, and utilize provenance records in order to make provenance available to data analysts, research subjects, privacy officers, and machines (machine readability). Each aspect is tackled separately, resulting in the implementation of a provenance toolbox. We aim to do so within the context of HiGHmed, a research consortium established within the medical informatics initiative in Germany. In this testbed of federated IT-infrastructures, the toolbox shall assist each stakeholder in answering domain-specific and domain-agnostic questions regarding the provenance of data sets. This way, we will improve data re-use, transparency, and reproducibility.

***Keywords*:**

Data Accuracy, Metadata, Research, Reproducibility of Results

Introduction

In addition to the variety of hospital information systems, university hospitals and academic medical centers operate specialized systems, each one serving specific use cases in a best-of-breed manner to support research [1]. Furthermore, a hospital runs multiple software systems in order to serve medical care, generating vast amounts of heterogeneous data. Using information that is contained in healthcare and research systems requires cleansing, harmonizing, and integrating the data sets – typically referred to as extract, transform, and load (ETL) processes. Finally, an integrated data pool can be generated for further analysis. As the numbers of data sets and systems grow, for example by federating systems from multiple sites, the complexity of performing ETL related tasks grows as well. This makes the assessment of data quality a tedious task. Moreover, data reuse is oftentimes difficult and then neglected, which results in recaptured and regenerated data sets per research project [2]. Thus, creating reusable data sets represents an important focus in current research [3]. The FAIR principles require data sets to be findable, accessible, interoperable, and reproducible in order to ensure high-quality data sharing and enable data reuse [4] both for humans and machines. Fine-grained documentation of processes, transformations, and influences on an object – also known as provenance – from data sources, data integration pipelines, and data repositories can be exploited to enable data reuse and improve data quality [5,6]. Provenance is used to enhance reproducibility, transparency, presentation, meta-analysis, and machine-readability [7].

Capturing, storing, and utilizing provenance have shown to provide valuable insights within biomedical research projects [8] – technical experts inspect provenance in order to assess data quality and gain insight into data processing across multiple sites. We aim to implement a generalized toolbox focused on provenance using our experience, providing instruments to capture, store, and utilize provenance for technicians and non-technicians alike. More discipline-specific experience on how to handle provenance in federated research projects is available from the geographic information systems (GIS) [9], bio-informatics [10] or the physics domain communities [11].

We will implement and test the provenance toolbox in the context of the HiGHmed platform, which aims to implement a federated platform of medical data sets in order to improve healthcare and medical research at multiple institutions [12]. Conforming to the HiGHmed principles, the toolbox needs to be scalable, compliant to local data safety and privacy regulations, sustainable, federated, and focused on improving patient healthcare. This toolbox should be applicable by different stakeholders in medical research projects.

We propose this implementation roadmap towards a provenance system architecture describing how we intend to develop solutions to capture provenance appropriately, store provenance securely embedded into a heterogeneous IT-infrastructure, and implement the means to utilize provenance for each relevant stakeholder in medical research.

Methods

A thorough literature review was conducted as a basis for our work [13]. Additionally, surveys of Moreau [14], Herschel et al. [15] as well as Pérez et al. [16] were considered. These works present background knowledge regarding provenance research as well as several solutions that implement provenance.

We will use the provenance definition by the W3C PROV-DM specification: “provenance is defined as a record that describes the people, institutions, entities, and activities involved in producing, influencing, or delivering a piece of data or a thing.” [6] Provenance is captured in several granularities, ranging from provenance meta-data to data provenance as depicted in Figure 1 [15]. Provenance meta-data refers to a generic, hardly utilizable type of provenance,

e.g. a free-text description of a method in a research paper. Workflow and information system provenance refer to more specific types of machine-usable data, for example, a definition and execution of a workflow defined in a workflow management system. Distinguishing between these two types of provenance is not always possible. Data provenance refers to the most specific type of provenance, which is usually tailored to the precise context the data describes. Dublin Core, for example, describes several elements of the provenance of digital resources [17].

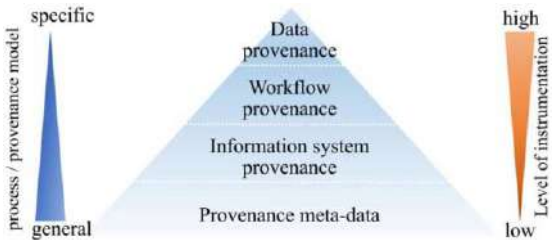


Figure 1 – Provenance Types Hierarchically Ordered by the Provenance Model and the Level of Instrumentation [15]

Capturing provenance manually is considered ineffective [18]. Workflow management systems enable capturing data based on predefined workflows, resulting in workflow-based provenance containing domain-specific insights on processing steps. Process-based (or activity-based) methods require a system or application itself to collect relevant provenance data. Operating-system-based (OS-based) methods allow to capture provenance based on OS functions, like file reads [13]. Similarly, captured provenance based on service-oriented architectures results in service-based provenance. Finally, provenance capturing in the context of relational database systems is considered as a specialized approach to collect provenance [16].

Tracing provenance is performed by two approaches: lazy and eager [16]. Lazy describes collecting provenance after the data processing took place, for example by scraping logs or reverse engineering database queries. The eager approach aims to collect provenance data during data processing (or immediately after), for example by wrapping data processing applications.

PROV-DM is a data model for provenance on the web, defined by the W3C in an effort to summarize and consolidate existing provenance models [6]. It models provenance as an acyclic directed graph, consisting of nodes and edges which represent provenance. Entities, Activities, and Agents represent objects, processes, and influences respectively.

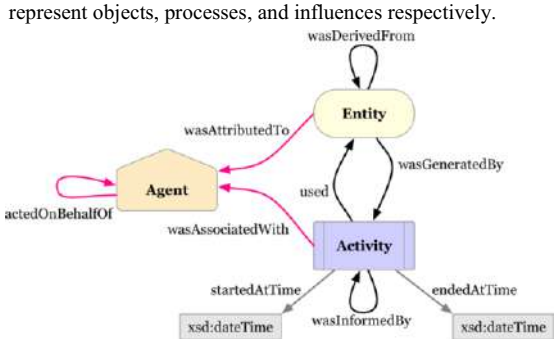


Figure 2 – W3C PROV-DM Depicted as an Acyclic Directed Graph, Showing the Three Starting Point Classes Entity, Activity, Agent in Yellow, Blue, Orange Respectively, Connected by the Relations Defined by PROV [21] and Two Activity-specific Properties are Visualized in Grey

These three nodes are connected by different relations as depicted in a general overview in Figure 2. PROV-DM is highly extensible, allowing to tailor general concepts to more specific constructs. The W3C also provides different serializations for PROV like XML [19] as well as general guidelines on how to use PROV-DM data, like access and query mechanisms [20].

Results

Using provenance from heterogeneous medical research information systems requires considering three aspects: extraction, storage, and utilization. Extraction (or capturing) of provenance is the documentation of data processing steps. Storage of provenance involves a data store, a data format and the means to query the stored provenance data. Finally, utilization generates knowledge from stored provenance data, enabling possible uses from provenance like meta-analysis or reproducibility of computing steps.

Extraction of provenance

As a first step, coordination-points – central computing systems which process high loads of data – will be tackled in order to extract provenance [22]. PROV@TOS has been implemented to capture provenance from data integration jobs based on Talend Open Studio for Data Integration [23]. In combination with version control systems, full-featured provenance documents for data integration pipelines will be extracted [24]. As a part of HiGHmed, the Medical Data Integration Center at the University Medical Center in Göttingen (UMG MeDIC) will use an IHE compliant infrastructure that will use Audit Record Repositories defined by IHE ATNA [25]. These will serve as sources to extract provenance besides data integration pipelines. Upon enabling provenance-awareness for those two coordination-points, existing IT-systems are tackled. In addition to wrapping, log-scraping or creating provenance templates [26], a concept to categorize different provenance sources regarding the quality of produced provenance data will be defined and used to prioritize the implementation of capturing techniques per system. Although the impact of different capturing techniques on provenance quality has been investigated [18], we intend to tailor a comparison to the medical domain in order to prioritize the implementation of per system provenance capturing mechanisms.

Solution to store provenance

Storing, accessing and querying provenance are made possible at a centralized provenance store [27] utilizing the standardized W3C PROV data model [6]. A neo4j NoSQL database (https://neo4j.com/) with an appropriate W3C PROV connector (https://github.com/DLR-SC/provneo4j) will serve as the provenance store. It will be a part of an IHE-compliant IT-infrastructure, hard linking provenance records to the actual data they represent [28]. User access rights are defined via the IHE Consent Profiles (BPPC/APPC) where applicable. The definition of appropriate access rights needs to consider data privacy (i.e. General Data Protection Regulation (EU) 2016/679 (GDPR)). Concepts analogous to the "Cross-Community" profiles defined by IHE will be created and implemented. As a basis, the PROV-AQ [20] mechanisms can be used and extended to fit the needs within a medical context. Furthermore, FHIR Resources for PROV (http://hl7.org/fhir/provenance.html) are available for usage, enabling captured provenance to be used across a whole data sharing community.

HiGHmed's architectural framework relies on the openEHR standard for semantic modeling of clinical information [12]. This standard provides the Feeder Audit (https://www.openehr.org/releases/RM/latest/docs/common/common.html#_feeder_system_audit) which enables linking or directly storing provenance information on different granularity levels even for single data entries where desired.

Visualization and exploration of provenance

On a per stakeholder basis, tools to access provenance records are – if no ready-to-use solutions are available – developed and evaluated, each serving the specialized needs of a stakeholder regarding provenance questions. As a starting point, three stakeholders are taken into account: domain experts (medical informaticians) that perform data analytics, data privacy officers that need to assess regulatory questions and laymen (patients) that seek informational value from provenance records. As intended, use case and domain specific knowledge varies greatly among these three stakeholders, therefore stakeholder-specific access and interaction interfaces to provenance must be designed. A stakeholder analysis needs to be performed prior to implementing the tools followed by validation of the implemented tools against the requirements previously gathered [29]. The PrIMe methodology can serve as a starting point for this task [30]. Tools visualizing such metadata to researchers preparing data analysis can benefit from the availability of standardized provenance information. In order to assist in tasks like data analysis effectively, the planned tools should be embedded into other applications (for example, tranSMART (http://transmartfoundation.org/) for data analysis). As other research fields have already implemented solutions for their respective fields (for example, DataONE [9]), an analysis of those implementations will be performed. Additionally, gaining insight from provenance records is not limited to visualizations, as network analytics are currently evaluated within the provenance research community [31]. To contribute to this research, the metrics that are defined by the authors can be put to use within controlling and monitoring systems.

Discussion

Solutions to capture, store, and use provenance data are scarce. DataONE [9] implements provenance capturing, storing, and usage to recall processing steps related to the data objects that are stored in DataONE, enabling reproducibility. Similarly, platforms like Galaxy [10] or CRISTAL [11] aim to make processing workflows reproducible and recallable. In the Software Evolution domain, the means to utilize provenance data within source code versioning solutions for domain specific needs exist, e.g. defect prediction [32] or simulation of software evolution processes [33]. Our aim will be to learn from these solutions and create the possibility to use provenance to answer domain-specific questions regarding reproducibility of results, assessing data quality, presentation of results, and meta-analysis of all processes within the context of medical research as Curcin et al. suggest [8].

Extraction of provenance

Several software solutions to capture provenance are available, most of them based on workflow management systems [16]. This makes it difficult to tailor them to the needs within the HiGHmed MeDICs. Moreover, the client-server architecture defined by the IHE profiles [25] makes the use of OS-based provenance capture mechanisms less viable. The majority of data sources are only available by proprietary software, urging us to choose an approach similar to PLUS [22]: we will focus on provenance from coordination-points first and tackle single systems second, employing PrIMe [24] to wrap or extend systems. Data privacy needs to be considered when extracting provenance from medical data sets due to GDPR. Where applicable, fine-grained domain-specific provenance data will be extracted from source systems in order to represent the whole data capturing process within the provenance store [11].

Audit logs compliant to the Good Clinical Practice (GCP) showed to yield valuable information regarding provenance [5]. Hence, the audit logs captured based on the ATNA IHE-profile specification may yield comparable data from a coordination-point. Source systems that forward audit logs compliant with ATNA could also be addressed this way. Further investigation is needed to validate this assumption, being an integral part of establishing provenance-aware medical research IT-infrastructures.

Solutions to store provenance

As W3C PROV-DM is modeled as an acyclic directed graph [6], the neo4j graph database is a primary candidate to store PROV data in contrast to commonly used relational database storage systems [8]. Neo4j provides a reliable and scalable solution to store and query graph data [34] and is also used by established provenance recording systems [22,35].

Data privacy within provenance is oftentimes regarded as the need to ensure access rights to provenance data [8,14]. Albeit an important consideration, restricting access to provenance is insufficient to meet both, data privacy regulations and interests of data analysts. Chapman et al. introduce the use of surrogates to mask sensitive data to unprivileged users [22]. This enables the use of sensitive data without violating privacy rights. Under the umbrella term "secure (data) provenance", Torra et al. summarize four requirements [36]: distribution of provenance, integrity of provenance, availability of provenance and privacy and confidentiality of provenance. Tailoring these requirements to the medical research domain will be an important part of future research and the implementation in the UMG MeDIC and other Data Integration Centers.

Provenance data is considered essential within big data applications [37]. Although medical research IT-infrastructures are not necessarily big data applications [1], some characteristics (variety, veracity and, to some degree, volume of data) remain important in data integration in translational research. Within the GIS community, distributing data with its provenance is common practice [38]. Hence, federating provenance across multiple sites in ways similar to the federation of medical data sets is an important focus which needs to be addressed when implementing data federation infrastructures. FHIR, as a standard for medical data exchange, incorporates W3C PROV and provides a starting point to focus on provenance federation.

Visualization and exploration of provenance

Utilizing provenance usually comes in two steps: visualization and analysis. Several solutions to visualize provenance exist, primarily an acyclic directed graph defined by PROV-DM [6]. Prov-O-Viz by Hoekstra and Groth is another graph-based visualization that focuses on the data flow within activities and entities [39]. Schreiber and Struminski present a solution to make provenance understandable for laymen using a representation with comics [40]. The latter solution will serve as a starting point to implement provenance visualization solutions for all stakeholders. In combination with the data

itself, provenance is able to assist visualization of medical records in a similar approach to EVLIN [41].

Due to the nature of PROV-DM, analysis of provenance – that is, computations utilizing provenance without visualizing it – are closely related to network analysis techniques. Huynh et al. tailored several metrics established within network analysis to provenance data, demonstrating the usefulness of such techniques [31]. The usage of provenance templates may add insight in combination with network analysis [26].

Provenance insight serves different purposes [7]. Tackling all purposes in a single application would result in a highly complex tool that requires not only a deep understanding of provenance itself but also of the data it describes. To circumvent this challenge, insights from research regarding data quality of medical data sets can be exploited. Kahn et al. recommend utilizing provenance to improve data quality and data reuse of observational and administrative data [42]. Also, the term "fitness for purpose" was coined to describe that every use case requires a specialized view on data sets to gain optimal performance [43]. Following this paradigm, we aim to make provenance usable for each stakeholder by selecting relevant provenance data sets and by using a specialized applications, tailored to a small set of provenance questions.

As an example, the GCP-compliant data capture software secuTrial (https://www.secutrial.com) implements a function to review audit trails within the application itself, bringing provenance directly into the context of the data itself.

Conclusions

The means to extract, store, and utilize provenance have been tackled individually by several tools within the provenance community. Solutions that grant all of these features are scarce but show great potential to improve reproducibility, recall, insight, presentation, and enable meta-analysis of processing steps within heterogeneous IT-infrastructures. We aim to learn from successful implementations in other disciplines (GIS [9], bio-informatics [10] and software evolution [32,33]), and harness this knowledge to jump start provenance applications in medical research implemented in a provenance toolbox that pools all features related to extraction, storage, and utilization of provenance. This will improve reproducibility, re-use, interoperability, and overall quality in medical research [8,44].

Acknowledgements

This work was supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the research and funding concepts of the Medical Informatics Initiative (01ZZ1802B/HiGHmed) and the BMBF e:Med program (01ZX1606C/sysINFLAME, 031L0024A/myPathSem).

References

[1] C.R. Bauer, N. Umbach, B. Baum, K. Buckow, T. Franke, R. Grütz, L. Gusky, S.Y. Nussbeck, M. Quade, S. Rey, T. Rottmann, O. Rienhoff, and U. Sax, Architecture of a Biomedical Informatics Research Data Management Pipeline, *Stud Health Technol Inform*. **228** (2016) 262–266. doi:10.3233/978-1-61499-678-1-262.

[2] C. Safran, Update on Data Reuse in Health Care, *Yearb Med Inform*. **26** (2017) 24–27. doi:10.15265/IY-2017-013.

[3] S.M. Meystre, C. Lovis, T. Bürkle, G. Tognola, A. Budrionis, and C.U. Lehmann, Clinical Data Reuse or Secondary Use: Current Status and Potential Future Progress, *Yearb Med Inform*. **26** (2017) 38–52. doi:10.15265/IY-2017-007.

[4] M.D. Wilkinson, M. Dumontier, Ij.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da S. Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data*. **3** (2016). doi:10.1038/sdata.2016.18.

[5] M. Parciak, C.R. Bauer, B. Baum, H. Kusch, and U. Sax, Technical Aspects of Data Provenance in Clinical Trials, in: 62. Jahrestagung Der Deutschen Gesellschaft Für Medizinische Informatik, Biometrie Und Epidemiologie e. V. (GMDS), German Medical Science GMS Publishing House, Oldenburg, Germany, 2017. doi:10.3205/17gmds155.

[6] L. Moreau, and P. Missier, PROV-DM: The PROV Data Model, (2013). http://www.w3.org/TR/2013/REC-prov-dm-20130430/ (accessed August 27, 2018).

[7] E.D. Ragan, A. Endert, J. Sanyal, and J. Chen, Characterizing Provenance in Visualization and Data Analysis: An Organizational Framework of Provenance Types and Purposes, *IEEE Transactions on Visualization and Computer Graphics*. **22** (2016) 31–40. doi:10.1109/TVCG.2015.2467551.

[8] V. Curcin, S. Miles, R. Danger, Y. Chen, R. Bache, and A. Taweel, Implementing interoperable provenance in biomedical research, *Futur. Gener. Comp. Syst.* **34** (2014) 1–16. doi:10.1016/j.future.2013.12.001.

[9] Y. Cao, C. Jones, V. Cuevas-Vicenttín, M.B. Jones, B. Ludäscher, T. Mcphillips, P. Missier, C. Schwalm, P. Slaughter, D. Vieglais, L. Walker, and Y. Wei, DataONE: A Data Federation with Provenance Support, in: Provenance and Annotation of Data and Processes, Springer-Verlag New York, Inc., McLean, VA, USA, 2016: pp. 230–234. doi:10.1007/978-3-319-40593-3_28.

[10] J. Goecks, A. Nekrutenko, and J. Taylor, Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences, *Genome Biology*. **11** (2010). doi:10.1186/gb-2010-11-8-r86.

[11] A. Branson, R. McClatchey, J.-M. Le Goff, and J. Shamdasani, CRISTAL: A practical study in designing systems to cope with change, *Information Systems*. **42** (2014) 139–152. doi:10.1016/j.is.2013.12.009.

[12] B. Haarbrandt, B. Schreiweis, S. Rey, U. Sax, S. Scheithauer, O. Rienhoff, P. Knaup-Gregori, U. Bavendiek, C. Dieterich, B. Brors, I. Kraus, C. Thoms, D. Jäger, V. Ellenrieder, B. Bergh, R. Yahyapour, R. Eils, H. Consortium, and M. Marschollek, HiGHmed – An Open Platform Approach to Enhance Care and Research across Institutional Boundaries, *Methods of Information in Medicine*. **57** (2018) e66–e81. doi:10.3414/ME18-02-0002.

[13] M. Parciak, Provenancekonzept für Datenbestände aus einer heterogenen Forschungsinfrastuktur (am Beispiel einer klinischen Forschergruppe), Masters Thesis, Georg-August-Universität Göttingen, 2017.

[14] L. Moreau, The Foundations for Provenance on the Web, *WEB*. **2** (2010) 99–241. doi:10.1561/1800000010.

[15] M. Herschel, R. Diestelkämper, and H.B. Lahmar, A survey on provenance: What for? What form? What from?, *The VLDB Journal*. **26** (2017) 881–906. doi:10.1007/s00778-017-0486-1.

[16] B. Pérez, J. Rubio, and C. Sáenz-Adán, A systematic review of provenance systems, *Knowl Inf Syst*. (2018) 1–49. doi:10.1007/s10115-018-1164-3.

[17] J. Kunze, and T. Baker, The Dublin Core Metadata Element Set, *RFC 5013*. (2007). doi:10.17487/RFC5013.
[18] D. Gammack, S. Scott, and A.P. Chapman, Modelling Provenance Collection Points and Their Impact on Provenance Graphs, in: Provenance and Annotation of Data and Processes, Springer, Cham, 2016: pp. 146–157. doi:10.1007/978-3-319-40593-3_12.
[19] L. Moreau, H. Hua, C. Tilmes, and S. Zednik, PROV-XML: The PROV XML Schema, (2013). http://www.w3.org/TR/2013/NOTE-prov-xml-20130430/ (accessed October 25, 2018).
[20] L. Moreau, O. Hartig, Y. Simmhan, J. Myers, T. Lebo, K. Belhajjame, S. Miles, S. Soiland-Reyes, G. Klyne, and P. Groth, PROV-AQ: Provenance Access and Query, (2013). http://www.w3.org/TR/2013/NOTE-prov-aq-20130430/ (accessed October 25, 2018).
[21] T. Lebo, S. Sahoo, and D. McGuinness, PROV-O: The PROV Ontology, (2013). http://www.w3.org/TR/2013/REC-prov-o-20130430/ (accessed October 25, 2018).
[22] A. Chapman, B.T. Blaustein, L. Seligman, and M.D. Allen, PLUS: A provenance manager for integrated information, in: 2011 IEEE International Conference on Information Reuse Integration, IEEE, Las Vegas, NV, USA, 2011: pp. 269–275. doi:10.1109/IRI.2011.6009558.
[23] M. Parciak, C.R. Bauer, R. Lodahl, C. Thoms, H. Kusch, S. Rey, and U. Sax, PROV@TOS, a Java Wrapper to capture provenance for Talend Open Studio jobs, *63. Jahrestagung Der Deutschen Gesellschaft Für Medizinische Informatik, Biometrie Und Epidemiologie e.V. (GMDS)*. (2018). doi:10.3205/18gmds096.
[24] T. Bender, C.R. Bauer, M. Parciak, R. Lodahl, and U. Sax, FAIR conform ETL processing in translational research, *63. Jahrestagung Der Deutschen Gesellschaft Für Medizinische Informatik, Biometrie Und Epidemiologie e.V. (GMDS)*. (2018). doi:10.3205/18gmds095.
[25] IHE International, Inc., Integration Profiles, IHE International, Inc., 2017. https://www.ihe.net/uploadedFiles/Documents/ITI/IHE_ITI_TF_Vol1.pdf (accessed October 25, 2018).
[26] V. Curcin, E. Fairweather, R. Danger, and D. Corrigan, Templates as a method for implementing data provenance in decision support systems, *Journal of Biomedical Informatics*. **65** (2017) 1–21. doi:10.1016/j.jbi.2016.10.022.
[27] P. Groth, J. Sheng, S. Miles, S. Munroe, V. Tan, S. Meacham, and L. Moreau, An Architecture for Provenance Systems, University of Southampton, 2006.
[28] D. Koop, E. Santos, B. Bauer, M. Troyer, J. Freire, and C.T. Silva, Bridging Workflow and Data Provenance Using Strong Links, in: Scientific and Statistical Database Management, Springer, Berlin, Heidelberg, 2010: pp. 397–415. doi:10.1007/978-3-642-13818-8_28.
[29] R. Bjugn, and B. Casati, Stakeholder Analysis: A Useful Tool for Biobank Planning, *Biopreservation and Biobanking*. **10** (2012) 239–244. doi:10.1089/bio.2011.0047.
[30] S. Miles, P. Groth, S. Munroe, and L. Moreau, PrIMe: A Methodology for Developing Provenance-aware Applications, *ACM Transactions on Software Engineering and Methodology*. **20** (2011) 8:1–8:42. doi:10.1145/2000791.2000792.
[31] T.D. Huynh, M. Ebden, J. Fischer, S. Roberts, and L. Moreau, Provenance Network Analytics, *Data Min Knowl Disc*. (2018) 1–28. doi:10.1007/s10618-017-0549-3.
[32] S. Herbold, A. Trautsch, and J. Grabowski, Global vs. local models for cross-project defect prediction, *Empir Software Eng*. **22** (2017) 1866–1902. doi:10.1007/s10664-016-9468-y.
[33] V. Honsel, D. Honsel, S. Herbold, J. Grabowski, and S. Waack, Mining Software Dependency Networks for Agent-Based Simulation of Software Evolution, in: 2015 30th IEEE/ACM International Conference on Automated Software Engineering Workshop (ASEW), 2015: pp. 102–108. doi:10.1109/ASEW.2015.9.
[34] D. Dominguez-Sal, P. Urbón-Bayes, A. Giménez-Vañó, S. Gómez-Villamor, N. Martínez-Bazán, and J.L. Larriba-Pey, Survey of Graph Database Performance on the HPC Scalable Graph Analysis Benchmark, in: H.T. Shen, J. Pei, M.T. Özsu, L. Zou, J. Lu, T.-W. Ling, G. Yu, Y. Zhuang, and J. Shao (Eds.), Web-Age Information Management, Springer Berlin Heidelberg, 2010: pp. 37–48. doi:10.1007/978-3-642-16720-1_4.
[35] A. Gehani, and D. Tariq, SPADE: support for provenance auditing in distributed environments, in: Middleware 2012, Springer, Berlin, Heidelberg, Montreal, Quebec, Kanada, 2012: pp. 101–120. doi:10.1007/978-3-642-35170-9_6.
[36] V. Torra, G. Navarro-Arribas, D. Sanchez-Charles, and V. Muntés-Mulero, Provenance and Privacy, in: V. Torra, Y. Narukawa, A. Honda, and S. Inoue (Eds.), Modeling Decisions for Artificial Intelligence, Springer International Publishing, 2017: pp. 3–11.
[37] B. Glavic, Big Data Provenance: Challenges and Implications for Benchmarking, in: Specifying Big Data Benchmarks, Springer, Berlin, Heidelberg, 2014: pp. 72–80. doi:10.1007/978-3-642-53974-9_7.
[38] G. Closa, J. Masó, B. Proß, and X. Pons, W3C PROV to describe provenance at the dataset, feature and attribute levels in a distributed environment, *Computers, Environment and Urban Systems*. **64** (2017) 103–117. doi:10.1016/j.compenvurbsys.2017.01.008.
[39] R. Hoekstra, and P. Groth, PROV-O-Viz - Understanding the Role of Activities in Provenance, in: Provenance and Annotation of Data and Processes, Springer-Verlag New York, Inc., Cologne, Germany, 2015: pp. 215–220. doi:10.1007/978-3-319-16462-5_18.
[40] A. Schreiber, and R. Struminski, Visualizing Provenance using Comics, in: 9th USENIX Workshop on the Theory and Practice of Provenance, USENIX Association, Seattle, WA, USA, 2017. https://www.usenix.org/conference/tapp17/workshop-program/presentation/schreiber.
[41] H.B. Lahmar, M. Herschel, M. Blumenschein, and D.A. Keim, Provenance-Based Visual Data Exploration with EVLIN., in: Proceedings of the 21th International Conference on Extending Database Technology, EDBT 2018, Vienna, Austria, March 26-29, 2018., 2018: pp. 686–689. doi:10.5441/002/edbt.2018.85.
[42] M.G. Kahn, J.S. Brown, A.T. Chun, B.N. Davidson, D. Meeker, P.B. Ryan, L.M. Schilling, N.G. Weiskopf, A.E. Williams, and M.N. Zozus, Transparent Reporting of Data Quality in Distributed Data Networks, EGEMs (Generating Evidence & Methods to Improve Patient Outcomes). 3 (2015) 7. doi:10.13063/2327-9214.1052.
[43] N.G. Weiskopf, and C. Weng, Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research, *J Am Med Inform Assoc*. **20** (2013) 144–151. doi:10.1136/amiajnl-2011-000681.
[44] B. Baum, C.R. Bauer, T. Franke, H. Kusch, M. Parciak, T. Rottmann, N. Umbach, and U. Sax, Opinion paper: Data provenance challenges in biomedical research, *It - Information Technology*. **59** (2017) 191–196. doi:10.1515/itit-2016-0031.

Address for correspondence

Corresponding author: marcel.parciak@med.uni-goettingen.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190232

What a Comprehensive, Integrated Data Strategy Looks Like: The Population Level Analysis and Reporting (POLAR) Program

Christopher Pearce[a,b], Adam McLeod[a], Natalie Rinehart[a], Jason Ferrigi[a], Marianne Shearer[c]

[a] *Outcome Health, East Burwood, Victoria, Australia,*
[b] *Department of General Practice, Monash University, Victoria, Australia*
[c] *Gippsland Primary Health Network, Victoria, Australia*

Abstract

In Australia, general practice (GP) acts as the gatekeeper to the rest of the healthcare system, and therefore the vast majority of the population have an electronic medical record. It follows that the largest database of the population is therefore on the distributed GP computers. Informed by a comprehensive system-wide data strategy, the Population Level Analysis and Reporting program extracts data from the GP electronic medical records and repurposes it for multiple uses. The program requires the data to be coded and then structured for multiple uses clinical care, clinical governance, research, and policy.

Keywords:

Primary Health Care, Electronic Health Records, Medical Record Linkage.

Introduction

With 85% of the population seeing a general pratice (GP) at least once a year [2], GP currently contains the most comprehensive database of health data in Australia. Hospital systems contain largely episodic data, and national projects contain limited project-specific data. For instance, the national schemes Medicare (the national health insurance scheme) and Pharmaceutical Benefits Schedules (subsidised medicines), contain only data collected for administrative purposes. Some organisations rely on registries – purpose-built data collections around specific diseases – and often based on hospital settings. Others use small, practice-based research networks.

The described project occurs in the context of a worldwide move to use the data from electronic medical records for research [4; 5], although significant barriers apply, including data extraction [9], and data quality.

Australia has meso-level organisations designed to assist general practice to deliver population-based care. Formerly called GP networks [15], these Primary Health Networks (PHNs) fulfil a critical role as the chosen delivery platform for elements of key health reforms and initiatives and are now embedded in the primary health care landscape as drivers of quality, efficiency, coordination and improved access to health services. PHNs have adopted a comprehensive program to supply their general practices with a multi-faceted approach to help them provide patient centred care, improve their data quality, and take a population-based approach to analytics, based on common principles.

The Population Level Analysis and Reporting (POLAR) program and its research arm, POLAR Data Space, represent an attempt to bridge the power of the former with the needs of the latter, by creating a useful environment to understand the health needs of the population using Australian GP derived data. POLAR is a platform designed primarily to improve quality of care for the Australian population. It does so at several levels – acknowledging a data hierarchy [14] designed to take the use of health data beyond the term 'secondary use of data' – to a position where all uses of data are important. Just because a piece of information is recorded for the care of a patient, does not mean uses of it beyond that are 'secondary'. Indeed, the most significant impact of the data could be from the use of data beyond the individual clinical record.

The program is informed by the principles of an 'organisational wide data quality management program' [8] which expands the data hierarchy to be:

- Identification of the patient
- Clinical care
- Coordinate and integrate services (clinical governance)
- Care of populations
- Research, evaluation, and monitoring of safety and quality
- Policy and strategy
- Administration and logistics

The objective of the POLAR program is to construct a data quality platform that allows data to be 'fit for purpose' across all of these possible uses.

Lacking often in consideration of a data strategy is a theoretical basis of the interaction of data with the business of health. While society has been extensively altered by the arrival of the Internet and the digitization of society – creating the term 'Digital Revolution' to mirror the 'Industrial Revolution' – its effects have been delayed in health, in part because health remains heavily reliant on human-to-human interactions.

Relatively under-theorised, the patient-doctor relationship has been characterised using grounded theory [1], and complexity theory [10]. More recently, the patient-doctor-computer relationship has been described using the frameworks of both dramaturgy [20] and the work of Habermas [11]. These frameworks were created at a time when the potential of data to alter the interaction was not fully appreciated. Early work focused on the role of the computer as an agent, whereas increasingly the computer is a conduit for data to influence care. The focus now needs to be on data as the active agent. Data from multiple sources now influences the patient in their healthcare interactions. In developing this integrated strategy, we looked to extend the use of Habermas.

Habermas framed the world according to the actions of the *lifeworld* and the *system* [6; 7]. *Lifeworld* represents the individual and cumulative actions of individuals – through what is known as communicative action. As the actions of individuals, communicative action represents their expression as aspects of personality, culture and even society. Importantly in our context, it is the result of communication that creates this.

By contrast, *system*, and the concomitant *strategic action* represents the actions of the created society on the individual. This creates for us a tension when considering a framework to examine data – for data is generated by the micro-interactions that represent communicative action, yet is also utilised by the system, and represents a means by which stratifications can be implemented. This is big data at work.

By example – a consultation (or other healthcare interactions) represents communicative action based on of the *lifeworld*. In past work, this would be understood as understanding the whole person in developing a plan [17]. Now, however, such an interaction may be influenced by the actions of the computer – which may be the conduit for the strategic action of the system – recommending specific actions based on national initiatives and allowing for the passage of information to inform the system [12]. The conduit for this activity is the data codes generated by the system.

So in describing the data hierarchy below, in effect it represents a ladder from communicative to strategic influence – from changing the communication between individuals though working between individuals and the system, and to the system itself. This creates the data as a mediator between the two worlds, and the organisation wide governance framework is a means of making that work.

Methods

At the initial level, GPs (and practice nurses, other practice staff) collect and record data in their Clinical Information System (CIS) primarily for the care of an individual patient. Thus a blood pressure, test result or social history recording helps an individual GP with an individual patient.

Data in the systems is collected in a variety of formats. Within the record, there is both structured and unstructured data. Unstructured data can be found in the clinical notes section, in incoming and outgoing correspondence, and other documents such as discharge summaries. At the time of writing, because of the potential of identifiability, unstructured narrative data is not collected.

Next is structured data – which may or may not be coded. Structured data includes elements such as diagnoses, medications, and measurements. This data is extracted and processed, de-identified, and placed in a data repository (see figure 1).

The POLAR extraction tool (called Hummingbird) extracts data into the POLAR database and then de-identified data is sent outside the practice to the POLAR data warehouse, where it is framed and processed for the various uses – including feedback to the practice.

At POLAR we now have a multi-layered approach to cleaning and rationalising data to be used at multiple levels. Medication data is organized according to the World Health Organizationav's Anatomic and Therapeutic Coding (ATC) system. Diagnoses are extracted and undergo an automated process to apply a SNOMED-CT-AU code. Use of this process allows a code to be applied to 95% of data extracted from the diagnosis field. Most of the remnants are not diagnoses, but either administrative or other notes recorded in the diagnosis section. Once the diagnoses are coded, there are further overarching groups created – all diabetes codes into a single diabetes category. Key chronic disease groups are utilised as a qualifier as well.

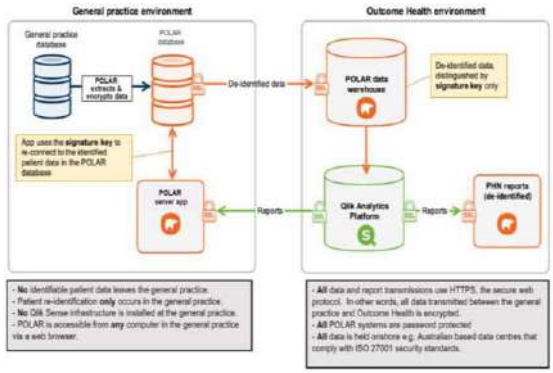


Figure 1– Extraction Environment

Pathology codes are recorded in the system generally using LOINC codes, but again there are many variations across the laboratories and inconsistent naming conventions. We have gone through the extractions and grouped them according to clinical utility (hepatitis testing, etc.) [18]. A similar process has been used for radiology investigations, where no current coding schema is available.

Governance

Governance is just as important as technical ability. Outcome Health provides the POLAR program to the PHNs, and the PHNs provide the service to the practice. As part of accreditation requirements, practices must inform patients of the potential uses of their data.

Ethics approval has been granted separately for the collection and storage of data for the POLAR program, and also for linkage activities. The ethics approval is not project-based, but standing ethics for the underlying processes.

Outcome Health then acts as the data custodian. Various groups then oversee the uses of the data. The research governance group consists of the representative PHN CEOs and reviews applications for research using specific de-identified data fields. The POLAR Data Governance Committee oversees the program as a whole and is made up of a range of internal and external experts. This group is guided by a Data Governance Framework through regular meetings to ensure that the data and processes are ethical, secure and beneficial for all stakeholders.

Other working groups that oversee the presentation of the data to the practices and the PHNs help ti ensure that the tools provided meet the needs of the end users.

Results

Using the data hierarchy described above, the data is repurposed for use at multiple levels. Importantly, this is not secondary use of data – each use of the data is as valid as any other.

Identification of the patient

At the core of the program is the necessary data to identify the individual concerned. At the practice level this includes basic demographics such as: date of birth, sex, address, government derived individual health identifier, etc. However, one of the

principles of POLAR is that no identifying information leaves the practice, for privacy and security issues. Therefore, Outcome Health has created a linkage process, called ORCA. For any given patient record ORCA creates 3 Hash keys based upon a combination of identifiable data. ORCA uses both AES encryption and SHA256 hashing combined with a salting process to protect the integrity of the hash key.

The linkage process allows patient data to leave the practice, but if necessary be returned to the practice and be re-identified during analytics conducted by the general practice staff. For example, a practice may want to identify all their patients with an active diagnosis of diabetes to ensure they are all participating in a Diabetes Cycle of Care.

In addition, through the use of the HASH keys, de-identified patient data can then be linked across practices and with other sources of data. This mechanism allows for innovative research projects across other data sets such as hospital, mental health, etc. to trace and better understand patient journeys through the health system.

Clinical care

The reason the data is collected is to provide a longitudinal view of the patient for all the treating clinicians in the practice. This EMR began as an electronic representation or reflection of the paper record – but has evolved beyond that. Digital health data has allowed the system to more efficiently manage clinical/administrative items such as recalls, tracing of results chronic disease management and many other items. [19] It also has made for more complete information on communications such as referrals [3].

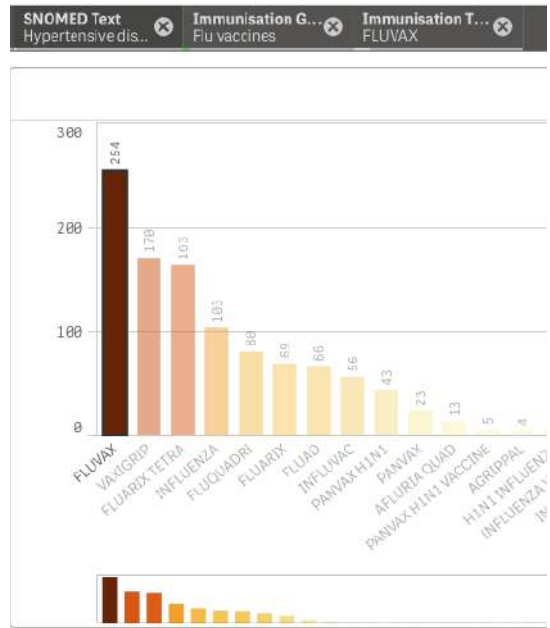
Coordinate and integrate clinical services (Clinical Governance)

Within the practice, the coding and classification of data allow GPs to easily identify cohorts of patients. POLAR then extracts the data from the GP system for clinical governance – allowing GPs to look at their practice population – how many patients have untreated blood pressure, or abnormal test results. Alternatively, sophisticated calculations such as CHADS2-VASC scores. This allows individual GPs to monitor their populations and ensure consistency of care. System-wide recalls (such as adverse reactions to new drugs, or ensuring over 65s receive an influenza vaccination) are easily handled in batches.

Care of Populations (Population Health)

De-identified data is presented at the Primary Health Network level to allow the PHN to look at it at the population level – both in clinical and geographical measures. This becomes crucial to allow both population planning and also benchmarking back to the practices. Such programs have been in place for many years [15], but the level of sophistication has improved with the creation of the organised and coded values. It is essential for PHNs to understand what issues are crucial and in what areas to best leverage and administer their health program funding to improve the health system and the health outcomes for their communities.

Figure 2, for instance, can be replicated at the PHN level, giving valuable information on individual practice performance (for feedback and development initiatives) or for the PHN to design population-based interventions.



ALL IMMUNISATIONS BY PATIENT & TYPE

Patient ID	Full Name	Immunisatio
327	Abanades, Abdellali	FLUVAX
10289	Abar, Abdelmalek	FLUVAX
30235	Abaygar, Abdenabi	FLUVAX
30235	Abaygar, Abdenabi	FLUVAX

Figure 2– All hypertensive patients immunisation status

Research, evaluation, and monitoring of safety and quality

One feature of the program is POLAR Data Space, essentially a governance framework involving the PHNs that allows for the data to be used for research. With the focus of the PHNs on making a difference in primary care – the emphasis is on collaborative research with an intent to create practical outcomes through better primary care but also better-integrated care. Quality data can inform many different modes of research, from traditional descriptive methods through to advanced analytics.

Descriptive studies currently being undertaken or completed include:

- Analysis of after hours-presentations [22]
- Prescribing and antibiotic patterns [23]
- Cardiovascular screening in musculoskeletal disorders. [21]

More advanced research involves using the GP data for real-time monitoring of immunisation adverse events, with a trial project underway. Even mora advanced is using machine learning/artificial intelligence to provide advice at the GP consultation on the risk of emergency deparment (ED) attendance in the next 30 days. Using both linked data (mapping the GP journeys of 5 years worth of patients who attended local ED's) and using machine learning on the GP data [16], this tool can accurately predict the risk 75% of the time [13].

Policy and Strategy

Good data is what should inform strategy and policy. It is impossible to plan if you do not know the landscape in which you are planning. The completeness of the data, its longitudinal nature, and the ever-increasing degree of linkage provide the PHNs significant opportunities to plan service delivery in their local areas, as well as influence policy at a state and national level.

Uniquely, the involvement of the PHNs as change agents allows the finding of the research groups to be rapidly implemented into practice, where appropriate, therefore closing the circle of learning. Examples of these are shown in figures 3, 4 and 5

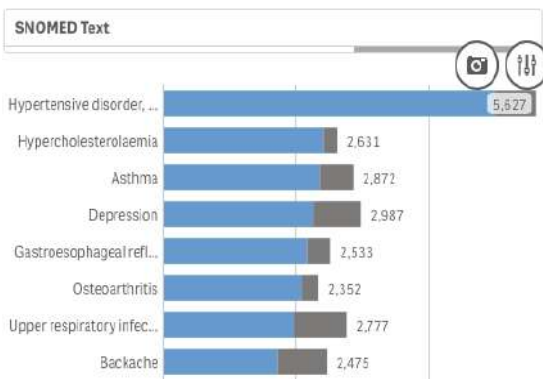


Figure 3– Top 8 diagnoses of active patients who have attended more than 12 times in a year.

Figure 4– Same patient cohort as figure 3, but counting the number of active diagnoses per patient.

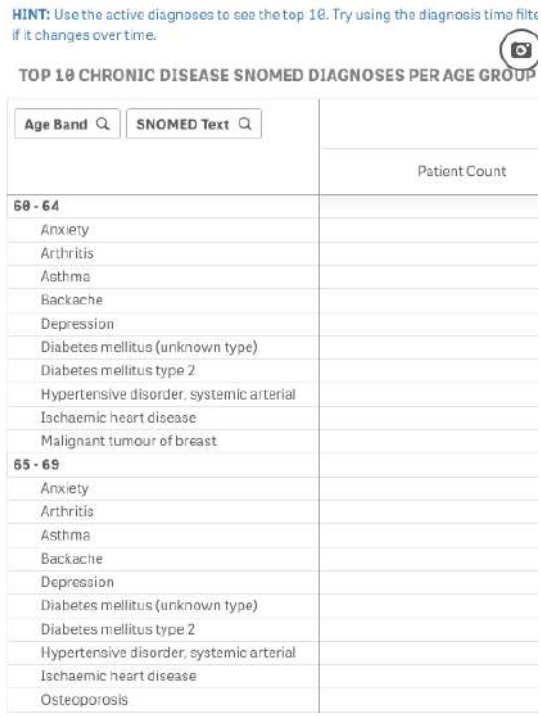


Figure 5– Example of PHN level data visualisation: Top 10 Active disease groups in specific age groups

Administration and Logistics

Within the practice, the data can be used for business and strategic planning to monitor appropriate billing for future service availability, or ensure clinician rostering is appropriate to meet the needs of their patient cohort, address waiting times or anticipate peaks in demand.

For the PHN, the de-identified data enables regional and local level planning to achieve better whole of system care. PHNs need data on health issues in order to understand the needs of communities, target and invest in services to address those needs, prioritise health system improvements, or evaluate performance and outcomes. With their emphasis on making a measurable difference, logistical analytics informs the journey of care so that fragmentation can be reduced by measuring integration and coordination of health services.

Discussion

That which is readable by humans is not by a computer, and the challenge here is to make the data easily exchangeable by the system. Data in an electronic medical record exists in many forms. Free text is the free-flowing writing that exists in many areas such as narrative notes, and many individual data types may be mixed. Advanced computing techniques such as Natural Language Processing (NLP) are required to interpret these notes. Next comes structured text – in which the free text is at least placed in a recognisable area. Writing in a designated 'diagnosis' field fits into this category. Finally comes coded text – in which a specific term has been used to link to a computer readable code.

In order to make the data useable beyond the GP environment, POLAR has chosen to concentrate on the structured text that exists in the system. Narrative data contains much identifying information, and the commitment is that no identifying

information shall leave the practice. Focusing on the structured data, we first concentrated on the diagnosis fields. Using a combination of automated and manual processing, we have been able to apply SNOMED-CT-AU codes to the vast majority of diagnoses. Once coded, we have been able to group relevant diagnoses into clinically meaningful areas. So all instances of diabetes, for instance – or a broader classification of chronic diseases. This makes manipulation of data for clinical governance and population health all the more manageable through the business intelligence analytics dashboards provided to general practice and PHNs for analysis.

Medications have been mapped to the Anatomic and Therapeutic Classifications (ATC) system, which applies five different levels, from 'all cardiovascular' down to individual drugs. This grouping, again, makes manipulation and interpretation of medications much simpler. A similar process is underway for pathology and radiology testing.

Conclusions

The key to making digital data understandable by the system, and useful for multiple purposes, is a multi-pronged strategy of coding, grouping and iterative analysis, underpinned by robust governance structure and a clear strategy.

Acknowledgements

We acknowledge the support and activity of the PHN's involved: Eastern Melbourne, Gippsland, South East Melbourne, Central East Sydney, South West Sydney.

References

[1] A. Als, The desk-top computer as a magic box: patterns of behaviour connected with the desk-top computer; GPs' and patients' perceptions, *Fam Pract* **14** (1997), 17-23.
[2] J. Beilby and J. Furler, General Practitioner Services In Australia, in: *General Practice In Australia: 2004,* Australian Department Of Health and Ageing, ed., Australian Government Publishing Service, Canberra, 2005.
[3] Crampton N, Reis S, and Shachak A, Computers in the clinical encounter: a scoping review and thematic analysis, *J Am Med Inform Assoc* **23** (2016), 654-665.
[4] S. de Lusignan and C. Pearce, What are the informatics barriers to conducting international research using routinely collected primary care data?, in: *Medinfo 2010*, Cape Town, South Africa, 2010.
[5] S. de Lusignan and C. van Weel, The use of routinely collected computer data for research in primary care: opportunities and challenges, *Fam Pract* **23** (2006), 253-263.
[6] Habermas, *The theory of communicative action, v 1: Reason and the rationalisation of society*, Beacon Press, Boston, 1984.
[7] Habermas, *The Theory of Communicative Action v2; Lifeworld and System: A Critique of Functionalist Reason.*, Polity Press, Cambridge, 1987.
[8] S.T. Liaw, C. Pearce, H. Liyanage, G.S. Liaw, and S. de Lusignan, An integrated organisation-wide data quality management and information governance framework: theoretical underpinnings, *Inform Prim Care* **21** (2014), 199-206.
[9] S.T. Liaw, J. Taggart, H. Yu, and S. de Lusignan, Data extraction from electronic health records - existing tools may be unreliable and potentially unsafe, *Aust Fam Physician* **42** (2013), 820-823.
[10] C.M. Martin and J.P. Sturmberg, General practice--chaos, complexity and innovation, *Med J Aust* **183** (2005), 106-109.
[11] C. Pearce, Computers, Patients and Doctors - Theoretical and Practical Perspectives, in: *Health Professionals Education in the Time of Clinical Information Systems,* A. Shachak, E. Borycki, and S. Reis, eds., Academoic Press, London, 2017, pp. 5-22.
[12] C. Pearce, M. Arnold, C. Phillips, S. Trumble, and K. Dwan, The patient and the computer in the primary care consultation, *Journal of the American Medical Informatics Association : JAMIA* **18** (2011), 138-142.
[13] C. Pearce, A. McLeod, N. Rinehart, J. Patrick, A. Fragkoudi, J. Ferrigi, E. Deveny, R. Whyte, and M. Shearer, POLAR Diversion: Using General Practice Data to Calculate Risk of Emergency Department Presentation at the Time of Consultation, *Appl Clin Inform* **10** (2019), 151-157.
[14] C. Pearce, M. Shearer, K. Gardner, and J. Kelly, A division's worth of data, *Aust Fam Physician* **40** (2011), 167-170.
[15] C. Pearce, M. Shearer, K. Gardner, J. Kelly, and T.B. Xu, GP Networks as enablers of quality of care: implementing a practice engagement framework in a General Practice Network, *Australian Journal of Primary HEalth* **18** (2012), 101-104.
[16] C.M. Pearce, A. McLeod, J. Patrick, D. Boyle, M. Shearer, P. Eustace, and M.C. Pearce, Using Patient Flow Information to Determine Risk of Hospital Presentation: Protocol for a Proof-of-Concept Study, *JMIR Res Protoc* **5** (2016), e241.
[17] S.J. Potter and J.B. McKinlay, From a relationship to encounter: an examination of longitudinal and lateral dimensions in the doctor-patient relationship, *Soc Sci Med* **61** (2005), 465-479.
[18] G. Sezgin, A. Georgiou, R.A. Hardie, L. Li, L.G. Pont, T. Badrick, G.S. Franco, J.I. Westbrook, N. Rinehart, A. McLeod, C. Pearce, M. Shearer, R. Whyte, and E. Deveny, Compliance with pathology testing guidelines in Australian general practice: protocol for a secondary analysis of electronic health record data, *BMJ Open* **8** (2018), e024223.
[19] D. Swinglehurst, T. Greenhalgh, and C. Roberts, Computer templates in chronic disease management: ethnographic case study in general practice, *BMJ Open* **2** (2012).
[20] D. Swinglehurst, C. Roberts, S. Li, O. Weber, and P. Singy, Beyond the 'dyad': a qualitative re-evaluation of the changing clinical consultation, *BMJ Open* **4** (2014), e006017.
[21] L.R. Turner, F. Cicuttini, C. Pearce, and D. Mazza, Cardiovascular disease screening in general practice: General practitioner recording of common risk factors, *Prev Med* **99** (2017), 282-285.
[22] L.R. Turner, C. Pearce, M. Borg, A. McLeod, M. Shearer, and D. Mazza, Characteristics of patients presenting to an after-hours clinic: results of a MAGNET analysis, *Aust J Prim Health* **23** (2017), 294-299.
[23] J. Yan, L. Hawes, L. Turner, D. Mazza, C. Pearce, and J. Buttery, Antimicrobial prescribing for children in primary care, *J Paediatr Child Health* (2018).

Address for correspondence

Dr. Christopher Pearce, drchrispearce@mac.com.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190233

Comparative Analysis of Topical Evolution Patterns and Temporal Trends of Hypertension Research

Yuxing Qian[a], Liqin Zhou[a], Rui Zhang[a], Zhiyuan Li[a], Chun Yan[a]

[a] *Center for Studies of Information Resources, Wuhan University, Wuhan, Hubei, China*

Abstract

Exploring the topical evolution patterns and temporal trends of hypertension can promote knowledge communication among experts, and is of great significance to understand the profile and frontiers of chronic disease. Current popular topic detection mainly focuses on two directions: one is based on social network analysis (SNA), the other is based on the topic models. Aiming at distinguishing their similarities and differences, this paper adopts the community detection method and expanded topic model Dirichlet-multinomial regression (DMR) respectively to detect the topic distribution and evolution trends of hypertension research. A total of 26,717 articles in the PubMed database were used as examples to construct the MeSH Terms co-occurrence matrix. It is found that hypertension literature is mainly concentrated on three communities and five research topics. MeSH Terms obtained from SNA are more specific and clearer, while the DMR has an advantage in exploring the evolution patterns of various themes.

Keywords:

Hypertension, Medical Informatics

Introduction

Hypertension is one of the most common chronic diseases and the main risk for cardio-cerebrovascular disease. Some 9.4 million people die of hypertension worldwide every year, in which the complications (such as stroke and heart disease) can lead to about half deaths. Thus, an increasing number of scholars have begun to devote themselves to the study of hypertension. As of May 12, 2017, 284322 articles related to hypertension were retrieved in the PubMed database. However, such large-scale medical literature has caused great trouble for researchers. It is difficult to present a comprehensive overview of the profile and frontiers of Hypertension. As the bibliometric method was widely used to analyze the medical literature, some scholars [1-3] have concentrated on the macro level (including major research countries, institutions, authors, etc.) of hypertension research, without in-depth analyzing their topic distribution and evolution patterns at the medium or micro levels.

Existing literature mainly focused on two directions to detect the topic distribution and evolution patterns: one is based on social network analysis (SNA), the other is based on the topic model [4]. The SNA is widely used in data mining, knowledge management, information dissemination, and knowledge network. It can show the relationship of subject terms clearly and visually and can provide effective support for analyzing the location and association of topic words in the network. M. L. Wallace et al.[5] utilized two case studies to demonstrate that the SNA method (such as community detection) helps in identifying research directions, and can reveal more structural details of the knowledge domain than traditional co-citation analysis.

Meanwhile, the topic model, which was widely used in natural language processing [6], information retrieval, text mining, and other fields, is based on the probability and statistics model in the machine learning field. D. Mimno and A. McCallum [7] proposed the Dirichlet-multinomial Regression (DMR) model extended and derived from the LDA model. M. Song et al. [8] utilized the DMR model to detect the topic distribution and evolution patterns of Alzheimer's disease and achieved good results. Compared with the co-word analysis and citation analysis, the topic model can better reflect the relationship between "words-topics-documents." It has great advantages in topic detection.

The purpose of this study is to detect the topic distribution and evolution patterns of hypertension research and distinguish the similarities and differences among topics obtained by the SNA method and Dirichlet-multinomial regression topic model. To illustrate the process of the methods, 26717 articles related to Hypertension in PubMed database are taken as an example to construct the co-occurrence knowledge network.

Methods

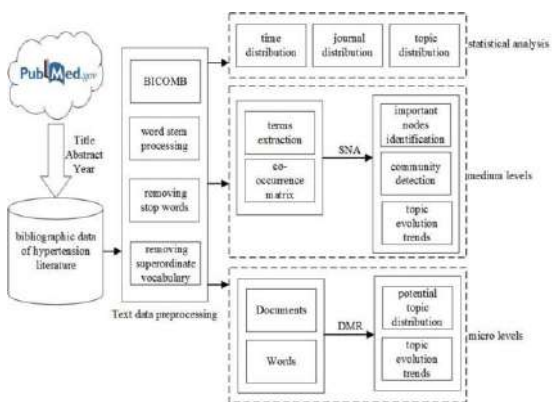


Figure 1– Research Framework

As shown in figure1, the process of our research includes data collection and preprocessing, basic statistical analysis, topic, and community detection by SNA and DMR, comparison of the analysis results.

Data Collection and Processing

The search strategy "Hypertension[MeSH Terms] AND ("2000/1/1"[PDat]: " 2017/5/1")" is used to retrieve the literature of hypertension in the PubMed database. 99252 articles were retrieved. After removing unrelated literature, we finally obtained 26717 articles containing both abstracts and full texts. Then this data was imported into the Bibliographic Items Co-occurrence Matrix Builder (BICOMB) to generate a co-occurrence matrix of bibliographic data. We can find that the 26717 articles were distributed in 1701 journals, involving 171637 authors and 9978 keywords through the extraction and statistics of journals, authors, and keywords.

In order to determine the high-frequency MeSH Terms, the frequency threshold of the subject words is obtained as 77 according to the formula of high and low-frequency word boundary [9]. The high-frequency MeSH subject terms are obtained from the bibliographic data, and the word co-occurrence matrix is constructed in the BICOMB.

Community Detection Based on Optimized Network Modularity

Social network refers to the collection of social actors and their relationships, mainly developed to measure, analyze, and predict the structure and attributes of relationships between various entities in the network [10]. In the academic literature, scholars always use the same or similar words to express the hot-spots in a certain field. The relationship between massive text data can be linked through the subject words of the text to form a huge network. The community is a common phenomenon in social networks composed of a group of highly aggregated and closely connected nodes. Nodes belonging to the same community are more likely to have similar functionality, and community structures can indicate the relationship between network structure and functionality. The most representative community identification algorithm is the optimized network modularity method proposed by M. E. J. Newman [11]. Module degree is an index that can measure the quality of network partitioning, also called Q value. In essence, the modularity-based algorithm performs community identification based on changes in the intermediaries and modularity of the edges.

Since the nodes in the word co-occurrence network are the topic words, the process of determining the community representative topics transformed into the process of finding the core nodes. A few core nodes represent the scientific research themes corresponding to the community. In complex networks, there are many important indicators of nodes such as centrality and PageRank values. These indicators consider the calculation of the number of edges, centrality, and connections with other nodes to determine the core nodes from the global level of the network.

Dirichlet-multinomial Regression Topic Model

The Dirichlet-multinomial regression topic model, as shown in figure 2, mainly obtains the distribution of topics under different conditions by adjusting the characteristics of the observed documents. This paper takes the time of publication of hypertension literature as a variable to explore the trend of the topic over time.

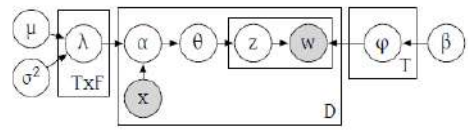


Figure 2– DMR topic model

In the document set D, for each document d, x_d represents the feature vector of the metadata, α is a function of the observable document feature, represents the prior probability distribution of the subject; given a prior probability distribution $\mathrm{N}(0,\Sigma)$, hyperparameters β , documents and words are generated as follows:

For each topic t, draw $\phi_t \sim Dir(\beta)$. $Dir(\beta)$ is a different topic-word distribution from the previous Dirichlet distribution.

For each document d, draw $\theta_d \sim Dir(\alpha_d) = \mathrm{Dir}(\exp(\tau_d))$, $\tau_d \in \tau$. For α_d of each document, the parameter of Dirichlet distribution and τ_d is covariance functions $\mathrm{f}(y_d, x_k)$, where y_d is the observed attribute vector of documents, and x_k is the vector of metadata.

For each word w, draw $z_{d,w} \sim Multi(\theta_d)$. $z_{d,w}$ is the subject allocation of the word t_w and θ_d is the proportion of the document d belonging to a certain topic, draw $T_{d,w} \sim Multi(\phi_{z_{d,w}})$. $T_{d,w}$ is the w-th word in document d, ϕ_t is the preference of topic t, $\Sigma_n \phi_{t,n} = 1$.

In the DMR topic model, three fixed parameters are set including the variance of the previously distributed parameter values σ^2, Dirichlet topic-word distribution β and the number of topics |T|.

Results

Topic Detection and Evolution Trend Analysis Based on SNA

In order to achieve better visualization, the nodes with the frequency no higher than 77, and the isolated nodes are deleted. 632 nodes are obtained to construct a co-occurrence matrix of high-frequency words. To reduce its complexity, the top 100 nodes are selected, and 4950 edges are included. A node is a biological entity derived from articles. The edge represents the relationship between the entities. The weight of the edge indicates the frequency at which the two entities co-occur in the specific sentence of the article. The data is imported into Gephi, and the community detection algorithm [12] is used for visualization, as shown in Figure 3.

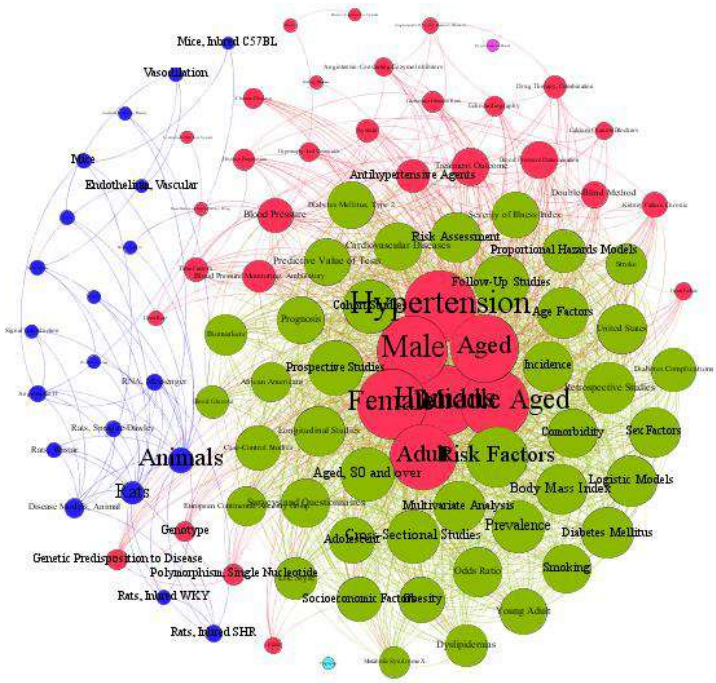


Figure 3–community detection (SNA)

To identify the most important biological entities in the literature of hypertension, four well-known centrality indicators including PageRank value, weighted centrality, closeness centrality, and betweenness centrality are calculated. Detailed centrality measurements are made by S. Wasserman [13] and S. Brin [14]. The top-ranked biological entities with PageRank values are similar to the top entities with weighted centrality rankings. Moreover, PageRank has unique biological entities, such as rates that only appear in the PageRank values top 10. The risk factor, prospective studies, and follow-up studies, only appear in the weighted centrality top 10, while Blood Pressure, Animals, and Body Weight only appear in the betweenness centrality top 10. The top 10 words in closeness centrality are exactly the same as the weighted centrality. In general, the nodes that are at a critical position in the entire network are mainly Hypertension, Male, Female, Age, Adult, Human, Risk factors, Body Weight, Blood Pressure, Animals, Prospective Studies, etc. These nodes are also at the core of the community distribution map.

The modular algorithm proposed by V. D. Blondel [12] is used for community detection and the resolution is set to 1 [15]. 5 modules are detected including 3 main modules (shown in red, blue, and green in Figure 3). The modular value Q is 0.187. The average path length is 2.645, and the diameter is 6. Choosing the Fruchterman Reingold layout, the two communities of light blue and purple in the picture account for only 1% of the total communities and are negligible. Therefore, communities of red, blue and green are mainly considered. The largest green community accounts for 42% of the network, including Risk Factors, Aged, Prevalence, Sex Factors, Prospective Studies, Follow-Up Studies. The second largest community (red part) accounts for 36% of the network, including Hypertension, humans, female, middle-aged, Treatment Outcome, Antihypertensive Agents, etc. The third largest community (blue part) accounts for 20% of the network, including animals, Rats, Inbred SHR, Messenger, Mice, etc. Through manual judgment and expert identification, the three communities can be initially divided into five topics.

Topic 1 mainly includes words related to risk factors for hypertension, such as age, pregnancy, gender, smoking, obesity, myocardial infarction, etc. Topic 2 mainly includes research methods and models related to hypertension, such as prospective studies, follow-up studies, Lateral studies, cohort studies, retrospective studies, etc., also includes some indicator parameters for hypertension research, such as morbidity, disease severity index, experimental expectations, and so on. Topic 3 mainly includes the basic elements of hypertension, such as gender, age, blood pressure, heart rate, renin, glomerular filtration rate, etc. Topic 4 mainly includes disease diagnosis and treatment, such as diagnostic effects, antihypertensive drugs, blood pressure measurement, drug dose response relationship and so on. Topic 5 mainly includes animal, rat, RNA, and Inbred SHR. That is to verify the indicators of hypertension through animal experiments. In general, the biological entities within each community are closely linked to form a variety of specific research topics relevant to the hypertension literature.

According to the time distribution, the hypertension literature was divided into three stages: 2000-2005, 2006-2010, 2011-2017. After data processing, they are imported into Gephi and visualized by a community detection algorithm. According to the visual distribution map and various parameters, the three stages of hypertension topics are distributed in three communities, and the proportion of each community is relatively average. The proportion of communities in the period from 2000 to 2005 was 38%, 33%, and 29 %, the proportion of each community in the period of 2006-2010 was 40%, 38%, 22%. In 2011-2017, the proportion of each community was 42%, 37%, 21%; compared with the MeSH Terms obtained in the three stages, the most frequent occurrences of the three stages are hypertension, male, and female. Most of the terms appear in three stages at the same time, but the distribution of the theme communities in each stage is slightly different. The topic community distribution parameters at each stage are shown in Table 1. The average degree, graph density, and modularity in these three stages are constantly increasing,

indicating that the number of MeSH erms in each stage is increasing with time and the distribution of subject communities is constantly changing. However, this method is very labor intensive, and it is difficult to accurately detect the proportion of each topic in each time period and the path that the theme evolves over time.

Table 1–Parameters at Each Stage

	2000-2005	2006-2010	2011-2017
Average degree	13.94	16.88	20.4
Graph density	0.141	0.171	0.202
Modularity	0.126	0.158	0.175
Average clustering coefficient	0.891	0.438	0.869

Topic Detection and Evolution Trend Analysis Based on DMR

The bibliographic data of 26717 articles were processed as follows: word stem processing; removing stop words, words of length 1 and words with a frequency less than 5 times; removing the superordinate vocabulary in hypertension field. Each bibliographic data generates a text file as a document for the DMR topic model. The data is then processed through an open source machine learning language processing package Mallet [16], according to the DMR model and algorithm. In order to contrast with the topics detected by SNA, the number of topics |T| is set to 5, while adjusting the σ^2 and β values, the resulting five topics and related subject terms. In order to make the identified topics and terms more meaningful, the numbers of terms in the topic-vocabulary distribution of each topic are set to 10, 20, and 30. 10 words that appear more frequent and meaningful are selected, as shown in Table 2.

Topic 1 contains mice, angiotensin, renin, vascular, response, effects, receptor, rats, etc., mainly to describe animal experiments related to hypertension; topic 2 contains risk, factors, age, obesity, gene, women and other words Mainly used to describe the risk factors of hypertension, including age, diabetes, obesity, gender, genes, etc. Topic 3 includes systolic, diastolic, group, compare, rate, invalid, significant, etc., mainly used to describe and hypertension related research methods. Topic 4 contains patients, blood, gene, treatment, results, and other words, mainly used to describe basic elements of hypertension. Topic 5 contains treatment, antihypertensive, therapy, coronary, mortgage, medication, care, control, etc., mainly used to describe the diagnosis and treatment of hypertension.

Table 2–Topic Distribution (DMR)

Topic 1 animal experiments	**Topic 2** risk factors	**Topic 3** research methods	**Topic 4** Basic elements	**Topic 5** Diagnosis Treatments
Mice	Risk	Systolic	Patient	Patient
Renal	Age	Group	Blood	Antihypertensive
Vascular	Prevalence	Significant	Treatment	Coronary
Effects	Blood	Rate	Results	medication
Expression	Disease	Pressure	Finding	clinical
Angiotensin	Factors	Diastolic	cardiovascular	Treatment
Proteins	Diabetes	Compare	Gene	Therapy
Response	Obesity	Invalid	Association	Mortality
Receptor	Gene	Term	Evidence	care

Then the time of publication of the hypertension literature is taken as a variable to explore the trend of the topic over time. The relative distribution of each topic from 2000 to 2017 is shown in Figure 4 below. In general, each topic is constantly evolving over time. In 2000, topic 1 (animal experiment) and topic 4 (basic elements) accounted for a relatively large proportion, and topic 5 (diagnosis and treatment) research was relatively weak; over time, topic 1 (animal experiment) The trend of gradual decline, the topic 4 (basic elements) first decline and then rise, but it has been in an important position; and the proportion of topic 5 (diagnosis and treatment) has increased year by year. It had a relatively important proportion in 2017. Topic 2 (risk factors) has developed relatively stable and has been in a relatively important position; topic 3 (research method) has slightly fluctuated and its proportion has increased year by year since 2007.

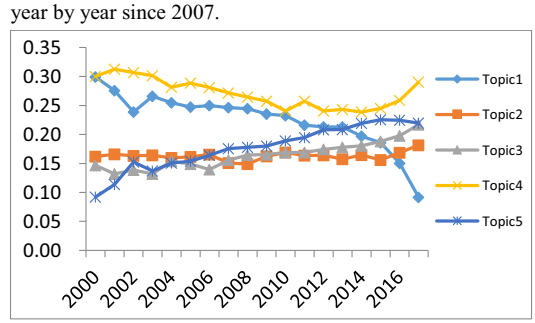


Figure 4– Trend of the Topic Evolution (DMR)

Discussion

The results show that the topics obtained by the SNA are almost the same with those detected by the DMR, both including risk factors, research methods, basic elements, diagnosis, and treatment as well as animal experiments. From the macroscopic view, the MeSH Terms obtained using the SNA method are more specific and more precise, while the DMR are broader. For example, in the topic risk factors, the MeSH Terms determined by SNA, including age factors, Diabetes Mellitus, sex factors, smoking, cardiovascular disease, obesity, and lifestyle can represent relatively concrete risk factors. On the contrary, the term identified by DMR, including age, Diabetes, obesity, and gene, are wider, only to illustrate each general categories of risk factors. In addition, in research methods part, SNA can not only recognize the terms covering prospective studies, logistic models, surveys and questionnaires which stand for research methods, but also include targets such as risk assessment, odds ratio, severity of illness index while DMR recognizes the terms including group, compare, rate and significant. This is because the dataset is too large. Those objects selected by SNA are the top 100 highest-frequency MeSH Terms, while DMR focuses on the entire document. Different research objects thus result in the dissimilarity in the outcome. The number of community topics is also different. SNA sets the number subjectively, and DMR adjusts the number of topics and terms in advance as well. Besides, the relation among 'terms-topics-documents' can be better displayed by DMR so that the probability distribution of the topics can be figured out in any document.

In the comparison in evolution trends, SNA can merely detect the distribution of topic communities for a specific period, hard to compare the evolution trends happening in every duration. However, DMR can detect the ratios of different topics in every period and the evolution trends of topics depending on time,

and it has advantages in exploring the process of topics evolution trends.

In general, the SNA method and DMR method pay different attention to detecting community and evolution trends. The MeSH Terms obtained by SNA are more concrete and precise while terms obtained by DMR are more widely, which need interpretations but are more advantageous in exploring the evolution trends of every topic. Given that combining these two methods together, exploring the community and evolution trends of knowledge network from both the medium and micro view. As a result, they can supplement each other.

Conclusions

Firstly, it is found that the hypertension literature is mainly concentrated on three communities, which can be divided into five research topics, such as risk factors, research methods, basic elements, diagnosis and treatment, and animal experiments. Secondly, the topic changes constantly with time going by. The basic situation of patients has always occupied a high proportion of research. Researches of animal experiments have decreased yearly. Development of risk factors analysis has accounted for a relatively important ratio steadily. The percentage of research topics have been increasing since 2007. Third, it is also found that the topic obtained from SNA and DMR are basically similar. But the MeSH Terms obtained using the SNA method are more specific and precise, while the DMR are broader and have an advantage in exploring the evolution of various themes. If they are applied jointly, the result will be better.

These investigations can help researchers who just start to explore hypertension study to understand the field overview, discover research hot-spots and predictive research frontiers in the field, and promote knowledge exchange within and between domains among experts to help decision makers follow up the flow of knowledge in the field of hypertension. At the same time, the analysis methods of community detection and topic evolution trends in this paper can be extended to other areas of chronic diseases, such as diabetes, coronary heart disease, etc.

Based on the results of this study, PubMed and other databases can provide such information services, helping researchers to obtain a global understanding of relevant fields in the literature search, and improve the relevance and efficiency of the literature search.

This paper also has some limitations. Some Mesh Terms are wildly used in the PubMed publications and are not specific to hypertension, such as “humans”, “female”, “middle-aged”. The accuracy of this research will be further improved by means of filtrating these terms by normalizing their frequency in the target corpus with their overall frequency in PubMed. The DMR topic model needs to preset the number of topics when detecting potential topics. But the number of topics identified by DMR in this paper is subjectively determined; In addition, “a direction for future work is analyzing the internal structures and correlations of communities and topics, which is the next step of our study.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Nos. 71661167007).

References

[1] Y.S. Oh and Z.S. Galis, Anatomy of success: the top 100 cited scientific reports focused on hypertension research, *Hypertension* **63** (2014), 641-647.
[2] C. Schreiber, C. Edlinger, S. Eder, M. Ausserwinkler, B. Wernly, I. Pretsch, C. Jung, U.C. Hoppe, and M. Lichtenauer, Global research trends in the medical therapy of pulmonary arterial hypertension 2000–2014, *Pulmonary Pharmacology & Therapeutics* **39** (2016), 21-27.
[3] M. Götting, M. Schwarzer, A. Gerber, D. Klingelhöfer, and D.A. Groneberg, Pulmonary hypertension: scientometric analysis and density-equalizing mapping, *PloS One* **12** (2017), e0169238.
[4] Y. Ding, Community detection: topological vs. topical, *Journal of Informetrics* **5** (2011), 498-514.
[5] M.L. Wallace, Y. Gingras, and R. Duhon, A new approach for detecting scientific specialties from raw cocitation networks, *Journal of the American Society for Information Science and Technology* **60** (2009), 240-246.
[6] A. Gross and D. Murthy, Modeling virtual organizations with Latent Dirichlet Allocation: a case for natural language processing, *Neural networks* **58** (2014), 38-49.
[7] D. Mimno and A. Mccallum, Topic models conditioned on arbitrary features with Dirichlet-multinomial regression, in: *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence (UAI2008)*, 2008, pp. 411-418
[8] M. Song, G.E. Heo, and D. Lee, Identifying the landscape of Alzheimer’s disease research with network and content analysis, *Scientometrics* **102** (2015), 905-927.
[9] Y. Yang, M. Wu, and L. Cui, Integration of three visualization methods based on co-word analysis, *Scientometrics* **90** (2012), 659-673.
[10] C.T. Butts, Social network analysis: A methodological introduction, *Asian Journal of Social Psychology* **11** (2008), 13-41.
[11] M.E. Newman and M. Girvan, Finding and evaluating community structure in networks, *Physical review E* **69** (2004), 026113.
[12] V.D. Blondel, J.L. Guillaume, R. Lambiotte, and E. Lefebvre, Fast unfolding of communities in large networks, *Journal of Statistical Mechanics* **2008** (2008), 155-168.
[13] A.W. Wolfe, Social network analysis : methods and applications, *American Ethnologist* **24** (1997), 219-220.
[14] S. Brin and L. Page, The anatomy of a large-scale hypertextual web search engine, *Computer networks and ISDN Systems* **30** (1998), 107-117.
[15] R. Lambiotte, J.-C. Delvenne, and M. Barahona, Laplacian Dynamics and multiscale modular structure in networks, *Physics* (2008), 1-29.
[16] H.M. Wallach, D.M. Mimno, and A. McCallum, Rethinking LDA: why priors matter, in: *Advances in Neural Information Processing Systems*, 2009, pp. 1973-1981.

Address for correspondence

Corresponding author: Liqin Zhou, PhD student. Center for Studies of Information Resources, Wuhan University, Wuhan, Hubei, China
E-mail: zhoulq92@163.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190234

IDOMEN: An Extension of Infectious Disease Ontology for MENingitis

W. R. Cédric BÉRÉ[a,d], Gaoussou CAMARA[b], Sadouanouan MALO[c], Moussa LO[d], Stanislas OUARO[a]

[a] *LAMI, Université Ouaga 1 Pr Joseph Ki-ZERBO, Burkina Faso,*
[b] *LIMA, Université Alioune Diop de Bambey, Sénégal*
[c] *Université Nazi BONI, Burkina Faso,*
[d] *LANI, Université Gaston Berger de Saint-Louis, Sénégal*

Abstract

In sub-Saharan African countries the prevention and control of epidemic diseases requires the improvement of the surveillance system for these diseases. Biomedical ontologies are a growing field that can improve health information systems. Indeed biomedical ontologies allow semantic support, data integration, automated reasoning. We are building a meningitis ontology to assist filtering messages relevant to meningitis domain on social media in order to predict a possible epidemic. Indeed, the messages filtered are used for data and event extraction that serve as input for a meningitis surveillance system. In this paper we focused on the modeling and formalization of different perspectives of the meningitis disease such as biological perspective, clinical perspective, epidemiological and public health perspective. This paper presents the three modules in the global Infection Disease Ontology for Meningitis (IDOMEN) and at the end, we illustrate a case of reasoning with our ontology.

Keywords: Knowledge, Infectious Diseases, Meningitis

Introduction

Meningococcal meningitis is an infectious disease that causes an inflammation of the meninges surrounding the brain and the spinal cord [1]. It is due to a gram negative bacteria called neisseria meningitidis. Our work focuses only on meningitis, called meningococcal meningitis.

Since two decades the situation of meningitis is not clearly under control. Nearly one million of meningitis suspected cases and over 100 000 meningitis death are reported in the *African meningitis belt* [2] made up of 26 countries from western Senegal to eastern Ethiopia, via Mali, Burkina Faso, Niger, Chad, Nigeria and other countries.

Unfortunately, in the sub-Saharan countries in general there is no real-time data collection tools for the meningitis surveillance systems. Therefore, the risk analysis for early detection of the meningitis epidemic is delayed.

In this work we propose an approach of early detection of epidemics based on social networks analysis. Our approach is justified by the fact that the work of [3] shows how social media analysis can improve better health monitoring in terms of availability of information for risk analysis and patient support. As stated in social media analysis, health monitoring is conducted through 3 main steps: data collection, semantic analysis, and results sharing with users or decision makers. In this work we use the social media platform Twitter as data source to improve early detection of meningitis epidemic risk in Burkina Faso. We are interested in Twitter because of it increasing use Africa. Indeed, in, the report 2015 “How Africa Tweets”, Portland Communications analyzed 1.6 billion tweets from Africa during this year. They show that African people adopt more and more social media. At December 2017 Internet users in Burkina Faso exceeded to 3 million, with internet growth more than 36 percent.

However, the tremendous of stream of data generated by social media requires filtering while collecting data. It requires the use of a domain-controlled vocabulary to assist data filtering. A domain ontology for meningitis could act as better candidate for this task since ontologies provide at first a domain terminology and secondly axioms describing relations across terms. Our ontology for meningitis is built as an extension of infectious diseases ontology (IDO) [4]. As existing domain ontologies extending IDO, the meningitis ontology is called IDOMEN as Infectious Disease Ontology of/for MENingitis. The scopes of IDOMEN cover the classic views on the disease such as the biologic aspect, clinical aspect, and public health aspect.

In the following of this paper we first present the global architecture of IDOMEN in Methods. Then we highlight the knowledge modeling and formalization of the different meningitis aspects covered by our ontology in the Results.

Finally, we conclude and draw the perspectives of our work.

Methods

Ontologies define a set of basic terms and relations which exist in the vocabulary of a given domain, as well as the rules and axioms that apply on these ones. Gruber has given a fairly comprehensive of "what is an ontology" : « An ontology is a formal, explicit specification of a shared conceptualization ». Depending on the level of abstraction and the object of conceptualization we distinguish three kinds of ontologies :

- The **top level** (upper-level) **ontologies** are characterized by ontologies that model universals, general knowledge;
- The **core ontologies** or generic ontologies describe generic knowledge that are not specific to one domain but are common to several domains ;
- The **domain ontologies** represent knowledge from a specific domain.

The IDOMEN is designed as domain specific ontology since it represents knowledge of the meningitis disease domain.

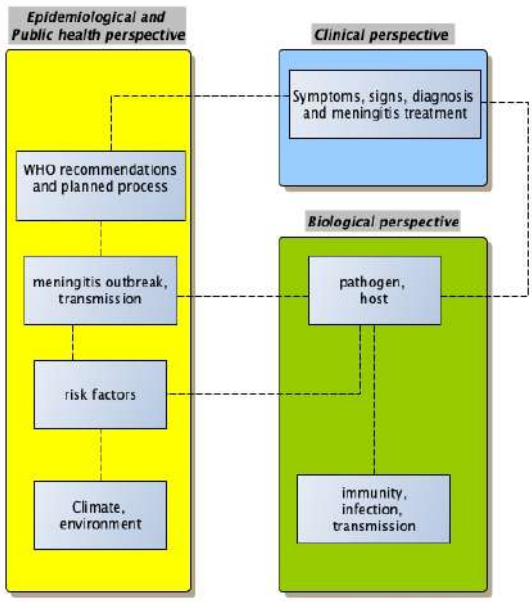


Figure 1 – Modular perspectives design of IDOMEN

The goal of IDOMEN is to assist extracting data and events from text messages published in social media, and to allow data integration, domain knowledge sharing and efficient communication between actors within the epidemiological surveillance system [10]. Therefore, to satisfy these requirements, the scope of IDOMEN covers the disease aspects related to the biology, clinical manifestation and care, public health, prevention and control, biomedical research, and epidemiological studies.

The different aspects have been packaged in three modular perspectives that are biological perspective, clinical perspective, epidemiological perspective and public health perspectives (figure 1) :

- The **biological perspective** focuses on biology and microbiology of organisms implied in the disease course. In this perspective we model immunity, virulence factors, pathogen and host biology, etc. ;
- The **clinical perspective** focuses on clinical manifestations (signs, symptoms), complementary examination (laboratory test, laboratory findings, etc.), diagnosis, treatment, etc. ;
- The **epidemiological and public health perspective** focuses on topics related to the disease spread/outbreak such as risk factor, transmission, climate, environnement, etc. ; and deals with standard strategies of preventing or fighting against meningitis outbreaks.

This organization of the knowledge allows us to design our ontology in modular perspectives. Modularity makes it possible to build ontologies of large or small sizes and offers the possibility of easily reusing existing ontological resources. In the field of biomedical ontologies the reuse of existing ontological components is a best practice that is strongly encouraged.

This leads to the principles of the Open Biomedical Ontologies (OBO) Consortium [6] which are best practices and recommendations that promote interoperability. There are many biomedical ontology repositories containing ontology of different types and level of abstraction that can be freely reused. IDOMEN is designed and built upon this principle.

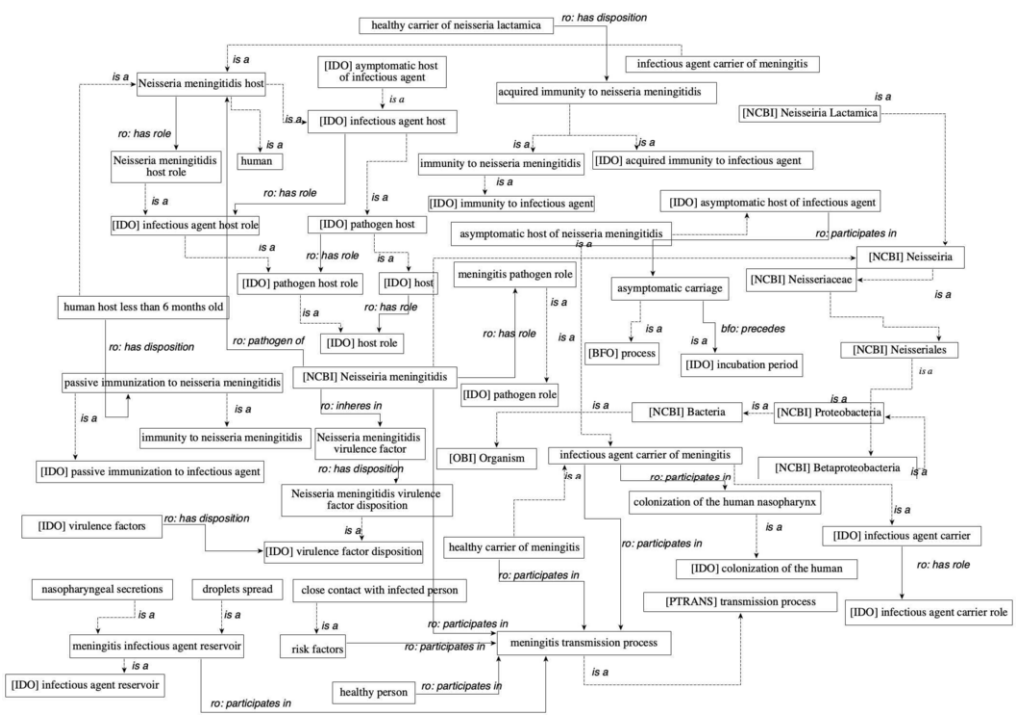


Figure 2– Conceptual model of the biological perspective of the meningitis disease

Results

Modeling biological perspective in IDOMEN

Pathogen and host

In figure 2, we present the knowledge modeling of the pathogen agent of meningococcal meningitis called *Neisseria meningitidis.* We use the classification of NCBI Taxonomy to provide a classification of the organism *neisseria meningitidis* in IDOMEN. The species *Neisseria meningitidis* in the taxonomic classification of NCBI is of the genus

Neisseria, and the *Neisseria* are from family of the *Neisseriaceae*. In the Figure 2 this is represented by *"[NCBI]Neisseria meningitidis"* is a *"[NCBI]Neisseria"*, followed by the relation *"[NCBI]Neisseria"* is a *"[NCBI]Neisseriaceae"*. Browsing further in the hierarchy of the NCBI Taxonomy, we found that *Neisseria Meningitidis* is also a subclass of *Proteobacteria*. *Proteobacteria* is a phylum of bacteria consisting of the purple bacteria *"[NCBI]Proteobacteria"* is a *"[NCBI]Bacteria"*. Bacteria is an organism. *Neisseria meningitidis* being the pathogen agent of meningitis, it plays a *"pathogenic role"* (*"meningitis pathogen role"*) in the human body : *"meningitis pathogen role"* ***is a*** *"[IDO]pathogen role"*. Thus, we have *"[NCBI]Neisseria meningitidis"* ***bearer of*** *"Meningitidis pathogen role"*. *Neisseria meningitidis* has the capability under specific conditions to trigger a pathogenic process, we say that it has a *"pathogenic disposition"*. A disposition is an intrinsic attribute of an entity that allows it to trigger itself a specific process or processes when particular conditions are satisfied : *"[NCBI]Neisseria meningitidis"* ***has disposition*** *"[IDO]pathogenic disposition"*.

Pathogenesis

The pathogenesis of meningitis refers to the process or processes responsible for triggering and developing meningitis. The pathogenesis of meningitis occurs in three (03) sequences that are: transmission, acquisition, colonization.

There are two (02) modes of transmission of meningococcal. The aerosol mode and the contaminated secretions one. A person who caught the pathogen agent Neisseria meningitidis but who has not any manifestations (symptom and sign) of the meningitis is called *"healthy carrier of meningitis"*. We translate this by *"healthy carrier of meningitis"* **is a** *"infectious agent carrier of meningitis"*. *"Asymptomatic carriage"* is a process during which the infected person develops no symptomatic manifestation of meningitis. This state of fact is described by 3 relations: : *"healthy carrier of meningitis"* ***is a*** *"[IDO]asymptomatic host of neisseria meningitidis"*. *"[IDO]asymptomatic host of neisseria meningitidis"* ***is a*** *"[IDO]asymptomatic host of infectious agent"*. "[IDO]asymptomatic host of infectious agent" **participates in** "asymptomatic carriage" and "asymptomatic carriage" **is a** "[BFO]process". This process asymptomatic carriage precedes the phase of colonization in the nasopharynx, what is transcribed by these two relations *"asymptomatic carriage"* precedes *"[IDO]incubation period"* and *"infectious agent carrier of meningitis"* ***participates in*** *"colonization of the human nasopharynx"*.

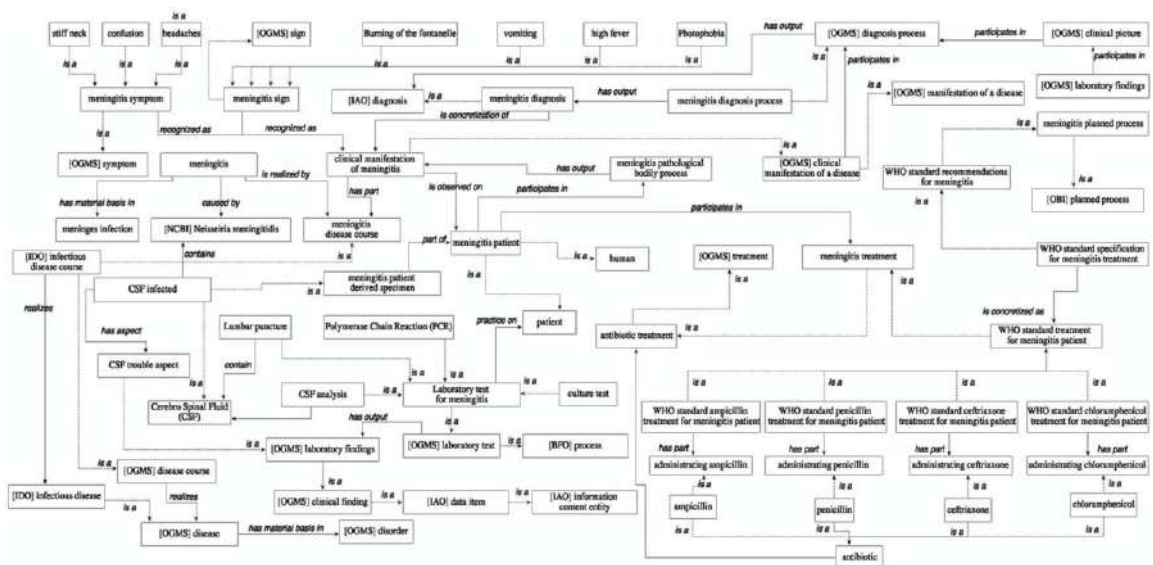


Figure 3– Conceptual model of the clinical perspective of the meningitis disease

Modeling clinical perspective in IDOMEN

Signs and symptoms

Meningitis symptoms and signs are clinical features of meningitis. The most common symptoms of acute meningitis are high fever, headache, stiff neck, vomiting, irritability or confusion, photophobia (sensitivity to light), Kerning signs, lethargy seizures. *"meningitis symptom"* and *"meningitis sign"* are perceived or noticed on patient. They are formalized as follow *"meningitis sign"* is a *"[OGMS]sign"* and *"meningitis symptom"* is a *"[OGMS] symptom"* by reusing existing concepts in OGMS. For instance, we can also infer the following relations *"headaches" is a "meningitis symptom"*, *"headaches" is a "meningitis symptom"*, *"photophobia" is a "meningitis sign", "Kernig sign" is a "meningitis sign"*.

Meningitis patient is a patient presenting clinical manifestations of meningitis. "Meningitis patient" is a subclass of *"patient"* and a *"clinical manifestation of meningitis"* ***is a*** *"[OGMS]manifestation of a disease"*. Therefore *"meningitis symptoms"* ***recognized as*** *"clinical manifestations of meningitis"* and *"meningitis signs"* ***recognized as*** *"clinical manifestation of meningitis"*.

Diagnosis

The first step in the diagnosis consists mainly in the recognition of clinical manifestations.

In terms of diagnosis, *"meningitis diagnosis" is concretization of "clinical manifestation of meningitis"*, *"meningitis diagnosis" is a "[IAO]diagnosis"*. *"meningitis diagnosis process" is a "[OGMS]diagnosis process"*. and *"meningitis diagnosis process" has output "meningitis diagnosis"*.

After seeing clinical signs, we can perform a CSF sample to further the diagnosis. In case of the observation of CSF ("*CSF analysis"*) reveals a cloudy or purulent appearance after macroscopic examination, this indicates a probable case of bacterial meningitis. CSF analysis can be extended by other laboratory tests such as Polymerase Chain reaction, antibiogram, etc., used to diagnose meningitis are organized in class called "Laboratory test for meningitis".

Finally, only the detection of the bacterial agent in the CSF at the laboratory level makes it possible to effectively define a *confirmed case of meningitis*.

"Laboratory test for meningitis" is a *"[OGMS]Laboratory test"*. The results of the laboratory tests are subclasses of *"[OGMS]Laboratory findings"*.

Treatment

Meningitis is a disease whose evolution is very fast in the patient, which necessitates a quick diagnosis in order to start the treatment. The different treatments for meningitis are specified in WHO standard protocols called "*WHO standard specifications for meningitis treatment*". Here are some treatments recorded in the standard protocol: "*WHO standard ampicillin treatment for patient meningitis*", "*WHO standard penicillin treatment for patient meningitis*", "*standard WHO ceftriaxone treatment for patient meningitis*", *"WHO standard chloramphenicol treatment for patient meningitis*". All of these different treatments are based on antibiotic administration [7], so we have for example "*WHO standard ampicillin treatment for meningitis patient*" ***has part*** "*ampicillin*" and "*ampicillin*" ***is a*** "*antibiotic*".

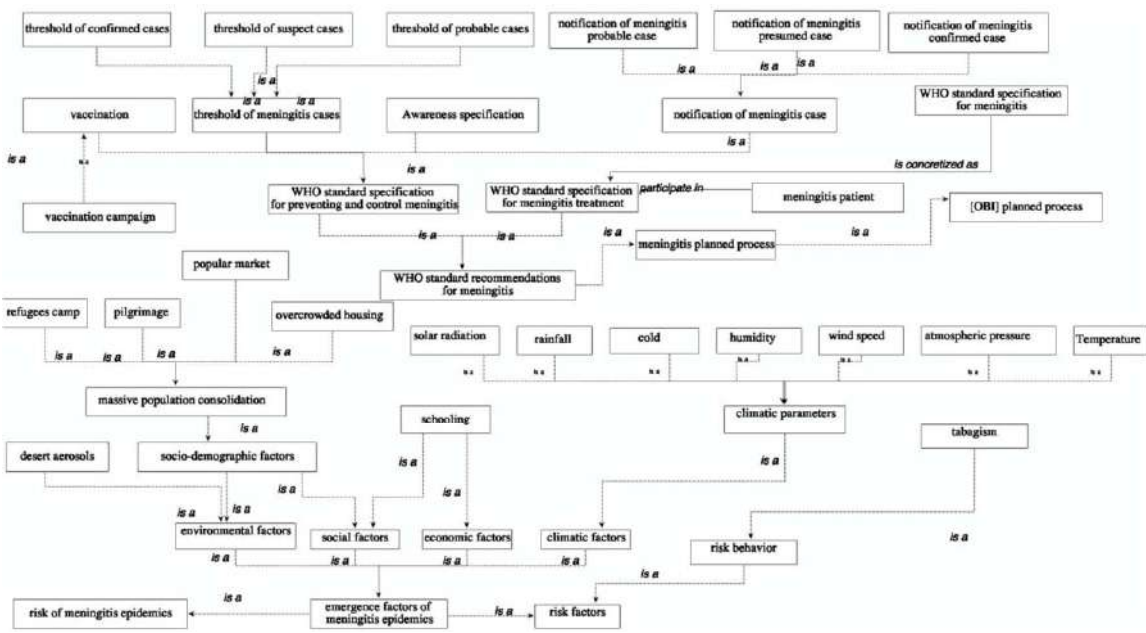


Figure 4– Conceptual model of the epidemiological and public health perspective of the meningitis disease

Modeling epidemiological and public health perspective in IDOMEN

In the figure 4, we describe different aspects related to disease spread/outbreak such as surveillance, prevention, epidemic emergence factors (risk behaviours, climate and environment).

Prevention

The prevention and control of meningitis is based on vaccination, vaccination campaigns and awareness : *"vaccination"*, *"vaccination campaigns"* and *"awareness"* are subclass of *"WHO standard specifications for preventing and control meningitis". "WHO standard specifications for preventing and control meningitis"* ***is a*** *"meningitis planned process". "meningitis planned process"* ***is a*** *"[OBI]planned process"*. To control and prevent the occurrence of an outbreak there is a mechanism called *"cases notification"* consisting in systematically notifying the different cases encountered in health facilities (*"presumed case of meningitis"*, the *"probable case of meningitis"* and the *"confirmed case of meningitis"*). We have represented this notification mechanism in IDOMEN as follow *"presumed case of meningitis notification"* ***is a*** *"meningitis case notification", "probable case of meningitis notification"* ***is a*** *"meningitis case notification"*, *"confirmed case of meningitis notification"* ***is a*** *"meningitis case notification".* As soon as one of the thresholds (*"threshold of presumed cases"*, *"threshold of probable cases"* and *"threshold of confirmed cases"*) are reached an alert that corresponds to the level of the situation is launched in order to bring appropriate measures.

Epidemic emergence factors

The recurrence of meningitis outbreaks is linked to a complex combination of several factors of kinds economic, social, climatic and environmental [8] [9]. We defined a class *"meningitis epidemic emergence factors*" which is a parent classes of following subclasses *"environmental factors*", *"social factors*", *"economic factors*", *"climatic factors*". The population dynamics (movements of populations and communities) are

considered to be factors influencing the intensity of meningitis epidemics in the African meningitis belt : *"massive population consolidation" is a "socio-demographic factors*". *"socio-demographic factors*" is a subclass of *"environmental factors*" and *"social factors*". For example of *"massive population consolidation"* in the meningitis belt we have : *"pilgrimage", "popular market"*, *"refugees camp"* and *"overcrowded housing"*. The "climatic parameters" are : *"atmospheric pressure", "wind speed"* , *"humidity", "cold", "rainfall" and "solar radiation"*.

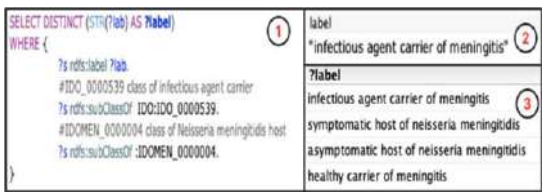


Figure 5– Use case reasoning

Conclusions

In this first stage of the IDOMEN ontology building, we provided formalization in OWL2 using the Protege editor which also allows us to experiment reasoning use cases on the ontology. For example, the Figure 5 presents (1) a SPARQL query about ***"Which are the infectious agent carriers that also are Neisseria Meningitidis hosts ?"***, (2) the result without launched inference engine, and (3) the results by activating the reasoner FaCT++ 1.6.5 engine in the Protégé editor. This explains that we obtain more complete results by deductive inferences on classes and subclasses. In the case of our query the deduction is based on the transitivity properties from the classes inheritance.

In this work we modeled the knowledge relating to the biological perspective, clinical, epidemiological and public health perspectives of the meningitis disease. A usable version is available for download at https://github.com/cedricbere/IDOMEN/blob/master/Ontology/idomen.owl.

Our future work will focus on validating and evaluating the ontology before using for filtering on social media.

Acknowledgements

This work is part of the Management and Access of Biomedical knowledge through Ontology network (MABO) project which is supported by the African Center of Excellence in Mathematics Computer Science and ICT (CEA-MITIC).

References

[1] Y.-L. Tzeng, and D.S. Stephens, Epidemiology and pathogenesis of Neisseria meningitidis, *Microbes and Infection*. **2** (2000) 687–700. doi:10.1016/S1286-4579(00)00356-7.

[2] G. WHO, GHO, GHO : Number of suspected meningitis cases and deaths reported, *WHO Global Health Observatory (GHO) Data*. (n.d.). http://www.who.int/gho/epidemic_diseases/meningitis/suspected_cases_deaths_text/en/ (accessed September 4, 2018).

[3] A. Farzindar, and D. Inkpen, Natural Language Processing for Social Media, *Synthesis Lectures on Human Language Technologies*. **8** (2015) 1–166. doi:10.2200/S00659ED1V01Y201508HLT030.

[4] L.G. Cowell, and B. Smith, Infectious Disease Ontology, in: V. Sintchenko (Ed.), Infectious Disease Informatics, Springer, New York Dordrecht Heidelberg London, 2010. http://dx.doi.org/10.1007/978-1-4419-1327-2_19.

[5] W.R.C. Béré, G. Camara, S. Malo, M. Lo, and S. Ouaro, Towards Meningitis Ontology for the Annotation of Text Corpora, in: C. M. F. Kebe, A. Gueye, and A. Ndiaye (Eds.), Innovation and Interdisciplinary Solutions for Underserved Areas, Springer International Publishing, 2018: pp. 322–331.

[6] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L.J. Goldberg, K. Eilbeck, A. Ireland, C.J. Mungall, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S.A. Sansone, R.H. Scheuermann, N. Shah, P.L. Whetzel, and S. Lewis, The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration, *Nat Biotechnol*. **25** (2007). doi:10.1038/nbt1346.

[7] OMS, Contrôle des épidémies de méningite en Afrique : Guide de référence rapide à l'intention des autorités sanitaires et des soignants, (2015). http://www.who.int/csr.

[8] A.M. Molesworth, L.E. Cuevas, S.J. Connor, A.P. Morse, and M.C. Thomson, Environmental Risk and Meningitis Epidemics in Africa, *Emerg Infect Dis*. **9** (2003) 1287–1293. doi:10.3201/eid0910.030182.

[9] P. Yaka, B. Sultan, H. Broutin, S. Janicot, S. Philippon, and N. Fourquet, Relationships between climate and year-to-year variability in meningitis outbreaks: a case study in Burkina Faso and Niger, *International Journal of Health Geographics*. **7** (2008) 34. doi:10.1186/1476-072X-7-34.

Address for correspondence

W. Régis Cédric BERE

cedric.bere@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190235

Improving Mechanical Ventilator Clinical Decision Support Systems with a Machine Learning Classifier for Determining Ventilator Mode

Gregory B. Rehm[a], Brooks T. Kuhn[b], Jimmy Nguyen[b], Nicholas R. Anderson[b], Chen-Nee Chuah[a], Jason Y. Adams[b]

[a] University of California Davis, Davis CA 95616, USA
[b] University of California Davis Medical Center, Sacramento CA 95817, USA

Abstract

Clinical decision support systems (CDSS) will play increasing role in improving quality of medical care for critically ill patients. However, due to limitations in current informatics infrastructure, CDSS do not always have complete information on state of supporting physiologic monitoring devices, which can limit input data available to CDSS. This is especially true in use case of mechanical ventilation (MV), where current CDSS have no knowledge of critical ventilation settings, such as ventilation mode. To enable MV CDSS make accurate recommendations related to ventilator mode, we developed a highly performant machine learning model that is able to perform per-breath classification of five of most widely used ventilation modes in USA with average F1-score of 97.52%. We also show how our approach makes methodologic improvements over previous work and is highly robust to missing data caused by software/sensor error.

Keywords:

Artificial respiration, clinical decision support systems, machine learning

Introduction

Mechanical ventilation (MV) is life-saving intervention delivered in intensive care unit (ICU) to patients with acute respiratory failure. When delivered properly, MV allow injured lungs heal while ventilator performs majority of work of breathing for patient. When delivered improperly, MV has been associated with variety of adverse clinical outcomes including patient discomfort, increased sedative dosing, longer ICU length of stay, increased chance of ventilator-induced lung injury, and lower survival [1,2]. New generation of clinical decision support systems (CDSS) promises to reduce chances of delivering improper MV by automating aspects of ventilator configuration, and providing clinically accurate and relevant alerts to providers. However, key detriment to these systems is lack of access to configured state of ventilator and therefore lack information that may improve efficiency of these CDSS [3].

One such piece of information that many MV CDSS lack is choice of ventilation mode (VM) that determines pattern of flow and pressure delivery with each breath (Figure 1 B-D). This information is generally unavailable to CDSS due to lack of interoperability and information exchange between CDSS and ventilator or electronic health record [3]. CDSS knowledge of VM is important because changing VMs may be a necessary procedure in course of patient care [4]. For example, if CDSS determines that patient is breathing asynchronously with ventilator, it may be able to make recommendation that providers choose a different VM that provides more comfort and flexibility in breathing to patients [5-8]. Another example would be that CDSS could provide alerts to clinicians if patients continually violate safe volumes of air to inhale. This would be especially important in cases where patients have acute respiratory distress syndrome and need limited tidal volumes [9,10]. In this case CDSS could recommend patients be placed on VM that limits tidal volumes such as volume-control.

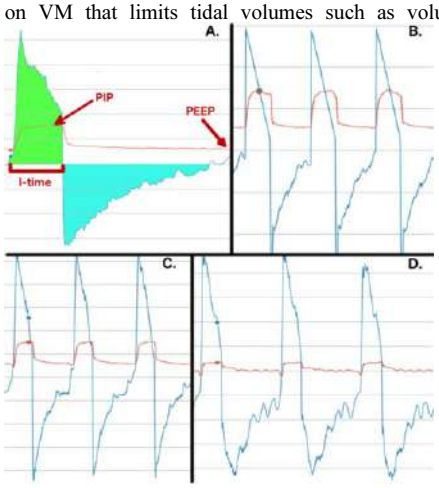


Figure 1- Displays visualizations of ventilator waveform data (VWD). Flow measurements represented in blue, and pressure in red. A.) Here we display examples of how to extract information from VWD. Positive End Expiratory Pressure (PEEP) is noted as minimum pressure supplied by ventilator, and peak inspiratory pressure (PIP) is maximum pressure supplied during inhalation. Inspiratory time (I-time) is amount of time patient breathes in. Total amount of air breathed in represented in green, and air breathed out shown in teal. B.) Shows canonical example of volume control (VC), mode patient receives fixed volume of air for each breath. C.) Shows example of pressure control (PC). In PC, pressure is fixed during inhalation. D.) Example of continuous positive airway pressure (CPAP). Here minimal pressure support is given, and all breaths initiated by patient.

If MV CDSS lacks knowledge of VM from more traditional methods, it may still be able to access it by utilizing information derived from streams of flow and pressure readings that comprise ventilator waveform data (VWD). To the best of our knowledge, only one previous effort has developed rule-based classifier using analysis of VWD for providing hourly VM classifications. However, its use of closed dataset, limited temporal resolution, and accuracy of model represent potential limitations both for research and decision support [3,5]. Having

highly granular temporal resolution VM classification results is important because in practice providers may change VM frequently based on changes in clinical state or patient tolerance of VM. These changes may cause specific VM to remain constant for as low as minutes of time. To improve upon previous work, we note that machine learning (ML) has proven capable of accounting for highly variable nature of physiologic data such as VWD on temporally granular time scales [11,12]. So we created a ML model that could identify different VMs on per-breath basis, with freely accessible dataset, using only VWD as input.

In this paper, we detail multiple important considerations for modeling ML classifier that can classify VM. First, we discuss how we created one of the largest datasets of per-breath labeled information, extraction of features from VWD, and performance of our resulting ML model that can determine five of most widely used ventilation modes in USA [4]. Second, we discuss experiments of how well our model performs in presence of missing training data. Finally, we discuss experimentation we conducted for reducing size of our training dataset by nearly 72% while maintaining generalizability of our classifier to our testing set. To allow reproducibility of our work, our code and dataset are publicly accessible and published on GitHub. Thus, we hope that our work will serve as catalyst for continuing to improve capabilities and efficiency of MV CDSS.

Methods

In this study, we used dataset of VWD collected from 103 subjects (IRB# 647002) within intensive care environments of University of California Davis Medical Center (UCDMC) consisting of MV flow and pressure measurements sampled at 50 Hz [13,14]. Ventilation mode was not recorded in course of VWD data collection. We then randomly selected 2-4 hour epochs of VWD from the 103 subjects. All VWD was stored in data files of 2 hours in length, and approximately 2,000 breaths were stored per data file. Each breath in these epochs was annotated by three expert clinicians (JYA, BTK, JN) for presence of one of five VMs: volume control (VC), pressure control (PC), pressure support (PS), continuous positive airway pressure (CPAP), and proportional assist ventilation (PAV) (Table 1). Many patients had 2-4 hour periods selected where VM was switched multiple times, other modes such as pressure regulated volume control (PRVC), volume support, and airway pressure release ventilation (APRV) were found, and annotated within these epochs, but were excluded in our final analysis because of their rarity of use at UCDMC.

Table 1-Descriptive statistics for our dataset for each ventilator mode analyzed. Also analyzed number of patient ventilator asynchrony (PVA), suction, and cough breaths found [14]. While these breaths do not represent normal breathing, they are typical in clinical practice.

	Volume Control	Pressure Control	Pressure Support	CPAP	PAV
Patients	23	37	55	28	22
Total Breaths	61,662	78,635	92,360	14,795	36,303
PVA Breaths	7,714	4,570	6,924	2,373	7,669
Suction Breaths	750	136	681	350	373
Cough Breaths	229	117	178	56	96

Given VWD is so heterogeneous it can be difficult for even expert clinicians to make consistent classification of breathing patterns [15]. Thus, in performing classification of VM we ensured that each breath was dual clinician adjudicated, meaning that two clinicians would independently annotate a single breath, and if classifications disagreed they would be resolved through communication between the two [14]. To further account for breathing heterogeneity, we included regions containing pathologic patient-ventilator interactions such as patient-ventilator asynchrony (PVA), routine clinical events such as suctioning and cough, and regions of noisy data caused by moisture/blood/mucus in ventilation circuit tubing [14].

Table 2-Set of proposed features for our model. Features were segmented into per-breath and multi-breath time frames.

Feature	Description
Inspiratory Flow Slope Variance (per breath)	This feature measures variance of successive, 0.08-second long slope measurements of inspiratory flow curve of a single breath. This feature was effective for classifying volume control.
Variance of Pressure (per breath)	This feature takes variance of all pressure measurements for a single breath. This feature was helpful for classifying CPAP which typically utilizes low pressures relative to PEEP on inspiration.
Variance of Per-Breath Inspiratory Flow Slope Variance	The inspiratory flow slope variance was found on per breath basis, and this feature takes variance of inspiratory flow slope variance across a 10 breath window.
Inspiratory Time (I-time) Variance (10 breath window)	The amount of time that patient inhales for single breath is called I-time. This feature calculated variance of 10 successive breaths.
Pressure-Based I-time Variance (10 breath window)	We defined pressure-based I-time as amount of time (seconds) that pressure is elevated by [0.4 * (PIP - PEEP)] above PEEP. This was an important variable to measure in pressure control and pressure support, where flow-based I-time can be shorter than ventilator's set I-time, which may occur in delayed cycling asynchrony.
N Plateau Pressures (20 breath window)	A plateau pressure is taken on ventilator when inspiratory flow is set to 0 for a certain amount of time, during which ventilator can read residual pressure in respiratory system. PAV will repetitively take plateau pressures in order to adjust ventilation to patient's needs.
Pressure-Based I-time Variance (100 breath window)	In this feature, pressure-based I-time statistic is also calculated for 100-breath window. This feature was necessary to provide capacity for differentiating between pressure control and pressure support in synchronously breathing patients.

With this dataset, we utilized 55 patients and 140,928 breaths for our training cohort, and 48 patients and 165,988 breaths for our testing cohort. There was no patient overlap between testing and training cohorts. Testing set was chosen to be approximately as large as training set because initial modeling

yielded strong results, and we wished to utilize large testing set as further validation for our approach. Using both Scikit-learn and Pytorch ML libraries [16,17], we evaluated use of multiple ML algorithms including: support vector machine (SVM) [18], multi-layer perceptron (MLP), long-short term memory recurrent neural network (LSTM RNN) [19], logistic regression, and random forest (RF) classifier [20]. All models performed classification on per-breath basis, highest possible level of granularity possible in VM classification. Based on model investigation, we settled on usage of RF with parameterization of 30 classifier trees for our final model (see online supplemental).

Our feature set is composed of 7 items of expert-guided information derived from raw VWD, and is described in Table 2. Our features are derived from both per breath and multi-breath analytic time frames. Per-breath time frames occur over single breath, while multi-breath time frames are composed of windows of short, medium, and long periods of breathing. Short window is 10 breaths long, medium window 20 breaths, and long window 100 breaths. Tuning of features and hyperparameters was guided by performing 10-fold cross-validation of our training data. After tuning model hyperparameters during the validation phase, we evaluated our model on our testing set. No additional changes to our feature set, or model hyperparameters were performed after model development was completed in training set. Model performance is primarily reported through F1-score because it is more representative of class-imbalanced classifier performance than accuracy is. F1-score is calculated as harmonic mean of precision (PPV) and recall (sensitivity):

$$\text{F1-score} = 2\frac{precision * recall}{precision + recall}$$

Limitation to using RF to classify ventilator mode is RF classifier assumes that all breaths are independent of each other. However, ventilator mode is a continuous setting that does not vary over time, unless it is manually changed by provider. Therefore, one breath's mode is often predictive of next breath's mode. This modeling incongruity causes RF classifier to sometimes perform incorrect VM classification even in periods where classifier correctly predicts correct VM for a majority of breaths. To smooth these incorrect predictions, we implement an algorithm we term "look-ahead smoothing" which operates as second pass heuristic on all per breath RF breath predictions. More specifically, once RF is finished, look-ahead smoothing examines each breath VM classification sequentially, and if it determines breath's classification is not in accordance with previous n breaths then it will look ahead at next n breaths in sequence. The breath will then be re-classified in accordance to majority x percent of subsequent n breaths. Both n and x are configurable parameters that we set at $n = 50$ and $x = 60$, parameters which were found via sensitivity analysis. In real-time classification, assuming average respiratory rate of 20 breaths per minute, this technique results in latency of at most 2.5 minutes between a given breath and availability of its final classification.

Finally, we implemented experiment to test how well our classifier would generalize to larger dataset if random breaths in our training dataset were missing due to some technical error. So we conduct experiment where we ablate (i.e. remove) data observations at random from our training dataset in equal proportion for VC, PC, PS, CPAP, and PAV. We do not perform any ablation on the testing set. We then report results of this experiment by recording F1-score for each class with respect to percentage of dataset that simulated as missing.

Results

Using RF model with feature set defined in Table 2, we initially performed 10-fold cross validation with our training set to test performance of our VM classifier. We found that during cross validation our model consistently performed within 98-99% for F1-score, recall, and specificity for all VMs. We then evaluated our model on withheld test set. CPAP suffered largest drop in performance because it confused PS for CPAP for an entire patient. VC/PAV suffered no drop in performance and PC/PS only suffered slight declines in performance (Table 3).

Table 3-Performance of our Random Forest model when applied to our withheld testing set.

Mode	F1-Score	Accuracy	Precision	Recall	Specificity
VC	0.999	1.0	0.998	1.0	1.0
PC	0.989	0.993	0.983	0.996	0.992
PS	0.975	0.981	0.993	0.958	0.996
CPAP	0.85	0.988	0.767	0.952	0.989
PAV	0.994	0.999	0.99	0.998	0.999

We hypothesized that since the model performed well on both training and testing sets that it would also be robust to scenarios in which breath data went missing due to reason of sensor or software failure. We report results for this experiment in Figure 2. We found model is robust to missing data until approximately 90% of data is removed. After this point PC and PS F1-score performance begins to decrease and other classifications begin to fluctuate. After 99% of data is removed our classifications lose clinical utility.

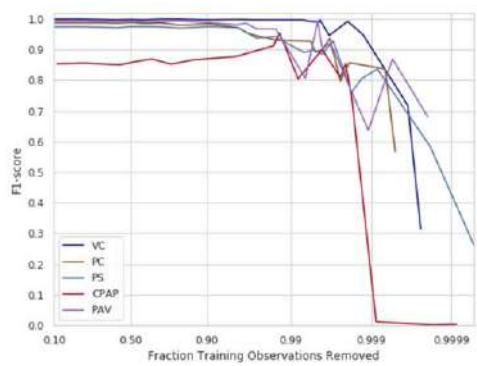


Figure 2-Here we simulate scenario where percentage of training observations is missing due to some kind of software/hardware error.

Given results of random ablation experiment, we hypothesized that we may have created too large a training set. To reduce the size of our training set in generalizable, non-random way, we hypothesized we only needed to keep the first of certain number of contiguous breaths from each VM per data file, and still maintain performance of our original model. In this respect, we could make recommendations to physicians to only annotate first $m$breaths in a series and just leave the rest alone. This could also decrease amount of time necessary to annotate VM on future patients. So, we performed a sensitivity analysis to determine what the optimal number of contiguous observations to keep per ventilator mode is. We do this by sequentially iterating over each VM in our training set and only picking first m breaths in a file while keeping number of observations from other VMs constant. Our analysis (Figure 3) showed that it was

most optimal to only use first 450 VC observations, first 120 PC, 1,200 PS, 160 CPAP, and 80 PAV observations in a file. Using this methodology, we ablated overall number of training observations by 71.41% from 140,928 to 40,285 observations, while still maintaining generalizability of our training set to our withheld test set, and largely improved CPAP performance (Table 4). By performing ablation we were able to boost average F1-score of our classifier to 0.9752 from 0.9614 that was reported in Table 3.

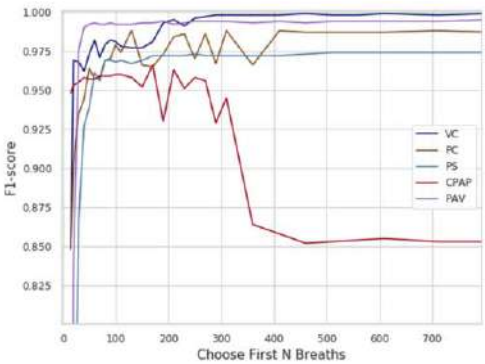


Figure 3-Results from our sensitivity analysis for choosing first N contiguous breaths for a given mode in a data file.

Table 4-Final results of our ablation experiment where we only keep first 450 VC, 120 PC, 1,200 PS, 160 CPAP, and 80 PAV observations in a data file. We note final number of training observations that we kept, and report how much of reduction that was in contrast to original training set. Performance improvements/degradation over results listed in Table 3 are bracketed alongside final performance metrics. E.g. performance increase of 2.0% is denoted as (+.02).

Mode	Training Observations	F1-Score
VC	6,079 (-83.65%)	0.998 (-0.001)
PC	2,154 (-92.77%)	0.964 (-0.025)
PS	27,892 (-26.81%)	0.967 (-0.008)
CPAP	3,040 (-73.55%)	0.955 (+0.105)
PAV	1,120 (-94.46%)	0.993 (-0.001)

Discussion

In this paper, we highlighted how we created dataset of 308,957 breaths annotated for VM on per-breath basis and how we developed highly accurate, ML-based VM classification model that only utilizes raw VWD to perform classifications. Our VM classifier was highly performant in detecting five of most widely used VMs in USA, even in presence of signal noise, episodes of PVA, and routine clinical events such as cough and suction [4]. Using our approach, we were able to achieve methodological and performance improvements in VM classification compared to current state of art [3]. In this regard, Murias reported 89% accuracy at detecting per-hour VM classification, and we report average accuracy of 98.05% of per-breath VM classification (Table 3). Finally, we examined how robust our model is to presence of missing training data, and additional experimental results that suggested how we can decrease the size of our dataset while still maintaining generalizability of our classifier.

We took multiple measures to ensure we were not overtraining our classifier. First, we utilized a relatively large and diverse sampling of patients to create both our testing and training sets. This created one of the largest available datasets of per-breath labeled VWD. Two to four hour epochs were chosen at random from each of these patients. Our testing set included almost as many patients as our training set, and was composed of more breaths than our training set. There was no overlap of patients between training and testing sets. Finally, all model features and hyperparameters were evaluated on the training set using K-fold validation, and were frozen after initial evaluation of our testing set.

Our ablation experiments deserve additional consideration. Results of the random ablation experiments highlight multiple things: 1) RF is extremely performant with our featurization approach, and is also able to perform VM classification with small amounts of data. 2) Our ablation results also illustrate that it may not be necessary to create very large training datasets of information to create performant ML classifiers for VM. 3) Our size reduction experiments did see some decreases in performance in PC and PS because of the manner in which we performed our sensitivity analysis. In our analysis we only modified observations from a single VM type while keeping other VM observations constant, so it was not possible to determine side effects from simultaneously ablating several modes at once. Future experiments could perform more computationally demanding task of ablating multiple modes at once to further explore the issue. 4) Our size reduction experiment showed that first 160 breaths seem to be most representative of CPAP breathing patterns. We hypothesize this can be explained by the fact that some patients tire quickly when on CPAP, and thus their breathing can become more irregular. In this case, later breaths in CPAP sequences may more closely resemble asynchronous breathing from other ventilator modes instead of CPAP.

This work had several limitations. Our use of "look-ahead smoothing" introduced small latency of 2.5 minutes to real-time ventilator mode predictions. This time delay in classification would not likely be of clinical consequence since CDSS recommendations requiring VM state information would rarely be executed over such short time frames to ensure that transient changes in waveforms do not trigger frequent false alarms. If latency is not desired then "look-behind smoothing" can be used as alternative approach. Our study was also confined to a single academic medical center and single ventilator type. There are also additional ventilator modes such as PRVC that we were unable to add to our model due to their paucity of use at UCDMC. We welcome additional collaboration and inclusion of multi-center data and have publicly released our code and dataset.

Conclusions

In conclusion, we created a highly-performant ML classifier for detecting five of most commonly used ventilator modes in USA, using only raw VWD as input. Our use case further demonstrates utility of ML analysis of physiologic waveform data to improve CDSS characterization of patient state when state is missing due to limitations of available informatics infrastructure. We also illustrated usage of dataset ablation to characterize how missing data affects generalization performance of our classifier, and how we can restrict size of our training set while maintaining model generalization to our test dataset. Our classifier will facilitate development of more advanced automated MV CDSS to improve management of patients experiencing respiratory failure.

Acknowledgements

This work was generously supported by National Heart, Lung, and Blood Institute (NHLBI) Individual Predoctoral Fellowship (award number 1F31HL144028-01), and NHLBI Emergency Medicine K12 Clinical Research Training Program (grant number K12 HL108964).

References

[1] A. Slutsky, and M. Ranieri, Ventilator-induced lung injury, *N Engl J Med* **369** (2013), 2126-2136.
[2] A. Bagchi, M.I. Rudolph, P.Y. Ng, F.P. Timm, D.R. Long, S. Shaefi, K. Ladha, M.F. Vidal Melo, and M. Eikermann, The association of postoperative pulmonary complications in 109,360 patients with pressure-controlled or volume-controlled ventilation, *Anaesthesia* **72** (2017), 1334-1343.
[3] G. Murias, J. Montanyà, E. Chacón, A. Estruga, C. Subirà, R. Fernández, B. Sales, C. de Haro, J. Lopez-Aguilar, U. Lucangelo, J. Villar, R.M. Kacmarek, and L. Blanch, Automatic detection of ventilatory modes during invasive mechanical ventilation, *Crit Care* **20** (2016), 258.
[4] T. Pham, L.J. Brochard, and A.S. Slutsky, Mechanical ventilation: state of the art, *Mayo Clin Proc* **92** (2017), 1382-1400.
[5] N. Rittayamai, C.M. Katsios, F. Beloncle, J.O. Friedrich, J. Mancebo, and L. Brochard, Pressure-controlled vs volume-controlled ventilation in acute respiratory failure: a physiology-based narrative and systematic review, *Chest* **148** (2015), 340-355.
[6] G. Murias, U. Lucangelo, and L. Blanch, Patient-ventilator asynchrony, *Curr Opin Crit Care* **22** (2016), 53-59.
[7] K.G. Mellott, M.J. Grap, C.L. Munro, C.N. Sessler, P.A. Wetzel, J.O. Nilsestuen, and J.M. Ketchum, Patient ventilator asynchrony in critically ill adults: frequency and types, *Heart Lung* **43** (2014), 231-243.
[8] D.L. Grieco, M.M. Bitondo, H. Aguirre-Bermeo, S. Italiano, F.A. Idone, A. Moccaldo, M.T. Santantonio, D. Eleuteri, M. Antonelli, J. Mancebo, and S.M. Maggiore, Patient-ventilator interaction with conventional and automated management of pressure support during difficult weaning from mechanical ventilation, *J Crit Care* **48** (2018), 203-210.
[9] Acute Respiratory Distress Syndrome Network, R.G. Brower, M.A. Matthay, A. Morris, D. Schoenfeld, B.T. Thompson, and A. Wheeler, Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome, *N. Engl J Med* **342** (2000), 1301-1308.
[10] L. Gattinoni, J.J. Marini, A. Pesenti, M. Quintel, J. Mancebo, and L. Brochard, The 'baby lung' became an adult, *Intensive Care Med* **42** (2016), 663-673.
[11] G. Rehm, J. Han, B. Kuhn, J.P. Delplanque, N. Anderson, J. Adams, and C.N. Chuah, Creation of a robust and generalizable machine learning classifier for patient ventilator asynchrony, *Methods Inf Med*, **57** (2018), 208-219.
[12] P.D. Sottile, D. Albers, C. Higgins, J. Mckeehan, and M.M. Moss, The association between ventilator dyssynchrony, delivered tidal volume, and sedation using a novel automated ventilator dyssynchrony detection algorithm, *Crit Care Med* **46** 2018, e151-e157.
[13] G.B. Rehm, B.T. Kuhn, J.P. Delplanque, E.C. Guo, M.K. Lieng, J. Nguyen, N.R. Anderson, and J.Y. Adams, Development of a research-oriented system for collecting mechanical ventilator waveform data, *J Am Med Inform Asso*, **25(** 2017), 295-299.
[14] J.Y. Adams, M.K. Lieng, B.T. Kuhn, G.B. Rehm, E.C. Guo, S.L. Taylor, J.P. Delplanque, and N.R. Anderson, Development and validation of a multi-algorithm analytic platform to detect off-target mechanical ventilation, *Sci Rep* **7** (2017), 14980.
[15] D. Colombo, G. Cammarota, M. Alemani, L. Carenzo, F.L. Barra, R. Vaschetto, A.S. Slutsky, F. Della Corte, and P. Navalesi, Efficacy of ventilator waveforms observation in detecting patient-ventilator asynchrony, *Crit Care Med* **39** (2011), 2452-2457.
[16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, Scikit-learn: machine learning in python, *J Machine Learn Res* 12 (2011), 2825-2830.
[17] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, Automatic differentiation in pytorch, In: *31st Conference on Neural Information Processing Systems*, Long Beach, CA, USA, 2017, 1-4.
[18] M.A. Hearst, S.T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, Support vector machines, *IEEE Intelligent Systems* **13** (1998), 18-28.
[19] F.A. Gers, J. Schmidhuber, and F. Cummins, Learning to forget: continual prediction with LSTM, In: *9th International Conference on Artificial Neural Networks: ICANN '99*, Edinburgh, UK, 1999, 850-855.
[20] L. Breiman, Random Forests, *Machine Learning*, **45** (2001), 5-32.

Address for correspondence

For correspondence: Gregory B. Rehm: grehm@ucdavis.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190236

Analyzing Social Media Data to Understand Consumer Information Needs on Dietary Supplements

Rubina F. Rizvi[a,b]*, Yefeng Wang[a] *, Thao Nguyen[c], Jake Vasilakes[a,b], Jiang Bian[d], Zhe He[e], Rui Zhang[a,b]

[a] *Institute for Health Informatics, University of Minnesota, Minneapolis, MN, USA,*
[b] *Department of Pharmaceutical Care & Health Systems, University of Minnesota, Minneapolis, MN, USA*
[c] *Data Science, University of Minnesota, Minneapolis, MN, USA*
[d] *Department of Health Outcomes & Biomedical Informatics, University of Florida, Gainesville, FL, USA,*
[e] *School of Information, Florida State University, Tallahassee, FL, USA,*
*(*Equal contributor first authors)*

Abstract

Despite the high consumption of dietary supplements (DS), few reliable, relevant, and comprehensive online resources could satisfy information seekers. This research study aims to understand consumer information needs on DS using topic modeling, and to evaluate accuracy in correctly identifying topics from social media. We retrieved 16,095 unique questions posted on Yahoo! Answers relating to 438 unique DS ingredients mentioned in sub-section, "Alternative medicine" under the section, "Health". We implemented an unsupervised topic modeling method, Correlation Explanation (CorEx) to unveil the various topics in which consumers are most interested. We manually reviewed the keywords of all the 200 topics generated by CorEx and assigned them to 38 health-related categories, corresponding to 12 higher-level groups. We found high accuracy (90-100%) in identifying questions that correctly align with the selected topics. The results could guide us to generate a more comprehensive and structured DS resource based on consumers' information needs.

Keywords:
Dietary supplements, social media, topic modeling.

Introduction

Dietary supplements (DS) usage has gained popularity in recent years with almost 52% of U.S. adults reporting the use of one or more supplement [1]. This high DS usage is especially common among adults aged ≥60 years, where 70% have reported using one or more DS [2]. In spite of this escalating trend in DS consumption across a wide range of consumers, there are not many online resources that consumers could refer to for DS information that is personalized, reliable, succinct, up-to-date, and in a language that is easily comprehensible by a lay-person.

In recent years, the internet has emerged as an important source of health-related information providing an opportunity for people to search online for free health information. According to a Pew Research Center report, 80% of internet users have looked for health information online [3, 4]. This would be especially true in the case of DS as its use is primarily self-initiated rather than based on clinicians' recommendations [5]. Existing online DS health information resources in the U.S. can range from open access, publicly available databases, e.g., Food and Drug and Administration (FDA) [6]; Office of Dietary supplements (ODS) [7]; Dietary Supplement Label Database (DSLD) [8], to commercial databases that often require a paid subscription, e.g., Natural Medicines (NM) [9]. When it comes to personalized queries from consumers, the information is often consolidated under online resources such as "Frequently Asked Questions". However, the information dissipated from such resources is often very basic, non-specific, and not very helpful.

The rapid growth of digital data in today's world, especially in the healthcare domain, offers great opportunities for secondary use in clinical research. Topic modeling [10] has been an area of great interest and to date, several studies have been conducted to make use of electronic data and utilize this novel methodology. The reason for topic modeling's growing popularity is the area's ability to reveal the latent structure and groupings of the underlying corpus without any prerequisite knowledge. Some of the applications of topic modeling in healthcare research include: analyzing clinical notes from Electronic Health Record (EHR) data; discovering and understanding health care trajectories [11]; identifying medication prescribing patterns [12]; mining adverse events of DS from product labels [13]; and discovering health topics in social media [14, 15] among various others.

There are various social Questions and Answers (Q&A) sites and online forums within health communities, e.g., Yahoo! Answers, allowing one to seek information through posting questions and receiving answers from others users (e.g., consumers, health professionals) [16]. Previously, we have used Yahoo! Answers data in several studies e.g., to investigate the terminology and language gap between health consumers and health professionals [17]; to mine consumer friendly medical terms to enrich consumer health vocabulary [16]; and to understand the information needs for diabetes patients about their laboratory results [18].

The purpose of this research study is to understand the information needs of DS consumers by analyzing questions coming directly from consumers and in their own language. The goal is achieved by using Correlation Explanation (CorEx) - a topic modeling algorithm on the title and body of each question under the Q&A section of the Yahoo! Answers database in order to unveil the "topics" around DS information needs. We generated a list of coherent topics that more accurately represent the areas of DS-related information and associated DS ingredients that consumers are most interested in. We will also evaluate the accuracy of the CorEx method in correctly identifying the topics from social media. In the future, the knowledge gained from this study could be used as a guide for developing more meaningful DS resources for consumers that are better aligned with their information needs.

Methods

Figure 1 illustrates the overview of the methods. We extracted and pre-processed questions retrieved from the Yahoo! Answers database, focusing on questions around DS. We performed topic modeling using CorEx in order to understand DS-related topics and categories that consumers are most interested in. We then evaluated the accuracy of the topic modeling methodology by manually reviewing a subset of top ranked questions. We further investigated the actual DS ingredients associated with all the questions under each topic.

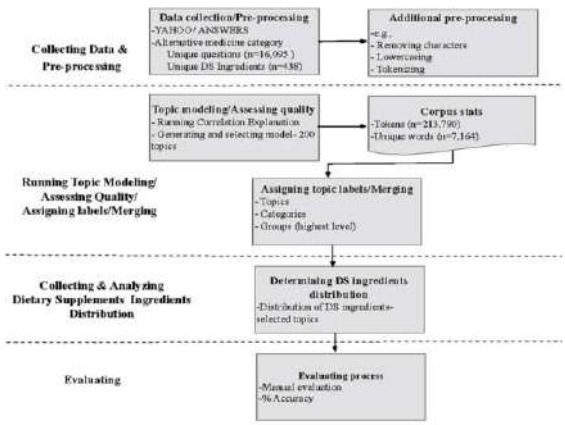


Figure 1— Process overview

Collecting and Pre-processing Data

We collected in total 2,820,179 Yahoo! Answer questions and the corresponding answers posted under 21 sub-categories belonging to the main category "Health". We further extracted 112,090 questions (including their titles and contents) from one of the sub-categories "Alternative Medicine". We then matched the preferred DS names in "iDISK", the first Integrated Dietary Supplements Knowledge base where DS related information is represented in a comprehensive and standardized form [19], with the DS ingredient name in the questions. After two assessors (YW, RR) had manually reviewed the matched preferred names, we cleaned up the DS ingredient names list based on the following rules: 1) only including ingredients with more than 5 matched questions; 2) excluding commonly consumed everyday food/drink items, e.g., fruits, vegetables, wine, caffeine, and water; 3) excluding body parts, e.g., adrenal cortex, brain, and stomach; and 4) excluding recreational drugs e.g., marijuana, poppy seed. Only the questions that exactly matched the DS ingredient names on this list were kept.

These questions were further pre-processed by subject matter experts (TN, JV) and used for topic modeling. We removed all ingredient mentions within the questions to understand the information needs non-specific to certain DS. Each question was then lower-cased and tokenized. Special characters, hyperlinks, and common stop-words (e.g., 'I', 'you', etc.) were removed, and each word was normalized using the normalized string generator (Norm) from the SPECIALIST NLP tool [20]. We only considered words that had at least 3 characters, since any word shorter than that was usually not meaningful. We also removed words that occurred fewer than five times, or more than 85% of the time, as they might not contribute much to the question.

Identifying Topics for DS Questions

In our preliminary investigation of different topic modeling strategies, we found that Correlation Explanation (CorEx) [21] discovered the most coherent topics compared to Latent Dirichlet Allocation (LDA). In contrast to LDA, which defines a generative model for inferring topics, CorEx discovers topics by maximizing the mutual information between words and topics. A subjective assessment of topic quality was performed by two assessors/co-authors and subject matter experts (YW, RR). A topic was considered "coherent" by the experts if assessors found a clear semantic criterion that unites the words under a particular topic. In total, we evaluated several results corresponding to various CorEx models on different numbers of topics (n = 100, 150, 200, 250). Comparing topic modeling results from 100 to 250 topics, we found the model with 200 topics yields the most coherent topic categories.

The selected model was further analyzed and assigned topic names after mutual agreement between two assessors (YW, RR). The "topics" with similar themes were then merged into "categories" (e.g., gastrointestinal disorders, psychiatric disorders) that were further condensed into higher level "groups" (e.g., "uses and symptoms"). For the group, "uses and symptoms", we utilized System Organ Classification (SOC) created by the Medical Dictionary for Regulatory Activities (MedDRA), a medical terminology used to classify adverse event information associated with the use of biopharmaceuticals and other medical products [22].

Topic Evaluation

To evaluate the accuracy of the topic modeling, we selected 15 topics and extracted their corresponding 10 questions with highest ranked probabilities. Manual review (RR, YW) was conducted to determine if the extracted questions correctly aligned with topics generated by the above topic modeling methods. The measure of correctness was reported as percentage accuracy. We also extracted the DS ingredient names corresponding to each topic in order to explore the distribution of ingredients names across various topics. We also reported the DS ingredients associated with most questions for selected topics.

Results

Question Data and Topic Analysis

The final list consisted of 438 unique DS terms in total associated with 16,095 unique matching questions. After data pre-processing, our corpus contained a total of 213,790 tokens, which made up of 7,164 unique words.

From the 200 topics generated by CorEx modeling method, the domain experts (RR, YW) identified topics with similar themes and classified them into 38 unique categories by (Table 1). The 38 unique categories were further summarized into the following 12 higher level groups: uses or adverse effects, product-related, healthy lifestyle, information resources/scientific evidence, addiction, time of use qualifier, sleep disorder, interventions, adverse effect in general, health benefits, mind and body, and population qualifier. The distribution of higher-level groups and number of their associated categories is provided below (Figure 2).

After evaluating the top 10 ranked questions for selected topics, we reported accuracy as number and percentage of questions that correctly aligns with the generated topic. Table 1 lists examples of selected topic groups, their associated categories along with the top 15 most probable words and common ingredients mentions.

Table 1— Selected Topic Categories with Associated Features and Accuracy of Topic Modeling

Topic Groups	*Topic categories (Topic Index)*	*Representative key words*	*Question example*	*Number of correctly matched (accuracy)*	*Representative supplements*
Uses & adverse effects	Gastrointestinal disorders (65)	constipation, laxative, fiber, enema, suppository, constipate, dulcolax, docusate, insoluble, poop, saline, fleet, along with, bum, fecal	Want to take magnesium for health, but it irritates my IBS? Does anyone have any advice for taking Mg?	10 (100%)	Magnesium, Senna, Castor, Probiotics, Glycerin
	Musculoskeletal disorders (93)	arthritis, bracelet, rheumatoid, knee, magnetic, anabolic, steroid, juvenile, sheet, gout, rheumatism, mattress, wonderfully, bangle, supplemental	Is hyaluronic acid supplement safe for children with juvenile rheumatoid arthritis?	10 (100%)	Copper, Fish Oil, Vitamin D
	Psychiatric Disorders (11)	john, anxiety, depression, wort, htp, antidepressant, calm, social, root, 5htp, sam, tryptophan, zoloft, ssri, withdrawal	Is there alternative solution for anxiousness. I'd like to know is if anybody has any feedback on Buspar, Kava root, Passion flower & St. John's wort?	10 (100%)	Valerian, St. John's Wort, SAM-e
	Respiratory, thoracic/ mediastinal disorders (1)	remedy, throat, sore, home, lemon, salt, hot, cure, natural, gargle, warm, epsom, voice, soothe, homeopathic	How can I soothe my sore throat? I have been drinking warm tea with honey & gargle some salt water but it doesn't work.	9 (90%)	Honey, Ginger, Vitamin C, Garlic
	Skin & subcutaneous tissue disorders (23)	skin, face, acne, wash, foot, hand, red, spot, itchy, facial, cream, wrinkle, mask, oily, swell	What are some natural ways to get rid of pimples (home remedies)? I have heard about Aloe	10 (100%)	Apple cider vinegar, Fish oil, Honey, Tea tree oil
	Cardio-vascular/blood & lymphatic system disorders (2)	blood, level, low, pressure, high, normal, cholesterol, , cause, heart, disease, diagnose, raise, thin, flow	Does anyone know about an alternative medicine way of treating high blood pressure? I have heard of l-arginine to be effective.	9 (90%)	Iron, Fish Oil, Sodium
	Endocrine disorders (68)	thyroid, hypothyroidism, synthroid, levothyroxine, hypothyroid, gut, patient, syndrome, leaky, radioactive, testimony, underactive, iodide, success, advisable	Are there natural alternatives after having radioactive iodine treatment for Graves' disease?	10 (100%)	Iodine, Kelp, Vitamin D
	Infections & infestations (31)	infection, yeast, bladder, antibiotic, treat, kill, bite, mannose, parasite, mosquito, coat, douche, poultice, intestinal, frequent	Home remedies for yeast infections? I have heard where diluted apple cider vinegar in a bath tub could provide temporary relief	10 (100%)	Cranberry, Garlic, Apple Cider Vinegar,
	Pregnancy, puerperium/ perinatal conditions (40)	control, birth, period, pregnancy, pregnant, miscarriage, irregular, headache, hormone, tension, 1st, abortion, fertility, defect, endometriosis	Is it safe to drink penny royal tea along with parsley tea to induce a miscarriage?	10 (100%)	Vitamin C, Iron, Dong Quai
	Dental & gingival conditions (124)	pot, tooth, eliminate, desperate, prune, brush, resistant, walnut, wisdom, piss, neti, dentist, dependency, swish, tonsillitis	I am getting my wisdom teeth out & I would like to know if clove oil will help in pain in my gums?	10 (100%)	Apple Cider Vinegar, Silver, Garlic, Clove
Addiction	Smokables (21)	smoke, weed, marijuana, cigarette, legal, pipe, bowl, roll, tobacco, illegal, bud, blunt, smoker, kush, hash	What is the difference between Salvia-A & Salvia divinorum? I smoked salvia divinorum	10 (100%)	Damiana, Catnip, Mullein, Salvia divinorum, Kratom, Clove
Product-related	Dose/dose form/ preparation (43)	mcg, 1000, complex, omega, 5000, 2000, 2500, multiple, adult, 2000iu, standardize, milligram, cla, strength, gnc	Is the dose of vitamin A 2500 IU? Why can I not find that in a supplement?	10 (100%)	Biotin, Vitamin D, Vitamin C, Fish Oil, Vitamin B12, Folic Acid, Vitamin A
Sleep disorder	Sleeping (50)	fall, asleep, 2am, cry, couch, 3am, deaf, sleeper, category, pulmonary, carpet, regardless, proportion, haven't, recreational	Need more info about Melatonin? I have trouble falling asleep	8 (80%)	Melatonin, Valerian
Frequency/ Time	Frequency/Time reference (9)	even, month, since, may, though, well, always, never, pretty, end, fine, mean, without, course, quite	Chromium supplements... do they work? I have been taking them for a few weeks now	7 (70%)	Iron, Vitamin D, Cranberry
Health life style	Weight control (19)	weight, lose, gain, loose, berry, cardio, primarily, weight, workout, lift, shed, cambogia, underweight, usda	Would it be safe to take a weight loss pill with the supplements?	9 (90%)	Acai, Apple Cider Vinegar, Honey, Garcinia,

We found the percent accuracy for most of the selected topics is between 90% - 100%, except sleep (80%) and frequency/time categories (70%). "Use and adverse effects" is the most dominant topic group and accounted for 50 topics out of 200. Under this topic, there were 15 categories classified based on MedDRA SOC (Figure 3).

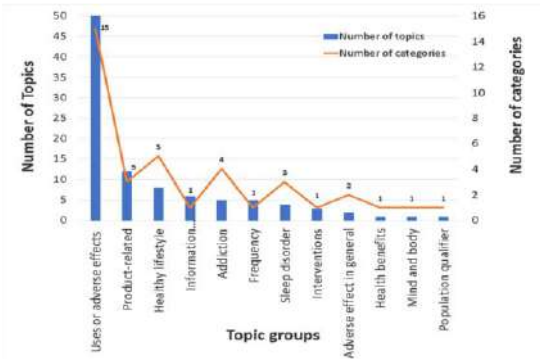


Figure 2— Distribution of TopicGgroups and the Associated Number of Categories.

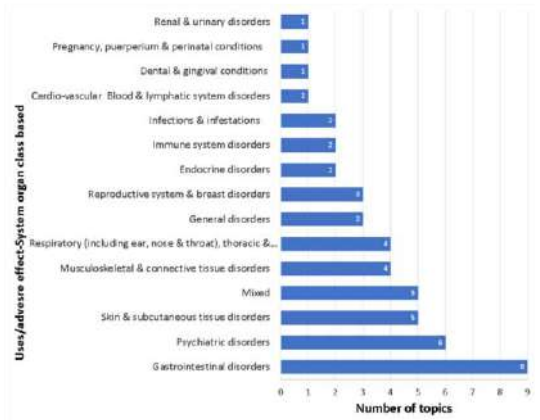


Figure 3— Distribution of uses/adverse effects related categories based on system organ class (SOC)

Dietary Supplements Associated with Most Questions

We also extracted the DS ingredient names associated with most questions corresponding to a particular topic in order to explore the distribution of most commonly discussed ingredients. Only DS ingredients associated with ≥10% of questions under a specific topic were reported (Table 2).

Discussion

In this study, we employed CorEx topic modeling over user-generated questions coming from the Yahoo! Answers data in order to better understand the information needs of consumers. We also discovered interesting information on the distribution of DS ingredients across topics of special interest to consumers. This research effort further validates the feasibility of topic modeling to extract important information hidden in large corpus of social media data.

Applying CorEx topic modeling methods, we were able to accurately identify 12 topic groups. The top three groups with the most number of respective assigned categories and topics, which can be regarded as the information most sought by consumers, are: "use and adverse effects", "product-related", and "healthy life style" (Figure 2). Extracted information pertaining to any symptom or sign could either be an indication or an adverse event of a DS, (e.g., diarrhea, abdominal pain, palpitations, headaches); therefore, uses and adverse effects were combined as one group, "use and adverse effects". We found a higher number of topics and the associated number of questions concerning: gastrointestinal system (specifically diarrhea and constipation); psychiatric (mainly anxiety and depression); and skin and subcutaneous tissues (primarily acne and UV protection). We also had a "mixed group", having keywords corresponding to more than one system. For "product-related groups", we merged categories like dose, dose from, preparation because of their co-occurrence under one topic (e.g., Topic #43). Under the "healthy life style" group, the topics were mostly around eating healthy and weight control/exercise.

Table 2— Total Number of Questions Matched for Each Topic The representative ingredient is the one that matched the most questions. The percentage of the questions that mentioned the representative ingredient is shown in the parenthesis. The text in bold represents the ingredient with high percentage of associated questions.

Topic categories Topic Index)	Number of questions matched	Representative Ingredient	Questions containing Representative Ingredient
Gastrointestinal disorders (65)	145	Magnesium	24 (16.55%)
Musculoskeletal disorders (93)	45	Copper	11 (24.44%)
Psychiatric disorders (11)	256	Valerian	45 (17.58%)
Respiratory (including ear, nose & throat), thoracic & mediastinal disorders (1)	**476**	**Honey**	**226 (47.48%)**
Skin & subcutaneous tissue disorders (23)	84	Apple Cider Vinegar	17 (20.24%)
Cardio-vascular/blood & lymphatic system disorders (2)	264	Iron	38 (14.39%)
Endocrine disorders (68)	48	Iodine	11 (22.92%)
Infections & infestations (31)	160	Cranberry	37 (23.12%)
Smokables (21)	134	Damiana	11 (8.21%)
Dose/dose form/preparation (43)	98	Biotin	35 (35.71%)
Sleeping (50)	**45**	**Melatonin**	**29 (64.44%)**
Weight control (19)	171	Acai	26 (15.20%)

We found high accuracy when we identified questions that correctly align with the topic categories/groups (Table 1). We found few low matching accuracy topics also having questions related to other topics, e.g., sleeping disorders topic with questions related to recreational drug, anxiety/depression. We also reported actual DS ingredient names associated with most questions for a particular topic (Table 2). We found a substantially higher percentage of questions for the ingredients "Honey" under respiratory disorders and "Melatonin" under sleep disorders. This information provides essential knowledge on the use of DS for various specific reasons and needs further exploration.

This research study had several limitations. We analyzed only questions belonging to alternative medicine sub-category under "health" section and might have missed dietary supplement occurrences under other sub-categories, e.g., mental health conditions, general health care. We only used preferred DS ingredient names and not their synonyms (e.g., scientific names, common names) to extract the corresponding questions. Also, there are inherent limitations to topic modeling e.g., topics were generated based on the statistical word distribution within the questions and thus topics with incoherent topic keywords were also generated.

Conclusions

This research provides essential insights on extracting and understanding the information needs of consumers around dietary supplements using CorEx-based topic modeling that could identify the relevant topics embedded in a large corpus of Yahoo! Answers data with high accuracy. The knowledge gained here could be used to generate a more comprehensive repository of resource for consumers around dietary supplements usage. Thus, this study is an important contribution in further accentuating the potential benefits of using social media data in the clinical research.

Acknowledgements

This research was supported by the U.S. National Institutes for Health (R01AT009457) (Zhang).

References

[1] E. D. Kantor, C. D. Rehm, M. Du, E. White, and E. L. Giovannucci, "Trends in Dietary Supplement Use Among US Adults From 1999-2012," (in eng), *JAMA,* vol. 316, no. 14, pp. 1464-1474, Oct 2016, doi: 10.1001/jama.2016.14403.

[2] J. J. Gahche, R. L. Bailey, N. Potischman, and J. T. Dwyer, "Dietary Supplement Use Was Very High among Older Adults in the United States in 2011-2014," (in eng), *J Nutr,* Aug 2017, doi: 10.3945/jn.117.255984.

[3] D. J. Amante, T. P. Hogan, S. L. Pagoto, T. M. English, and K. L. Lapane, "Access to care and use of the Internet to search for health information: results from the US National Health Interview Survey," *Journal of medical Internet research,* vol. 17, no. 4, 2015.

[4] S. Fox. "Health topics: 80% of internet users look for health information online. Washington, DC: Pew Internet & American Life Project." http://www.pewinternet.org/2005/05/17/health-information-online/ (accessed 2018).

[5] P. Institute of Medicine . Committee on the Use of Complementary and Alternative Medicine by the American, *Complementary and alternative medicine in the United States*. Washington, D.C. National Academies Press, 2005.

[6] "FDA 101: Dietary Supplements." https://www.fda.gov/ForConsumers/ConsumerUpdates/ucm050803.htm (accessed 2018).

[7] "National Institute Of Health (NIH), Office of Dietary Supplements (ODS)." https://ods.od.nih.gov (accessed 2018).

[8] "Dietary Supplement Label Database (DSLD)." https://www.dsld.nlm.nih.gov/dsld/index.jsp (accessed 2018).

[9]"Natural Medicine (NM)." https://naturalmedicines.therapeuticresearch.com (accessed 2018).

[10] D. M. Blei, "Probabilistic topic models," *Communications of the ACM,* vol. 55, no. 4, pp. 77-84, 2012.

[11] L. Wang, J. Lakin, C. Riley, Z. Korach, L. N. Frain, and L. Zhou, "Disease Trajectories and End-of-Life Care for Dementias: Latent Topic Modeling and Trend Analysis Using Clinical Notes," AMIA 2018 Annual Symposium, 2018.

[12] A. L. Beam *et al.*, "Predictive modeling of physician-patient dynamics that influence sleep medication prescriptions and clinical decision-making," *Scientific reports,* vol. 7, p. 42282, 2017.

[13] Y. Wang, D. R. Gunashekar, T. J. Adam, and R. Zhang, "Mining Adverse Events of Dietary Supplements from Product Labels by Topic Modeling," *Studies in health technology and informatics,* vol. 245, p. 614, 2017.

[14] M. J. Paul and M. Dredze, "Discovering health topics in social media using topic models," *PloS one,* vol. 9, no. 8, p. e103408, 2014.

[15] K. W. Prier, M. S. Smith, C. Giraud-Carrier, and C. L. Hanson, "Identifying health-related topics on twitter," in *International conference on social computing, behavioral-cultural modeling, and prediction*, 2011: Springer, pp. 18-25.

[16] Z. He, Z. Chen, S. Oh, J. Hou, and J. Bian, "Enriching consumer health vocabulary through mining a social Q&A site: A similarity-based approach," *Journal of biomedical informatics,* vol. 69, pp. 75-85, 2017.

[17] M. S. Park, Z. He, Z. Chen, S. Oh, and J. Bian, "Consumers' use of UMLS concepts on social media: diabetes-related textual data analysis in blog and social Q&A sites," *JMIR medical informatics,* vol. 4, no. 4, 2016.

[18] Z. Zhang, Y. Lu, Y. Kou, D. Wu, J.Huh-Yoo, and Z. He. Understanding Patient Information Needs about their Clinical Laboratory Results: A Study of Social Q&A Site. Studies in Health Technology and Informatics. In press.

[19] X. He *et al.*, "Prototyping an Interactive Visualization of Dietary Supplement Knowledge Graph," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2018: IEEE, pp. 1649-1652.

[20] "Lexical tools." https://lexsrv3.nlm.nih.gov/LexSysGroup/Projects/lvg/2013/docs/userDoc/tools/norm.html (accessed 2018).

[21] G. Ver Steeg and A. Galstyan, "Discovering structure in high-dimensional data through correlation explanation," in *Advances in Neural Information Processing Systems*, 2014, pp. 577-585.

[22] "Medical Dictionary of Regulatory Activities (MedDRA)." https://www.meddra.org/about-meddra/quality-controls (accessed 2018).

Address for correspondence

Rui Zhang, PhD, Email: zhan1386@umn.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190237

Semantic Provenance Graph for Reproducibility of Biomedical Research Studies: Generating and Analyzing Graph Structures from Published Literature

Satya S. Sahoo[a], Joshua Valdez[b], Michael Rueschman[b], Matthew Kim[b]

[a] *Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH, USA*
[b] *Department of Medicine, Brigham and Women's Hospital and Beth Israel Deaconess Medical Center, Harvard Medical School, Harvard University, Boston, MA, USA*

Abstract

***Objective**: To characterize the scientific reproducibility of biomedical research studies by query and analysis of semantic provenance graphs generated from provenance metadata terms extracted from PubMed articles. **Methods**. We develop a new semantic provenance graph generation algorithm that uses a provenance ontology developed as part of the Provenance for Clinical and Health Research (ProvCaRe) project. The ProvCaRe project has processed and extracted provenance metadata from more than 1.6 million full text articles from the PubMed database. **Results**. The semantic provenance graph generation algorithm is evaluated using provenance terms extracted from 75 selected articles describing sleep medicine research studies. In addition, we use eight provenance queries to evaluate the quality of semantic provenance graphs generated by the new algorithm. **Conclusion**. The ProvCaRe project has created a unique resource to characterize the reproducibility of biomedical research studies and the semantic provenance graph generation algorithm enables users to effectively query and analyze the provenance metadata in the ProvCaRe knowledge repository.*

Keywords:

Knowledge Bases; Biomedical Ontologies; Health Information Systems

Introduction

Provenance metadata describes the history or origin of data, and therefore plays a critical role in supporting the reproducibility of results from scientific experiments. Scientific reproducibility has received increasing attention in the biomedical research community, including clinical research, drug development, and basic science research, with the publication of the U.S. National Institutes of Health (NIH) "Rigor and Reproducibility guidelines" [6; 15]. Reproducibility of results from biomedical research studies is critical to ensure that valid results are used when designing new research studies, developing new drugs with less likelihood of failing during clinical trials, and ensuring the allocation of limited funding resources to rigorously designed research studies [13; 21]. Provenance metadata therefore represents essential information about research studies, including protocols, data acquisition methods, and statistical analysis models.

However, there has been limited work in the use of provenance metadata to support reproducibility in biomedical research. Therefore, as part of the Provenance for Clinical and Health Research (ProvCaRe) project, we are developing a framework to use semantic provenance metadata for characterizing the reproducibility of biomedical research studies [23]. We have developed a provenance-focused natural language processing (NLP) pipeline to extract provenance metadata terms from more than 1.6 million full-length articles in the National Center for Biotechnology Information (NCBI) PubMed database consisting of 28 million citations [1]. The extracted provenance information is stored in a structured form in the ProvCaRe knowledge repository.

The ProvCaRe knowledge repository is a unique resource for sudying and characterizing the reproducibility of biomedical research studies using a systematic and exhaustive approach. Provenance metadata is intuitively modeled as a graph consisting of nodes and edges, for example *ResearchStudy* → *hadCohortSizeOf* → *5861(participants)*, where *ResearchStudy* and *5861(participants)* are nodes and *hadCohortSizeOf* is an edge [8]. The World Wide Web Consortium (W3C), which is the organization for Web technology standards, developed the PROV specifications to facilitate interoperability and interchage of provenance metadata in data management and Web-based information systems [16; 19]. The PROV specifications can be extended to biomedical domain to model provenance as a graph with formal semantics and terms mapped to ontologies. We used this approach to generate semantic provenance information in the ProvCaRe knowledge repository with terms mapped to a provenance ontology to support queries related to scientific reproducibility.

In this paper, we investigate the generation of provenance graphs from the structured provenance information stored in the ProvCaRe knowledge repository. In particular, we describe a new method to transform structured provenance information into a W3C Resource Description Framework (RDF)-based semantic provenance graph. We also describe an approach to query and analyze the provenance graph using a test case of provenance information from 75 selected articles.

Background

Resources

W3C PROV Specifications

The W3C PROV specifications consists of: (1) the core model called PROV-DM, which defines a common terminology system for representing interoperable provenance metadata [19]; (b) the PROV Ontology (PROV-O) that represents PROV-DM in the W3C Web Ontology Language (OWL2) [12; 16]; and a set of constraints (PROV-Constraints) to define valid forms of provenance conforming to the PROV specifications [4]. We extended PROV-O in the ProvCaRe project to model provenance terms associated with biomedical research studies. In particular, we extended the three core concepts of PROV-O,

namely "prov:Agent", "prov:Entity", "prov:Activity"; and its properties, namely "prov:wasDerivedFrom", "prov:wasGeneratedBy", and "prov:used". The namespace *prov* refers to "*http://www.w3.org/ns/prov#*".

ProvCaRe Knowledge Repository

The provenance terms extracted by the ProvCaRe NLP pipeline are currently stored in the ProvCaRe knowledge repository as a set of three related provenance terms (for a detailed description of the NLP pipeline, please refer to [23]). For example, we stored provenance terms extracted from an article describing a study on sleep-disordered breathing and hypertension by O'Connor et al. [20] as "hypertension status of participants", "was ascertained at", "second follow-up exam" in the ProvCaRe knowledge repository. However, this representation of provenance metadata has some limitations, for example (1) it is verbose (making it difficult to query efficiently); and (2) it cannot be modeled as a RDF graph structure. Given a user query, "*How was the hypertension status of participants determined in this research study?*", the ProvCaRe query interface uses keyword-based matching together with ontology-based "expansion" of query expression for retrieving query results. Therefore, we address these challenges by developing a new algorithm to map the provenance information in the ProvCaRe repository to a semantic provenance graph.

Related Work

Provenance metadata has long been used in fine arts, library management, and computer science to track art items, literary resources, and data elements. In particular, provenance metadata plays a critical role in ensuring data quality [3] and reproducibility of experiment results [17]. Provenance terms describing the context of biomedical research experiments, for example sample size of the experiment, randomization techniques used, protocol used for selection of participants in a research study cohort, are critical for enabling the reproducibility of research study results [15].

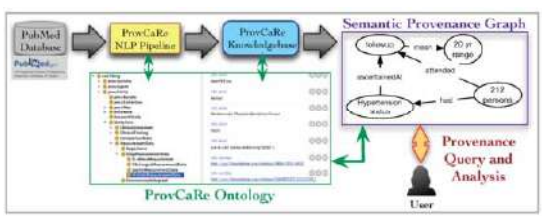


Figure 1: Architecture of the ProvCaRe framework and its components, including the ProvCaRe ontology

Although previous work in the biomedical research domain had limited or no focus on provenance metadata for reproducibility, a number of guidelines and best practices have been adopted to improve reporting of research studies to support reproducibility, for example the NIH "Rigor and Reproducibility guidelines" [10] and the Ontology for Clinical Research (OCRe) project, which developed a formal model of clinical research studies [24]. Projects that have explored the role of provenance in scientific reproducibility in other domains include work by Chirigati [5] and [18]. Many projects have also explored querying of provenance graphs using query language for workflow provenance [2] and Semantic Web technologies such as SPARQL [17; 22].

To the best of our knowledge, ProvCaRe is first project to generate a semantic provenance graph from unstructured text to support queries for scientific reproducibility.

Methods

An overview of the ProvCaRe platform and its components is shown in Figure 1 illustrating a provenance-focused NLP pipeline, storage of provenance triples in the knowledge repository, and the creation of semantic provenance graphs. The ProvCaRe ontology plays a central role as the reference knowledge model during both the extraction of provenance metadata and generation of provenance graphs.

ProvCaRe ontology for biomedical provenance metadata

The ProvCaRe ontology was developed by extending the classes and properties of the W3C PROV ontology using OWL2 description logic [16]. The two key design objectives of the ProvCaRe ontology are: (1) modeling the essential components of a research study, and (2) incorporating terms from existing domain ontologies for terms specific to different medical disciplines to provide appropriate level of granularity for provenance terms. Example terms modeled in the ProvCaRe ontology to meet the first objective are: the method used to conduct the study, the instruments used to record or analyze data in the study, and the different types of data used in the research study. Similarly, to meet the second objective, the ProvCaRe ontology currently models terms specific to sleep medicine, endocrinology, and neurological disorders. Example terms in the ProvCaRe ontology include: polysomnogram (sleep medicine), electroencephalogram (neurology), and Hyperaldosteronism (endocrinology). These terms are mapped to existing biomedical ontologies, such as SNOMED CT, National Cancer Institute (NCI) Thesaurus, and Gene Ontology (GO).

These mappings allow the provenance terms mapped to ProvCaRe ontology classes to reference terms in existing biomedical ontologies, which supports interoperability of the ProvCaRe knowledgebase with existing biomedical databases, such as NCBI Gene and Protein database. The ProvCaRe NLP pipeline uses the ProvCaRe ontology to identify and extract provenance terms from full-text articles in the PubMed database. At present, the ProvCaRe ontology consists of about 300 classes and 40 object properties. The ProvCaRe ontology plays a key role in the generation of the semantic provenance graphs from the provenance terms as it provides the schematic model to represent the nodes and edges of the graph structure.

Generating semantic provenance graph

The W3C Resource Description Framework (RDF) is a widely used Semantic Web technology for representing graph structures [11]. A RDF graph consists of "subject", "predicate", and "object". For example, *ResearchStudy* → *wasApprovedBy* → *EthicsCommittee*, where *ResearchStudy* is the "subject", *wasApprovedBy* is the "predicate", and *EthicsCommittee* is the "object" of the RDF triple. Two or more RDF triples can be aggregated to form a RDF graph. The RDF graphs can be queried using the W3C SPARQL query language that also supports reasoning to infer new knowledge from RDF graphs.

We use RDF syntax to generate the semantic provenance graph in the ProvCaRe project. Figure 2 illustrates a new algorithm developed in the ProvCaRe project to transform structured provenance metadata extracted from unstructured text into graph structures with "subject", "predicate", and "object" mapped to ontology terms. The algorithm uses the output of the ProvCaRe NLP pipeline, which consists of the constituents of the sentence such as nouns, verb ("chunked sentence"), and output of the named entity recognition (NER) as listed in Line 1. The provenance-focused NER module has an important feature that uses the ProvCaRe ontology together with existing

biomedical NLP resources such as the Unified Medical Language System (UMLS)-based MetaMap and the National Center for Biomedical Ontologies (NCBO) annotator for entity recognition [14]. This comprehensive approach to provenance NER ensures that the ProvCaRe NLP pipeline is able to effectively identify and extract provenance terms from unstructured text.

In Line 3, 4, and 5 of the graph generation algorithm (Figure 2), we use mappings between structured provenance and ProvCaRe ontology classes to generate concise representation of provenance terms, which can be used to generate

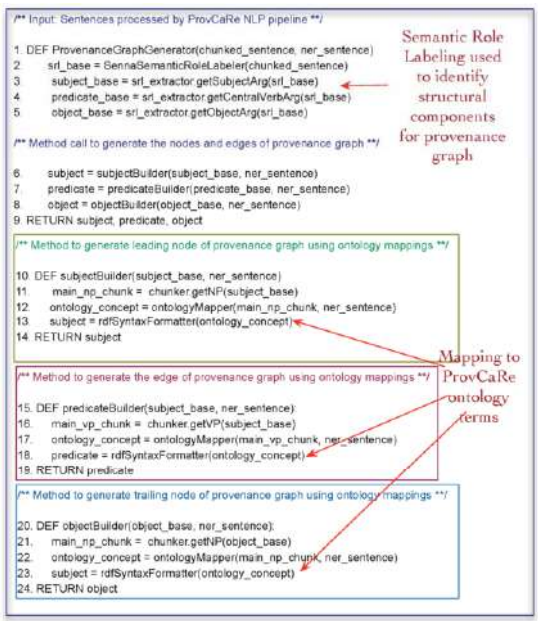


Figure 2: the algorithm used to generate a sematic provenance graph from terms stored in the ProvCaRe knowledgebase

components of the provenance graph. For example, the sentence from an article by O'Connor et al. describing a prospective cohort study on sleep disordered breathing and hypertension [20], "*The institutional review boards of all participating institutions approved the study, and participants signed a consent form*" describes provenance of ethical approval and participant consent for the research study. The graph generation algorithm following the steps described in Figure 2 generated a semantic provenance graph (Figure 3).

It is important to note the differences between the verbose structured provenance information generated by the ProvCaRe NLP pipeline and the concise graph representation generated by the new algorithm. For example, the provenance RDF graph supports querying by graph query languages, including SPARQL and the use of reasoning techniques based on RDF semantics [9]. An important feature of the semantic provenance graph generation algorithm is its ability to correctly identify different constituents of a sentence containing provenance terms and generate one or more provenance RDF triples that accurately represents the information content of a sentence. As shown in Figure 3, the algorithm correctly generates two provenance triples representing: (1) the formal approval for the study; and (2) the consent forms signed by the study participants.

These individual provenance triples can be linked or aggregated together to form the corresponding semantic provenance graph for analysis. This approach allows the ProvCaRe query feature to support graph traversal-based queries. For example, using the predicate "recruited" to link the two provenance RDF triples in Figure 3, a graph traversal query can identify that a research study used consent forms as part of the recruiting protocol.

Query and analysis of provenance graphs

As we discussed earlier, a key objective of the ProvCaRe

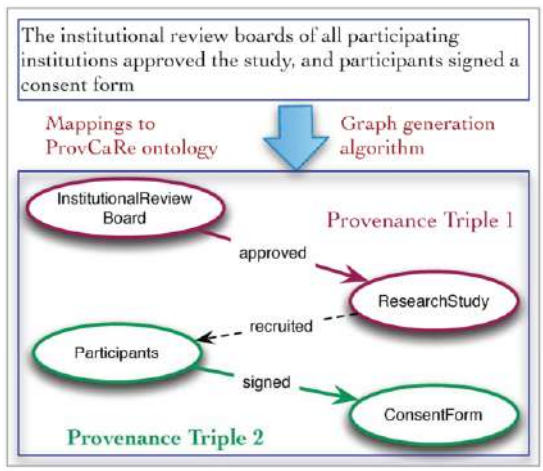


Figure 3: An example semantic provenance graph generated from provenance terms extracted from a sentence in an article describing a prospective cohort study by O'Connor et al.

project is to query and analyze these semantic provenance graphs to characterize the reproducibility of research studies. Therefore, the ProvCaRe project developed a Web-based user interface to support user queries that search for research studies associated with their research hypothesis. The query results include the provenance metadata extracted from the research studies associated with the user query. This provenance query and analysis interface is accessible at https://provcare.case.edu, which allows users to explore the ProvCaRe knowledge repository.

The current version of the ProvCaRe query module uses a "keyword-based search" together with ontology-driven "query expansion" approach to query the structured provenance information in the ProvCaRe knowledgebase. As part of our ongoing work, we are implementing a SPARQL query engine in ProvCaRe to query the semantic provenance graph generated by the new graph generation algorithm described in this paper. In the next section, we evaluate the results of the graph generation algorithm.

Results

We used a corpus of 75 articles describing sleep medicine research studies for evaluation of the graph generation algorithm. This corpus of articles was selected from sleep medicine reearch studies that made their data available as part of the National Sleep Research Resource (NSRR) [7]. However, we note that as part of the ProvCaRe project, we have processed more than 1.6 million full text articles available in the PubMed database. For example, Figure 4 illustrates the distribution of 35 million terms extracted from the 1.6 million articles categorized according to a select number of ProvCaRe

ontology classes (terms are organized by their frequency in the figure). Provenance terms describing the design of research studies, for example "comparative" research studies have the highest frequency, whereas provenance terms describing the statistical data analysis term "variance" has the lowest frequency.

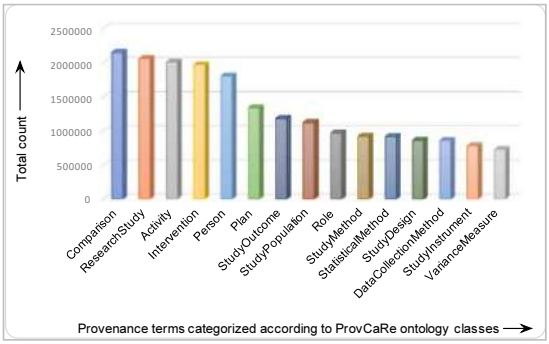


Figure 4: Distribution of provenance terms from 1.6 million full text articles corresponding to ProvCaRe ontology classes

Evaluation of the graph generation algorithm and properties of the semantic provenance graph

In Table 1, we describe the characteristics of the semantic provenance graph generated from the corpus of 75 papers with 10,875 provenance RDF triples generated from more than 30,000 sentences. The 33,783 nodes of the provenance RDF triples were mapped to various biomedical ontologies, for example the term *Population* was mapped to SNOMED CT and the term *Study* was mapped to the ProvCaRe ontology. It is interesting to note that more than 6900 sentences generated more than 1 provenance RDF triple, which is a key feature of the semantic provenance graph generation algorithm as it accurately reflects the information content of the sentence.

Table 1– Properties of semantic provenance graph generated from 75 full-text research articles

Type of characterization	Count
Number of provenance RDF triples	10,875
Provenance terms mapped to ontology classes	33,783
Total number of sentences processed from 75 papers	30,450 sentences
Total number of sentences with more than 1 provenance RDF triple generated	6917 sentences

An important feature of the semantic provenance graph generation algorithm is scalability in terms of time required to process large number of provenance terms. Figure 5 shows the performance of the algorithm for five sets of articles with increasing size. The results show that the algorithm scales almost linearly as the number of articles increases. As part of our ongoing work, we are further optimizing the performance of the algorithm as we aim to process all 35 million provenance terms extracted from the 1.6 million articles in ProvCaRe .

Qualitiative evaluation of the semantic provenance graph

In addition to the quantitative analysis of the provenance triples generated by the semantic provenance graph generation algorithm, we qualitatively evaluated the graph generated for the research study described by O'Connor et al. (this article was selected at random from the corpus of 75 articles). We used a set of eight queries to represent important contextual information about not only this specific research study reported by O'Connor et al., but also other biomedical research studies. The queries are listed below:

Q1. What were the inclusion and exclusion criteria?

Q2. How were subjects assigned to groups?

Q3. How was blood pressure measured?

Q4. What was the definition of hypertension?

Q5. What method was used to perform sleep studies?

Q6. How were apnea-hypopnea indices categorized?

Q7. When were repeat blood pressure measurements checked?

Q8. How was the data analyzed?

The objective of evaluating these queries over the semantic provenance graph was to evaluate the quality of the graph. The query results were manually reviewed and were found to closely correspond to the provenance information available in the original article by O'Connor et al. We propose to perform a systematic evaluation of the provenance graph query results using the sematic provenance graph generated from the 1.6 million articles in future.

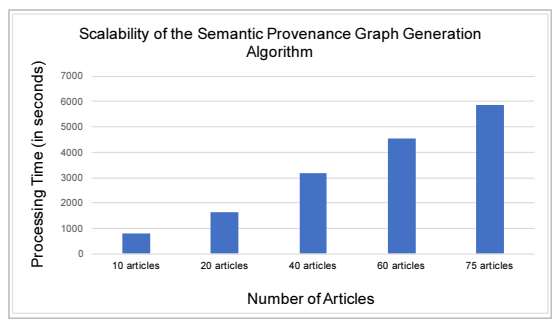


Figure 5: The performance of the semantic provenance graph generation algorithim over five different sets of articles

Discussion

The use of the semantic provenance graph structure to underpin the ProvCaRe knowledge repository is expected to significantly improve query and analysis features available to users. This will in turn enable the biomedical research community to query and analyze provenance metadata reported by various research studies in the context of scientific reproducibility. However, there are several practical challenges that need to be addressed before the ProvCaRe knowledge repository can play a greater role in characterizing the reproducibility of research studies. These challenges include the lack of detailed description of a research study in published articles, for example protocol related to data collection, data analysis, and validation of study hypothesis.

This challenge can be addressed by making available the provenance information associated with research studies in public metadata repositories. Together with existing data repositories, these provenance metadata repositories can support reproducibility of published studies. An important challenge in the ProvCaRe project is the scalability of graph query operations as we progressively convert all the provenance

information in the ProvCaRe knowledgebase into an RDF provenance graph. We note that the ProvCaRe query interface includes a novel provenance-based ranking feature that allows users to assign weight values to ProvCare ontology terms to rank query results. Together with the provenance-based ranking feature and time complexity of graph operations, the current performance of the ProvCaRe query interface requires significant optimization to support real-time user interactions. We are exploring the use of customized indexing techniques to address this challenge.

Conclusions

We demonstrated the practical use of provenance metadata extracted from full-text articles in the PubMed database to characterize the reproducibility of research studies. In particular, we described the development and application of a new graph generation algorithm to transform verbose provenance metadata extracted by the ProvCaRe NLP pipeline into semantic provenance graphs, which can support the query and analysis of provenance metadata. The ProvCaRe knowledgebase, which includes provenance data extracted from more than 1.6 million full-text PubMed articles, is a unique resource to characterize the scientific reproducibility of biomedical research studies and its representation as a semantic provenance graph will allow users to perform knowledge discovery tasks using ontology-driven reasoning techniques.

Acknowledgements

This work is supported by the NIH-NIBIB Big Data to Knowledge 1U01EB020955 grant, and NSF grant# 1636850.

References

[1] PubMed, https://www.ncbi.nlm.nih.gov/pubmed/.

[2] M.K. Anand, Bowers, S., Ludäscher, B., Techniques for efficiently querying scientific workflow provenance graphs. , in: *Proceedings of the 13th international Conference on Extending Database Technology,* S.S. I. Manolescu, J. Teubner, M. Kitsuregawa, A. Leger, F. Naumann, A. Ailamaki, F. Ozcan, ed., ACM, New York, NY, Lausanne, Switzerland, 2010, pp. 287-298.

[3] P. Buneman, Davidson, S., Data provenance - the foundation of data quality, in, 2010.

[4] J. Cheney, Missier, P., Moreau, L., Constraints of the PROV Data Model, in: *W3C Recommendation*, World Wide Web Consortium W3C, 2013.

[5] F.S. Chirigati, Shasha, D.E., Freire, J., ReproZip: Using Provenance to Support Computational Reproducibility, in: *Theory and Practice of Provenance (TaPP)*, 2013.

[6] F.S. Collins, Tabak, L.A., Policy: NIH plans to enhance reproducibility, *Nature* **505** (2014), 612-613.

[7] D.A. Dean, Goldberger, A.L., Mueller, R., Kim, M., Rueschman, M., Mobley, D., Sahoo, S.S., Jayapandian, C.P., Cui, L., Morrical, M.G., Surovec, S., Zhang, G.Q., Redline, S., Scaling up scientific discovery in sleep medicine: the National Sleep Research Resource., *SLEEP* **39** (2016), 1151–1164.

[8] Y. Gil, Cheney, J., Groth, P., Hartig, O., Miles, S., Moreau, L., Pinheiro da Silva, P., Coppens, S., Garijo, D., Manuel Gomez, J., Missier, P., Myers, J., Sahoo, S.S., Zhao, J., Provenance xg final report, in: *W3C Technical Report*, W3C, 2010.

[9] P. Hayes, RDF Semantics, in: *W3C Recommendation,* B. McBride, ed., World Wide Web Consortium, 2004.

[10] N.I.o. Health, Principles and Guidelines for Reporting Preclinical Research, in, 2016.

[11] I. Herman, Adida, B., Sporny, M., Birbeck, M., RDFa 1.1 Primer - Second Edition, in: *W3C Working Group Note*, World Wide Web Consortium (W3C), 2013.

[12] P. Hitzler, Krötzsch, M., Parsia, B., Patel-Schneider, P.F., Rudolph, S., OWL 2 Web Ontology Language Primer, in: *W3C Recommendation*, World Wide Web Consortium W3C, 2009.

[13] J.P.A. Ioannidis, Why Most Published Research Findings Are False, *PLoS Med* **2** (2005), e124.

[14] C. Jonquet, Shah, N.M., Musen, M.A., The Open Biomedical Annotator, in: *AMIA Summit on Translat Bioinformatics*, AMIA, San Francisco, 2009, pp. 56-60.

[15] S.C. Landis, Amara, S.G., Asadullah, K., Austin, C.P., Blumenstein, R., Bradley, E.W., Crystal, R.G., Darnell, R.B., Ferrante, R.J., Fillit, H., Finkelstein, R., Fisher, M., Gendelman, H.E., Golub, R.M., Goudreau, J.L., Gross, R.A., Gubitz, A.K., Hesterlee, S.E., Howells, D.W., Huguenard, J., Kelner, K., Koroshetz, W., Krainc, D., Lazic, S.E., Levine, M.S., Macleod, M.R., McCall, J.M., Moxley, R.T. 3rd, Narasimhan, K., Noble, L.J., Perrin, S., Porter, J.D., Steward, O., Unger, E., Utz, U., Silberberg, S.D., A call for transparent reporting to optimize the predictive value of preclinical research, *Nature* **490** (2012), 187-191.

[16] T. Lebo, Sahoo, S.S., McGuinness, D., PROV-O: The PROV Ontology, in: *W3C Recommendation*, World Wide Web Consortium W3C, 2013.

[17] P. Missier, Sahoo, S.S., Zhao, J., Goble, C., Sheth, A., Janus: from Workflows to Semantic Provenance and Linked Open Data, in: *IPAW 2010*, Troy, NY, 2010.

[18] L. Moreau, Provenance-based reproducibility in the semantic web., *Web Semantics: Science, Services and Agents on the World Wide Web* **9** (2011), 202-221.

[19] L. Moreau, Missier, P., PROV Data Model (PROV-DM), in: *W3C Recommendation*, World Wide Web Consortium W3C, 2013.

[20] G.T. O'Connor, Caffo, B., Newman, A.B., Quan, S.F., Rapoport, D.M., Redline, S., Resnick, H.E., Samet, J., Shahar, E., Prospective study of sleep-disordered breathing and hypertension: the Sleep Heart Health Study, *Am J Respir Crit Care Med* **179** (2009), 1159-1164.

[21] F. Prinz, Schlange, T., Asadullah, K., Believe it or not: how much can we rely on published data on potential drug targets?, *Nature Reviews Drug Discovery* **10** (2011), 712.

[22] S.S. Sahoo, *Semantic Provenance: Modeling, Querying, and Application in Scientific Discovery*, Ph.D., Wright State University, 2010.

[23] S.S. Sahoo, Valdez, J., Kim, M., Rueschman, M., Redline, S., ProvCaRe: Characterizing Scientific Reproducibility of Biomedical Research Studies using Semantic Provenance Metadata, *International Journal of Medical Informatics* (2018).

[24] I. Sim, Tu, S.W., Carini, S., Lehmann, H.P., Pollock, B.H., Peleg, M., Wittkowski, K.M., The Ontology of Clinical Research (OCRe): an informatics foundation for the science of clinical research., *Journal of Biomedical Informatics* **52** (2014), 78-91.

Address for correspondence

Satya S. Sahoo, PhD

2103 Cornell Road, Case Western Reserve University School of Medicine, Cleveland, OH 44106. Phone: 216 368 0532.

Email: satya.sahoo@case.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190238

Towards Automating Location-Specific Opioid Toxicosurveillance from Twitter via Data Science Methods

Abeed Sarker[a], Graciela Gonzalez-Hernandez[a], Jeanmarie Perrone[b]

[a] *Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, U.S.A.,*
[b] *Department of Emergency Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, U.S.A.*

Abstract

Social media may serve as an important platform for the monitoring of population-level opioid abuse in near real-time. Our objectives for this study were to (i) manually characterize a sample of opioid-mentioning Twitter posts, (ii) compare the rates of abuse/misuse related posts between prescription and illicit opiods, and (iii) to implement and evaluate the performances ofsupervised machine learning algorithms for the characterization of opioid-related chatter, which can potentially automate social media based monitoring in the future.. We annotated a total of 9006 tweets into four categories, trained several machine learning algorithms and compared their performances. Deep convolutional neural networks marginally outperformed support vector machines and random forests, with an accuracy of 70.4%. Lack of context in tweets and data imbalance resulted in misclassification of many tweets to the majority class. The automatic classification experiments produced promising results, although there is room for improvement.

Keywords:
Social media, Opioids, Surveillance

Introduction

The problem of opioid (prescription and illicit) addiction and overdose is having lethal consequences all over the United States [1]. The 2015 National Survey on Drug Use and Health (NSDUH) estimated that 11.5 million adults misused/abused prescription opioids, and among adults with prescription opioid use, 12.5% reported misuse [2]. The number of opioid overdose deaths continue to rise alarmingly, with 174 people dying from drug overdoses daily [3], and the current rate of opioid prescriptions is three times higher than in the 90s. Between 2014 and 2015, opioid related death rates increased by 15.6%, continuing a trend from 1999, and this increase was driven by illicit opioids other than methadone [4]. Despite the significant acceleration of the crisis in recent years, surveillance measures are slow, and deriving estimates from surveys, such as the NSDUH, is belated. There is almost a two-year lag between the occurrence of overdose related deaths and the time by which the statistics are publicized.[†] Such a lag in the process of data collection and synthesis makes it impossible to determine the trajectory of the epidemic or identify geographic areas that are more greatly impacted by the crisis at a specific point of time. Whether its illicit or prescription opioids, the vast numbers of people affected means that a comprehensive public health approach is needed to curb the crisis, going beyond simply changing patterns of prescribing [5]. Kolodny and Frieden [1] recommended 10 steps that the federal government should take to reverse the opioid epidemic, and, as their first point, the authors outlined the need for real-time assessment of the numbers, patterns, or trends of opioid misuse/addiction.

In this paper, we explore the possibility of using social media, namely Twitter, as a resource for performing real time surveillance of opioid abuse, including both prescription and illicit opioids. Past studies have shown that users post information related to drug abuse on social media [6]–[8]. However, there is a lack of analysis of the differences in abuse-related chatter versus other types of chatter, such as consumption, although it is well known that not all drug-related chatter represents abuse [9]. There is also a lack of understanding regarding the differences between the chatter associated with prescription and illicit opioids (*e.g.*, what proportions of illicit *vs*. prescription opioid mentioning chatter represent abuse?). Unsupervised methods that primarily rely on the volume of data do not take into account the large amounts of noise that is present in social media data (*e.g.*, [10]). There are currently no prototype end-to-end, automated pipelines that can enable the real time surveillance of opioid abuse/misuse via social media. In this paper, we take the first steps in addressing these gaps. We present (i) data collection strategies from Twitter, including the use of automatically generated misspellings and geolocation metadata, (ii) an analysis of the contents of tweets mentioning prescription and illicit opioids, and (iii) a comparison of several supervised classification approaches. Our experiments show that opioid chatter on Twitter can vary significantly between prescription and illicit opioids, with some illicit opioid keywords being too ambiguous to be useful for data collection. We also show that using annotated data, we can train supervised learning algorithms to automatically characterize tweets. We suggest that such a supervised classification system, paired with geolocation metadata from Twitter, can be used to perform localized surveillance of opioid abuse/misuse. We present our pilot methods using the state of Pennsylvania as example.

[†]Available at: https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates. Accessed: October 22, 2018.

Methods

Data Collection

We collected data from Twitter using names of prescription and illicit opioid keywords (*e.g.*, *Percocet*® and *heroin*), including street names (*e.g.*, *china white*, *tar*, *skag, percs*) and common misspellings (*e.g.*, *percoset*, *heorin*). We used the list of drug slang terms recently released by the Drug Enforcement Agency (DEA) of the United States to create an initial list of possible slang terms for different prescription and illicit opioids [11]. We manually reviewed the terms and removed those we were sure to be too ambiguous. For example, some of the slang terms associated with heroin, as per the document, are '*basketball*', '*coffee*', '*lemonade*' and '*whiskey*'. Through manual searches of the Twitter web interface, we could not find any instances where these terms were used to refer to opiods. Therefore, we removed these to reduce the retrieval of noise. This strategy led us to use a total of 56 unique names of opioids. Since drug names are often misspelled on social media, we automatically generated misspellings for these keywords using a misspelling generation system [12]. Table 1 presents some sample opioid-related keywords and their automatically generated misspellings. After collecting an initial set, we analyzed samples of retrieved tweets for each keyword. We discovered that despite our initial filtering of keywords, approximately 85% of the tweets were retrieved by 4 keywords—*tar* (~6.5%), *dope* (~54%), *smack* (~20.5%) and *skunk* (~4%)—and in these tweets, these keywords were almost invariably unrelated to opioids and represented something else. We, therefore, removed these keywords for our final data collection. In this manner, we collected tweets between the years 2012 to 2015, only including those geolocated from within Pennsylvania and excluding retweets.

Table 1– Sample of opioid-related keywords and their automatically-generated frequently occurring misspellings

Keyword	Generated Misspellings
Tramadol	trammadol tramadal tramdol tramadols tramado tramedol tramadoll tramadole tramidol tamadol tranadol tramodol tremadol
Heroin	herione herroine heroins heroine heroin heorin herion
Methadone	methadones methadose methodone mehtadone metadone methadon methdone
Oxycontin	oxicontin oxcotin oycotin oxycotins oycontin oxycontins oxycoton oxicotin ocycotin oxycodin oxycottin oxycotine ocycontin
Codeine	codiene coedine codine codene codein
Dilaudid	delaudid dialudid dilaudad diluadid diaudid dilaudin dilauded dilauid dillaudid

Annotation

Our intent was to compare the distributions of tweets for prescription and illicit opioids and to attempt to train supervised learning algorithms—both of which require manual annotation/labeling of a sample of tweets. Based on manual inspection of the collected data,. we decided to manually code the tweets into 4 broad categories—self-reported abuse, information sharing, non-English, and unrelated. Details about these categories are as follows.

Abuse-related (A)

Abuse-indicating or possible abuse by the poster or by someone the user knows or is communicating with. This category also includes admissions of abuse in the past. For illicit opioids, any consumption is considered to be abuse. For prescription opioids, consumption is considered to be abuse only when there is evidence that the user is taking the drug without a prescription, through a non-standard route (e.g., injecting, snorting) or in combination with other substances in order to experience certain sensations.

Information Sharing/Seeking/Providing (I)

Tweets in which the poster is asking for information or providing information about an opioid. This category also includes expressions of medical use (e.g., mentions of having a prescription or taking painkillers after surgery), and sharing of news articles or other media that contain information about opioids. General statements about the drug may are also put into this class.

Non-English (N)

Tweets that are not written in English belong to this category.

Unrelated (U)

This category includes tweets that are not about the drug or opioid, but about something else. This category also includes tweets that make metaphorical comparisons (e.g., I am addicted to X like heroin). Some examples of tweets belonging to this category: handle related (@codeine_CXXX), heroine (hero), cooking (brown sugar). This category also includes tweets about movies or lyrics of songs that mention opioids, but don't have any information value. Table 2 presents examples of tweets belonging to these four categories.

Table 2– Sample tweets and their categories; opioid keywords shown in bold

Tweet	Category
@username naa, i just popped a few **percs** at 2, i drink, sip lean. Wbu?	*A*
Sooooo **heroine** addicts robbed the house 3 houses away from me...makes me feel safe	*I*
Ok I thought that it was just a really funny **oxy** clean commercial but turns out it was just the Spanish channel	*U*
Te quieroo muchito mi hermana **negra**	*N*

We iteratively annotated a set of 100 tweets and discussed the disagreements between pairs of annotators. The disagreements on the initial set were resolved via discussion, and the same process was executed twice until an acceptable level of agreement was reached. In the final set, disagreements for overlapping tweets were resolved by a third annotator—the first author of this article.

Analysis and Supervised Learning

Prescription versus Illicit Opioids

Using the annotated dataset, we compared the volumes of prescription and illicit opioids in the sample to better understand which of these two broad classes of opioids were

more frequently discussed on Twitter. Since the sample for annotation was drawn randomly, we assumed that the distributions of prescription and illicit opiod mentions represented their natural distribution in publicly available Twiter chatter. We also assessed the differences in the distributions of the four tweet categories for these two types of opioids by comparing their proportions. The results of these comparisons are presented in the Results section.

Supervised Machine Learning

To train and evaluate several machine learning algorithms, we first split the annotated data into training (~80%) and test (~20%) sets. We used the training set for analysis, algorithm training and feature analyses, and held out the test set for evaluation. Our intent was primarily to assess the utility of the annotated corpus for supervised machine learning, with the assumption that if supervised classification produced adequate performance, they can be employed in the future for real time monitoring. We trained and optimized three different classification algorithms over the dataset—support vector machines (SVMs), random forests (RFs) and deep convolutional neural network (d-CNN), and compared their performances with a naïve bayes (NB) baseline classifier. SVMs and RFs have been shown in the past to perform well for text classification tasks, particularly because of their suitability for handling large feature spaces. Meanwhile, CNN based classifiers have become popular in the recent past, and they work particularly well in the presence of large annotated data sets. For the SVMs, RF and NB classifiers, we performed basic feature engineering based on our findings from past work on the topic of automatic prescription medication abuse detection from social media [13]. As features, we used preprocessed n-grams (n=1—3), word clusters or generalized representations of words, and the presence or absence of abuse-indicating terms. We used 10-fold cross validation over the training set for the RF and SVM classifiers to find optimal parameter values. For the SVMs, we optimized the kernel and the cost parameter. For the RF classifier, we optimized the number of trees. For the d-CNN classifier, we used dense word vectors, or word embeddings as input. We obtained pre-trained word embeddings from our past work [14]. We used a three-layer convolutional neural network, and for optimizing the various hyperparameters, we split the training set further into two sets and used the larger set for training and the smaller set for validation. For NB, SVM and RF classifiers, we used implementations provided by the python scikit-learn library [15], and for the d-CNN classifier, we used the TensorFlow library [16]. Figure 1 summarizes our entire processing workflow for this study—from spelling variant generation through to supervised classification of tweets.

Results

A total of 9006 tweets mentioning both prescription and illicit opioids were annotated by 4 annotators. Among 550 overlapping tweets, average inter-annotator agreement was 0.75 (Cohen's kappa [17]). The final data set consisted of 1748 abuse tweets, 2001 information tweets, 4830 unrelated tweets, and 427 non-English tweets. The majority of the tweets mentioned illicit opioids—7038 illicit and 2257 prescription.‡ Figure 2 shows the distributions of illicit and prescription opioid mentioning tweets in our annotated set, illustrating that although the relative volume of illicit opioid tweets is much higher, a significantly larger proportion of these tweets are unrelated to opioids. The significantly higher number of unrelated tweets for illicit opioid mentioning posts suggests that such tweets have higher amounts of noise associated with them, and may be more difficult to mine knowledge from despite the large volume.

Table 3 presents the performance of the three classifiers and the NB baseline over the test set. In total, we used 7204 tweets for trainining and 1802 tweets for evaluation. For the d-CNN classifier, the training set was further split into 6304 for training and 900 for validation. It can be seen that, in terms of overall accuracy, macro-averaged recall and precision, the d-CNN classifier marginally outperforms the two traditional benchmark classification approaches (SVMs and RF) despite the relatively small amount of annotated data that was used. All the three classifiers perform significantly better than the NB baseline. The high performance of the d-CNN classifier is encouraging because such deep neural network based classifiers have more room for improvement, compared to their traditional counterparts, as more data is annotated.

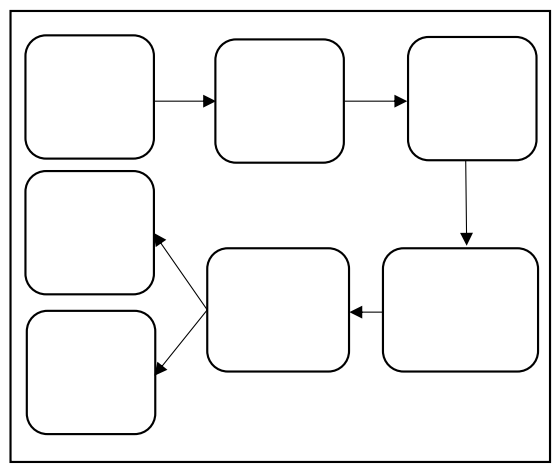


Figure 1 The Twitter data processing workflow for this study

Discussion

Our experiments produced very promising results and showed that automatic machine learning based approaches may in fact provide a possible mechanism for monitoring opioid abuse in near real time for targeted geographic locations (e.g., at the state leve). By combining geolocation information and manually annotated data, we were able to automatically characterize opioid-mentioning chatter from Pennsylvania with moderate accuracy. Table 4 shows three sample tweets, their automatic classifications, location by county and timestamp.

Our manual categorization efforts revealed the difficulty of annotating tweets with high inter annotator agreement. Creating a specific annotation guideline and several iterations of discussions over small sets of overlapping annotations helped improve agreement, although in many cases, due to the lack of context in the tweets, the assigned category depended on the subjective assessement of the annotator. This suggests that

‡ Note that the sum is of these two numbers is greater than the total number of tweets annotated (9006) since some tweets mention both prescription and illicit opioids.

thorough annotation guidelines and such an iterative approach to annotation are very important for achieving acceptable agreement levels for complex annotation tasks such as this.

We found that illicit opioid mentioning tweets were particularly highly noisy, with references to song lyrics or movie quotes, which led to a large proportion of them to be labeled as unrelated. The high proportions of unrelated tweets for both types of opioids, and particularly for illicit opioids, illustrate the importance of a supervised classification system for automatic surveillance. Keyword-based surveillance methods, which rely on the volume of data collected using specific keywords, are evidently not suitable for opioid toxicosurveillance since most of the data retrieved by the keywords will be unrelated noise. The amount of noise may increase or decrease based on events publicized over media outlets. In addition, as our initial analysis of the retrieved data showed, if ambiguous keywords are to be used, the vast majority of tweets collected via the ambiguous keywords (e.g., dabs) can be noise, and this noise may mask the real abuse related signals. Thus, when designing surveillance strategies for similar tasks via social media, care must be taken to identify noisy keywords that may invalidate the surveillance process by bringing in too much noise.

The automatic classification experiments produced acceptable performances, suggesting that automated, real-time opioid toxicosurveillance may be a possibility. In the future, we will explore additioal classification strategies for further improving performance. A brief error analysis revealed that lack of context in tweets caused our learning algorithms to often misclassify tweets to the majority class (U). To better understand the characteristics of the missclassified tweets, more analyses are required.

In the future, we will also apply supervised classifiers trained using our annotated data to automatically characterize unlabeled posts collected over a longer time period to better understand how opioid abuse related tweets are distributed over time and more fine-grained geolocations. Such an analysis may reveal specific time periods that are associated with higher rates of abuse. We will also explore how the opioid abuse rates reported on Twitter correlate, if at all, with real-world data regarding the opioid crisis, such as geolocation-centric opioid overdose death rates.

Conclusions

Our study suggests that Twitter is a promising platform to perform real-time surveillance of opioid abuse/misuse. Although we have only used geolocation data to identify the origins of tweets at the state level, it may be possible to further narrow down to the county or city level, particularly as the volume of data grows over time. Our manual categorization of the data and analyses shows that keyword based data collection from Twitter results in the retrieval of significant amounts of noise. Therefore, studies attempting to use streaming Twitter data for surveillance must be wary of the amount of noise retrieved per keyword and only use keywords that are unambiguous. The same protocol should be followed for research involving data from other social networks. Our annotation also showed that even when using keywords with high signal-to-noise ratios, the number of unrelated tweets is significantly higher for illicit opioids compared to prescription opioids. Thus, the total volume of opioid related chatter may not be indicative of the real abuse or misuse of opioids, but may be driven by other factors such as news articles or the release of movies/songs. To overcome this problem, we employed a supervised classification approach to automatically categorize the tweets, and we found a deep convolutional neural network to produce the best performance with an overall accuracy of 70.4%. In the future, we will try to improve on this classification performance by employing more advanced strategies, and also use the output of the classifiers to perform downstream geolocation-centric analyses.

Table 3– Classifier accuracies over the test set

Classifier	Recall	Precision	Accuracy (%)	95% CI
Naïve Bayes	0.61	0.58	53.9	51.6-56.3
Random Forest	0.66	0.70	70.1	67.9-72.2
Support Vector Machines	0.68	0.70	69.9	67.8-72.1
Deep Convolutional Neural Network	0.70	0.71	70.4	68.2-72.5

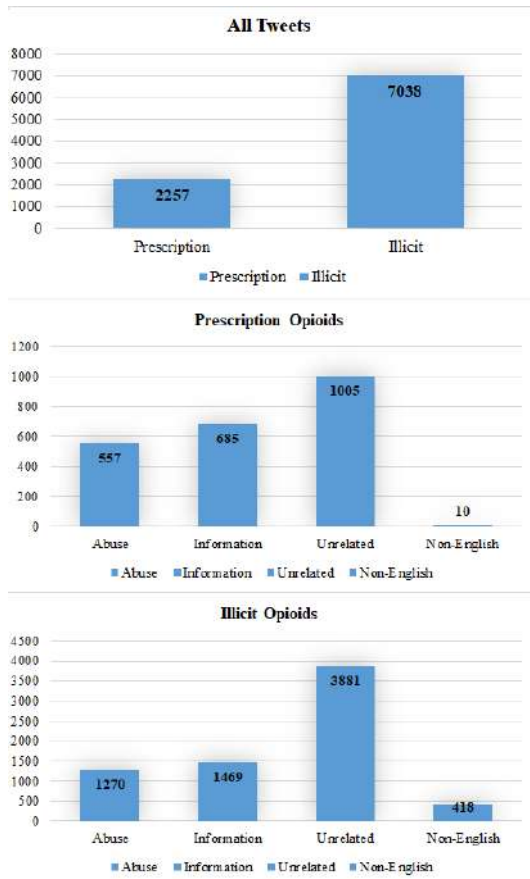


Figure 2– Distributions of tweets belonging to each category for illicit and prescription opioid mentioning tweets. The charts show that aabout 75% of the tweets in the sample mention illicit opioids, and that illicit opioid mentioning tweets have much higher proportions of unrelated information (including non-English tweets), while prescription opioid mentioning tweets have higher proportions of misuse/abuse and information oriented tweets.

Table 4– Sample tweets and classification in real-time with geolocation information (county level) and timestamps[§]

Tweet	Class	County	Timestamp
Enjoying this healthy breakfast recommendation frm @username. Oatmeal w/raisins/walnuts/brown sugar frm @username	Unrelated	Philadelphia	12:37:11 XX-XX- 2015
@username i shouldnt have done all that heroin this morning	Abuse	Allegheny	13:32:34 XX-XX-2015
I know everyone is socialized different and wired uniquely. I still want to smack a ******* for not staying in their lane.	Unrelated	Philadelphia	13:54:21 XX-XX-2015
its on the news.. kensington oxys on the loose	Information	Philadelphia	15:01:55 XX-XX-2015

Acknowledgements

Research reported in this publication was supported in part by the National Institute on Drug Abuse of the National Institutes of Health under Award Number R01DA046619. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The data collection and annotation efforts were partly funded by a grant from the Pennsylvania Department of Health. The Titan Xp used for this research was donated by the NVIDIA Corporation. The authors would like to thank Karen O'Connor, Alexis Upshur and Annika DeRoos for performing the annotations. This study was approved by the institutional review board at the University of Pennsylvania.

References

[1] A. Kolodny and T.R. Frieden, Ten steps the federal government should take now to reverse the opioid addiction epidemic, *JAMA* **318** (2017), 1537-1538.
[2] B. Han, W.M. Compton, C. Blanco, E. Crane, J. Lee, and C.M. Jones, Prescription opioid use, misuse, and use disorders in U.S. adults: 2015 National Survey on Drug Use and Health, *Ann Intern Med* **167** (2017), 293-301.
[3] H. Jalal, J.M. Buchanich, M.S. Roberts, L.C. Balmert, K. Zhang, and D.S. Burke, Changing dynamics of the drug overdose epidemic in the United States from 1979 through 2016, *Science* **361** (2018), 1184.
[4] R.A.R. Rudd, P. Seth, F. David, and L. Scholl, Increases in drug and opioid-involved overdose deaths — United States, 2010–2015, *Morb Mortal Wkly Rep* **65** (2016), 1445-1452.
[5] A. Schuchat, D. Houry, and G.P. Guy, New data on opioid use and prescribing in the United States, *JAMA* **318** (2017, 425-426.
[6] L. Shutler, L.S. Nelson, I. Portelli, C. Blachford, and J. Perrone, Drug use in the Twittersphere: a qualitative contextual analysis of tweets about prescription drugs, *J Addict Dis* **34** (2015), 303-310.
[7] M. Chary, N. Genes, C. Giraud-Carrier, C. Hanson, L.S. Nelson, and A.F. Manini, "Epidemiology from tweets: estimating misuse of prescription opioids in the USA from social media, *J Med Toxicol* **13** (2017) 278-286.
[8] D. Cameron *et al.*, PREDOSE: a semantic web platform for drug abuse epidemiology using social media, *J Biomed Inform* **46** (2013), 985-997.
[9] A. Sarker *et al.*, Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from Twitter, *Drug Saf* **39** (2016), 231-240.
[10] R.L. Graves, C. Tufts, Z.F. Meisel, D. Polsky, L. Ungar, and R.M. Merchant, Opioid discussion in the Twittersphere, *Subst Use Misuse* **53** (2018), 2132-2139.
[11] DEA Houston Division, *Slang Terms and Code Words: A Reference for Law Enforcement Personnel,* US Drug Enforcement Administration, Washington, DC, 2018.
[12] A. Sarker and G. Gonzalez-Hernandez, An unsupervised and customizable misspelling generator for mining noisy health-related text sources, *J Biomed Inform* **88** (2018), 98-107.
[13] A. Sarker *et al.*, Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from Twitter, *Drug Saf* **39** (2016), 231-240.
[14] A. Sarker and G. Gonzalez, A corpus for mining drug-related knowledge from Twitter chatter: language models and their utilities, *Data in Brief* **10** (2017), 122-131.
[15] F. Pedregosa *et al.*, Scikit-learn: machine learning in Python, *J Mach Learn Res* **12** (2011), 2825-2830.
[16] M. Abadi *et al.*, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*, Google Research, Mountain View, CA, 2016.
[17] J. Cohen, A coefficient of agreement for nominal scales, *Educ Psychol Meas* **20** (1960), 37-46.

Address for correspondence
Abeed Sarker, Ph.D.
Mailing Address: Level 4, 423 Guardian Drive, Division of Informatics, Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.
Email: abeed@pennmedicine.upenn.edu
Phone: +1-215-746-1700

[§] The tweets and their metadata have been modified to protect the anonymity of the actual users.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190239

Urban Disadvantage, Obesity, and Underweight in 31 Lower-Income Countries

Lincoln R. Sheets[a], Henok G. Woldu[a], Rina Swart[b], Eduardo Simoes[a]

[a] *Department of Health Management and Informatics, University of Missouri, Columbia, Missouri, United States*
[b] *Department of Community and Health Sciences, University of Western Cape, Cape Town, Western Cape, South Africa*

Abstract

Although some studies have shown that obesity and other non-communicable diseases are more common in more disadvantaged areas, no publications to date have examined the interaction of obesity with urban and rural disadvantage in lower-income countries. This study analyzed the rates of obesity and underweight in disadvantaged urban women and disadvantaged rural women in 31 lower-income countries, and calculated the age-adjusted odds ratios of urban vs. rural obesity and underweight. The odds of obesity were significantly ($p<0.05$) higher for urban populations in 16 of the 31 countries and in all aggregated regions; the evidence that underweight is also associated more with urban populations was mixed. Because obesity is a rapidly-growing threat to the public health and financial strength of lower-income countries, and urban disadvantage is associated with more obesity than rural disadvantage, policymakers should work to understand, predict, and prevent obesity in urban populations specifically.

Keywords:

Obesity Epidemiology, Healthcare Disparities, Geographic Factors

Introduction

An Inequitable Burden

Obesity is growing rapidly in low- and middle-income countries (LMIC), and is becoming a greater threat to global health than undernutrition and infectious diseases [1]. Obesity is a potentially preventable risk factor for many types of cancer and for heart disease [2], which are the leading causes of death around the world [3]. The global cost to treat excess morbidity due to obesity is already billions of dollars annually and the associated burdens of mortality, morbidity, economic costs, and social stigma are rising fastest in the countries least able to bear them [4].

What Is Known: Economic Disadvantage and Obesity

Some studies have shown that obesity and other non-communicable disease are more common in neighborhoods with more poverty, pollution, and crime, and less common in neighborhoods with access to healthy foods, exercise opportunities, and public transportation [5, 6]. These geographic associations can guide population-level interventions such as community-based preventive health programs and national health policy decisions [7]. While most of the published literature comes from the United States and Europe, there has been a recent increase in the number of high-quality global studies of obesity and geographic disadvantage [8, 9], including retrospective studies of the the rich, multi-national, multi-year data of the Demographic and Health Surveys (DHS) Program [10].

What Is Not Known: Urban and Rural Disadvantage

Although studies have shown urban-rural differences in the association of obesity with economic disadvantage in both high-income countries [11] and low-income countries [12], no publications to date have examined the interaction of obesity with urban and rural disadvantage in countries with diverse economic conditions. This study examined DHS data from multiple low-income and lower-middle-income countries to examine the association of obesity with urban disadvantage and rural disadvantage. Because many lower-income countries bear a double burden of underweight and obesity [13], the prevalences of both these conditions were analyzed.

Hypothesis

Both obesity and underweight are more strongly associated with urban disadvantage than with rural disadvantage among adults in lower-income countries.

Methods

Population

Thirty-one countries (shown in Table 1) met all four inclusion criteria: (a) DHS survey results were available within the last ten years (i.e. in 2008 or later) [10], (b) the survey results included body-mass index (BMI) values, (c) the survey results included responses from both urban and rural areas, and (d) the country was classified by the World Bank as a "Low Income" or "Lower-Middle Income" economy [14].

Data Sources

Results were included from the most recent DHS survey in each country [10]. DHS surveys only include weight-and-height measurements on women of child-bearing age, which DHS defines as 15 to 49 years old. Our study included those subjects but excluded subjects who were younger than 18 years old, and excluded women who were pregnant at the time of their BMI measurement Because independent studies [15, 16] have detected qualitative differences in health outcomes between adults in the lower three quantiles and the upper two quantiles of the DHS Wealth Index, the lower three quantiles were used to define economic disadvantage and only subjects in those quantiles were included in the analysis.

Data Elements

The independent variables collected for each included individual were country, urban/rural category [10], DHS Wealth Index [10], and age. Because the body-mass index or BMI classifications of the World Health Organization [17] have been shown to have predictive value in health outcomes including mortality [18], BMI categories were used as nominal dependent variables.

Statistical Methodology

The prevalences of obesity (BMI ≥ 30) and underweight (BMI < 18.5) were calculated for disadvantaged urban women and disadvantaged rural women in each region. The age-adjusted odds ratios of urban vs. rural obesity and urban vs. rural underweight were calculated for each region, using normal weight (18.5 ≤ BMI < 25) as the reference category. The 95% confidence interval was calculated for each odds ratio.

Results

Table 1 shows the survey characteristics and population characteristics of each region.

Table 1 – Survey and Population Characteristics by Region

Region	Income Class	Survey Year	Women (n)	Mean Age (SD)	Urban %	Mean BMI (SD)
Africa						
Benin	Low	2012	7,791	31 (8)	22	23 (4)
Burundi	Low	2017	3,692	31 (9)	3	20 (2)
Cameroon	Lower-Middle	2011	3,380	31 (9)	20	23 (4)
Comoros	Low	2012	2,368	30 (9)	31	25 (5)
Congo	Lower-Middle	2012	3,517	32 (9)	18	22 (4)
Congo Democratic Republic	Low	2014	4,531	31 (9)	12	21 (3)
Cote d'Ivoire	Lower-Middle	2012	2,359	31 (9)	15	22 (3)
Ethiopia	Low	2016	6,107	31 (9)	3	20 (3)
Gambia	Low	2013	2,330	30 (9)	15	22 (4)
Ghana	Lower-Middle	2014	2,471	31 (8)	27	23 (4)
Guinea	Low	2012	2,048	31 (9)	4	22 (3)
Kenya	Lower-Middle	2014	2,025	31 (9)	20	22 (4)
Lesotho	Lower-Middle	2014	1,541	30 (9)	9	25 (5)
Malawi	Low	2017	3,558	30 (9)	5	22 (3)
Mozambique	Low	2015	4,772	32 (9)	12	22 (3)
Rwanda	Low	2015	3,056	31 (9)	7	23 (3)
Sao Tome and Principe	Lower-Middle	2009	1,184	32 (9)	40	24 (5)
Tanzania	Low	2016	5,335	31 (9)	6	22 (4)
Togo	Low	2014	2,302	32 (9)	3	22 (4)
Uganda	Low	2016	2,719	31 (9)	8	22 (3)
Zambia	Lower-Middle	2014	7,375	31 (9)	22	22 (3)
TOTAL: Africa	--	--	79,711	31 (9)	14	22 (4)
Asia						
India	Lower-Middle	2016	359,117	32 (9)	13	21 (4)
Kyrgyz Republic	Lower-Middle	2012	3,907	33 (9)	5	25 (5)
Myanmar	Lower-Middle	2016	6,579	33 (9)	9	22 (4)
Nepal	Low	2016	3,376	31 (9)	55	21 (3)
Pakistan	Low	2013	2,311	33 (9)	24	23 (5)
Tajikistan	Low	2012	4,413	31 (9)	10	23 (4)
Timor-Leste	Lower-Middle	2016	5,560	32 (9)	11	20 (3)
TOTAL: Asia	--	--	385,263	32 (9)	13	21 (4)
Latin America						
Bolivia	Lower-Middle	2008	7,809	32 (9)	36	26 (4)
Haiti	Low	2017	4,532	31 (9)	15	23 (4)
Honduras	Lower-Middle	2012	12,026	31 (9)	22	26 (5)
TOTAL: Latin America	--	--	24,367	31 (9)	25	25 (5)
Worldwide						
Low Income	--	--	65,241	31 (9)	13	22 (4)
Lower-Middle Income	--	--	364,821	32 (9)	14	21 (4)
TOTAL: Worldwide	--	--	489,341	32 (9)	14	21 (4)

n = number of subjects; SD = standard deviation

As shown in Table 2, the odds of obesity were significantly ($p < 0.05$) higher for disadvantaged urban women than for disadvantaged rural women in 16 of the 31 countries and in all larger regions that were analyzed. The odds of obesity were not significantly different in the other regions.

The odds of underweight were significantly higher for disadvantaged urban women than for disadvantaged rural women in two of the 31 countries, but significantly lower in one country (India), in the aggregate of all lower-middle income countries, and in the aggregate of all 31 countries.

Table 2 – Urban and Rural Prevalences and Odds Ratios by Region

	Obesity			Underweight		
Region	**Urban Prevalence**	**Rural Prevalence**	**Urban/ Rural OR (CI)**	**Urban Prevalence**	**Rural Prevalence**	**Urban/ Rural OR (CI)**
Africa						
Benin	4.2	4.0	*2.7 (2.3 - 3.1)	6.3	5.6	1.0 (0.9 - 1.2)
Burundi	0.8	0.2	3.1 (0.4 - 5.4)	22.8	20.6	1.1 (0.7 - 1.7)
Cameroon	10.0	4.6	*2.6 (1.9 - 3.5)	6.7	9.2	0.8 (0.6 - 1.1)
Comoros	18.4	12.8	*1.8 (1.4 - 2.3)	5.5	5.5	1.2 (0.8 - 1.7)
Congo	8.4	3.1	*3.7 (2.6 - 5.3)	14.1	14.0	1.2 (0.9 - 1.5)
Congo Democratic Republic	1.0	0.9	1.2 (0.5 - 3.0)	15.0	16.6	0.9 (0.7 - 1.2)
Cote d'Ivoire	7.5	2.4	*4.4 (2.7 - 7.2)	11.7	6.8	2.1 (0.4 - 3.0)
Ethiopia	1.2	0.5	2.3 (0.5 - 9.6)	25.9	29.6	0.9 (0.6 - 1.2)
Gambia	7.0	4.3	*2.0 (1.2 - 3.3)	16.6	16.9	1.0 (0.8 - 1.4)
Ghana	10.7	6.9	*1.9 (1.4 - 2.6)	6.2	6.0	1.2 (0.8 - 1.8)
Guinea	2.5	2.1	1.2 (0.3 - 4.9)	7.5	14.1	0.6 (0.2 - 1.3)
Kenya	5.0	4.9	1.1 (0.9 - 1.5)	12.0	14.4	0.9 (0.7 - 1.0)
Lesotho	11.1	15.3	0.7 (0.4 - 1.3)	5.9	3.1	1.8 (0.8 - 4.0)
Malawi	6.6	2.6	*3.2 (1.6 - 6.1)	9.6	6.2	*1.8 (1.1 - 3.2)
Mozambique	0.9	0.8	1.2 (0.5 - 3.2)	9.8	8.5	1.2 (0.9 - 1.7)
Rwanda	3.2	1.3	*2.6 (1.1 - 5.8)	6.4	6.2	1.1 (0.6 - 1.9)
Sao Tome and Principe	12.5	9.4	1.3 (0.9 - 2.0)	6.8	7.3	1.0 (0.6 - 1.6)
Tanzania	4.6	3.9	1.2 (0.7 - 2.1)	6.7	9.3	0.8 (0.5 - 1.2)
Togo	5.3	4.2	1.5 (0.5 - 4.2)	10.5	8.7	1.3 (0.6 - 2.8)
Uganda	4.3	2.4	2.0 (1.0 - 4.1)	12.3	11.0	1.2 (0.8 - 1.9)
Zambia	3.2	2.1	*1.8 (1.3 - 2.4)	9.8	11.2	0.9 (0.8 - 1.1)
TOTAL: Africa	6.2	3.3	*2.0 (1.8 - 2.2)	9.7	12.2	0.9 (0.7 - 1.0)
Asia						
India	4.1	1.8	*2.5 (2.3 - 2.6)	20.3	25.8	*0.8 (0.8 - 0.8)
Kyrgyz Republic	14.0	14.8	0.8 (0.5 - 1.3)	4.0	4.5	0.9 (0.4 - 1.9)
Myanmar	5.7	3.8	*1.7 (1.1 - 2.4)	14.3	15.2	1.0 (0.8 - 1.3)
Nepal	2.0	1.4	1.4 (0.8 - 2.4)	18.1	16.8	1.1 (0.9 - 1.3)
Pakistan	11.3	7.5	*1.6 (1.1 - 2.2)	11.3	15.6	0.8 (0.6 - 1.0)
Tajikistan	10.9	8.0	1.6 (1.0 - 2.3)	8.5	8.6	1.1 (0.8 - 1.6)
Timor-Leste	1.0	0.8	1.4 (0.6 - 3.2)	25.3	26.1	1.0 (0.8 - 1.2)
TOTAL: Asia	4.2	2.1	*2.2 (2.1 - 2.3)	20.0	25.1	0.8 (0.8 - 0.8)
Latin America						
Bolivia	21.0	16.2	*1.8 (1.5 – 2.0)	1.1	0.7	*1.8 (1.1 - 2.8)
Haiti	8.9	6.4	*1.4 (1.1 - 1.9)	10.0	10.5	1.0 (0.7 - 1.3)
Honduras	26.6	17.2	*2.0 (1.8 - 2.3)	3.3	3.4	1.2 (0.9 - 1.5)
TOTAL: Latin America	22.0	14.6	*2.0 (1.8 - 2.2)	3.1	4.2	0.9 (0.7 - 1.0)
Worldwide						
Low Income	5.7	3.3	*1.9 (1.7 - 2.1)	11.5	12.7	1.0 (0.9 - 1.1)
Lower-Middle Income	6.2	2.8	*2.6 (2.5 - 2.7)	17.5	23.6	*0.8 (0.8 - 0.8)
TOTAL: Worldwide	6.2	2.9	*2.4 (2.3 - 2.5)	16.7	22.1	*0.8 (0.8 - 0.8)

OR = odds ratio; CI = 95% confidence interval; * = 95% confidence interval which excludes the null hypothesis

Urban Disadvantage and Obesity

Figure 1 shows the geographic distribution of urban-rural disparities in the prevalence of obesity. Darker reds indicate a higher urban-rural odds ratio of obesity, ranging from 1.0 in countries with no significant urban-rural difference to an odds ratio of 4.4 in Cote d'Ivoire.

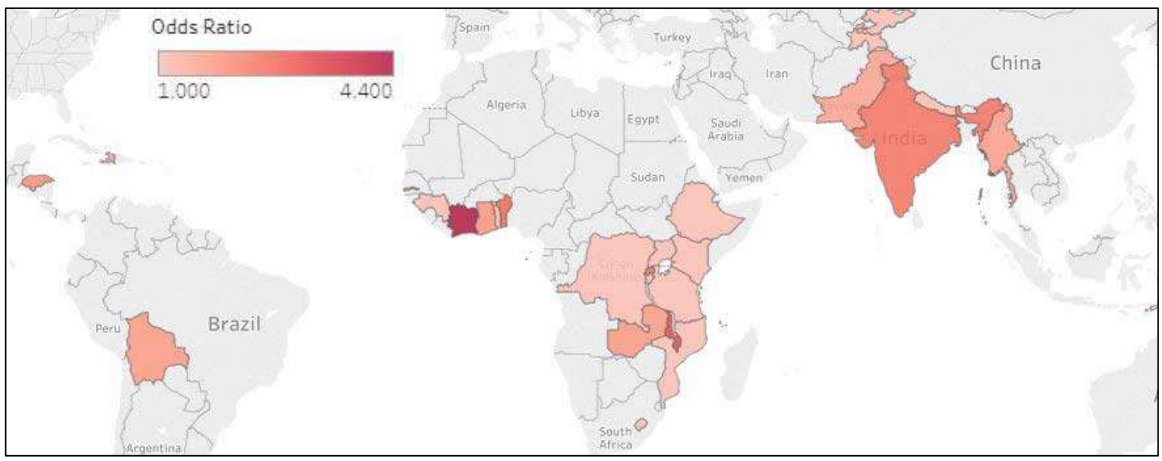


Figure 1– Odds Ratio of Urban vs. Rural Obesity by Country

Urban Disadvantage and Underweight

Figure 2 shows the geographic distribution of urban-rural disparities in the prevalence of underweight. Darker browns indicate a higher urban-rural odds ratio of underweight, ranging from 0.8 in India, through 1 in countries with no significant difference, to an odds ratio of 1.8 in Bolivia and Malawi.

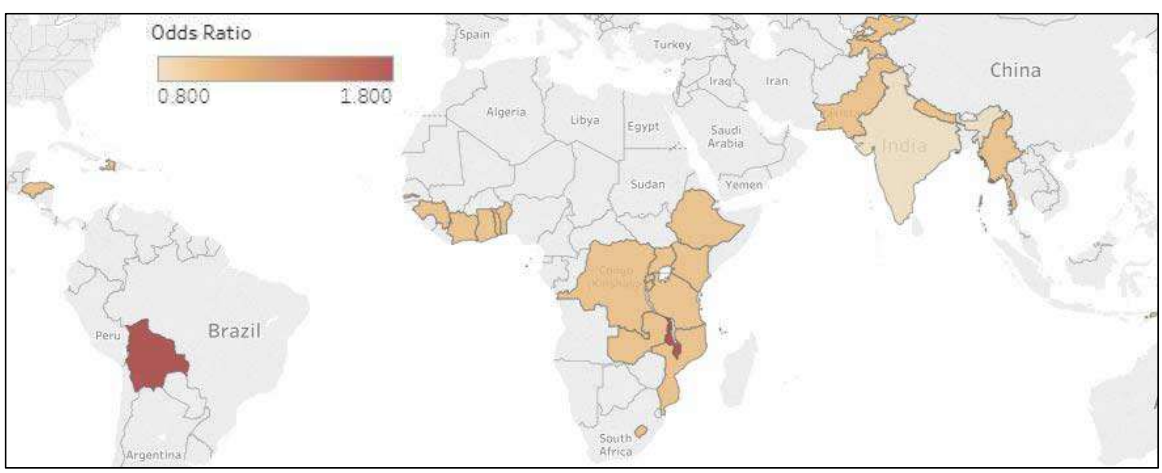


Figure 2– Odds Ratio of Urban vs. Rural Underweight by Country

Discussion

The hypothesis that obesity is associated more with urban disadvantage than with rural disadvantage was found to be true in about half of the lower-income countries for which data were available, in the aggregates of each continent, and in the worldwide aggregate of all 31 countries. However, the evidence that underweight is also associated more with urban disadvantage than with rural disadvantage was very weak, and was contradicted by the finding that underweight is actually associated less with urban disadvantage than with rural disadvantage in some regions and in the worldwide aggregate.

Limitations

The study was limited by the relatively few countries for which data was available, and by the restriction of BMI measurements to women in the survey data. The findings of this study cannot be generalized to higher-income countires, may or may not be generalizable to other lower-income countries, and cannot be be generlized to men in lower-income countries.

Further study

This preliminary study used a relatively simple statistical analysis on the more disadvantaged part of each country's population, as defined by quintiles of the Wealth Index. In order to better understand the interaction of disadvantage with urban residence in the rates of obesity and underweight, a multinomial regression against the Wealth Index of the entire population would provide further insights.

Broader impacts

Because obesity is a rapidly-growing threat to the public health and financial growth of lower-income countries, and urban disadvantage is associated with more obesity than rural disadvantage, policymakers should work to understand, predict, and prevent obesity in urban populations specifically.

Conclusions

The evidence that obesity is more strongly associated with urban disadvantage than with rural disadvantage is compelling, although no clear relationship could be established between underweight and urban or rural disadvantage. As obesity becomes more prevalent in lower-income countries, it is becoming more deadly and more expensive than underweight, as is already true in higher-income countries. More resources are needed for the scientific study and public-policy responses to this threat.

Acknowledgements

The authors wish to thank the USAID for making their DHS data available publicly. The authors have no conflicts of interest to declare.

References

[1] World Health Organization (WHO), *Obesity: Preventing and Managing the Global Epidemic*, World Health Organization, Geneva, Switzerland, 2000.

[2] R. S. Padwal, A. M. Sharma, Prevention of cardiovascular disease: obesity, diabetes and the metabolic syndrome, *Can J Cardiol* **26** (2010), 18C–20C.

[3] M. Naghavi, A. A. Abajobir, C. Abbafati, et al., Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016, *Lancet* **390** (2017), 1151–1210.

[4] M. Tremmel, U. G. Gerdtham, P. M. Nilsson, and S. Saha, Economic burden of obesity: a systematic literature review, *Int J Environ Res Public Health* **14** (2017), 435-452.

[5] Y. Wang, M. A. Beydoun, The obesity epidemic in the United States--gender, age, socioeconomic, racial/ethnic, and geographic characteristics: a systematic review and meta-regression analysis, *Epidemiol Rev* **29** (2007), 6-28.

[6] L. Letarte, A. Lebel, E. O. D. Waygood, et al., Longitudinal designs to study neighbourhood effects on the development of obesity: a scoping review protocol. *BMJ Open* **8** (2018), e017704.

[7] O. M. Meyer, C. A. Caburnay, R. S. Graves, L. R. Sheets, The effect of geographic disadvantage on metabolic health: a scoping review, unpublished data.

[8] F. Razak, D. J. Corsi, S. V. Subramanian, Change in the body mass index distribution for women: analysis of surveys from 37 low- and middle-income countries, *PLoS Med* **10** (2013), e1001367.

[9] R. Kim, I. Kawachi, B. A Coull, S. V. Subramanian, Contribution of socioeconomic factors to the variation in body-mass index in 58 low-income and middle-income countries: an econometric analysis of multilevel data, *Lancet Glob Health* **6** (2018), e777–e778.

[10] United States Agency for Economic Development (USAID), The DHS Program (2018), https://dhsprogram.com (accessed November 9, 2018).

[11] J. A. Johnson 3rd, A. M. Johnson, Urban-rural differences in childhood and adolescent obesity in the United States: a systematic review and meta-analysis, *Child Obes* **11** (2015), 233-241.

[12] R. Ofori-Asenso, A. A. Agyeman, A. Laar, D. Boateng, Overweight and obesity epidemic in Ghana-a systematic review and meta-analysis, *BMC Pub Health* **16** (2016), 1239.

[13] World Health Organization (WHO), *The Double Burden of Malnutrition*, World Health Organization, Geneva, Switzerland, 2017.

[14] The World Bank Group, World Bank Country and Lending Groups (2018), https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups (accessed November 9, 2018).

[15] B. K. Atsu, C. Guure, and A. K. Laar, Determinants of overweight with concurrent stunting among Ghanaian children, *BMC Pediatr* **17 (**2017), 177.

[16] I. Tareque, A. Koshio, A. D. Tiedt, and T. Hasegawa, Are the rates of hypertension and diabetes higher in people from lower socioeconomic status in Bangladesh? Results from a nationally representative survey, *PLoS One* **10** (2015), e0127954.

[17] World Health Organization (WHO), BMI Classification (2006), http://www.who.int/bmi/index.jsp?introPage=intro_3.html (accessed November 9, 2018).

[18] K. M. Flegal, B. I. Graubard, D. F. Williamson, et al., Excess deaths associated with underweight, overweight, and obesity, *JAMA* **293** (2005), 1861-1867.

Address for correspondence

Lincoln R. Sheets, MD, PhD
204 Medical Science Building
University of Missouri
Columbia MO 65212
SheetsLR@health.missouri.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190240

Multimodal Behavior Analysis Towards Detecting Mild Cognitive Impairment: Preliminary Results on Gait and Speech

Kaoru Shinkawa[a], Akihiro Kosugi[a], Masafumi Nishimura[b], Miyuki Nemoto[c], Kiyotaka Nemoto[c], Tomoko Takeuchi[c], Yuriko Numata[c], Ryohei Watanabe[c], Eriko Tsukada[c], Miho Ota[c], Shinji Higashi[c], Tetsuaki Arai[c], Yasunori Yamada[a]

[a] *IBM Research, Tokyo, Japan*
[b] *Shizuoka University, Shizuoka, Japan*
[c] *University of Tsukuba, Ibaraki, Japan*

Abstract

Behavioral analysis for identifying changes in cognitive and physical functioning is expected to help detect dementia such as mild cognitive impairment (MCI) at an early stage. Speech and gait features have been especially recognized as behavioral biomarkers for dementia that possibly occur early in its course, including MCI. However, there are no studies investigating whether exploiting the combination of multimodal behavioral data could improve detection accuracy. In this study, we collected speech and gait behavioral data from Japanese seniors consisting of cognitively healthy adults and patients with MCI. Comparing the models using single modality behavioral data, we showed that the model using multimodal behavioral data could improve detection by up to 5.9%, achieving 82.4% accuracy (chance 55.9%). Our results suggest that the combination of multimodal behavioral features capturing different functional changes resulting from dementia might improve accuracy and help timely diagnosis at an early stage.

Keywords:

Alzheimer's disease, linguistic features, motion capture

Introduction

As the world's elderly population increases, the number of people living with dementia is rising rapidly, making dementia an increasingly serious health and social problem. According to a previous survey, around 47 million people globally were living with dementia as of 2015, corresponding to about 7.6% of the world's over-65-year-olds [1]. While dementia affects individuals within many domains, including cognition, neuropsychiatric symptoms, activities of daily living, and usually comorbid physical illnesses, it also affects their supporters, including relatives and even wider society because people with dementia require constant and costly care for years [2]. In fact, approximately 85% of the healthcare costs of dementia are related to family and society, and global annual costs reached over 818 billion USD in 2015 [2]. In contrast, there have been numerous attempts to develop an effective drug to combat dementia, but the results of trials have all been negative [2]. One of the remaining possible way to improve this situation is thought to intervene at earlier disease stages such as mild cognitive impairment (MCI) or preclinical Alzheimer's disease. Although there is no validated effective intervention, longitudinal studies suggested that early intervention at stage of MCI might reduce the progression to dementia [3]. An intervention that delays the onset of Alzheimer's disease (AD) dementia by 5 years is estimated to result in a 57% reduction in the number of AD patients and to reduce 45% of the projected Medicare costs [4]. Thus, early detection of dementia has been increasingly important. However, diagnostic coverage of dementia worldwide remains low, and even in high-income countries, only 40–50% of dementia sufferers have received a diagnosis [5,6]. Thus, detection at earlier disease stages should be a more challenging issue.

One of the most promising ways to detect dementia at early disease stages is identifying the evolution of behavioral change over the course of dementia's progression. Instead of medical examinations including brain imaging as well as in-clinic neuropsychological assessment, being capable of inferring dementia and MCI from behavioral features that can be measured in various everyday situations holds promise for increasing the opportunity for timely detection.

Among the behaviors, the most investigated might be speech and gait [7]. Speech has been used for characterizing language dysfunction resulting from cognitive changes [8-12]. For example, memory impairment causes difficulties with word-finding and word-retrieving, which has been measured by speech features such as fillers, including non-words and short phrases (e.g., "umm" or "uh") [9]. The reduction in speech expressiveness is another language dysfunction typically observed in both MCI and AD. This reduction is measured by the decrease in adjectives and indicators related to vocabulary richness (such as type-token ratio and Brunet's index) [10]. Gait disturbances are also common across the dementia spectrum, although the main clinical hallmark of dementia is cognitive impairment especially related to memory impairment [13]. Over the past decade, large cohort studies on dementia have shown the relationship between the severity of cognitive impairment and increased gait abnormalities [13-16]. For example, dementia is associated with a decrease in gait velocity and an increase in stride variability [17]. While there are limited studies evaluating gait in MCI, some but not all studies have suggested that gait dysfunction can be observed in patients with MCI even under normal walking conditions [13]. Although speech and gait features have been suggested as behavioral biomarkers for AD and possibly early in the course of dementia, including MCI, no studies have investigated whether combining both behavioral features could improve detection accuracy for MCI and AD. If speech and gait each could capture different aspects of subtle changes related to physical and cognitive functioning, the multimodal behavioral approach seems to be promising for building a model enabling detection at an earlier disease stage.

In this study, we investigate whether combining behavioral features of speech and gait could improve detection accuracy for patients with MCI. We collected speech and gait behavioral data from Japanese seniors consisting of cognitively healthy

adults and patients with MCI. Specifically, speech data were collected while performing an picture description task by microphones, while gait performance was assessed during a 5-meter walk at normal speed by a marker-based motion capture system. Through the analyses, we demonstrate that exploiting the combination of multimodal behavioral data to identify MCI outperforms that based on each set of single-modality data alone by up to 5.9%, achieving 82.4% accuracy (chance 55.9%). The results demonstrate how our multimodal behavioral analysis could identify the early stages of dementia by exploiting the combination of subtle changes.

Methods

Participants

We collected data from 34 elderly individuals (20 females and 14 males, between 64 and 82 years, i.e., 73.06 ± 4.76). Nineteen participants were grouped as healthy controls (HCs) and fifteen as having MCI. There was no significant difference in age among the groups. Table 1 shows the number of participants (number of female participants), mean age, and mean Mini-Mental State Examination (MMSE) score for the HC and MCI groups. None of the participants in the HC group were diagnosed as having MCI or dementia before the experiment. The definitions of the MCI groups were based on diagnosis by psychiatrists through medical examinations including structural magnetic resonance imaging, blood tests, and neuropsychological tests. More specifically, the doctors followed the guidelines and criteria of the study by Petersen et al [18]. This study was conducted under the approval of the Ethics Committee, University of Tsukuba Hospital.

Speech data collection and feature extraction

We collected speech data from the participants while they performed the Cookie Theft picture description task. The Cookie Theft picture description task, adapted from the Boston Diagnostic Aphasia Examination, is a task used to test the production of free speech in a structured context [19]. A picture is provided to the participants, showing a mother and two children (a boy and girl). Participants are asked to tell everything they see in the picture. This task is used to test the ability of the participants to describe the characters and events in the scene.

The task was administered using a 2nd generation iPad Air through a web-based application using a Wizard of Oz experiment method. This method is efficient for examining user interaction with computers and facilitating rapid iterative development of dialog wording and logic. The method requires two machines linked together, one for the subject and one for the experimenter. In this implementation, the experimenter (the "Wizard"), pretending to be a computer, "operates" using complete replies to user queries or presses function keys to which common messages have been assigned. The software automatically records the dialog and its timing. This was done as a step towards a fully automated system to first assess the ability to achieve similar results using tablet devices as in traditional assessment styles.

Table 1 – Demographics of participants

Status	No. of Participants (Female)	Mean Age (SD)	Mean MMSE (SD)
Control	19 (12)	71.63 (4.39)	28.42 (1.47)
MCI	15 (8)	74.87 (4.73)	25.53 (3.89)

Voice recordings were collected using three microphones: a throat microphone (NANZU SH-12iK), a lavalier microphone (SONY ECM-CS3), and the iPad's internal microphone. The throat and lavalier microphones were fitted onto the participants' necks and connected to a USB recording device (ZOOM H1/MB) for voice recording (wav format, 44.1 k/stereo) (Figure 1A). Analysis was done on voice recordings gathered from the lavalier microphone, which were selected after synchronization with the throat microphone showing which portions of the recordings contained actual speech from participants. The throat microphone does not record the sound of open-air, so participants' speech area can be detected by extracting the sections above a certain volume level in the recorded audio. Recordings from the iPad's internal microphone were collected but not analyzed in this study. These recordings will be used in later analysis after the proper features have been selected and a model is developed from higher quality audio. After assessment, voice recordings from the lavalier microphone were used for preprocessing. The recorded audio was preprocessed by automatic speech recognition (ASR), which automatically transcribes audio data into text format. Then the experimenter manually corrected the errors of the ASR by listening to the recorded audio. The experimenter also annotated fillers and false starts during the transcribing procedure. For the preprocessing, we used the Japanese morphological analyzer MeCab [20].

Once all the transcribing and annotatince were completed, we collected speech features on the basis of the previous studies.

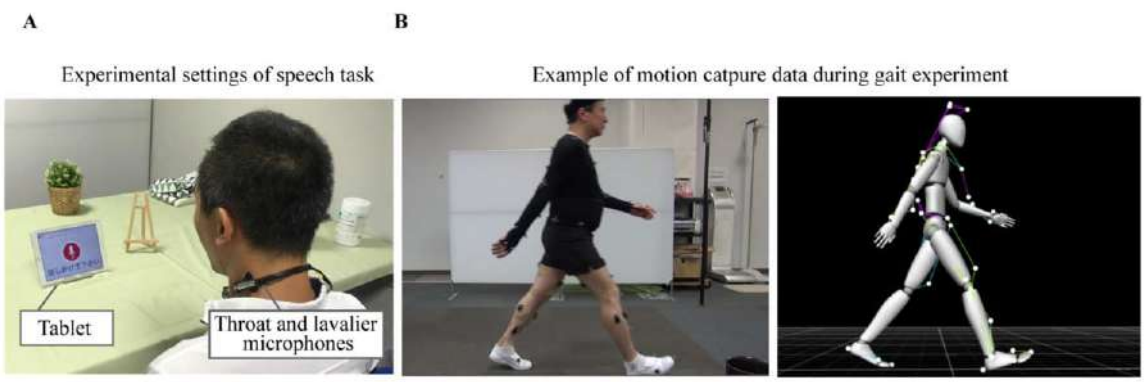


Figure 1 – Overview of speech and gait experiments. (A) Example of the speech experimental setup. (B) Example of motion capture data during gait experiment.

Previous study reported that syntactic complexity is closely associated with the incidence of dementia [21]. Syntactic complexity was measured in various ways, such as the mean length of sentences, number of sentences, ”part-of-speech” frequency, and dependency distance [10,22]. Dependency distance infers the number of intervening words between two syntactically related words in a sentence [23]. Another feature that must be considered is vocabulary richness. This measures lexical diversity, which tends to reduce in dementia cases [9]. Repetitiveness is also reported as an important factor of capturing the linguistic dysfunctions of the patients with dementia [24]. Some matrixes measure the frequency of repeated words and phrases, and others estimate sentence similarities by calculating the cosine distance between two sentences [10]. The feature which widely used in image description tasks is semantic density. It was calculated on the basis of ”informational units” that are predefined objects or text segments that might refer to important information. For example, in the Boston Cookie Theft picture description task, information units consist of objects such as ”Woman”, ”Cookies”, and ”Boy taking the cookie”. With the information units, semantic density can be defined as the number of information units divided by the total number of words [10,22]. A number of previous studies have found that individuals with dementia tend to produce speech with lower information, defined as semantic density, than with healthy controls [22, 25].

In this study, we collected 49 speech features from the transcribed and annoteated text of picture descrtion task. It includes twenty-seven features related to parts-of-speech (POS), four features related to information units (e.g. frequency, ratio), six features for syncatcic complexity (e.g. maximum dependency level), six different measures of sentence similarities (e.g. cosine-disntance), and three measures of vocabulary richness (e.g. type-token-ratio). Number of false starts and number of fillers are extracted from the annotation in the transcribed text. Addition to the text features, the speech rate (audio length/syllables) was also caluclated.

Gait data collection and feature extraction

The gait experiment was conducted on the same day as the speech experiment. The participants took enough rest between experiments. We collected motion data of the participants walking five meters in a lab area at their usual speed. To ensure that gait features were collected during steady-state walking, participants started walking at least two meters before the target zone and completed their walk at least two meters beyond it, which makes nine meters walk in total. Start and end points were marked on the floor with tape. Before each experiment, a trained experimenter gave verbal task instructions to the participants as “Please walk up to the tape at your usual speed”.

Participants’ body positions during their walk were captured by the motion capture system, OptiTrack Flex 13, with eight cameras allocated in a 6 meters × 12 meters walking area (Figure 1B). Three-dimensional kinematic data were measured at 120 Hz using a full body skeleton model of 50 markers. 50 markers includes head (4 markers), torso (6 markers), waist (4 markers), shoulder (4 markers), arm (6 markers), hand (8 markers), leg (8 markers), foot (6 markers), and toe (4 markers). Skeleton is calibrated for each participant before the gait experiment.

In a previous study, a broad range of characteristics is used to describe gait performance. Gait speed is applied widely as an evaluative and a predictive measure of various diseases [13,14,26]. Other measures such as gait velocity and stride variability has also been suggested as a sign of dementia [17]. Step width and step width variability reflect the postural control of gait [27].

In this study, we collected 13 gait features from the positional data obtained from each markers during a gait task. We calculated the gait cycle from four markers on each foot (toe in, toe out, toe tip, and heel). Once all the data preprocessing was completed, gait speed (walk length/time), step length (mean and standard deviation), stride (mean and standard deviation), left-right stride variability, toe angle (mean and standard deviation), left-right toe angle variability and step width (mean and standard deviation) were extracted from gait cycle. We also

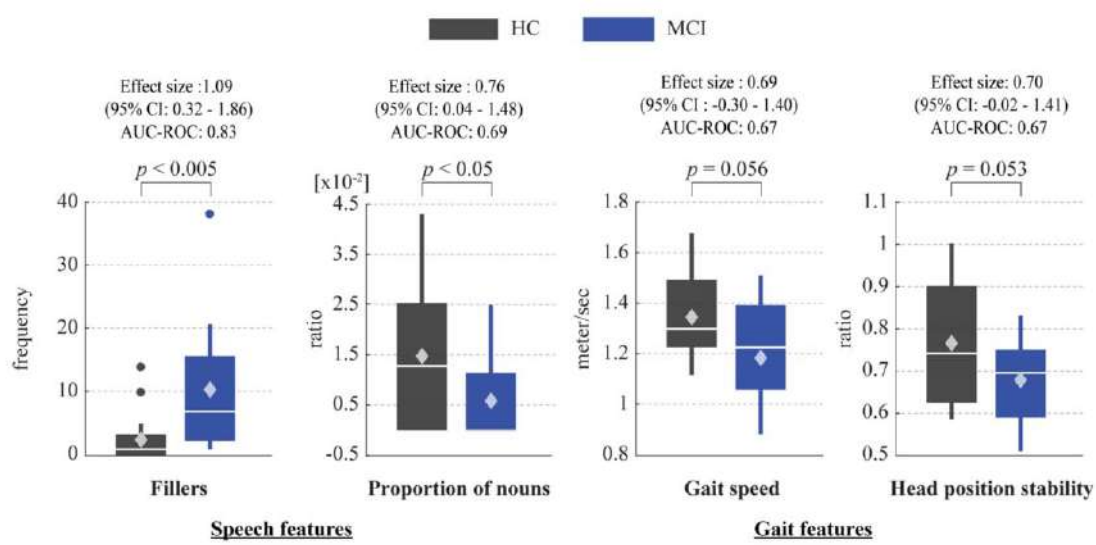


Figure 2 – *Distributions for speech and gait features which may have tendancy between HC and MCI groups. Boxes denote 25th (Q1) and 75th (Q3) percentiles. Line within box denotes 50th percentile, while whiskers denote upper and lower adjacent values that are most extreme within Q3+1.5(Q3-Q1) and Q1-1.5(Q3- Q1), respectively. Filled circles show outliers, and squares represent mean values.*

extracted the position of head top marker and calculated the variablity of the distance from the direction of travel to see the stability of the head position.

Feature anaysis & classification models

After linguistic features and gait features were extracted, we investigated if each feature differed significantly between the HC and MCI groups. We performed Student t-test for statistical examination, and considered to be significantly different when the p value was less than 0.05.

Classification was then done using support vector machine (SVM) models with a linear kernel function with a two-class classification model for testing the strength of features for differentiating patients with MCI from HCs in an automated fashion [28]. We used the algorithm for the SVM models implemented in MATLAB (MathWorks Inc., Natick, MA). Features were selected on the basis of the receiver operating characteristic (ROC) score for each feature.

We evaluated the model's accuracy by leave-one-out cross-validation, where classifers were trained using data collected from all participants expect one and then were tested on data of the one participant left out of the training data set. After obtaining loss estimate using cross-validation, accuracy, sensitivity, specificity and F-measure were calaulated for the model.

Results

We investigated if these features differed significantly between the HC and MCI groups for each task. Through the analysis, two features in speech, fillers and proportion of nouns showed siginificant difference between HC and MCI groups ($p < 0.005$ and $p < 0.05$, respectively; Figure 2). We calculated the effect size (Cohen's d) of each feature as discriminative power [29]. For Cohen's d, the 0.8 effect size is thought to be large, while the 0.5 effect is medium, and the 0.2 effect size is small [29]. We found large and medium effect size with fillers (effect size of 1.09, 95% CI: 0.32-1.86; Figure 2) and medium effect size with proportion of noun (effect size of 0.76, 95% CI: 0.04-1.48; Figure 2). For the gait features, gait speed and head postition stability were higher with HC than MCI, although this differences were not statistically significant (p=0.056 and p=0.053, respectively; Figure 2). These changes in the speech and gait features consistent with the results of previous studies that investigated the differnces in these measures with HC and dementia [10,13,22].

Next, we evaluated classification models for each speech and gait task between HC and MCI groups. As a result, for the speech features, three features were selected to be used for the classification model, which achieved 76.5% accuracy (sensitivity: 73.3%, specificity: 78.9%, F-measure: 0.733). As for gait features, three features were also selected for the classification model, which achieved 76.5% accuracy (sensitivity: 88.9%, specificity: 72.0%, F-measure: 0.667).

Finally, we analyzed the classification model by combining speech and gait features. We combined speech and gait features that were selected in each classification model. The accuracy was 82.4% (sensitivity: 76.5%, specificity: 88.2%, F-measure: 0.813) which improved by 5.9% from that of the classification models using the features of individual tasks.

Table 2 – Classification model performance for HC vs. MCI. Performance is measured after leave-one-out cross-validation (Acc: accuracy, Spe: specificity, Sen: sensitivity).

Features	Acc (%)	Spe (%)	Sen (%)	F-measure
MMSE (baseline)	76.5	72.0	88.9	0.667
Speech	76.5	78.9	73.3	0.733
Gait	76.5	72.0	88.9	0.667
Speech & Gait	82.4	88.2	76.5	0.813

Discussion

In contrast to previous studies focusing on detecting MCI and AD from a single modality of behavioral data, we aimed to argue the effectiveness of multimodal behavioral analysis. To this end, we collected and investigated speech data during the image description task and gait data during the 5-meter gait test with normal speed from cognitively healthy controls and MCIs.

We first investigated the significant differences of each feature and found significance only in speech features with a medium to large effect size. We did not find differences in gait features when comparing people with MCI to controls. Gait disturbance in MCI is still controversial [30], but a reason for the absence of significant differences can be explained by the following: gait patterns differ between MCI subtypes, and then the use of a heterogeneous MCI group, including people with single-domain and multi-domain MCI, increases the intra-group variability of gait features in MCI groups [31,32]. In our experiment, such a heterogeneous MCI group in addition to the small number of participants might be a possible reason. Therefore, we could not find features that differed significantly between the MCI and controls with a larger effect size in both sets of behavioral data.

We then built a classification model by combining these features. When we built models by using speech and gait data separately, the models each showed the same accuracy of 76.5%, which was the same accuracy of the baseline model by using MMSE scores. In comparison with these models, we showed that the model using multimodal behavioral data could improve detection by up to 5.9%, achieving 82.4% accuracy. The results indicate that exploiting the combination of multimodal behavioral data could improve the detection accuracy of MCI. Though no gait features differed significantly between the MCI and controls, these gait features could contribute to improving detection performance by combining speech features. These results might be made possible by exploiting the combination of multimodal behavioral features capturing different functional subtle changes resulting from MCI. If this hypothesis is true, our approach focusing on the multimodal behavioral analysis might be important especially (i) in targeting at earlier stages such as MCI or preclinical AD stages and (ii) in using data collected in non-controlled environments such as free living situations with a lot of noise.

One of the limitation in this study is the relatively small number of participants.. For future work, we will need to confirm our results on a larger number of participants. Another limitation is the limited number of task trials. The speech and gait data we analyzed were collected from a single trial from a single task. We need to conduct further research with data from multiple trials to verify our results.

To the best of our knowledge, this is the first empirical study to demonstrate that multimodal behavioral analysis on speech and gait data could improve detection accuracy for patients with MCI and might be useful for early detection of AD. We hope the results of our study will help promote future efforts towards timely diagnosis at an early stage such as MCI.

References

[1] M. J. Prince, M. Prina and M. Guerchet, World Alzheimer Report 2013: journey of caring: an analysis of long-term care for dementia. *Alzheimer's Disease International,* (2013).

[2] G. Livingston, A. Sommerlad, V. Orgeta, S. G. Costafreda, J. Huntley, *et al.,* Dementia prevention, intervention, and care. *The Lancet*, 390(10113) (2017), 2673-2734.

[3] T. Horr, B. Messinger-Rapport, and J. A. Pillai, Systematic review of strengths and limitations of randomized controlled trials for non-pharmacological interventions in mild cognitive impairment: focus on Alzheimer's disease. *J Nutr Health Aging 19*(2) (2015), 141-153.

[4] R. A. Sperling, P. S. Aisen, L. A Beckett, D. A. Bennett, *et al.*, Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia*, 7(3) (2011), 280-292.

[5] M. Prince, A. Comas-Herrera, M. Knapp, M. Guerchet, *et al.,*World Alzheimer report 2016: improving healthcare for people living with dementia: coverage, quality and costs now and in the future (2016).

[6] V. Kotagal, *et al*., Factors associated with cognitive evaluations in the United States. *Neurology* 81 (2015), 64–71.

[7] A. Alberdi, A. Aztiria, and A. Basarab, On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey. *Artificial Intelligence in Medicine*, *71* (2016), 1-29.

[8] I. Hoffmann, D. Nemeth, C. D. Dye, M. Pákáski, *et al.,* Temporal parameters of spontaneous speech in Alzheimer's disease. *Int J Speech Lang Pathol*, 12(1) (2010), 29-34.

[9] C. I. Guinn and A. Habash, Language Analysis of Speakers with Dementia of the Alzheimer's Type. In *AAAI Fall Symposium: Artificial Intelligence for Gerontechnology* (2012), 8-13.

[10] K. C. Fraser, J. A. Meltzer, and F. Rudzicz, Linguistic features identify Alzheimer's disease in narrative speech. *J Alzheimers Dis 49*(2) (2016), 407-422.

[11] K. Shinkawa and Y. Yamada, Word Repetition in Separate Conversations for Detecting Dementia: A Preliminary Evaluation on Data of Regular Monitoring Service. *Summit Transl Bioinform* (2018), 206-215.

[12] K. Shinkawa and Y. Yamada, Topic Repetition in Conversations on Different Days as a Sign of Dementia. *Studies in health technology and informatics*, *247* (2018), 641-645.

[13] V. Valkanova and K.P. Ebmeier, What can gait tell us about dementia? Review of epidemiological and neuropsychological evidence. *Gait & posture*, 53 (2017), 215-223.

[14] L. M. Waite, D. A. Grayson, O. Piguet, H. Creasey, *et al.,* Gait slowing as a predictor of incident dementia: 6-year longitudinal data from the Sydney Older Persons Study. *J Neurol Sci*, *229* (2005), 89-93.

[15] L. M. Allan, C. G. Ballard, D. J. Burn, and R. A Kenny, Prevalence and severity of gait disorders in Alzheimer's and non-Alzheimer's dementias. *J Am Geriatr Soc*, *53*(10) (2005), 1681-1687.

[16] E. Scherder, L. Eggermont, J. Sergeant, and F. Boersma, Physical activity and cognition in Alzheimer's disease: relationship to vascular risk factors, executive functions and gait. *Rev Neurosci*, *18*(2) (2007), 149-158.

[17] M. B. van Iersel, W. Hoefsloot, M. Munneke, B. Bloem, *et al*., Systematic review of quantitative clinical gait analysis in patients with dementia. *Z Gerontol Geriatr*, *37*(1) (2004), 27-32.

[18] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, et al., Aging, memory, and mild cognitive impairment. *International psychogeriatrics*, 9(S1) (1997), 65–69.

[19] E. Kaplan, The assessment of aphasia and related disorders (Vol. 2). *Lippincott Williams & Wilkins* (1983).

[20] T. Kudo, Mecab: Yet another part-of-speech and morphological analyzer. *http://mecab. sourceforge. net/* (2005).

[21] S. Kemper, M. Thompson, and J. Marquis, Longitudinal change in language production: Effects of aging and dementia on grammatical complexity and semantic content. *Psychol Aging* (2001), 600–614.

[22] S. Ahmed, A. M. F. Haigh, C. A. de Jager, and P. Garrard, Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain*, 136(12) (2013), 3727-3737.

[23] S. Pakhomov, D. Chacon, M. Wicklund, and J. Gundel, Computerized assessment of syntactic complexity in Alzheimer's disease: A case study of Iris Murdoch's writing. *Behavior research methods*, *43*(1) (2011), 136-144.

[24] Tomoeda CK, Bayles KA, Trosset MW, Azuma T, and McGeagh A, Cross-sectional analysis of Alzheimer disease effects on oral discourse in a picture description task. *Alzheimer Dis Assoc Disord* 10 (1996), 204-215.

[25] B. Croisile, B. Ska, M. J. Brabant, A. Duchene, *et al.*, Comparative study of oral and written picture description in patients with Alzheimer's disease. *Brain Lang,* 53(1) (1996), 1–19.

[26] S. Lord, B. Galna, J. Verghese, S. Coleman, *et al.*, Independent domains of gait in older adults and associated motor and nonmotor attributes: validation of a factor analysis approach. *J Gerontol A Biol Sci Med Sci*, 68(7) (2012), 820-827.

[27] E. Nordin, R. Moe-Nilssen, A. Ramnemark, and L. Lundin-Olsson, Changes in step-width during dual-task walking predicts falls. *Gait & posture*, 32(1) (2010), 92-97.

[28] B. E. Boser, I. M. Guyon and V. N. Vapnik, A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory, ACM* (1992)*,* 144-152

[29] S. Nakagawa and I. C. Cuthill, Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological reviews*, 82(4) (2007), 591-605.

[30] A.F. Pettersson, E. Olsson, and L.O. Wahlund, Motor function in subjects with mild cognitive impairment and early Alzheimer's disease, *Dement Geriatr Cogn Dis,* 19 (5–6) (2005), 299–304.

[31] M. Montero-Odasso, A. Oteng-Amoako, M. Speechley, K. Gopaul, *et al.,* The motor signature of mild cognitive impairment: results from the gait and brain study, *J Gerontol A Biol Sci Med Sci,* 69 (11) (2014), 1415–1421.

[32] J. Verghese, M. Robbins, R. Holtzer, M. Zimmerman, *et al.,* Gait dysfunction in mild cognitive impairment syndromes, *J Am Geriatr Soc*, 56 (7) (2008), 1244–1251

Address for correspondence

Kaoru Shinkawa (e-mail address: kaoruma@jp.ibm.com)

Yasunori Yamada (e-mail address: ysnr@jp.ibm.com)

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190241

Finding Reasons for Vaccination Hesitancy: Evaluating Semi-Automatic Coding of Internet Discussion Forums

Maria Skeppstedt[a,b], Andreas Kerren[a], Manfred Stede[b]

[a] *Department of Computer Science and Media Technology, Linnaeus University, Växjö, Sweden,*
[b] *Applied Computational Linguistics, University of Potsdam, Potsdam, Germany*

Abstract

Computer-assisted text coding can facilitate the analysis of large text collections. To evaluate the functionality of providing an analyst with a ranked list of suggestions for suitable text codes, we used a data set of discussion posts, which had been manually coded for reasons given for taking a stance on the topic of vaccination. We trained a logistic regression classifier to rank these reasons according to the probability that they would be present in the post. The approach was evaluated for its ability to include the expected reasons among the n top-ranked reasons, using an n between 1 and 6. The logistic regression-based ranking was more effective than the baseline, which ranked reasons according to their frequency in the training data. Providing such a list of possible codes, ranked by logistic regression, could therefore be a useful feature in a tool for text coding.

Keywords:
Vaccination Refusal, Text Mining, Supervised Machine Learning

Introduction

Vaccination hesitancy has led to outbreaks of vaccine-preventable diseases in several parts of the world [1]. To study reasons that are given for vaccination refusal and vaccination hesitancy in different types of user-generated text resources, e.g., Internet discussion forums, might be one method for increasing our knowledge about this phenomenon.

There are previous studies in which these types of texts have been manually analysed for reasons for vaccination hesitancy [2, 3]. There are also studies which have applied topic modelling for automatically extracting vaccination-related information from text collections too large for a fully manual analysis [4, 5]. Previous research has shown topic modelling to be an efficient text-mining method for selecting and topically sorting texts, but a manual coding of the selected texts is also recommended for a deeper understanding of their content [6].

We have, in a previous study, performed such a manual coding of texts selected through topic modelling. We had applied topic modelling on Internet discussions of the subject of vaccination, and the topic modelling algorithm automatically extracted six topics present in the collection. We then manually coded 50 texts associated with each one of these topics, through identifying the reasons that the authors had given for taking a stance *for* or *against* vaccination. We detected 242 different unique reasons in total, some only occurring once and some reoccurring [7]. Although the text selection and sorting that was provided by the topic modelling facilitated the coding, it was still a demanding task. It was particularly difficult to determine whether a reason identified in the text was a new one, not yet included in the analysis, or whether this reason had occurred in any of the previously analysed texts.

The coding procedure might, therefore, be simplified if the user is assisted in the task of locating suitable text codes among the codes previously identified. That is, when an analyst selects a text for a manual coding of reasons, a system could automatically suggest which previously identified reasons that might be found in the text that is being coded. The system could, for instance, rank previously identified reasons according to how likely it is that they are present in the text. Such a ranked list could support the analyst in the task of determining whether a reason identified in the text that is being coded has occurred in any of the previously analysed texts.

The aim of the study described in this paper is to design and evaluate a method for providing such a ranked list of previously identified reasons.

Previous Studies on Reason Classification and Identification

Among the large body of recent research on stance detection and argumentation mining, the ones most relevant for the present study investigate a task that is referred to as argument identification or recognition [8, 9] or as reason identification and classification [10]. These studies have either (i) analysed a text collection for all reasons occurring in the collection, and then not included the infrequently occurring ones in the automatic classification experiments [10], or (ii) used external resources for finding a few important reasons that are likely to occur in the text collection, and then labelled the texts with these reason categories [8, 11]. When training classifiers on four different text collections to recognise 11, 12, 15, and 18 reason categories, respectively, F-scores in the low 50s were achieved [10]. When instead training classifiers on two different text collections to recognise six and seven reason categories, respectively, micro-average F-scores between 0.7-0.8 were achieved [8]. Topic-independent classifiers were created by using training data that was constructed through measuring the similarity between the opinionated text and the text describing the reason. Finally, an F-score of 0.49 was achieved for the task of classifying texts into 16 reason categories, by using the method of clustering texts according to 16 topics derived from topic modelling based on Non-negative Matrix Factorization (NMF) [11]. The automatically

extracted topics were manually mapped to 16 pre-defined, prominent reasons. There were also previous attempts to identify prominent reasons in texts in a fully unsupervised fashion, which however have yielded low results [9].

In contrast, for the use case studied here, also the less frequently used reasons must be included. A final report of the outcome of the text coding might only include frequently occurring reasons; however, during the process in which the analyst carries out the text coding through identifying reasons in the texts, it is still unknown which reasons that occur frequently in the text collection. Both the prominent and non-prominent ones among previously identified reasons are therefore relevant to identify in the text that is in the process of analysis. Consequently, both prominent and non-prominent reasons should be provided as coding suggestions to the user, and should therefore also be included in the experiment.

Methods

The experiments consisted of training a classifier to rank the reasons that had previously been identified in the coding, according to the likelihood that they are present in the text that is currently coded.

The Previously Constructed Data Set

Our previous study consisted of applying the NMF topic modelling algorithm on posts from vaccine-related Internet discussions from the British parental website Mumsnet [7]. We chose the data based on the fact that the discussions were taking place on a general discussion forum for parents. Although views expressed in the Mumsnet forum by no means could be claimed to be representative for a population other than its actively participating users, the opinions expressed on Mumsnet are likely to be more general than, e.g., opinions expressed on an anti-vaccination website. In addition, Mumsnet informs their users of the fact that the site's discussion threads are published as publicly available posts, for which no login is required. Debaters are also asked to anonymise their postings, e.g., by using a chat nickname. Therefore, according to the data classifications provided by Eysenbach and Till [12], the content of the forum belongs to a public sphere rather than a private one.

Before applying topic modelling, we pre-processed the texts by removing stop words and performing a concept cluster-detection. That is, we concatenated into one term word collocations that occurred frequently in the text collection, and we replaced different term instantiations of the same concept,

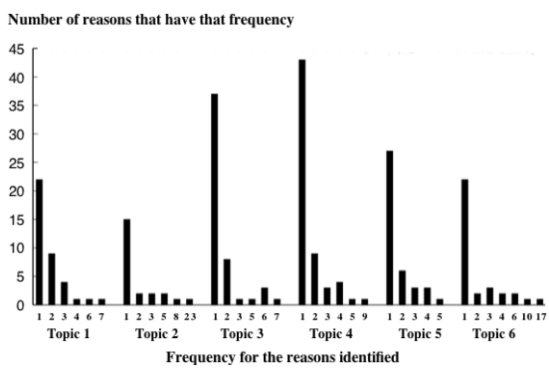


Figure 1 – The distribution of the frequency of the unique reasons identified for the six topics.

e.g., synonyms and different inflections of a word, with a common string representing this concept. The term-replacement was achieved by the use of word embedding vectors associated with the terms in the corpus. These vectors were clustered using DBSCAN clustering [13], and terms whose vectors were assigned to the same cluster were considered as belonging to the same concept. We chose DBSCAN, since it uses the properties of the data set that is to be clustered to determine how many clusters to create, and the user thereby does not have to provide an a priori specification of the number of appropriate clusters. A clustering algorithm with this property is a requirement for this clustering task, as the number of clusters of inflections and synonyms that occur in the text collection is unknown beforehand. We used a final cluster set of 402 clusters, which we obtained by a manual quality control of the automatically constructed clusters. Through the quality control, 165 terms were added to a list of terms to exclude in the cluster construction process.[1]

The topic modelling algorithm automatically extracted six topics from the text collection. The subject of Topic 1 and 4 was Measles, Mumps, and Rubella (MMR) vaccination; for Topic 1, it was in the form of reasons related to fear of and research on adverse vaccine reactions, and for Topic 4, it related to vaccine immunity duration, disease severity, and single vaccines. Topic 2 was about the vaccinology professor Paul Offit, Topic 3 was about eradication of diseases through vaccinations, and Topic 5 contained reasons related to trust/distrust in the medical profession and industry, as well as reasons related to attitudes towards vaccination among medical professionals. Finally, texts for Topic 6 discussed risk assessments for vaccination, for diseases, and for infecting others.

For each one of the topics extracted by the topic modelling algorithm, we manually coded 50 texts for reasons that were given *for* and *against* vaccination. This analysis consisted of reading the selected texts and identifying reasons mentioned in the text. A reason identified could either be (i) a reason that had not previously occurred in the text collection, or (ii) a reason previously identified in the text collection. In the first case, we wrote a new description of this reason in the coding sheet, and in the second case, we noted the exact same description as previously used in the coding sheet. We performed the analysis in Microsoft Excel, with one column for the texts and additional columns for the descriptions of the reasons identified in the texts.

Most of the coded descriptions of reasons were formulated as an argumentative statement, e.g., "*There is no proven link between MMR and autism, despite many studies*", "*That small pox vaccination has been successful does not mean that there are no problems with other vaccines*", "*Not only the MMR combination should be offered, but also single vaccinations*", and "*The risk of catching the vaccine-preventable disease or that the disease will result in complications is higher than the risk of the vaccination*". Some were, however, formulated as a meta-description of reasons, e.g., "*Expression of distrust in government/pharmaceutical industry*". The original study reported on the detailed results of the coding [7].

The number of reasons identified varied between the different topics, as did the frequency of occurrence for the reasons (Figure 1). For a few texts, no reasons were detected, and these were removed from the data set used in the experiments for the present study. This resulted in a final data set for the six topics that contained 42, 47, 50, 49, 49 and 50 texts,

[1] Examples of word pairs clustered together are: worry/concern, problem/difficulty and autism/autistic

respectively, and where each one of them had one or several associated reasons.

Classifier Training

We carried out six separate experiments for each one of the extracted topics. That is, classifiers were trained and evaluated on texts and reasons that belonged to one topic at a time. In addition, we also performed an experiment on a data set in which the six topic-divided data sets were combined into one set. This evaluation used leave-one-out cross-fold validation.

The type of machine learning classifier chosen for the experiment was a logistic regression classifier. This choice was based on the inherent data scarcity of the task evaluated. That is, the task of assisting a text coder in locating the appropriate reason category among a set of previously identified reasons required a classifier suitable for small training data sets. This inherent data scarcity is due to that, in the typical use case, most of these previously indentified reasons will have a limited number of associated texts that can be used for forming the training data set.

Moreover, the logistic regression classifier returns a probability estimate for each one of the possible classification categories in the data set. This probability estimate is not only suitable to use for ranking the previously created reasons that the coding process suggests to the user, but it also forms an output that is human interpretable. The probability estimate of a logistic regression classifier could therefore be shown to the user as an indication of the likelihood that a previously identified reason is present in the text that is currently being coded. This would not be the case for classification outputs such as the distance to the separating hyperplane of a support vector machine, which would make little sense to a human text coder.

The situation simulated in the experiments was that all texts associated with a topic would have been coded for reasons, except one of them. That is, a situation in which the left out data point in the leave-one-out scheme would not yet have been coded. We used the rest of the data, i.e., all data points except the one left out, for training a logistic regression classifier. Features were extracted from the texts, and their associated reasons were used as classification categories. Texts with several associated reasons were added multiple times to the training data set, once for each of its associated reasons, with that reason as the classification category. The describing text for each one of the coded reasons was also added to the training data, to form one additional training data point for each one of the classification categories.

Tokens that occurred at least twice in the data set were used as features, and the same stop word list as had been used for constructing the topic models was applied to remove stop words. Due to the scarcity of data, no n-grams were used. The probability estimate for each classification category that is returned by the logistic regression classifier was used for producing a ranked list of reasons, which thus formed the output of the classification. That is, the larger the probability that a text would be classified as containing a certain reason, the higher ranking did this reason achieve in the ranked list produced for this text.

Apart from the standard method of using tokens as features, we also performed an experiment with the same term-cluster replacement as used for the topic modelling. That is, different terms that belong to the same concept were replaced with a unique string representing this concept. This concept-string was then treated as a normal token by the feature extraction procedure.

The logistic regression classifier available in scikit-learn was used with default parameters [14], as we considered the data available too small for parameter tuning. However, we ran all experiments with the default inverse L2 regularisation strength of 1 as well as with a strength of 10. We used the Gensim library [15] for accessing embedding vectors and used an out-of-the-box word2vec model,[2] trained on Google news.

Evaluation

Results for the logistic regression classifier with standard features, as well as for the same classifier with cluster features, were recorded for both L2 settings. We applied leave-one-out cross-fold validation on all evaluations.

The situation that the analyst needs to find a reason among previously identified reasons only arises when a reason contained in the text already has been identified in at least one of the previously analysed texts, i.e., not when a text is going to be assigned to a *new* reason. For reasons that have exactly one text association in the data set, the only situation that arises is this one where a new reason is created. A ranking of existing reasons thereby does not make any sense for this situation. Therefore, this evaluation of the reason ranking did not include these text-reason associations as left out data points. The associations were however included in the training data, and thereby as possible classification categories. This is in accordance with the authentic situation, in which it would be unknown to the classifier that the reasons that only occur once in the data are not to be associated with the held-out data point. When performing the classification for the left out data point, these reasons are therefore just as valid as any of the other reasons to include in the ranking. The inclusion of these data points makes the evaluation more realistic, as the task would have been simplified if these reasons had been removed altogether from the experiment.

The ranking was evaluated on the criterion of the proportion of cross-validation folds for which the expected reasons were among the *n* top-ranked reasons. An *n* ranging from 1 to 6 was used, as up to around six reasons would be a reasonable number of coding suggestions to scan through when coding a text.[3]

Baseline and Ceiling

The aim of the study was to find out whether the use of machine learning would be an appropriate approach for solving the task of providing the user with coding suggestions. We therefore focused the study on evaluating the performance of the machine learning classifier we deemed most suitable for this task, rather than on comparing different machine learning algorithms. We compared this classifier, i.e., a logistic regression classifier, to a baseline strategy that did not employ machine learning. This baseline strategy instead ranked the reasons according to their frequency, i.e., according to the number of texts they were assigned to in the training data.

Note that the baseline strategy ranked the texts according to the reason frequency in the training data set, not the entire data set. This means that the reason frequencies that the held-out data point contributed to were not a part of the frequency count, as these would not be known in a realistic setting. This, in turn, results in that the baseline rankings sometimes differ depending on which data point is used as held-out data. It is

[2] code.google.com/archive/p/word2vec/ (March 27 2019)

[3] The experiments can be replicated by running the Python script *run_classifier.py*, found at https://github.com/mariask2/topics2themes (March 27 2019).

therefore possible for the baseline to, for each evaluation fold, fail to produce a correct ranking.

Since the same text often is associated with several reasons, it is not possible to achieve a score of 100 percent correct inclusion of reasons when the evaluation criterion only takes the most top-ranked elements into account. That is, not all *a* reasons that are associated with a held-out text can be included among the *n* top-ranked elements for the cases when $a > n$. For each of the six topics and each of the top *n* rankings for which results were evaluated, we therefore calculated an upper ceiling. If *n=1*, and there is only one associated reason (*a*) that belongs to the text (*a=1*), an optimal ranking will give that reason the highest-ranking position and contribute to one correct classification. However, if there, e.g., are three reasons associated with a text (*a=3*), two of these reasons can *not* be ranked highest, and an optimal ranking of reasons for that text will therefore contribute with one correct classification and two incorrect ones, or generally: for each text, if $a > n$, the number of incorrect classifications of an optimal ranking are equal to $a - n$.

Results

Table 1 below shows results for each one of the six topics as well as for the combined data set. The results show the proportion of times that the expected reason was included among the *n* top-ranked reasons (where *n* ranges from 1 to 6). Four numbers are presented, which includes the standard and cluster features for logistic regression (for an inverse L2 of 10) and the frequency-based baseline ranking and ceiling results, for each one of the points of measure. The boldfaced font indicates the best result among the three ranking methods.

Both of the logistic regression-based methods performed better than the baseline method, except for three points of measure (shown by underscore in Table 1). For one of these points of measure, the lower-performing of the two logistic regression-based methods performed equally to the baseline method, and for the other two points of measure, the lower-performing of the two logistic regression-based methods achieved a result that was two percentage points lower than the result achieved by the baseline. Despite these few exceptions, the logistic regression-based methods clearly outperformed the baseline method with general and large performance differences.

The difference was less evident between standard logistic regression and logistic regression with concept clusters. Which of the two methods performed best varied between different topics and between different values of *n*. Also the regularisation strength had a very limited effect on the overall results. For half of the topics, the results decreased a few percentage points with a stronger regularisation, while it led to a minor increase in results for the other half of the topics.

Discussion

There is a large variation among the six topics in the data set. That is, a variation (i) in terms of the number of identified reasons (from 23 to 61), (ii) in the frequency distribution of these reasons (see Figure 1), (iii) to what extent several reasons were associated with the same text (as indicated by the ceiling calculations), and (iv) in terms of what classification results were achieved. That the logistic regression methods generally performed better than the frequency-based ranking for this diverse data set indicates some level of generalisability of the results, despite the limitation that the experiment was performed on one single data set. Although the actual efficiency of the ranking might differ between different text collections and different manual coding strategies, the results achieved encourage us to implement text coding support by using logistic regression-based ranking of previously identified reasons.

In contrast, since the use of concept-cluster features does not yield any obvious advantage, there is no point in carrying out the more complex feature extraction process that the use of these features entail.

It should be noted that the situation evaluated here is when all but one of the texts already have been assigned reasons, and that other results might have been achieved if an earlier point in the analysis process had been simulated. To perform such a simulation is a task to include in future work. It might also be interesting to, similar to previous work [8], evaluate the performance of a ranking that relies entirely on the analyst's description of the identified reason; however, such an approach might be less suitable for detecting reasons that are formulated in the form of meta-descriptions.

We have previously constructed an interactive visualisation tool whose graphical user interface displays the output of topic modelling applied to a text collection [16]. The tool displays the texts selected by the topic modelling algorithm, and includes methods for searching and sorting among these texts. The tool also provides the functionality of attaching user-defined *themes* to the texts. That is, for the application described here, these themes would correspond to reasons identified when analysing texts. The interactive tool does, however, not yet include any functionality for suggesting which themes to associate to a text that is being analysed. Instead, the tool shows themes in a list previously identified by the user, which is simply sorted according to the creation time of the themes. The next step in developing the tool will therefore be to add the text analysis support investigated here.

*Table 1 – The proportion (%) of evaluated texts for which the expected arguments was found among the n top-ranked reasons (**log**: a logistic regression classifier with standard features, **clu**: a logistic regression classifier with cluster features, and **bas/cei**:baseline/upper ceiling).*

% found	Topic 1 (38 reasonss)			Topic 2 (23 reasons)			Topic 3 (51 reasons)			Topic 4 (61 reasons)			Topic 5 (40 reasons)			Topic 6 (33 reasons)			All topics (242 reasons)		
Rank	*log*	*clu*	*bas/cei*	*log*	*clu*	*bas/cei*	*log*	*clu*	*bas/cei*	*log*	*clu*	*bas/cei*	*log*	*clu*	*bas/cei*	*log*	*clu*	*bas/cei*	*log*	*clu*	*bas/cei*
n=1	**28**	21	15/61	**53**	**53**	45/71	35	**39**	0/58	32	**33**	16/49	**34**	**34**	0/74	**32**	28	28/61	29	**30**	8/61
n=2	**34**	32	28/93	**65**	**65**	61/91	45	**49**	14/81	**47**	46	16/76	**50**	47	13/91	43	**48**	45/89	**41**	**41**	13/86
n=3	**45**	34	28/100	**76**	**76**	61/100	57	**61**	14/90	**63**	60	16/88	**58**	55	16/97	**63**	53	50/98	**54**	51	16/95
n=4	**49**	40	30/100	**80**	78	**80**/100	67	**69**	22/95	**74**	70	28/95	63	**66**	24/100	**75**	70	65/100	**61**	57	16/98
n=5	**51**	**51**	36/100	**84**	82	80/100	**73**	71	61/98	**77**	74	30/100	**71**	68	24/100	80	**82**	68/100	**64**	62	16/100
n=6	**57**	**57**	36/100	**88**	**88**	80/100	**82**	76	61/99	**79**	**79**	44/100	**74**	68	26/100	**83**	82	72/100	**68**	66	20/100

That is, to use a logistic regression classifier to rank the previously identified themes in the list according to the likelihood that they are present in the text that is currently being analysed. Such a classifier could be re-trained, whenever the user creates a new text-theme association, and thereby be able to continuously adapt and improve the theme ranking when the user extends and refines the manual analysis.

We hope that such a tool, with the extended coding support investigated here, could be useful for assisting coding of recurring content in large text collections. For instance, for coding discussion forums for recurring arguments used in vaccination debates. The semi-automatic approach of such a tool, i.e., to extract texts typical to frequently occurring topics and then assist the user in the task of manually coding the extracted texts, does not guarantee that all important subjects discussed in the text collection are detected. However, the approach offers the possibility of accessing important content in text collections that are too large to make a fully manual coding feasible. The output of the semi-automatic coding could, for instance, be used for forming hypotheses on reasons why different types of health decisions are made, e.g., health decisions related to vaccination.

Conclusions

We have investigated the efficiency of ranking previously identified reasons according to the likelihood that they are present in a text that is being manually coded. The performance of a logistic regression classifier that ranked reasons according to the probability that they would be present in the text that is being coded was compared to a ranking based on the frequency of the reasons in the data set. When evaluating the approaches for their ability to include the expected reason among *n* top-ranked reasons, using an *n* between 1 and 6, we could conclude that the logistic regression-based ranking outperformed the frequency-based one. These results therefore encourage us to construct a system that provides the human text coder with a ranked list of suggestions for possible reasons that might be present in a text, and to implement this ranking by the means of a logistic regression-based classifier.

Acknowledgements

We would like to thank the Swedish Research Council (Vetenskapsrådet) that funded this study through the project "Navigating in streams of opinions: Extracting and visualising arguments in opinionated texts" (No. 2016-06681). We are also very grateful for the useful input provided by the reviewers.

References

[1] H.J. Larson, I. Ghinai, Lessons from polio eradication, *Nature* **473** (7348) (2011), 446–7.

[2] L. Grant, B.L. Hausman, M. Cashion, N. Lucchesi, K. Patel, J. Roberts, Vaccination persuasion online: A qualitative study of two provaccine and two vaccine-skeptical websites, *J Med Internet Res* **17**(5) (2015), 133.

[3] K. Faasse, C.J. Chatman, L.R. Martin, A comparison of language use in pro- and anti-vaccination comments in response to a high profile Facebook post, *Vaccine* **34**(47) (2016), 5808–5814.

[4] T.R. Tangherlini, V. Roychowdhury, B. Glenn, C.M. Crespi, R. Bandari, A. Wadia, M. Falahi, E. Ebrahimzadeh, R. Bastani, "Mommy Blogs" and the Vaccination Exemption Narrative: Results From A Machine-Learning Approach for Story Aggregation on Parenting Social Media Sites, *JMIR Public Health Surveill* **2**(2) (2016), 166.

[5] D. Surian, D.Q. Nguyen, G. Kennedy, M. Johnson, E. Coiera, A.G. Dunn, Characterizing Twitter Discussions About HPV Vaccines Using Topic Modeling and Community Detection, *J Med Internet Res* **18**(8) (2016), 232.

[6] E.P.S. Baumer, D. Mimno, S. Guha, E. Quan, G.K. Gay, Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence?, *Journal of the Association for Information Science and Technology* **68**(6) (2017), 1397–1410.

[7] M. Skeppstedt, A. Kerren, M. Stede, *Vaccine Hesitancy in Discussion Forums: Computer-Assisted Argument Mining with Topic Models*. In Proceedings of Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth, Studies in Health Technology and Informatics, IOS Press, (2018), 366–370.

[8] F. Boltužic and J. Šnajder, *Back up your Stance: Recognizing Arguments in Online Discussions*. In Proceedings of the First Workshop on Argumentation Mining, Association for Computational Linguistics, Stroudsburg, PA, USA, (2014), 49–58.

[9] F. Boltužic, J. Šnajder, *Identifying Prominent Arguments in Online Debates Using Semantic Textual Similarity*. In Proceedings of the 2nd Workshop on Argumentation Mining, Association for Computational Linguistics, Stroudsburg, PA, USA, (2015), 110–115.

[10] K.S. Hasan, V. Ng, *Why are You Taking this Stance? Identifying and Classifying Reasons in Ideological Debates*. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Stroudsburg, PA, USA, (2014), 751–762.

[11] P. Sobhani, D. Inkpen, S. Matwin, *From Argumentation Mining to Stance Classification*. In Proceedings of the 2nd Workshop on Argumentation Mining, Association for Computational Linguistics, Stroudsburg, PA, USA, (2015).

[12] G. Eysenbach, J.E. Till, Ethical issues in qualitative research on internet communities, *BMJ (Clinical research ed.)*, **323**(7321) (2001), 1103-5.

[13] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *A density-based algorithm for discovering clusters in large spatial databases with noise*. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, AAAI Press, Palo Alto, California, USA, (1996), 226–231.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research* **12** (2011), 2825–2830.

[15] R. Rehurek, P. Sojka, *Software Framework for Topic Modelling with Large Corpora*. In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, European Language Resources Association (ELRA), Paris, France, (2010), 45–50.

[16] M. Skeppstedt, K. Kucher, M. Stede, A. Kerren, *Topics2Themes: Computer-Assisted Argument Extraction by Visual Analysis of Important Topics*. In Proceedings of the LREC Workshop on Visualization as Added Value in the Development, Use and Evaluation of Language Resources, European Language Resources Association (ELRA), Paris, France, (2018), 9–16.

Address for correspondence

Maria Skeppstedt: maria.skeppstedt@lnu.se.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190242

Improving the Prescription Process Information Support with Structured Medical Prospectuses Using Neural Networks

Oana Sorina Chirila[a], Ciprian Bogdan Chirila[b], Lăcrămioara Stoicu-Tivadar[a]

[a] *Department of Automation and Applied Informatics, University Politehnica Timişoara, Timişoara, Timiş, Romania,*
[b] *Department of Computers and Information Technology, University Politehnica Timişoara, Timişoara, Timiş, Romania*

Abstract

To provide the best treatment, a physician needs information about both the patient and the medicines matching the patient status and improving it. In this article, we present three methods for structuring the sections of medical prospectuses using neural networks. To structure the information from a medical prospectus we use 3 web sources with structured data from sections (with names sections from prospectuses and with uniformized names of sections) to train as input for neural networks. The tests were conducted on Romanian prospectuses. After running the three algorithms, the prospectuses were compared in terms of accuracy and execution time for each source. It was concluded that the accuracy is higher in convolutional networks and in the case of uniform name sections. The output data is used in applications with decision support for the treatment, matching best treatment with the patient's status.

Keywords:

Prescriptions, neural networks (computer), drug information services

Introduction

Prescribing medicines for certain illnesses in as correct manner as possible is a challenge for all doctors and healthcare providers worldwide. The number of illnesses and medicines is higher and higher, and new treatments for new diseases are rare. The lack of sufficient information on treatments and the lack of uniformity of existing data on medication, as well as the lack of tools to compare and verify interactions between patient medication, creates a vulnerable situation. Prior to providing a specific treatment, a physician needs data such as: medical history, diagnosis, and complete information about the appropriate medication for the diagnosis.

The drug prospectus contain very important information for the doctor. For example, in a Nurofen 200 mg prospect, there are several sections, such as "What is Nurofen 200 mg and What is used for," "Warnings and precautions," "What should you avoid when you are taking this medicine?", "Nurofen 200 mg with other medicines", "Pregnancy, breast-feeding and fertility", "How to take Nurofen 200 mg", "Possible side effects", "Containing 200 mg nurofen". These data are useful for the physician to not prescribe wrong medication that may interact with another drug / disease / allergy / condition of the patient. For example, in the prospectus of Nurofen 200 mg, it is emphasized that a patient with gastro-duodenal ulcer cannot take this medicine. The same data is found in any medical prospect, but each manufacturer names the sections differently or locates them in different order. Any doctor has access to online prospectuses, but they need extended knowledge about the medication and for this knowledge they have to read the full prospectus for the right information each time they consult the information about the drug, especially for new ones. This means a lot of invested time for this process.

Information and medical data are both structured and unstructured. Most structured information is available in English. Various databases with structured information have been created and are easily included in medical applications or used by physicians to provide effective treatment for patients. Many researchers have begun to develop algorithms or to use new technologies to create the largest possible structured information. Structured information is easier to read and requires less time to be found. The researchers also seek to extract important data from various documents to use them in specific areas or to build for physicians structured databases with the extracted information. .

In this age of technology, we are confronted with a great deal of medical information coming from different sources. For physicians to have access to this ocean of information, structuring and compaction of data is needed. Important data can come from our daily activities, from the internet or even from clinical staff. Figure 1 shows the diversity of sources from which important clinical information can be obtained to be used later to treat and improve the medical condition of a patient. The goal is to integrate and structure information from these sources as to make it more easily available to medical units [1].

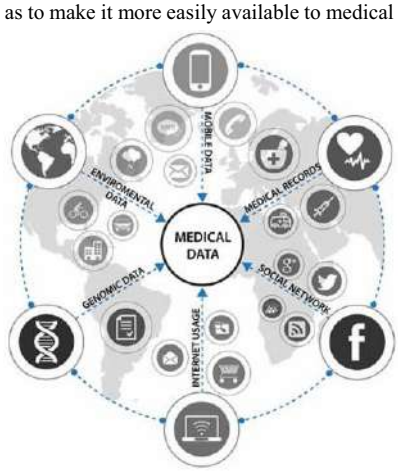


Figure 1– Sources of Medical Data [1]

A lot of work has been done on natural language processing (NLP) in English. This also applies to the processing of medical texts where the most robust data structuring and processing is for English texts. Deléger, Grouin, and Zweigenbaum [2] describe the implementation of a system for extracting

medicines and adjacent information from French texts based on an originally algorithm used for English medical text. The system is based on special dictionaries of medical terms and extraction rules. The texts resulted from 17,412 French electronic health records (EHRs) from the Cardiology Unit of the French University Hospital registered between 2004 and 2006. This dataset was divided into two data bodies: a corpus of development of 17,362 documents used to implement the system and a test body of 50 documents. The test body contains 253 drugs plus associated information items. The extraction rules were designed using English rules and examples of the development corpus. The dosing, form, frequency, duration and reason for which the drug was administered were extracted from the documents. The evaluation of the algorithm on the 50 documents obtained an F-measure score of 86.7%.

In Xu and Wang [3], a simple and accurate learning algorithm was developed to extract drug-disease pairs from 20 million biomedical summaries available on MEDLINE. The authors' process for extracting these data consists of the following steps:

1. Obtaining and analyzing the MEDLINE corpus
2. Creating lexicons for diseases and medicines
3. Correlation of MEDLINE sentences with disease and drug entities
4. Find patterns of treatments
5. Extract extra pairs from MEDLINE using selected patterns
6. Perform a semantic analysis of extracted drug-disease pairs.

The authors have used around 100 million sentences extracted from MEDLINE abstracts published between 1965 and 2010. A total of 34,305 pairs of single-disease treatments were extracted, most of which were not included in the existing structured databases. The algorithm of the authors reached a precision of 0.904 in extracting all pairs.

Extracting information about medicinal products from clinical texts is very important for EHR research. Jiang *et al.* [4] present the implementation of Java MedEX, an existing Unstructured Information Management Architecture (UIMA) data extraction system. In addition to showing earlier developments, the authors included new coding modules in the MedEx-UIMA system that map the data of drugs extracted with the RxNorm concepts. The MedEx-UIMA system consists of two main components: the clinical text extraction module and a standardization module encoding RxNorm (RxNorm concept unique identifiers) and normalizes TIMEX3 [5] frequency information. The information extraction module is a Java implementation of an older MedEx version then implemented in Python. The authors processed 826 documents with both systems (MedEx-UIMA and MedEx-Python) reaching similar results with both systems. Using the 300 annotated drug entries manually, the MedEx-UIMA system obtained the F-measure score for accuracy of 98.5%.

Casillas *et al.* [6] present a model for extracting allergic drug reactions from medical electronic records. The purpose of the paper is to extract this data from medical texts into Spanish. The authors developed two methods: a rule-based and a machine-based method. Both methods include semantic knowledge derived from FreeLing-Med, a software explicitly developed for medical text analysis. The corpus of text used in this research is composed of EHR documents written by doctors from Galdakao-Usansolo Hospital. First, a simple approach has been attempted to co-ordinate the terms of the EHR with SNOMED CT terms for allergy detection. This approach is effective but has not proven appropriate for medical texts written by physicians in which the type of allergy writing is different from the standard. Follow a rule-based algorithm has better results (70% accuracy) than the first simple approach (30% accuracy). The best approach with great results was the machine learning-Inferred Classifier method that gave a precision score of 88%.

Despotou *et al.* [7] uses NLP techniques to interpret UK NICE BNF drug recommendations that are provided as free text. The NLP component, MetaMap, identifies and interprets the semantic meaning of concepts in medical texts. NICE BNF provides structured definitions for drug-related issues but uses raw text for instructions on indications for drugs such as "500 mg three times a day" or "300-900 mg every 4-6 hours; maximum 4g per day." The authors analyze the dosing instructions and using NLP techniques identify types of semantic expressions and investigate how they can be generalized and used in specific rules to be applied in medical IT systems that contribute to e-prescribing.

Over the last few years, as presented in this section, a lot of applications were developed to support structuring and embedding information in medical applications. Databases have been created with a wealth of drug-related information, but most research has been made in and for English-speaking countries.

In this study, we use neural networks and the associated algorithms to structure Romanian medical prospectuses in order to extract the information from each section for further use in medical applications. We use the algorithms to train the neural networks since there are no other alternatives that we can use for Romanian text to extract certain sections from the prospectuses. This kind of processing works in cases where translate coding or databases is missing. In the following we present the methods, the results, and conclusions of the study.

Methods

Deep learning methods or neural networks have recently influenced many areas, including the processing of natural language. These methods are constantly improved with algorithms and increased performance compared to what exists in each field. A number of tools have been developed to enable the implementation of deep learning such as: Caffe, DeepLeaning4J, Eblearn, Keras, Neon, Scikit-learn, TensorFlow, Theano, etc. These tools attempt to optimize different aspects in learning or developing deep learning algorithms. These deep learning software tools have begun to receive a great deal of attention from the research community and are being increasingly developed to allow the formation of deep networks with thousands of parameters. Developers are trying to continually improve these tools to attract as many users as possible and to promote research [8].

In the current study we use medical prospectuses in Romanian language extracted from three online sources and structured on sections by extraction algorithms that we propose for this action [9].

In this study we use a method of training the neuronal networks with the names of selected already structured sections through various neural network algorithms and test them by providing other texts to which the trained network predicts the correct section name. The purpose of categorizing the text is to categorize the texts into categories that are easier to access. Each text can be categorized into one or more categories. We use neural networks to learn the classifiers in the examples and automatically categorize other documents in the same categories. In order to obtain this structure, we use three neural network models: Vector Support Machine Classifier from Scikit-learn library, Naïve Bayes Classifier from Scikit-learn

library, and 1D Sequence-Model Convolution Networks with sequential model from the Keras library. Figure 2 presents the workflow obtaining this structure.

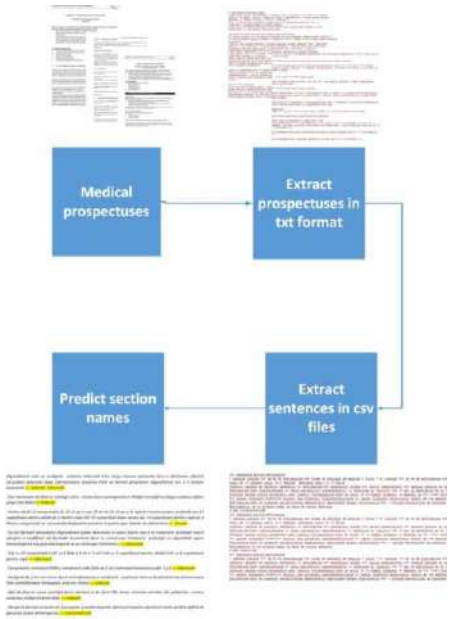


Figure 2 – Workflow for Structuring Prospectuses

Support Vector Machine (SVM) is a supervised classification algorithm that is extensively and successfully used for text classification task. An SVM represents the examples as points mapped in space so that the examples of the separated categories are divided by a clear gap that is as broad as possible [10].

Naive Bayesian is a simple and efficient classifier to implement NLP because it supposes that all the words of the documents are independent one of each other. The Naïve Bayes Classifier is the simplest probabilistic classifier used to classify text documents. The Naïve Bayes method is a module classifier based on probability and the probability of known conditionality. The basic idea is to use common words and category probabilities to estimate the class of a particular document [11].

Convolutional Neural Network (CNN) is a class of artificial neural networks that use convolutive layers to filter inputs for useful information. Convolution involves combining input data (a feature map) with a convolution kernel (filter) to form a map of the transformed features. Convolutional filters are modified based on the learned parameters to extract the most useful information. Text convolution is used for convolutional 1D networks [12].

To learn these three types of neural network models we use the structured data from the three medical prospecting websites: HelpNet, Pharmacists' Page and CSID. Table 1 presents the figures found in each data file.

After creating the files for each source as well as for mixed sources, we ran the three previously specified algorithms one by one to train the network section names for each text and then predict the section names for other text files. For training and testing, we use combinations between the above-mentioned sources. In order to run the algorithms, we use a computer with Intel Core i5-6400 processor of 2.70 GHz and 8 GB RAM and the processing time was quite good. Each run calculates the accuracy of the results and the running time. The outcomes of the process are presented in the Results section.

Table 1 – Number of Data Used in Neural Network Structuring

Name of the source	Number of prospectuses	Number of sections
CSID	3814	44834
Help Net	2820	18336
Pharmacists' Page	1513	22851

For all the algorithms, the first step consists of creating the learning and testing files containing the information of interest, specifically the label and the text associated with that tag. The learning files contain 70% of this information, and the test files 30%. For each solution we use specific items, as follows. The SVM (Support Vector Machine) algorithm uses the LinearSVC (Linear Support Vector Classification) classifier to train the network, and the prediction function is called on the test file. CNN uses the Sequential model in the Keras library to train the network, using the 'Dense layer' and calling the 'relu' activation function. We use 3 iterations (epochs) per learning file. Since we use textual classification, we only need a layer of convolution, so we use 1D Convolution. In the Naive Bayes algorithm, the text is transformed firstly as an array of elements, followed by a normalized expression of frequent terms and calling the NB multinomial classifier for the data previously obtained.

Each classifier is evaluated by calculating the accuracy and F1 score.

One of the main issues we had to deal with was the unevenness of the section names for the same type of section, different names for each drug may appear. For example, in the warning section, the following names may be found: caution, special warnings, warnings, warnings and precautions, precautions for use, warnings and special precautions, precautions, warning. To overcome this problem, we began aligning these sections creating files with similar section names and reference names. In [13] we propose a method of refining prospectuses sections to make them uniform.

To refine section names, we identified the most important section names: 'Indications', 'Contraindications', 'Dosage', 'Pregnancy and Breastfeeding', 'Active Substance', 'Composition', and any other section in the 'Information' section. We restored prospectus sources using just these section names and we ran the three algorithms again for all sources and their combinations. The outcomes of the process will be presented in the Results section.

Following the SVM algorithm running on data from the three sources, accuracy scores range from 1.99% in different training and testing sources and up to 56.78% on the same data source both for training and for testing. Execution time depends on the size of the text and the number of section names in the training source so that in the sources where the data from the CSID site is used, the running time is longer, the section names being varied and the prospectuses being more.

When running the Naive Bayes algorithm on the same data, the running time has dropped a lot, and accuracy is higher for all tests. For the algorithm, when referring to data with many different section names (CSID), we train the network only for fewer sections, choosing the most common ones. In this case, the results were more accurate, between 45.41% and 68.44%.

When running the convolutional network, the accuracy increased even more, but running time increased significantly as compared with the previous methods. The best accuracy was

recorded for training and test data for the CSID source (77.47%). With this method we had to lower more the number of training sections for an optimal algorithm operation. When we use CSID source to train the neural network we select only the most common 350 names from sections since the array that is created in the algorithm is very large and the algorithm cannot be run on a usual computer.

The percentage of accuracy is lower when providing new sources of information different from those with which the network was trained. Section names are very different from one source to another. Thus, in the CSIDs there are approximately 11,300 unique names for different sections, in the Pharmacists page there are around 450 unique names of sections, and in Help Net there are around 350 unique names of sections.

After the sections were reduced and uniformized, the accuracy increased greatly after running the SVM algorithm. The algorithm run times have also decreased, obtaining an accuracy of 38.26% for two different sources and an accuracy of up to 93.08% for training and testing data from the same source.

After running the Naive Bayes algorithm for data sources with uniform sections the accuracy increased significantly, and the running time of the algorithm decreased. The accuracy in this case ranges between 71.86% and 85.05%, both of which are this time for sections of the same data source.

Following the run of the convolution algorithm, the accuracy is highest, but the times stay high. Accuracy ranges between 77.62% and 91.31%. As can be seen, the accuracy of all source combinations is over 77%, which means that the correct prediction rate of the sections is the best.

As can be seen in the results presented with data uniformity, the accuracy result increased. Training the algorithms with the most complex and complete databases, the accuracy will increase. For what we need in medical applications, namely as accurate as possible, the best algorithm tested in this paper for accuracy in both non-uniform and uniform sections is the convoluted neural network. Due to the fact that it is working with multiple layers to drive information, the accuracy is the highest.

Discussion

The first algorithm that was run was the Support Vector Machine on the three sources and their combinations. The first run was on data with non-homogeneous sections, and the second run was on data with homogeneous sections. Figure 3 shows a comparative diagram of the two results in terms of accuracy.

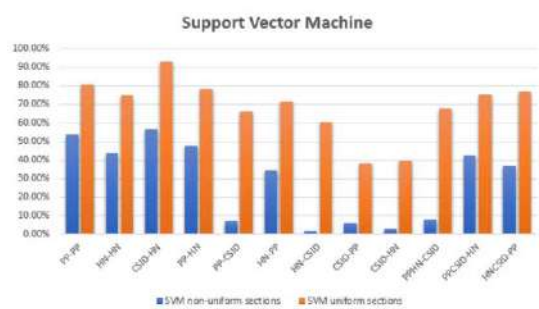


Figure 3 – SVM Accuracy

The first comparison is between the type of source used for the SVM algorithm, non-uniform sources in the blue version on the graphs, and sources with uniformized sections in the orange version on the graphs. As can be seen, the accuracy in the second case has increased significantly, especially if different data sources were used for training and testing and the running time decreased due to the number of low sections. For this algorithm, a relatively higher accuracy was obtained only for uniform sections.

The average accuracy for SVM on non-uniform sections is 28.52%, and for SVM on uniform sections the average increased to 68.58%, with an increase of 40.06%.

The median of accuracy for SVM on non-uniform sections is 35.65%, and for SVM on median uniform sections is 73.20%.

The second comparison was made between data used, with a different number of sections for the Naive Bayes algorithm. Figure 4 shows a comparative diagram of the two results in terms of accuracy.

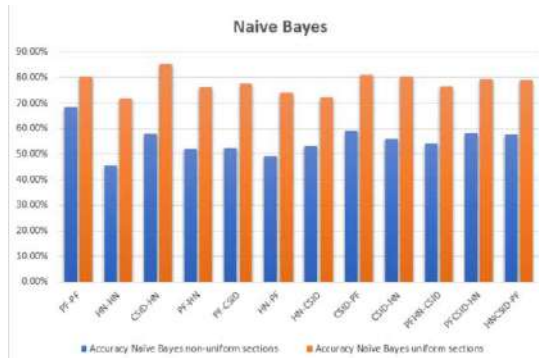


Figure 4 – Naïve Bayes Accuracy

In this case, the accuracy between the two types of sources used differed less. And in this case, the accuracy was bigger and more constant, as can be seen in the graph for the uniform sections. The average accuracy for Naive Bayes on non-uniform sections is 55.28%, and for Naive Bayes average uniform sections increased to 77.76%, with an increase of 22.48%.

The median of accuracy for Naive Bayes on non-uniform sections is 54.93%, and for Naive Bayes on uniform sections is 78.17%.

The last comparison between the accuracy of the different three algorithms on data sources was done for the convolutional network algorithm. Figure 5 presents a comparative diagram of the two results in terms of accuracy.

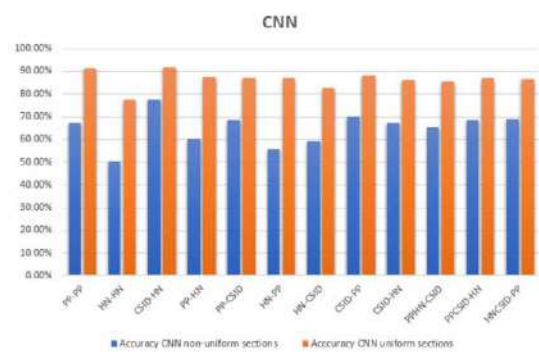


Figure 5 – CNN Accuracy

In this case, the accuracy was also good for non-uniform sections, but increased over 85% in most cases to uniform sections. The average of CNN accuracy for non-uniform

sections is 64.87%, and for CNN on uniform sections the average grew to 86.55%, with an increase of 21.68%.

The median of accuracy for CNN on non-uniform sections is 67.10%, and for CNN on uniform sections is 86.97%.

As can be seen, the best accuracy was in the case of convolutional neural networks and on uniformized sections. In Figure 6, we can see comparatively the accuracy of the algorithms on the non-uniform databases, and in Figure 7 we can see comparatively the accuracy of the algorithms on the uniform databases.

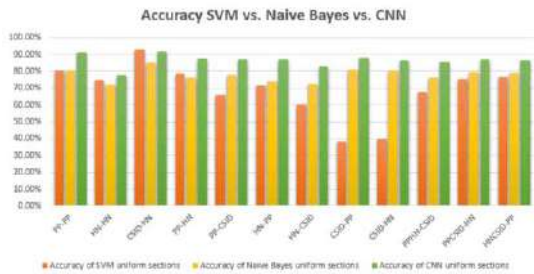


Figure 6 – Comparative Accuracy on Non-Uniform Sections

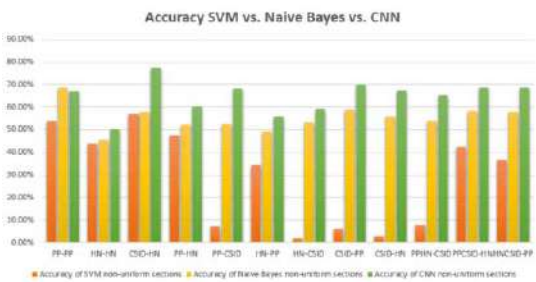


Figure 7 – Comparative Accuracy on Uniform Sections

Conclusions

Structuring medical information is very important to help doctors find the best treatments for patients and to create systems with alerts when drugs can interact with certain diseases / allergies or other medicines that the patient already has. In this study, we have chosen to structure the medical prospectuses in Romanian because in this language there are no databases with full structured information about the drugs. The extracted information can then be used to create databases that are added to the prescription modules. In the future, we will test algorithms on medical prospectuses from other languages in order to create a general model for structuring medical prospectuses. Future work also focuses on more solutions to be multilingual and adapt easy to situations in which there are no structured databases or translated codes with drug information.

The results are encouraging and the solution is easy to implement as a module in healthcare information systems.

References

[1] M. Ghassemi, L. A. Celi and D. J. Stone, State of the art review: the data revolution in critical care, *Critical Care (2015) 19:118*

[2] L. Deléger, C. Grouin, P. Zweigenbaum, Extracting Medication Information from French Clinical Texts, *Studies in Health Technology and Informatic*s, MEDINFO 2010, 160:949-953

[3] R. Xu and Q. Wang, Large-scale extraction of accurate drug-disease treatment pairs from biomedical literature for drug repurposing, *BMC Bioinformatics*, 2013:14:181

[4] M. Jiang, Y. Wu, A. Shah, P. Priyanka, J. C. Denny, H. Xu, Extracting and standardizing medication information in clinical text – the MedEx-UIMA system, *AMIA Jt Summits Transl Sci Proc. 2014* :37-42

[5] J. Pustejovsky, J. Castaño, R. Ingria, et al. TimeML: Robust Specification of Event and Temporal Expressions in Text, *Fifth International Workshop on Computational Semantics*. 2003

[6] A. Casillas, K. Gojenola, A. Perez and M. Oronoz, Clinical text mining for efficient extraction of drug-allergy reactions, *IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2016*

[7] G. Despotou, I. Korkontzelos, N. Matragkas, E. Bilicia and T. N. Arvanitis, Structuring Clinical Decision Support Rules for Drug Safety Using Natural Language Processing, *Data, Informatics and Technology: An Inspiration for Improved Healthcare, IOS Press*, 2018

[8] S. Bahrampour, N. Ramakrishnan, L. Schott, M. Shah, Comparative Study of Deep Learning Software Frameworks, *Cornell University Library*, Computer Science, 2016

[9] O. S. Lupşe and L. Stoicu-Tivadar, Extracting and Structuring Drug Information to Improve e-Prescription and Streamline Medical Treatment, *Applied Medical Informatics, Vol. 40, No. 1-2* /2018: 7-14

[10] Scikit-learn, http://scikit-learn.org, accessed in July 2018

[11] R. M. Chezian, C. Kanakalakshmi, Performance Evaluation of Machine Learning Techniques for Text Classification, *Proceedings of the UGC Sponsored National Conference on Advanced Networking and Applications*, 27th March 2015

[12] N. Ackermann, Introduction to 1D Convolutional Neural Networks in Keras for Time Sequences, Good Audience, https://blog.goodaudience.com, accessed in August 2018

[13] O. S. Lupşe and L. Stoicu-Tivadar, Supporting Prescriptions with Synonym Matching of Section Names in Prospectuses, *Stud Health Technol Inform*. 2018;251:153-156

Address for correspondence

Chirila Oana Sorina, email: oana.chirila@aut.upt.ro

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190243

Modeling Multi-View Dependence in Bayesian Networks for Alzheimer's Disease Detection

Parvathy Sudhir Pillai[a], Tze-Yun Leong[a], the Alzheimer's Disease Neuroimaging Initiative

[a] School of Computing, National University of Singapore, Singapore

Abstract

Early detection of Alzheimer's disease is important for deploying interventions to prevent or slow disease progression. We propose a multi-view dependence modeling framework that integrates multiple data sources to distinguish patients at different stages of the disease. We design interpretable models that can handle heterogeneous data types including neuro-images, bio- and clinical markers, and historical and genotypical characteristics of the subjects. We learn the dependence structure from data with guidance from domain knowledge in Bayesian Networks, visualizing and quantifying the conditional probabilistic dependence among the variables. Our results indicate that the hybrid dependence models also improve prediction performance.

Keywords:

Alzheimer Disease, Bayesian networks, Classification

Introduction

Alzheimer's Disease (AD) is a neurodegenerative disorder that leads to cognitive decline of the elderly and renders them incapable of performing routine activities of daily living. The neuronal degeneration is often irreversible [1]; common clinical symptoms such as memory loss or speech impairment may only appear at a later stage. It is important to identify the relevant factors that may individually or collectively impact the cognitive condition or state to assist in accurate and early diagnosis. Such factors may be categorized based on demographics, physical examinations, clinical tests, cognitive assessments, etc. Each category presenting information about an individual froms a distinct *view* or perspective. We propose a framework for detecting AD at different stages by fusing multiple views of patient-related data and explicating the dependence among the factors or variables.

In practice, AD severity is usually assessed by psychometric tests such as Mini Mental State Examination (MMSE) and Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog). Disease markers acquired from neuroimaging and protein studies are considered better indicators of AD in the early stages of Mild Cognitive Impairment (MCI) [1]. Genotype and non-behavioral background factors from demographics, family and medical history have been associated with cognitive decline. Linking potential diagnoses to background factors helps to identify the individuals at risk of developing AD and indicates ways to plan for its prevention and intervention.

We aim to build a disease model that accommodates the *correlational*, *causal* and *complementary* semantics of dependence and uncertainty from heterogeneous, multi-view data, possibly from different modalities. Clinical data is heterogeneous and variables relevant to AD coud be discrete (e.g., medical history of diabetes is binary) or continuous (e.g., individual's age). Knowledge discovery from medical data can be broadly categorized into correlation-based and causality-based. Correlation identifies how close two variables are to having a linear relationship with each other and indicates a predictive relationship. Most multi-view disease models identify correlations of bio- and clinical markers for predicting the clinical status. On the other hand, a variable A causally influences variable B if we manipulate A to different values, measure the effects on B, and observe changes in the probability distribution of B under different values of A. Incorporating the causal contributions from background factors in the model helps us understand their interactions that manifest as the individual's cognitive state. The cognitive state cannot be causally established from the background factors in a fully data-driven approach using only observational data [4]. Knowledge about disease epidemiology can help identify beneficial or risky causal factors. Complementarity separates the unique knowledge in a view, leading to better predictions through combining different views [2].

We propose a hierarchical probabilistic graphical model that incorporates two types of views: *markers* including MMSE, neuro-images that measure the cognitive state, etc., and *background* including genotypic, demographic variables that could possibly impact the cognitive state. The variables in each high-dimensional view are linear transformations of the underlying continuous, low-dimensional latent traits. The dependence between the predictor variables and the cognitive state is depicted in Bayesian Networks (BNs). We learn relevant evidence-based relations among the variables using prior knowledge of domain rules elicited from experts. We examine the key classification metrics: accuracy, precision and recall to determine the disease stage of a set of individuals from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [5].

Related Work

Multiple linear/logistic regressions are the most popular correlation-based approaches used by clinicians and epidemiologists. Classifiers such as naive Bayes (NB), decision tree, back-propagation neural network (NN), and support vector machine (SVM) [6] have also been used for disease prediction from clinical data. With the recent progress in multi-view machine learning, combinations of markers distinguish AD patients from cognitively normal (CN) controls with high accuracy. Multiple kernel learning [2], canonical correlation [15] and shared subspace learning [3] are applications of multi-view learning that target the correlations between views of bio-

or clinical markers. These methods treat all views uniformly as results of the clinical state.

To identify risky/beneficial factors from background views, it is necessary to understand how changes in the variables affect the clinical state. Previous studies that explore associations between background factors and diseases are mostly hypothesis-driven; the validity of a hypothesis is tested with the available data. Structural equation models and dynamic causal networks [8] operate on causal assumptions given by domain experts about the disease. However, since the data may be from different sources, it is hard to include all the assumptions that may be valid in one source but not in another. Also, those relationships that are not included a priori are possibly missed.

Jin et al. [9] analyzed 16 multimodal features from ADNI, including age, sex, education, hippocampal volume, 2 average Positron Emission Tomography intensity measures, 7 Single Nucleotide Polymorphisms (SNP), MMSE and ADAS-Cog for cognitive score prediction in a BN. They included the notion of causality in determining the conditional probabilsitic dependence. However, feature or variable selection was done manually and solely based on domain knowledge, without considering other potentially important markers from the corresponding modalities. Hence, they do not address correlations among the variables within a modality.

Methods

A Bayesian Network (BN) is a probabilistic model that consists of two parts: a graphical model visualizing the dependencies among the variables and a probability model quantifying the dependencies. Learning the structure of a BN is done either purely from domain knowledge or the dependencies learnt from the data or a combination of both.

We learn probabilistic dependence models from multi-view clinical data using a hybrid approach i.e., score-based BN structure learning [9] guided by inputs from experts about the direction of dependencies. We adopt the prior knowledge of a causal approach in which genetic variables like SNPs and demographic variables such as age and sex are fixed before other variables and are not influenced by them. We assume that i) the markers assume certain values as manifestations of the clinical status and ii) non-behavioral background variables may influence the occurrence of AD and not vice-versa. We use a multi-level BN [10] to simulate this hypothetical data generation hierarchy: background traits ⇒ cognitive state ⇒ marker measures. We summarize the procedure to learn the dependency model, i.e., the BN of multi-view data in Figure 1.

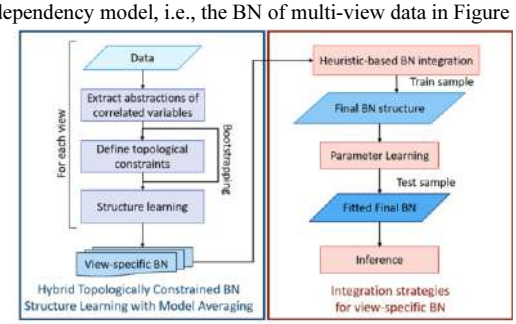


Figure 1. Modeling multi-view dependence via BN integration

We consider a supervised setting, where $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(M_{mark})}$ represent the multi-view markers, $\mathbf{B}^{(1)} \ldots \mathbf{B}^{(M_{bg})}$, the background views and $\mathbf{Y}$ the response variable, i.e., clinical status of AD, the intermediate stage of Mild Cognitive Impairment (MCI) or normal controls (NC). $\boldsymbol{M_{mark}}$ is the number of marker views, while $\boldsymbol{M_{bg}}$ is the number of background views and $\boldsymbol{N}$ the number of subjects. We divide the variables in the dataset into mutually exclusive views based on their modality and the features they describe. The target variable, i.e., the clinical status of the subject is included in each view; it connects the background variables and the disease markers during view integration.

Extracting abstractions of correlated variables

In using multi-view patient data, we face the 'curse of dimensionality', which results in complex models that overfit and are not very "interpretable". There is also the issue of multi-collinearity among the variables within a view. We overcome these by abstracting out latent factors/traits from the variables which simultaneously explain their correlations and achieve dimensionality reduction. The latent factors are continuous-valued, follow a normal distribution with zero mean and unit covariance and, emulate a continuum: low to high, sick to healthy, etc. We use Bayesian matrix factorization (BMF) [14] to extract latent factors from continuous-valued views. We represent markers, $\mathbf{X}^{(j)}$, and background variables, $\mathbf{B}^{(k)}$ as linear transformations of uncorrelated low-dimensional latent factors, $\mathbf{Z}^{(j)}$ and $\mathbf{U}^{(k)}$ with $\boldsymbol{l}^{(j)}$ and $\boldsymbol{l}^{(k)}$ numbers of latent factors respectively, as depicted in Equation 1.

$$\boldsymbol{X}_{pq}{}^{(j)} = \sum_{r=1}^{l^{(j)}} \boldsymbol{Z}_{pr}^{(j)} \boldsymbol{V}_{rq}{}^{(j)} + \boldsymbol{\varepsilon}^{(j)} \qquad (1)$$

$$\mathbf{Z}_{\boldsymbol{p}}^{(\mathrm{j})} \sim \boldsymbol{\mathcal{N}}_{\boldsymbol{l}^{(j)}}(\mathbf{0}, \mathbf{I}_{\mathbf{l}^{(\mathrm{j})}}), \mathbf{V}_{\boldsymbol{q}}^{(j)} \sim \boldsymbol{\mathcal{N}}(\mathbf{0}, \boldsymbol{\lambda}_{\mathbf{q}}^{(\mathrm{j})^2}), \boldsymbol{\varepsilon}^{(j)} \sim \boldsymbol{\mathcal{N}}(\mathbf{0}, \boldsymbol{\sigma}_{\mathbf{q}}^{(\mathrm{j})^2})$$

where $\mathbf{V}^{(\mathrm{j})}$ is the weight matrix and $\boldsymbol{\varepsilon}^{(j)}$ is the noise. Further, we specify a Gamma prior for the inverse variances, $\boldsymbol{\sigma}_{\mathbf{q}}^{(\mathrm{j})^{-2}}$ of the noise term, $\boldsymbol{\varepsilon}^{(j)}$ i.e., $\boldsymbol{\sigma}_{\mathbf{q}}^{(\mathrm{j})^{-2}} \sim$ Gamma $(\boldsymbol{a_\sigma}, \boldsymbol{b_\sigma})$, where $\boldsymbol{a_\sigma}$, $\boldsymbol{b_\sigma}$ are hyperparameters. The impact of each factor in $\mathbf{Z}^{(j)}$ on the actual variables in $\mathbf{X}^{(j)}$ can be derived by inspecting $\mathbf{V}^{(j)}$. The higher a variable weighs, higher is the corresponding factor's ability in capturing its correlations with other variables. Specifically, the variables with higher weights in a factor are interpreted as a cluster because they are similar due to their high correlations. To make the clusters interpretable, we need to reduce the number of variables with higher weights on many factors. Thus we apply sparsity constraints on the weights to have fewer non-zero weights and seggregate the strongest signals. We apply the constraint as a prior probability to the weight matrix. However, with the multiple heterogenenous views of data, our model needs to handle outliers and unknown sparsity structures of variables. For this we apply the horseshoe prior on the weight matrix, $\mathbf{V}^{(\mathrm{j})}$, with a shrinkage parameter, $\boldsymbol{\lambda}_{\mathbf{q}}^{(\mathrm{j})} \sim$ Half- Cauchy (0,1), that is robust at handling unknown sparsity and leaves out large outliers [12].

For categorical-valued views, we use multi-dimensional Item-Response Theory (IRT) models [16] to express the observed variables as resulting from continuous latent traits. For example, the probability of a person achieving a certain score in a test is a consequence of their ability, the test's difficulty and the discrimination power of a question. It is then possible to assume that individuals with the same ability answer a question correctly with a certain probability. Cognitive capability is an interpretation of this ability with respect to MMSE. IRT relates the latent ability, $\boldsymbol{\theta_i}$, to the probability of getting the correct answer to a question through the Item Characteristic Curve (ICC). Difficulty, $\boldsymbol{\delta_j}$, is practically the point on the ICC where the probability for answering correct is 50%, whereas, discrimination, $\boldsymbol{\alpha_j}$, is the slope of the ICC. An

ICC with a steep slope imparts sudden value jumps even for a small change in ability. This in turn implies that the particular question's discrimination power is very high. Meanwhile, if the item achieves a median (50%) probability at higher values of cognitive ability, the item is difficult. Equation 2 shows the probability of choosing the right option, $\boldsymbol{j}$, for a question, $\boldsymbol{i}$.

$$\boldsymbol{P(X_{ij} = 1|\theta_i,\ \alpha_j,\ \delta_j) = \frac{exp(\alpha_j(\theta_i - \delta_j))}{1+exp(\alpha_j(\theta_i - \delta_j))}} \quad (2)$$

where $\boldsymbol{\theta_i \sim \mathcal{N}(0, 1)}$. For ordinal-valued views, where there are more than two possible values of responses, we use the Partial Credit Model (PCM). The above equations apply for both marker and background views.

Learning the structure of dependence in a view

Directed edges in a BN represent the probabilistic dependence between variables, with the child node being conditionally dependent on the parent. Therefore, learning the structure of dependence among the latent factors from a view translates to learning the edges of the BN. However, as the number of variables in a view increases, an exhaustive search in the space of possible graph structures becomes infeasible. In clinical data, there my be many variables of interest in each view of different data types. We use the Conditional Gaussian BN (CGBN) framework [9] with multinomial and Gaussian distributions to model discrete and continuous nodes respectively. The discrete nodes are parameterized by conditional probability tables, with their values depending on their discrete parents. The mean of a continuous node is deduced as a linear regression of the continuous parents on each configuration of its discrete parents. Variance, however, depends only on the discrete parents. CGBN disallows discrete nodes with continuous parents. We incorporate the following techniques in the CGBN structure learning procedure to reduce the search space.

Incorporating topological constraints

Certain constraints which represent either scientific laws, common sense, expert opinions, accumulated personal experiences, etc., help to build better structures. These are usually qualitative constraints and do not signify the strength of dependence between variables. Following the convention in Li and Leong [3], we define the domain constraints in Table 1. The values of demographic (e.g., "Age", "Sex") and genetic variables (e.g., SNPs) are fixed before other variables and are not influenced by them; these should be roots in the BN. On the contrary, the values of bio- or clinical markers are dependent on other variables and do not affect other variables; these should be leaves in the network. Older age is known to be linked to AD and hence there is a directed edge from "Age" to the clinical status. We also restrict edges from clinical status and latent markers to latent background and, latent markers to status.

Table 1. Types of topological constraints in a BN

Constraints	Description
Roots	Cannot be children of other nodes
Leaves	Cannot be parents of other nodes
Known links	Links that domain experts know exist
Forbidden links	Links that domain experts know don't exist
Ordering	Chronological or logical ordering of variables

BN structure learning with search space reduction

We achieve further search space reduction by trying to build edges only between variables that are correlated. For each node the candidate neighbors with a definitive correlation, we create the dependency structure of a view using a search and score algorithm. We use a greedy hill climbing strategy that searches for a possible distribution of edges, while trying to maximize the Bayesian information criterion (BIC) score [10]. BIC measures the goodness of fit of the structure given the data, but also penalizes complicated structures with many parameters to learn. The search begins from an empty graph. At each step, the greedy algorithm performs an edge operation (adds, deletes or reverses a directed edge) that increases BIC maximally.

Model averaging

We use the bootstrap approach to assess the confidence of network structures learnt from a few hundreds of instances. Accuracy is a measure of confidence on the presence of an edge and its direction in the BN structure [9]. For a number of iterations, the algorithm re-samples the same number of instances, N as the dataset, D, with replacement. It further learns the network structure from this re-sampled dataset. The structures resulting from each iteration are averaged to help identify the nodes and edges which appear in at least half of them. We use only those edges which have edge strengths more than 0.5 and probability of edge direction more than 0.5. The edge strength signifies the confidence on the direct dependence relationship between two nodes and is estimated as its empirical frequency over the set of networks learned from bootstrap samples. The probability of the edge direction is computed conditional on the edge being present in the network.

Integration of multiple views

We learn the dependence model of the multi-view data by selectively combining nodes and edges from the view-specific BNs. In the second phase of structure learning, edges between latent variables from different views are established. From each view-specific BN, we segregate the nodes that form the Markov blanket (MB) of the clinical status node. The set of children, parents, and spouses of the node X is its MB [9]. In a BN, a variable is conditionally independent of all the other variables given its MB. Since the bootstrap approach iterates a number of times while resampling of the training data with replacement, we have as many candidate structures as there are iterations. If the edge between nodes X and Y is not present in any of the candidates, the edge strength of $E_{X,Y}$ is 0. If it appears in all the candidates, the strength is 1. Since the edge strength, Γ is a positive value between 0 and 1, it can be regarded as a probability of the validity of the edge in the BN, i.e., $\Gamma_{E_{X,Y}} = \frac{|E_{X,Y}|}{n_{iterations}}$. Thus, we follow a Bayesian approach with only the MB nodes from view-specific BNs and edges across views to be learned afresh with a default uniform prior and edges within views learned with a prior probability.

Parameter learning and inference

We learn the parameters of the integrated BN structure through maximum likelihood estimation with Laplace smoothing over the n_{train} training samples. We compute the predictions of the clinical status by averaging likelihood weighting simulations. The predicted value is the one with the highest conditional probability for a discrete target, and the highest expected value of the conditional distribution for a continuous target.

Experiments

Data: We work with a dataset of 589 subjects from ADNI which includes their background variables: demographic (4), genotypic (SNP~900) and medical history (25), and markers: grey matter volumes from baseline MRI (90), and cognitive measures: Mini Mental State Examination (MMSE) and

Clinical Status (AD/MCI/CN). Table 2 shows the demographics of the subjects included in our study.

Setup: We use the Stan probabilistic programming language to build view-specific abstractions and decide on the dimensions through multiple cross-validations [14]. We build and represent the BNs programmatically using the bnlearn R-package [9]. The integrated BN is used to predict the clinical status (2- AD, 1-MCI and 0-CN). For all the classification problems, we learn one BN structure per view with disease status as the target variable. This is followed by structure and parameter learning for the integrated BN for different view combinations. We report the performance of all models averaged over 5-fold cross validations. We evaluate the classification of a concatenated data vector comprising data from the three modalities, viewing it as a baseline study.

Table 2. Subject Demography

	AD (n = 128) 66 M, 62 F		**MCI (n = 287)** 179 M, 108 F		**CN (n = 174)** 98 M, 76 F	
	Mean	Range	Mean	Range	Mean	Range
Age	75.4 (±7.3)	58.4-87.7	75.3 (±7.2)	55.1-88.8	75.2 (±5.2)	62-84.7
MMSE	23.77 (±1.9)	20-26	27.12 (±1.6)	24-30	29.04 (±1.2)	25-30

Assumptions: We learn view-specific BNs using only the constraints that apply to the variables within. As the disease state is a categoric variable in classification, we discretize its possible parent variables such as "Age", "Years of education" according to the rules of CGBN. Only views with more than 5 variables are subject to the extraction of low-dimensional abstractions. We use a subset of SNPs (~900) as previously reported in [3]. Each SNP is matched to the gene/s located within a distance of 20kb. We learn a latent score for each gene per person using PCMs and use these for further processing.

Results

BMF captures the dense correlations among 90 regions of interest (RoI) captured into 25 latent factors. Figure 2 depicts the weights of the top 10 features (y axis) on the 25 factors, with longer bars indicating higher absolute weight in the corresponding color-coded factor. Left Amygdala has the longest bar, and hence the highest absolute weight. The latent factors are interpreted based on the hemisphere, lobe and area of the brain surface where the constituent RoI belongs. In Figure 3, the discrimination and difficulty parameters of the ICC for MMSE questions are elaborated in the left and right subplots. The five most discriminating questions are related to attention in computing alphabets in a series (MMO, MMR, MMW), language (MMWATCH), reading (MMREAD) and orientation (MMSTATE) while the five most difficult questions are alphabets in a series (MMD, MMO, MMR), orientation (MMONFLR) and immediate recall (MMTREE).

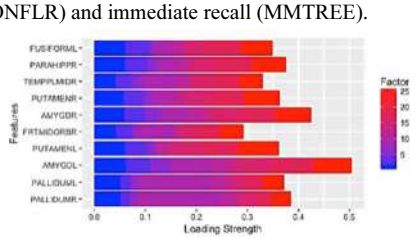


Figure 2. RoIs with highest absolute weights on factors

Table 3 shows the genes abstracted from their corresponding SNPs and factored brain RoIs that report the 5 highest associations with AD (p-values <= 0.05). Table 4 compares the performance of our approach for the multiclass classification of AD vs. MCI vs. CN with the baseline, single views, and the state-of-the-art classification for AD [7]. Figure 4 presents the integrated BN structure of demography and the abstractions from marker views of MRI, MMSE and background views of genes and medical history. The MRI factors are color-coded in navy, genes in yellow, clinical status (DX) in red, demographic variables in pink, medical history in ivory and MMSE in green.

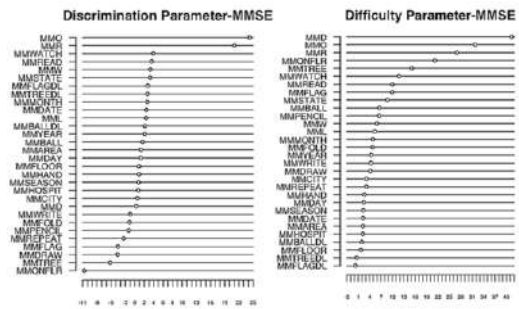


Figure 3. Discrimination and difficulty of MMSE (low-high)

Table 3. Genes, Brain areas identified by our model as associated with AD

Gene	p-value	Brain Area	p-value
APOE	<.001	Left Hippocampal	<.001
TOMM40	<.001	Right Hippocampal	<.001
SLC6A11	<.001	Left Amygdala	<.001
IL16	<.001	Right Amygdala	<.001
FGD6	.002	Left Insula Medial	<.001

Table 4. Key performance metrics for classifying AD/MCI/CN

Study/Data	**Accuracy%**	**Precision**	**Recall**
MMSE	71.7	0.58	0.91
Baseline	59.44	0.64	0.58
(Our work)	**80.18**(±8.89)	**0.73**(±.06)	0.68(±**.09**)
[7] Used MRI, PET	72.9	NA	NA

Discussion

Most clinicians use only total MMSE scores to assess AD staging. This reports an accuracy of 71.7%. Our method achieves an accuracy of 80.18% along with a precision of 0.73 and a recall of 0.68. The reported measures are averaged over predictions of AD, MCI and HC subjects by the model. The performance of the baseline (vertical concatenation of views) is not very impressive due to the inclusion of more features compared to training samples, i.e., *'large p small n'* that leads to model overfit. It underperforms grossly while working with 589 individuals and ~1030 features. The results in Table 4 show that our method fares better than the state-of-the-art approach to identify the stage of AD [7]. By categorizing the data views and defining edge directions, we are also able to visualize and interpret the model structure (refer to Figure 4). We notice that MRI factors are correctly modeled as interacting mostly through the disease variable and form a connected network. As we include only the variables from the Markov blanket (MB) of the disease variable, the less relevant ones are filtered out from each view due to the conditional independence property. The inclusion of the background features from genes, medical history and demography in the final BN in the MB of the clinical status variable (DX) implies that they are not

overshadowed by the markers. We also compare our results with a recent work [8] which used BN structure learning to predict the MMSE scores using a highly selective list of 16 biomarkers from various modalities. They report a mean square error of 2.81, while we achieve 2.44 ($\pm$0.29). Feature selection from each modality is done manually in their prediction model. Our approach also facilitates relevant biomarker discovery. The latent factors extracted from the marker and background views and visualization of correlated features lead to hitherto unexplored associations with the clinical status. Finally, in Figure 5, we report the impact of significant clinical features in the MB of the clinical status variable in the network we learnt. We examine the probabilities of a clinical status of AD or MCI conditioned on different values of the variables (scaled/standardized for continuous variables) in the MB of the clinical status. The results show that lower grey matter volume of the left hippocampus and insula region are indicative of a higher probability for cognitive impairment. The presence of APoE and TOMM40 SNPs strongly increase the chance of cognitive impairment and so does a lower education, MMSE score and higher age. These results are reasonable based on published evidence.

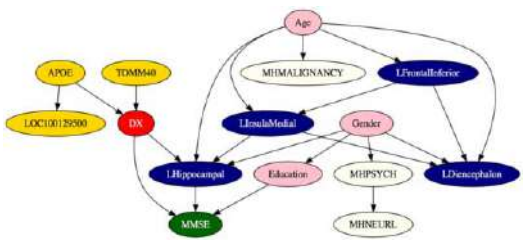
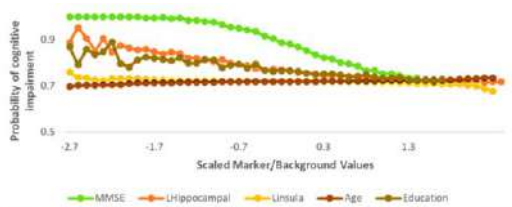


Figure 4. Multi-view integrated BN



Categories	APOE	TOMM40
0	0.62	0.42
1	0.76	0.69
2	0.92	0.72

Figure 5. Conditional probabilities for cognitive impairment

The CGBN framework, however, does have several restrictions on the nodes and their distributions. In particular, only direct conditional dependencies can be viewed. There could be hidden or non-linear associations that are not explicitly captured.

Conclusion

We have introduced a multi-view disease staging framework that takes into account the dependence semantics among the variables from disparate data sources. Our model achieves comparable and sometimes better prediction performance in identifying individuals at different stages of AD as compared with the state-of-the-art approaches. Moreover, our framework is scalable, takes into account the heterogeneity and the multitude of the data types, modalities, and quantities. It also serves as the basis of a general approach to clinical decision support with interpretable recommendations from multi-view data.

Acknowledgements

We would like to thank Dr. Feng Lei and his team at the Department of Psychological Medicine, NUS Yong Loo Lin School of Medicine, for providing the medical background and guidance. This project was supported by an Academic Research Grant No. MOE2016-T2-2-068. Data used in the preparation of this article was obtained from the ADNI database [5].

References

[1] Jack Jr CR, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. The Lancet Neurology. 2013 Feb 1;12(2):207-16.

[2] Zhang D, Wang Y, Zhou L, Yuan H, Shen D. Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage. 2011 Apr 1;55(3):856-67.

[3] Xu Z, Zhe S, Qi Y, Yu P. Association discovery and diagnosis of Alzheimer's Disease with Bayesian multiview learning. Journal of Artificial Intelligence Research. 2016 Jun 23;56:247-68.

[4] Li G, Leong TY. Biomedical knowledge discovery with topological constraints modeling in Bayesian networks: A preliminary report. In: Proceedings of the 12th World Congress on Health (Medical) Informatics. 2007 (p. 560). IOS Press.

[5] Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack C, Jagust W, et al. The Alzheimer's disease neuroimaging initiative. Neuroimaging Clinics. 2005 Nov 1;15(4):869-77.

[6] Williams JA, Weakley A, Cook DJ, Schmitter-Edgecombe M. Machine learning techniques for diagnostic differentiation of mild cognitive impairment and dementia. In: Workshops at the twenty-seventh AAAI conference on artificial intelligence 2013 Jun 29.

[7] Zhu X, Suk HI, Lee SW, Shen D. Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification. IEEE Transactions on Biomedical Engineering. 2016 Mar;63(3):607-18.

[8] Jin Y, Su Y, Zhou XH, Huang S. Heterogeneous multimodal biomarkers analysis for Alzheimer's disease via Bayesian network. EURASIP Journal on Bioinformatics and Systems Biology. 2016 Dec 1;2016(1):12.

[9] Scutari M. Learning Bayesian Networks with the bnlearn *R* Package, Journal of Statistical Software. (2010). 35(3):1-22.

[10] Pradhan M, Provan G, Middleton B, Henrion M. Knowledge engineering for large belief networks. In: Proceedings of the Tenth international conference on Uncertainty in artificial intelligence 1994 Jul 29 (pp. 484-490).

[11] Nakajima S, Sugiyama M. Theoretical analysis of Bayesian matrix factorization. Journal of Machine Learning Research. 2011;12(Sep):2583-648.

[12] Carvalho CM, Polson NG, Scott JG. Handling sparsity via the horseshoe. In: Artificial Intelligence and Statistics 2009 Apr 15 (pp. 73-80).

[13] Hambleton RK, Swaminathan H, Rogers HJ. Fundamentals of item response theory. Sage; 1991.

[14] Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A probabilistic programming language. Journal of statistical software. 2017 Jan 1;76(1):1-37.

[15] Zhu X, Suk HI, Lee SW, Shen D. Canonical feature selection for joint regression and multi-class identification in Alzheimer's disease diagnosis. Brain imaging and behavior. 2016 Sep 1;10(3):818-28.

Address for correspondence

Parvathy Sudhir Pillai (parvathy@comp.nus.edu.sg)

Medical Computing Laboratory, School of Computing, National University of Singapore, 13 Computing Drive, Singapore-117417.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190244

Cardiac Tissue Engineering as Use Case to Connect Biomedical Research Laboratories to an Emerging Global Data Infrastructure

Markus Suhr[a], Nadine Umbach[a,c], Tim Meyer[b,c], Wolfram-Hubertus Zimmermann[b,c], Ulrich Sax[a]

[a] *Department of Medical Informatics, University Medical Center Göttingen, Göttingen, Germany,*
[b] *Institute of Pharmacology and Toxicology, University Medical Center Göttingen, Göttingen, Germany*
[c] *DZHK (German Center for Cardiovascular Research), partner site Göttingen*

Abstract

Methods for cardiac tissue engineering and application in experiments are core technologies developed at the Institute of Pharmacology and Toxicology in Göttingen. As is the case in many academic research laboratories data capture and documentation may be improved to latest methods of digital research. A comprehensive information system infrastructure is the foundation of further advances toward automation of lab processes. A data management system concept is proposed and prototypically deployed that enables traceability of assets within the lab and reproducibility of published assays and results. The prototype integrates existing electronic lab notebook, experiment result database, and a newly introduced research data management system by means of a custom developed portal and integration component. The architecture concept and developed integration tools explore connection of routine experimental work in a biomedical research lab to a universal infrastructure of data.

Keywords:

Automation, Laboratory; Data Curation; User-Computer Interface;

Introduction

Biomedical research laboratories in industry and academic environments are challenged with an ever-increasing complexity and quantity of data to be handled. Engineered human tissue models are, for example, pushing into high-throughput drug screening [1], precision medicine [2], and clinical tissue replacement therapy applications [3]. Laboratory automation technology and data science methods are introduced into the biomedical lab domain. Complementary to this bottom-up need for digital transformation, regulatory and political initiatives to improve data management in research demand adherence to quality standards and promote infrastructure development to facilitate an interconnected ecosystem of data [4]. Both the potential benefits and the methodological pitfalls of "big data" applications in biomedical research are well documented [5] with an overall expectation that given expert handling and realistic interpretation of computational models and predictive methods, significant scientific progress is possible. Driven by everyday needs in the development of biotechnological methods for generation and application of engineered myocardium, we investigate the prerequisites and best practices for capturing laboratory process and experimental data, and enabling integration with cross-organizational data-driven research.

Background

The Institute of Pharmacology and toxicology of the University Medical Center Göttingen (UMG) has a long-standing history in the generation of engineered human myocardium (EHM) induced pluripotent stem cells following a well-defined protocol [6]. The primary functional parameter of EHM samples are force and frequency of contraction (FoC, R-R) which directly report on quality and maturity of tissues and are evaluated during *in vitro* compound screening experiments, see Figure 1. The technological challenge currently faced is to develop scalable methods for generation, cultivation and analysis of EHMs that are compliant with Good Manufacturing Practice (GMP). Our key approach is the automation of the entire process by removing human interventions in handling and documentation stepwise one after another.

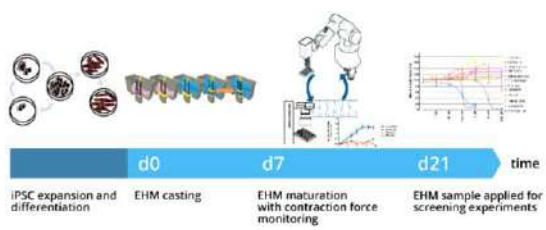


Figure 1 – Simplified process of engineered heart muscle generation and substance screening. (CC-BY-SA 4.0 T. Meyer, M. Suhr)

Automation of complete drug screening experiments depends on the same atomic laboratory processes as EHM generation, routine change of culture media and FoC measurement. A video-optical method for simultaneous measuring of 48 EHM samples on a microtiter culture-plate is developed on site, enabling major gains in process efficiency. Currently, EHM generation protocol steps are executed manually by lab personnel and are documented in paper notebooks and the electronic lab notebook (ELN) RSpace [7]. The video-optical FoC measurement is performed by a high-resolution camera attached to a hardware frame-grabber for near real-time transformation of the optical signal into time series data. Raw output from the camera system is processed using a custom MATLAB program to calculate functional parameters. The tool stores results in a MySQL database along with experiment parameters and metadata like EHM sample source materials (base cell line, generation protocol, batch ID, creation date, type and geometry of culture-plate, etc.). The database schema has been designed and frequently expanded to enable fine-grained documentation and planning of EHM generation and

experiments with process automation in mind. Data is stored describing assets like culture-plates, cell lines, chemical substances, and personnel involved as well as the execution of experimental steps.

These laboratory processes initiate a typical scientific data life cycle routine. Beginning with the planning and creation of primary data, several stages of processing, publication, and archiving are traversed before allowing re-use of result data for new investigations [8]. Such research data management (RDM) activities are supported by sophisticated infrastructures and systems for ETL processes, data publication, and data sharing [9]. Goals for technical implementation have been coined the FAIR principles, calling for research data that is findable, accessible, interoperable, and reusable [10]. Advanced RDM tools achieve compliance with the FAIR criteria but usually on a per-project basis and not as a basic tool for everyday documentation and data capture in the laboratory. Considering the overall trajectory towards interconnected systems of FAIR data across organizational and discipline boundaries [4], on-site data management infrastructure is needed that allows for user-friendly, FAIR-compatible, data provenance preserving, and possibly regulation compliant everyday experimental work in a constantly changing ecosystem of experimental settings and lab automation.

Objectives

Based on the EHM use case, we explore the preconditions and available tools for a laboratory RDM infrastructure enabling acquisition, storage, and sharing of raw data and metadata created during the process. Essential quality requirements for a resulting system architecture are usability in a laboratory environment and software sustainability in a highly specific area of biomedical research, i.e. drawing on lean human and financial resources for information technology. Ideally, the concept should be flexible enough to be reused for different experimental scenarios in biomedical research laboratories.

Methods

Literature search is conducted to (a) determine mandatory quality standards for data processing systems in the given experimental context, and to (b) collect reported best practices for RDM in similar biomedical contexts and tools available for prototypical implementation. Through iterative manual screening of articles' title, abstract, and full text, the result set is refined. Based on frequent interviews and discussion sessions with the involved biomedical researchers, laboratory processes are modeled as basis for formation of a system requirements catalog following recommendations from the requirements engineering field [11]. Combining the insights from the literature search, we propose a concept for an architecture of RDM tools to meet the requirements of a biomedical research lab and implement a prototype supporting the EHM generation use case.

Results

Goals and Requirements

Taking into account content and scope of application of multiple "GxP" guidelines [12], we include implications for information systems from GLP in the requirements engineering process, namely: user access control mechanism, audit trailing, metadata storage, data integrity and quality checks, as well as validation of interfaces, archival, and backup procedures [12]. From the analysis of data management best practices and available tools we derive six goals for a comprehensive system: (CG1) to enable research process, (CG2) to maximize usability, (CG3) to ensure information security, (CG4) to favor dynamic extension, (CG5) to maximize sustainability, and (CG6) to increase reporting quality of results. Based on the conducted literature and requirements analyses as well as the defined goals, we defined primary system requirements that imply general usage scenarios for information systems in the given context. This set represents the functionality of a socio-technical system to operate core processes at a biomedical research laboratory.

Tools

The literature search results revealed 10 distinct software tools and frameworks for data management in a research laboratory. Multiple of these are no longer actively developed since original publication. Among open source projects, two at least partially closed-source commercial products are reported (LabMatrix, LabKey Server) which have not been further considered for application in this exploratory case study. We exclude tools that are tailored only for minor or too specific aspects of the system requirements. Final candidate tools that enable both experimental data capture and FAIR-compatible data sharing are openBIS [13] and SEEK [14] from the "FAIRDOM" ecosystem.

Architecture Concept

Laboratory information systems are commonly divided into three categories: user-centric ELN, operative data- and process-centric laboratory information management system (LIMS), research result-centric scientific data management system (SDMS) [15]. The requirements specified exceed the functionality offered by any single system especially with regard to the need for extensibility: constantly changing experimental settings and analytical processes require a system that can be flexibly integrated and extended with new modules and functionality while at the same time meeting quality requirements like interoperability and sustainability with given resources.

Machina and Wild discuss three models for system integration in drug discovery research [15] which is very similar to the given environment: "ELN-centric" integration is driven by the idea to unify all user interaction with different subsystems through the interface provided by an electronic lab notebook. Modular integration describes the interconnection of sub-modules from neighboring systems like ELN and LIMS enabled by a distinct integration system. The service-oriented integration approach relates closely to the Service-Oriented-Architecture (SOA) paradigm of software design [16] with multiple independent modules (services) providing encapsulated functionality and communicating through messages. We combine aspects from these integration models into an architecture concept as follows: The available tools for the research lab data management system are self-enclosed software systems with internal modular structure as presumed in the modular-integration approach. To ensure operative flexibility and longevity of the systems, extensibility is maximized by applying the SOA paradigm. We propose a service-oriented integrated system architecture constructed around five logical components:

- **Electronic laboratory notebook** (ELN). A system for manual laboratory work documentation and collaboration in GxP-compliant manner.
- **Laboratory information management system** (LIMS). A leading system for laboratory asset registration and management.

- Multiple **experiment-specific subsystems** providing services for custom settings that cannot be implemented with ELN and LIMS.
- **Data integration and analysis system** (DIAS). A system for integration of data from disparate sources, visualization and analysis.
- **Research data management system** (RDMS). A system for research project planning, high-level documentation, result and data sharing.

To achieve integration of these components, three auxiliary building blocks complete the concept:

- **Portal**. A (web-based) central access point for all functionality that provides a unified graphical user interface.
- **Service integration layer**. The control center for service-based integration, connecting services and subsystems and exposing functionality via API or relaying GUI to the portal service. Central shared services like user management, authentication, and authorization should be realized within the integration layer to harmonize subsystem behavior.
- **External services**. An arbitrary number of external services and systems relayed to the overall system through connector services in the integration layer.

Communication between services and subsystems should be realized through application programming interfaces as the SOA paradigm proclaims. Subsystem user interfaces can be integrated into the portal-based unified GUI. If subsystems do not provide API or GUI, it should be possible to customize services within the integration layer to provide surrogate functionality. Figure 2 depicts a graphical representation of the architecture concept.

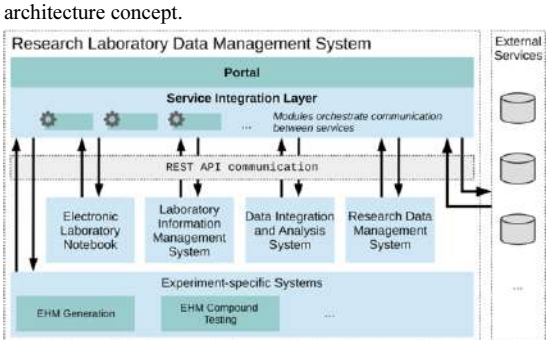


Figure 2 – Architecture concept for a comprehensive research laboratory data management system. (CC-BY-SA 4.0 M. Suhr)

Prototype Implementation

With regard to the goal of enabling especially the EHM experiment-specific workflows, ELN and RDMS are the components to be implemented with priority: ELN because it enables coherent documentation of manually executed workflow tasks; RDMS because the planning and publishing capabilities enable improvements for both organization of work and FAIR-compliant publication of results. The existing RSpace instance is to be integrated and a local instance of SEEK will be set up as part of the prototype system.

The MySQL database system developed for the EHM compound testing experimental data capture and storage and the MATLAB program for analysis of video-optical contraction force measurements are heavily interwoven in the current form. Instead of replacing these operational systems in an early phase of the project, they are to be integrated as a first experiment-specific subsystem.

Finally, portal and service integration layer components have been implemented as custom solution for demonstration purposes: The “LabHub” in its first iteration is intended as a central platform for user interaction, resource management, data entry, and access to data from past experiments in the EHM context. Aiming for usability, extensibility, and sustainability we utilize the widespread content management system Drupal as a software development framework to create a set of logically independent modules for user interaction and data processing. Service integration functionality is demonstrated by offering an administrative user interface for service and resource registration. Drupal modules in the LabHub context define service types which again are used by other modules as a resource. Instances of services are registered through GUI using internet protocol. This allows for example, to register multiple instances of the SEEK platform, which in turn can be connected to by an arbitrary number of modules that communicate with the SEEK REST API. “Connector” modules that define service types and “consumer” modules that interact with the respective web-service API are implemented for the SEEK and RSpace systems. API and application layer modules for the existing EHM MySQL database are drafted, enabling role-based management of laboratory asset data and access to experimental result data. Using the integration layer functionality, this experiment-specific system communicates with both RSpace and SEEK instances allowing for user-centric documentation of workflow steps in the ELN and FAIR-compliant publication of FoC measurement data in the RDMS component.

At the current stage of the project, access to both the Drupal and SEEK web servers is restricted to the local network. Developmental stage source code of the LabHub prototype is available through a public code repository (http://hdl.handle.net/21.11101/0000-0007-CADE-C).

Discussion

A challenge implicit in the research question is how to find candidate software for implementing an infrastructure satisfying all the diverse requirements of operative and research data processing and managing in a biomedical lab. This is mostly due to the circumstance that no canonical descriptive name for the category of software exists. “ELN” and “LIMS” are labels well-defined and attached to a multitude of open source, service-based, and commercial products. The term “SDMS” is mentioned in articles and per definition comes very close to the here proposed requirements for research lab data management [17] but is neither topical nor actively promoted. Of the discovered software tools that in description satisfied a considerable subset of our requirements, some are no longer available or maintained. The most challenging task for this project is to not follow these projects into a dead end but to aim for sustainability. Remarkably, some software tools initiated in the years 2007-2011 have succeeded due to innovative concepts and future-proof development processes.

Architecture Concept

Inefficiency of user interfaces is a central barrier of adoption of IT systems in the lab [19]. Overcoming usability barriers in practice is the actual challenge and must be made a focal point of infrastructure design and implementation. Four of our declared six main goals emphasize software quality aspects, addressing this need to improve quality of research data handling while not negatively interfering with operative processes [18].

Appropriate models of laboratory information system architecture and the integration of subsystems exist for decades already with similar challenges pronounced as [19]. Despite this understanding of the problems, formulation of solution models, and the obvious progress of technology, seamless integration of data producing and data processing entities is an ongoing effort. The presented analysis updates the modeling with explicit focus on sustainable operation of the technical infrastructure. Strength of the proposed concept is the extensible integration of existing software systems as well as service modules into an architecture accessible through a central portal component. Combination of integration methods [15] serves to integrate the existing systems as components, as demonstrated here with the RSpace instance. This approach allows for re-use of the framework at other laboratories as it is agnostic of whether or not components have been deployed before. "ELN-centric integration" [15] follows a similar reasoning but for adoption of the approach an extensible and modifiable ELN component is required. In comparison, the framework presented here does not require any existing system to function as the central access and integration component but encourages a modular solution.

Considering the broad spectrum of specialized systems discovered in the literature search and the indicated system requirements, only an integrated architecture of multiple systems and services is probable to succeed in the long run. Classic LIMS solutions are reported to expand in functionality and aim towards becoming comprehensive information systems for all laboratory needs [15] but the dynamic nature of method and experiment development at a research lab is not suitably addressed by traditional monolithic systems. Notably, the product STARLIMS is named as an example for an expanding LIMS [15] and is already deployed at UMG service units Biobank and Stem Cell Unit. Investigation of synergistic effects and integration opportunities is planned.

Among RDM projects in medical informatics and biomedicine, the framework discussed here is remarkable for not focusing infrastructure within a collaborative project. Typically, systems are developed and proposed for project-specific use cases on grant-based funding. Here, persistent organizational infrastructure is considered with research projects and the associated data management systems located within the outer system context. Sustainability and extensibility have been exclaimed goals so that the resulting architecture should be more long-lived than project-centric infrastructures. The modularity embedded in the concept is meant to ensure that subsystems can be exchanged and replaced to address changing technical and structural requirements. The more systems are integrated into a complex data management architecture however, the higher becomes the need for capturing data provenance records to prevent data quality from deteriorating [9,20].

LabHub Prototype

The goal dimensions initially declared are addressed by the prototype. Still, the achievements mark only a beginning when assessed against the business goals of enabling research process and improving reporting quality. Most problematic is the question of sustainability. The topic has been at the core of considerations leading to the architecture concept and implementation plan. The prototype created is not inherently sustainable though. Some of the approaches reported in the scientific literature have vanished quickly in spite of support by a professional team of researchers or developers. Re-evaluation of openBIS as a data management system may thus be of interest. Major drawback ascertained in assessment of openBIS for the EHM use case was the user interface design not suited for cross-platform operation. Given the resources to integrate openBIS into a modern, usable, platform-independent front end, reuse of the system's wide spectrum of data management features could be an elegant way to create an integrated platform, as already demonstrated by other projects [21].

Promoting LabHub development to become a data-centric counterpart for efforts to standardize control interfaces for lab automation devices may be considered. We plan cooperation with the Standardization in Lab Automation (SiLA) to advance the automation of EHM generation processes. To our knowledge, no platform for management, control, and scheduling of lab processes using SiLA compatible devices does exist yet, indicating an opportunity to investigate.

Systemic Challenges

Sustainability of a system is less a question of technology than of available resources. Project grants for specialized innovative research data processing systems may be easier to allocate than permanent funding for infrastructures needed for basic operation. Ideally, the operative infrastructure would support FAIR data sharing to promote secondary use of data internally for quality control and meta-analyses, and externally in research projects. Above all, implementing and maintaining a complex integrated system in such a way requires capable personnel. A small dedicated team of software engineers may become an integral asset.

Laboratories and institutes will need to adopt gateway services for inter-organizational collaborative data-driven research projects on (inter-) national scale. Here, the SEEK instance operates as such by allowing FAIR-compatible sharing of data. Projects like the Medical Informatics Initiative in Germany [22] or the European Open Science Cloud (EOSC) [23] promote infrastructural integration and cross-site data exchange. Independent of how the challenge of privacy preserving data sharing will be solved technically, providing internal primary data in FAIR-compliant ways enables integration with the emerging global data infrastructure [4]. Developing and operating a FAIR-compatible infrastructure at an institute or laboratory may prove to be of both scientific and strategic advantage in technological advancement of biomedical research in the coming years.

Conclusions

We describe challenges and requirements that transcend the specific setting of cardiac tissue engineering. The proposed architecture framework promotes extensibility and usability for highly dynamic, constantly evolving operative realities. The topic of innovative research lab data infrastructure is open for collaborative advancements and investigation. FAIR data publication and reporting are explored for the presented use case but are not yet operational and require further investigation and refinement especially with regard to application of semantics. Achieving sustainability of the LabHub software or any integrated infrastructure is a major challenge and requires commitment regarding resources and effort. The development of an integration and control framework for laboratory process automation is promising in terms of scientific and operative potential.

In conclusion, a framework for comprehensive research lab data management is proposed for discussion and possibly adaptation. Our findings are a foundation for diverse subsequent investigation and possibly software development. The core challenges require close cooperation and collaboration of biomedical researchers and computer scientists to pave the way for dynamic and persistent socio-technical information systems connected to a global infrastructure of data.

Acknowledgments

This work was funded by the BMBF in the project Data and Information Management (DZHK (German Center for Cardiovascular Research)) (grant 81Z7300173) as well as by the DFG for the Collaborative Research Centers (CRC) 1002 on Modulatory Units in Heart Failure, subproject INF, and CRC 937 on Collective Behavior of Soft and Biological Matter.

References

[1] B. Fermini, S.T. Coyne, and K.P. Coyne, Clinical Trials in a Dish: A Perspective on the Coming Revolution in Drug Development, *SLAS Discov. Adv. Life Sci. R D*. (2018) 12. doi:10.1177/2472555218775028.

[2] E.M. Antman, and J. Loscalzo, Precision medicine in cardiology, *Nat. Rev. Cardiol.* **13** (2016) 591–602. doi:10.1038/nrcardio.2016.101.

[3] M. Tiburcy, J.E. Hudson, P. Balfanz, S. Schlick, T. Meyer, M.-L. Chang Liao, E. Levent, F. Raad, S. Zeidler, E. Wingender, J. Riegler, M. Wang, J.D. Gold, I. Kehat, E. Wettwer, U. Ravens, P. Dierickx, L.W. van Laake, M.J. Goumans, S. Khadjeh, K. Toischer, G. Hasenfuss, L.A. Couture, A. Unger, W.A. Linke, T. Araki, B. Neel, G. Keller, L. Gepstein, J.C. Wu, and W.-H. Zimmermann, Defined Engineered Human Myocardium With Advanced Maturation for Applications in Heart Failure Modeling and Repair, *Circulation*. **135** (2017) 1832–1847. doi:10.1161/CIRCULATIONAHA.116.024145.

[4] P. Wittenburg, and G. Strawn, Common Patterns in Revolutionary Infrastructures and Data, (2018). doi:10.23728/b2share.4e8ac36c0dd343da81fd9e83e72805a0.

[5] H. Fröhlich, R. Balling, N. Beerenwinkel, O. Kohlbacher, S. Kumar, T. Lengauer, M.H. Maathuis, Y. Moreau, S.A. Murphy, T.M. Przytycka, M. Rebhan, H. Röst, A. Schuppert, M. Schwab, R. Spang, D. Stekhoven, J. Sun, A. Weber, D. Ziemek, and B. Zupan, From hype to reality: data science enabling personalized medicine, *BMC Med.* **16** (2018) 150. doi:10.1186/s12916-018-1122-7.

[6] M. Tiburcy, T. Meyer, P. Soong, and W.-H. Zimmermann, Collagen-Based Engineered Heart Muscle, in: M. Radisic, L.D. Black III, and L.D.B. III (Eds.), Card. Tissue Eng., Springer New York, 2014: pp. 167–176. doi:10.1007/978-1-4939-1047-2_15.

[7] S. Macdonald, and R. Macneil, Service integration to enhance research data management: RSpace electronic laboratory notebook case study, *Int. J. Digit. Curation*. **10** (2015) 163–172.

[8] L. Corti, V. Van den Eynden, L. Bishop, and M. Woollard, Managing and sharing research data: a guide to good practice, Sage, Los Angeles, 2014.

[9] C.R. Bauer, N. Umbach, B. Baum, K. Buckow, T. Franke, R. Grütz, L. Gusky, S.Y. Nussbeck, M. Quade, S. Rey, T. Rottmann, O. Rienhoff, and U. Sax, Architecture of a Biomedical Informatics Research Data Management Pipeline, *Stud. Health Technol. Inform.* **228** (2016) 262–266. doi:10.3233/978-1-61499-678-1-262.

[10] M.D. Wilkinson, M. Dumontier, Ij.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, *Sci. Data*. **3** (2016) 160018. doi:10.1038/sdata.2016.18.

[11] K. Pohl, Requirements Engineering Fundamentals, 2nd Edition: A Study Guide for the Certified Professional for Requirements Engineering Exam - Foundation Level - IREB compliant, Rocky Nook, Inc., 2016.

[12] S.H. Segalstad, International IT Regulations and Compliance: Quality Standards in the Pharmaceutical and Regulated Industries, John Wiley & Sons, 2008.

[13] A. Bauch, I. Adamczyk, P. Buczek, F.-J. Elmer, K. Enimanev, P. Glyzewski, M. Kohler, T. Pylak, A. Quandt, C. Ramakrishnan, C. Beisel, L. Malmström, R. Aebersold, and B. Rinn, openBIS: a flexible framework for managing and analyzing complex data in biology research, *BMC Bioinformatics*. **12** (2011) 468. doi:10.1186/1471-2105-12-468.

[14] K. Wolstencroft, S. Owen, F. du Preez, O. Krebs, W. Mueller, C. Goble, and J.L. Snoep, The SEEK: a platform for sharing data and models in systems biology, *Methods Enzymol.* **500** (2011) 629–655. doi:10.1016/B978-0-12-385118-5.00029-3.

[15] H.K. Machina, and D.J. Wild, Laboratory informatics tools integration strategies for drug discovery: integration of LIMS, ELN, CDS, and SDMS, *J. Lab. Autom.* **18** (2013) 126–136. doi:https://doi.org/10.1177/2211068212454852.

[16] T. Erl, Service-oriented architecture: concepts, technology, and design, Prentice Hall Professional Technical Reference, Upper Saddle River, NJ, 2005.

[17] H.K. Machina, and D.J. Wild, Electronic laboratory notebooks progress and challenges in implementation, *J. Lab. Autom.* **18** (2013) 264–268.

[18] K.T. Taylor, Evolution of Electronic Laboratory Notebooks, in: Collab. Comput. Technol. Biomed. Res., Wiley-Blackwell, 2011: pp. 301–320. doi:10.1002/9781118026038.ch19.

[19] R.D. McDowall, and D.C. Mattes, Architecture for a comprehensive laboratory information management system, *Anal. Chem.* **62** (1990) 1069A-1076A.

[20] B. Baum, C.R. Bauer, T. Franke, H. Kusch, M. Parciak, T. Rottmann, N. Umbach, and U. Sax, Opinion paper: Data provenance challenges in biomedical research, *It - Inf. Technol.* **59** (2017). doi:10.1515/itit-2016-0031.

[21] C. Mohr, A. Friedrich, D. Wojnar, E. Kenar, A.C. Polatkan, M.C. Codrea, S. Czemmel, O. Kohlbacher, and S. Nahnsen, qPortal: A platform for data-driven biomedical research, *PLOS ONE*. **13** (2018) e0191603. doi:10.1371/journal.pone.0191603.

[22] S.C. Semler, F. Wissing, and R. Heyder, German Medical Informatics Initiative, *Methods Inf. Med.* **57** (2018) e50–e56. doi:10.3414/ME18-03-0003.

[23] B. Mons, C. Neylon, J. Velterop, M. Dumontier, da S. Santos, L.O. Bonino, and M.D. Wilkinson, Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud, *Inf. Serv. Use*. **37** (2017) 49–56. doi:10.3233/ISU-170824.

Address for correspondence

Markus Suhr, University Medical Center Göttingen, Department of Medical Informatics, von-Siebold-Straße 3, 37075 Göttingen, +49 551 39 61519, markus.suhr@med.uni-goettingen.de, https://orcid.org/0000-0002-6307-3253

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190245

Early Prediction of Acute Kidney Injury in Critical Care Setting Using Clinical Notes and Structured Multivariate Physiological Measurements

Mengxin Sun[a], Jason Baron[b], Anand Dighe[b], Peter Szolovits[c], Richard G. Wunderink[d], Tamara Isakova[d], Yuan Luo[d]

[a] *Independent Researcher, Chicago, IL, USA*
[b] *Department of Pathology, Massachusetts General Hospital, Boston, MA, USA*
[c] *Computer Science and Artificial Intelligence Lab, MIT, Cambridge, MA, USA*
[d] *Feinberg School of Medicine, Northwestern University, Chicago, IL, USA*

Abstract

The onset of acute kidney injury (AKI) during an intensive care unit (ICU) admission is associated with increased morbidity and mortality. Developing novel methods to identify early AKI onset is of critical importance in preventing or reducing AKI complications. We built and applied multiple machine learning models to integrate clinical notes and structured physiological measurements and estimate the risk of new AKI onset using the MIMIC-III database. From the clinical notes, we generated clinically meaningful word representations and embeddings. Four supervised learning classifiers and mixed-feature deep learning architecture were used to construct prediction models. The best configurations consistently utilized both structured and unstructured clinical features and yielded competitive AUCs above 0.83. Our work suggests that integrating structured and unstructured clinical features can be effectively applied to assist clinicians in identifying the risk of incident AKI onset in critically-ill patients upon admission to the ICU.

Keywords:

Clinical Decision Support, Natural Language Processing, Acute Kidney Injury.

Introduction

Acute kidney injury (AKI) is commonly seen in adults in the intensive care unit (ICU). AKI is one of the major diagnoses among ICU patients and a leading factor associated with a prolonged hospital stay and subsequent morbidity or early mortality post discharge [1, 2]. Acute renal failure is a complex disorder that presents itself in a variety of settings with clinical manifestations, ranging from a minimal elevation in serum creatinine to anuric renal failure [3]. Unfortunately, the main biomarker of AKI, serum creatinine (SCr), is a late marker of injury, which delays diagnosis and treatment [4]. However, the efficacy of intervention greatly relies on the early identification of AKI [5]. Early recognition is critical in that AKI usually occurs over the course of a few hours to days and is potentially reversible if detected and managed early [5, 6].

In this study, we used the definition of AKI from the Kidney Disease Improving Global Outcomes (KDIGO) [7], in order to standardize the published diagnostic criteria. The diagnostic criteria are defined as an acute increase in the absolute level of serum creatinine of more than 0.3 mg/dl or 50% higher change in serum creatinine (SCr) from baseline within a 48-hour period or decreased glomerular filtration rate (GFR) to less than 0.5 ml/kg/hour for more than six hours [7, 8]. These criteria were based on accumulating evidence that even small alterations in SCr are associated with serious consequences. Therefore, an accurate creatinine forecast may enable prediction of AKI risk. In this study, we focus on predicting AKI using first-day measurements of a multivariate panel of physiologic variables, in order to elucidate early, subclinical deterioration of patient's physiologic baselines that are predictive of AKI.

Many factors including nephrotoxic medications, insufficient effective circulating fluid volume, and intrinsic renal disease can cause or contribute to AKI [9, 10]. In addition, epidemiology studies show multiple comorbidities, including diabetes mellitus, cardiovascular disease, chronic liver disease, cancer, and complex surgery have been associated with the development of AKI [5, 10, 11]. Thus, a comprehensive understanding of ICU patients is essential to predict the development of AKI.

The increasing use of electronic health records (EHRs) allows access a comprehensive and extensive amount clinical data to develop models to predict AKI in ICU [3]. There are several previous studies using EHR data to predict AKI [12-21]. Most studies achieved a modest performance with area under the receiver operation curve (AUC) close to 0.75. However, majority of these studies focus on specific patient population such as the elderly and cardic surgery patients. In addition, many prior studies rely on various static scoring algorithms and/or do not incorporate the temporal progression of the clinical, laboratory information that is shown to be effective for the prediction. Last but not least, some studies incoporate a limited set of predictors that are not inclusive enough to capture changes in clinical care that may impact AKI risk; similarly, some studies rely on non-routine biomarkers such as neutrophil gelatinase-associated lipocalin (NGAL) that are not always tested for general patients. Data-driven predictive modeling for AKI has recently gained traction, but is focused on using either structured EHR data [22] or unstructured clinical notes alone [23]. In summary, these models are not optimally suited for a relatively general patient population and are not set up as adaptive tools to assist the clinical decision-making process.

Our approach, in contrast, involves the careful modeling of a wide array of predictor data including clinical treatments and the temporal aggregation of predictor data. Specifically, we include not only structured data such as laboratory test results, but also unstructured data, the ICU clinical notes (e.g., physician notes and nurse notes) to provide a comprehensive picture of the patients' current pathophysiologic condition and help develop a powerful model to predict AKI.

From a pratical standpoint, our approach focuses on the early prediction of AKI on patients who do not meet AKI criteria on admission to the ICU, thus targeting a population that could

benefit from early intervention which might reverse the development of AKI or minimize its clinical impact. This is especially important since prior studies on automated AKI detection (as opposed to prediction) show limited effectiveness of therapeutic interventions in patients already meeting AKI criteria [24]. We expect the early prediction of AKI in this study to have a wide range of clinical applications.

Methods

Dataset

Data for this study was acquired from Medical Information Mart for Intensive Care III (MIMIC-III). MIMIC-III captures de-identified health information for more than 46,000 patients admitted to the critical care units at Beth Israel Medical Center between 2001 and 2012. We developed a SQL script to extract data of patients for who had a creatinine measured at 72 hours following ICU admission from the MIMIC-III database [25]. Only patients 18 years of age or older were included, and patients with the pre-existing condition of Chronic Kidney Disease (CKD), who have an estimated glomerular filtration rate GFR (eGFR) < 60 mL/min/1.73 m^2 were excluded [26, 27]. Data extracted include patients' age, gender, ethnicity, 72-hour serum creatinine (with only the creatinine value from the first 24 hours used as predictor, the value from the 2nd and 3rd days are used to identify AKI), vital signs and lab values during the first day of ICU admission, whether the patient was mechanically ventilated during the first day of ICU admission, the hourly rate of urine output during the first day of ICU admission, and the clinical notes during the first 24 hours of ICU admission. Only clinical notes from the first 24 hours of the ICU stay are included given the commonly seen lag between the provide-patient encounter and the time the notes were performed.

Based on the criteria from KDIGO [7], AKI is defined as either of the following two conditions being met: (1) greater than or equal to 50% increase from the baseline creatinine value to the current creatinine value and (2) greater than or equal to 0.3 mg/dL change in creatinine from the baseline creatinine to the current creatinine value. Since our study focuses on AKI developed after the ICU admission, patients who got AKI on admission (day 1) were excluded. After this step, the remaining patients' AKI status were determined by comparing day 2 and day 3 maximum creatinine to the day 1 minimum creatinine level (the baseline).

A total of 16,558 ICU stays of 14,469 patients met the inclusion criteria for this study. The dataset was first split into training and test sets by an 7:3 ratio. Then, reassignment was conducted to ensure multiple ICU stays of the same patient stayed in the same set. The AKI prevalence rate overall and in both sets are approximately 17% as shown in Table I below. Table II presents the predictor variables used in this study, along with their statistical characteristics such as mean and standard deviation for continuous variables and count and percentage for categorical or discrete variables.

Table I – AKI Status Distribution Overall and in Training and Test Sets.

Set	All	AKI	Non-AKI	AKI Prevalence
Overall	16,558	2,785	13,773	16.82%
Training	11,558	1,927	9,631	16.67%
Test	5,000	858	4,142	17.16%

Natural language processing of the clinical notes

To better interpret clinical notes, some preprocessing steps were needed. Masked protected health information (PHI) in the notes were removed first. Then, in order to make the information usable in machine learning classifiers, clinical notes were converted to structured features. Our first set of features consists of unigram bag-of-words [1], which identified and normalized lexical variants from the unstructured text content. Those features with document-frequency under 10 were removed to reduce noise. A total of 313 stop words were applied according to NCBI guide. Term frequency-inverse document frequency (tf-idf) weighting adjustment was also applied [28].

Table II - Univariate Characteristics for Predictors of Interest, N= 16,558

Variable	Mean	SD
Age (year)	60.80	16.16
Gender - N, %		
Female	6,735	40.68%
Male	9,823	59.32%
Ethnicity - N, %		
African-American	1,048	6.33%
White	11,981	72.36%
Hispanic	532	3.21%
Other	2,997	18.10%
Heart rate maximum (bpm)	105.43	19.76
Heart rate mean (bpm)	87.10	15.03
Systolic BP minimum (mmHg)	92.45	17.04
Systolic BP mean (mmHg)	119.06	15.85
Diastolic BP minimum (mmHg)	44.60	10.95
Diastolic BP mean (mmHg)	61.33	10.00
Temperature maximum (Celsius)	37.65	0.74
SpO2 minimum (%)	92.20	7.24
SpO2 mean (%)	97.53	1.83
Glucose level maximum (mg/dL)	173.82	72.33
Bicarbonate level minimum (mg/dL)	23.86	4.26
Creatinine level minimum (mg/dL)	0.73	0.20
Creatinine level maximum (mg/dL)	0.80	0.22
Hemoglobin level minimum (g/dL)	10.34	2.07
Platelet count minimum (K/μL)	208.83	113.13
Potassium level maximum (mg/dL)	4.42	0.80
Partial thromboplastin time minimum (s)	32.49	11.49
Partial thromboplastin time maximum (s)	40.39	23.63
International normalized ratio minimum	1.33	0.43
International normalized ratio maximum	1.47	0.68
Prothrombin time minimum (s)	14.48	3.00
Prothrombin time maximum (s)	15.42	4.23
Blood urea nitrogen level maximum (mg/dL)	16.54	8.34
White blood cell count maximum (K/μL)	12.94	8.14
Calcium level minimum (mg/dL)	8.08	0.75
Average urine output (mL)	2,249.40	1419.51
Estimated glomerular filtration rate (eGFR)	110.33	50.20
Mechanical Ventilation - N, %		
No (0)	7,649	46.20%
Yes (1)	8,909	53.80%

Missing data imputation for structured data

There are no missing values among age, gender, and race. Other structured clinical and laboratory data obtained from clinical settings contained missing values, which indicates certain tests were not performed during the patient's ICU stay. For example, the variables having a large proportion of missing values include minimum albumin level (74.1%), maximum bilirubin level (67.2%), maximum lactate level (55.8%), maximum c-reactive protein level (99.0%), maximum asparate aminotransferase level (66.8%), maximum pH level (36.6%), and minimum base excess level (64.8%).

We employed a two-step process to handle missing values. First, we removed the variables with missing values greater

than 20%. We thenfilled in the values for predictors (e.g., labs not performed or recorded) using Multivariate Imputation by Chained Equations (MICE) [29] for those variables with less than 20% missing values. MICE estimates a conditional model for each variable to be imputed, with the other variables as possible predictors [30]. The term chained equation comes from the adoption of a Gibbs sampler, which is an iterative Markov Chain Monte Carlo algorithm for obtaining a sequence of observations that are approximated from a joint probability distribution. As MICE closely tracks the conditional interdependencies among variables, we expect MICE to produce more accurate imputation. Then, as the second step, we used the measured and imputed values for these predictors plus age and gender to predict maximum creatinine results during day 2 and day 3. In this step, we predicted both numerical results for creatinine (linear regression) and whether creatinine increase would be classified as AKI (logistic regression). Although no creatinine results were actually missing from our dataset per the inclusion criteria, we assessed model performance and creatinine predictability by masking creatinine results from a test fold during five-fold cross validation and then compared predicted creatinine results to the masked (measured) values. The masked-measured values were treated as the "ground truth" in assessing model performance. The imputation stage was required because the prediction algorithms used in the second stage of our procedure could not directly accommodate missing data in predictors.

Machine learning classifiers

Logistic regression (LR), random forest (RF), multinomial naïve Bayes (NB), and supported vector machine (SVM) classifiers were implemented in scikit-learn to find the best prediction model. As the class ratio is imbalanced, we set the class_weight parameter to "balanced" for logistic regression, random forest and supported vector machine classifiers to down weight the more popular class. In order to tune parameters and reduce over-fitting and instablity, a grid search with 3-fold cross-validation was performed on the training set. The best parameters for each classifier were then applied to the test set to assess the prediction performance of each model.

Mixed-feature Convolutional Neural Networks

We also explored deep learning models for AKI prediction. We used mixed-feature Convolutional Neural Networks (CNN) to combine word features and structured clinical features. The architecture of mixed-feature CNN model is shown in Figure 1. It used pre-trained word embeddings and structured clinical data as the input. A one-dimensional convolution layer was built on the input embeddings. We used max pooling to select the most important feature with the highest value in the convolutional feature map and then concatenated the max pooling results of word embeddings and structured clinical features. The concatenated hidden features were fed into a fully connected layer, followed by a dropout and ReLU activation layer. Finally, a fully connected layer was fed to a softmax output layer, whose output is the probability distribution over labels (AKI or Non-AKI).

We adopted the following parameter settings for the mixed-feature CNN after further separating 10% from training data as validation set for parameter tuning, the number of convolution filters: 32, the convolution kernel size: 5, the dimension of hidden layer in the fully connected layer: 64, dropout keep probability: 0.8, learning rate: 0.001, batch size: 64. We used weight blancing in mixed-feature CNN training to address the class imbalance issue. In particular, we used the labels of classes to automatically adjust weights inversely proportional to class frequencies using the training data. The weights are then used in calcualting the weighted cross entropy from logits in the softmax layer.

Evaluation

We adopted area under the receiver operating characteristic (AUC) to evaluate the performance of imbalanced binary classifiers. Precision, recall, F-measure of positive AKI status were also calculated for reference.

Results

In this section, we report the evaluation results on the held-out test set using four supervised learning classifiers (LR, RF, NB, SVM) over three different feature configurations (structured feature only, unstructured clinical notes only, and structured features combined with unstructured clinical notes), as well as results by mixed-feature CNN. Table III presents the results from the above model configurations.

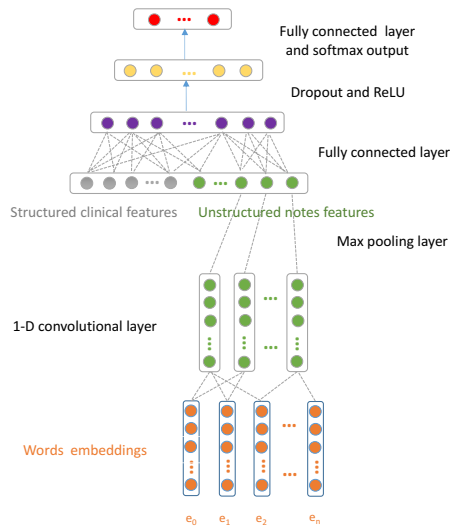


Figure 1. Mixed-feature Convolutional Neural Networks.

We first investigated the combinations of three feature representations and four supervised learning classifiers. The baseline predictor of each representation used Naive Bayes algorithm yielded AUC in the range from 0.65 to 0.75, F-measure from 0.17 to 0.37. In particular, combining structured and unstructured features in NB does not lead to better AUC. Logistic regression with L2-regularization over the mixed features yielded the best F-measure of 0.5423 on AKI status prediction, while calibrated SVM with L1-regularization gave the best AUC of 0.8352. Overall, all classifiers, except for NB as the baseline, yielded a competitive AUC score over 0.79, given that most previous models had modest AUCs around 0.75 (see Related Work section). In addition, the best LR model favored high recall (i.e., sensitivity: 0.7284) over precision (i.e., positive predictive value: 0.4319). This serves well for the clinical application of AKI onset alarm as we want to capture as many future AKI onsets as possible, while tolerating modest false alarms.

For the mixed-feature CNN, using ratio balancing class weights, yielded the best performance of 0.8167 AUC and 0.3559 F-measure. Despite heavy parameter tuning, all configurations of parameters of mixed-feature CNN did not outperform the best non-CNN classifiers. Based on our previous experiments on general domain text corpora [31], CNN-based architectures generally work well for datasets with

short texts, but may not outperform bag-of-word on corpus with long texts such as the AKI clinical note corpus. In addition, due to the fact that training CNN models is usually time-consuming, in our case, well-calibrated non-CNN classifiers seem to be a more suitable choice.

Feature analysis

We further examined important features by ranking coefficients in the L2-regularized logistic regression (best F-measure) over bag-of-words and structured clinical features. Table IV presents top 10 positive structured clinical features that contribute to AKI onset. The maximal creatinine level in day 1 carries a weight almost 10 times higher than the predictor with the second highest weight. The information critical to predict AKI is concentrated in the top three predictors by weight. These predictors are consistent with the known pathophysiology of AKI that older patients have higher incidence of AKI. Mechanical ventilation and coagulopathy (prolonged prothrombin times) are also known risk factors of AKI and also might represent patients with higher severity of illness and/or sepsis. The elevated potassium level likely represents early electrolyte disturbances in the setting of injured kidneys likely to meet AKI definition in the subsequent days. Finally, the elevated creatinine and BUN levels, while not meeting AKI criteria in these patients given the exclusion criteria of the study, likely represents an early elevation indicative of injured kidneys which has not peaked yet.

Table III – Machine Learning Model Results. Str: structured features. Unstr: unstructured features. ALG: algorithm. MCNN: Mixed-feature CNN. Best AUC and F-measure are in bold.

Features	ALG	AUC	Precision	Recall	F-meausre
Str	LR	0.8336	0.4319	0.7284	**0.5423**
+	RF	0.7914	0.6774	0.0490	0.0913
Unstr	NB	0.6728	0.2681	0.6166	0.3737
	SVM	**0.8352**	0.7274	0.2340	0.3541
	MCNN	0.8167	0.7292	0.2354	0.3559
Str	LR	0.8117	0.3782	0.7401	0.5006
	RF	0.8132	0.8125	0.1364	0.2335
	NB	0.6574	0.2560	0.6247	0.3631
	SVM	0.8097	0.7439	0.2133	0.3315
Unstr	LR	0.7735	0.4009	0.6457	0.4946
	RF	0.7582	0.5769	0.0350	0.0659
	NB	0.7495	0.5658	0.1002	0.1703
	SVM	0.7727	0.5306	0.0606	0.1088

Figure 2 presents top 50 positive bag-of-words features with theirs font sizes proportional to their coefficients in the model. For the selected features, in most cases, these features appear to be clinically meaningful. For example, 'lasix' with highest coefficient in bag-of-words is one of the diuretics that can treat fluid retention and edema that might be caused by kidney dysfunction. Also, 'co' and 'ci' in bag-of-words is short for 'cardiac output and cardiac index', which assess whether a patient's heart is pumping enough blood and delivering sufficient oxygen to cells. Patients with abnormal cardiac output and even heart failure often have a higher risk of AKI [36]. The words 'insulin' and 'incisional' indicate diabetes mellitus comorbidity and procedural risks that may predispose AKI onset. Feature examination confirms that clinically meaningful key words in clinical notes can be used to predict AKI onset and the models we built do capture those words.

Discussions, Limitations and Future Work

Due to data accuracy concern, this study only uses the increase in creatinine to determine AKI. Should data that accurately record the decrease of GFR for a prolonged period exist, future studies will include both criteria to better capture AKI.

Table IV – Top physiologic variables that are associated with increased AKI risk by L2-regularized logistic regression.

Physiologic Variable	Weight
Creatinine level maximum (mg/dL)	4.8374
Mechanical Ventilation	0.4578
International normalized ratio maximum	0.2315
Potassium level maximum (mg/dL)	0.0589
Prothrombin time minimum (s)	0.0267
Estimated glomerular filtration rate (eGFR)	0.0199
Age (year)	0.0192
Diastolic BP mean (mmHg)	0.0101
Partial thromboplastin time minimum (s)	0.0095
Blood urea nitrogen level maximum (mg/dL)	0.0063

Figure 2 Ranked top 50 positive features in Bag-of-words with its coefficients as font size in AKI onset prediction by L2-regularized logistic regression.

When working with structured clinical variables, we have used MICE imputation to fill in missing entries [30]. In principle, MICE assumes the missing-at-random pattern, an assumption that almost certainly will not hold in real clinical practice, since clinicians often order tests under certain expectations about the likely results. However, our previous study shows that even when the missing-at-random assumption may not hold, in practice, MICE may still be used as an effective way and baseline for comparing other multiple imputation methods due to its simple implementation [32, 33]. On the other hand, missingness may represent no indication for having the test performed, hence may offer clinical information. Thus, we plan to investigate missingness patterns as additional predictors for AKI prediction in future studies.

When working with clinical notes, we only explored the bag-of-words model in conventional machine learning models and the word embeddings in mixed-feature CNN model. Though this yields promising performance, there are other options for using medical concepts as features such as through NLP pipelines including MetaMap [34], clinical Text Analysis and Knowledge Extraction System (cTAKES) [35]. Further investigation is necessary in order to identify which pipeline is the most suitable tool to generate medical concepts in disease prediction with clinical notes. In addition, we only explored mixed-feature CNN as one deep learning framework. Other deep learning architecture that is suitable for clinical dataset with mixed feature types and long text content are of future interest.

The improved model performance implies the potential clinical application in the near future. However, it requires further study and discussion to determine the model performance that is needed before it could be intergreted in clincal decision-making process.

Conclusions

Our study demonstrates that a supervised learning method that integrates both structured physiologic variables and unstructured clinical notes can be effective in early prediction of AKI onset in the first 72 hours following ICU admission for general adult patient population. We showed that carefully selected physiologic variables and well-represented clinical notes as features can predict new AKI onset with an AUC greater than 0.83, competitive with previous studies focused on specific patient groups or on novel biomarkers. Our work suggests that prospective trials with independent model training and validation cohorts are needed to further evaluate the clinical utility of this approach for identifying at risk patients early in their hospital course and potentially instituting interventions to decrease the likelihood of developing AKI.

Acknowledgements

This study was supported in part by grant R21LM012618.

References

[1] T. Ali *et al.*, "Incidence and outcomes in acute kidney injury: a comprehensive population-based study," (in eng), *J Am Soc Nephrol,* vol. 18, no. 4, pp. 1292-8, Apr 2007.

[2] C. V. Thakar, A. Christianson, R. Freyberg, P. Almenoff, and M. L. Render, "Incidence and outcomes of acute kidney injury in intensive care units: a Veterans Administration study," (in eng), *Crit Care Med,* vol. 37, no. 9, pp. 2552-8, Sep 2009.

[3] H. E. Wang, P. Muntner, G. M. Chertow, and D. G. Warnock, "Acute kidney injury and mortality in hospitalized patients," (in eng), *Am J Nephrol,* vol. 35, no. 4, pp. 349-55, 2012.

[4] H. R. de Geus, M. G. Betjes, and J. Bakker, "Biomarkers for the prediction of acute kidney injury: a narrative review on current status and future challenges," (in eng), *Clin Kidney J,* vol. 5, no. 2, pp. 102-108, Apr 2012.

[5] J. A. Kellum, N. Lameire, and K. A. G. W. Group, "Diagnosis, evaluation, and management of acute kidney injury: a KDIGO summary (Part 1)," (in eng), *Crit Care,* vol. 17, no. 1, p. 204, Feb 2013.

[6] P. K. T. Li, E. A. Burdmann, R. L. J. A. j. o. n. Mehta, and transplantation, "Acute kidney injury: global health alert," vol. 6, no. 2, pp. 75-81, 2013.

[7] K. D. I. G. O. C.-M. W. Group, "KDIGO clinical practice guideline for the diagnosis, evaluation, prevention, and treatment of Chronic Kidney Disease-Mineral and Bone Disorder (CKD-MBD)," *Kidney international. Supplement,* no. 113, p. S1, 2009.

[8] R. L. Mehta *et al.*, "Acute Kidney Injury Network: report of an initiative to improve outcomes in acute kidney injury," (in eng), *Crit Care,* vol. 11, no. 2, p. R31, 2007.

[9] R. Bellomo, J. A. Kellum, and C. Ronco, "Acute kidney injury," (in eng), *Lancet,* vol. 380, no. 9843, pp. 756-66, Aug 2012.

[10] G. M. Chertow, E. Burdick, M. Honour, J. V. Bonventre, and D. W. Bates, "Acute kidney injury, mortality, length of stay, and costs in hospitalized patients," (in eng), *J Am Soc Nephrol,* vol. 16, no. 11, pp. 3365-70, Nov 2005.

[11] R. Bellomo, J. A. Kellum, and C. J. T. L. Ronco, "Acute kidney injury," vol. 380, no. 9843, pp. 756-766, 2012.

[12] R. J. Kate, R. M. Perez, D. Mazumdar, K. S. Pasupathy, and V. Nilakantan, "Prediction and detection models for acute kidney injury in hospitalized older adults," (in eng), *BMC Med Inform Decis Mak,* vol. 16, p. 39, Mar 2016.

[13] M. Bedford *et al.*, *Development of risk models for the prediction of new or worsening acute kidney injury on or during hospital admission: a cohort and nested study*. Southampton (UK): NIHR Journals Library, 2016.

[14] M. A. Perazella, "The urine sediment as a biomarker of kidney disease," (in eng), *Am J Kidney Dis,* vol. 66, no. 5, pp. 748-55, Nov 2015.

[15] K. F. Kerr, A. Meisner, H. Thiessen-Philbrook, S. G. Coca, and C. R. Parikh, "Developing risk prediction models for kidney injury and assessing incremental value for novel biomarkers," (in eng), *Clin J Am Soc Nephrol,* vol. 9, no. 8, pp. 1488-96, Aug 2014.

[16] W. H. Kim *et al.*, "Simplified clinical risk score to predict acute kidney injury after aortic surgery," (in eng), *J Cardiothorac Vasc Anesth,* vol. 27, no. 6, pp. 1158-66, Dec 2013.

[17] H. D. Kiers *et al.*, "Comparison and clinical suitability of eight prediction models for cardiac surgery-related acute kidney injury," (in eng), *Nephrol Dial Transplant,* vol. 28, no. 2, pp. 345-51, Feb 2013.

[18] T. Ploug and S. Holm, "Informed consent and routinisation," (in eng), *J Med Ethics,* vol. 39, no. 4, pp. 214-8, Apr 2013.

[19] L. Z. Zhou *et al.*, "Development and Validation of a Risk Score for Prediction of Acute Kidney Injury in Patients With Acute Decompensated Heart Failure: A Prospective Cohort Study in China," (in eng), *J Am Heart Assoc,* vol. 5, no. 11, Nov 2016.

[20] T. Wilson *et al.*, "Risk prediction models for acute kidney injury following major noncardiac surgery: systematic review," (in eng), *Nephrol Dial Transplant,* vol. 31, no. 2, pp. 231-40, Feb 2016.

[21] L. N. Sanchez-Pinto and R. G. Khemani, "Development of a Prediction Model of Early Acute Kidney Injury in Critically Ill Children Using Electronic Health Record Data," (in eng), *Pediatr Crit Care Med,* vol. 17, no. 6, pp. 508-15, Jun 2016.

[22] L. Zimmerman *et al.*, "Early Prediction of Acute Kidney Injury Following ICU Admission," *BMC Medical Informatics and Decision Making (accepted),* no. S1, pp. 5-16, 2019.

[23] Y. Li, L. Yao, C. Mao, A. Srivastava, X. Jiang, and Y. Luo, "Early Prediction of Acute Kidney Injury in Critical Care Setting Using Clinical Notes," in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2018.

[24] K. Colpaert *et al.*, "Impact of real-time electronic alerting of acute kidney injury on therapeutic intervention and progression of RIFLE class," (in eng), *Crit Care Med,* vol. 40, no. 4, pp. 1164-70, Apr 2012.

[25] A. E. Johnson *et al.*, "MIMIC-III, a freely accessible critical care database," (in eng), *Sci Data,* vol. 3, p. 160035, May 2016.

[26] A. S. Levey *et al.*, "A new equation to estimate glomerular filtration rate," (in eng), *Ann Intern Med,* vol. 150, no. 9, pp. 604-12, May 2009.

[27] L. A. Stevens, J. Coresh, T. Greene, and A. S. Levey, "Assessing kidney function--measured and estimated glomerular filtration rate," (in eng), *N Engl J Med,* vol. 354, no. 23, pp. 2473-83, Jun 2006.

[28] T. Joachims, "A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization," Carnegie-mellon univ pittsburgh pa dept of computer science1996.

[29] S. Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in R," *Journal of statistical software,* vol. 45, no. 3, 2011.

[30] S. van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate Imputation by Chained Equations in R," *2011,* vol. 45, no. 3, p. 67, 2011-12-12 2011.

[31] L. Yao, C. Mao, and Y. Luo, "Graph Convolutional Networks for Text Classification," in *AAAI*, 2019.

[32] Y. Luo, P. Szolovits, A. S. Dighe, and J. M. Baron, "Using Machine Learning to Predict Laboratory Test Results," (in eng), *Am J Clin Pathol,* vol. 145, no. 6, pp. 778-88, Jun 2016.

[33] Y. Luo, P. Szolovits, A. S. Dighe, and J. M. Baron, "3D-MICE: integration of cross-sectional and longitudinal imputation for multi-analyte longitudinal clinical data," *Journal of the American Medical Informatics Association (JAMIA),* vol. 25, no. 6, pp. 645-653, Nov 30 2017.

[34] A. R. Aronson, "Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program," in *Proceedings of the AMIA Symposium*, 2001, p. 17: American Medical Informatics Association.

[35] G. K. Savova *et al.*, "Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications," *Journal of the American Medical Informatics Association,* vol. 17, no. 5, pp. 507-513, 2010.

Address for Correspondence

Yuan Luo, email: yuan.luo@northwestern.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190246

A Privacy-Preserving Infrastructure for Analyzing Personal Health Data in a Vertically Partitioned Scenario

Chang Sun[a], Lianne Ippel[a], Johan van Soest[b], Birgit Wouters[a], Alexander Malic[a], Onaopepo Adekunle[a], Bob van den Berg[c], Ole Mussmann[c], Annemarie Koster[d], Carla van der Kallen[e], Claudia van Oppen[a], David Townend[f], Andre Dekker[b], Michel Dumontier[a]

[a] *Institute of Data Science, Maastricht University, Maastricht, The Netherlands,*
[b] *Department of Radiation Oncology (MAASTRO), GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre+, Maastricht, The Netherlands,*
[c] *Statistics Netherlands (Centraal Bureau voor de Statistiek), Heerlen, The Netherlands,*
[d] *Department of Social Medicine, CAPHRI Care and Public Health Research Institute, Maastricht University, The Netherlands,*
[e] *Department of Internal Medicine, CARIM School for Cardiovascular Diseases, Maastricht University, Maastricht, The Netherlands,*
[f] *Department of Health, Ethics and Society, CAPHRI Research School, Maastricht University, Maastricht, The Netherlands*

Abstract

It is widely anticipated that the use and analysis of health-related big data will enable further understanding and improvements in human health and wellbeing. Here, we propose an innovative infrastructure, which supports secure and privacy-preserving analysis of personal health data from multiple providers with different governance policies. Our objective is to use this infrastructure to explore the relation between Type 2 Diabetes Mellitus status and healthcare costs. Our approach involves the use of distributed machine learning to analyze vertically partitioned data from the Maastricht Study, a prospective population-based cohort study, and data from the official statistics agency of the Netherlands, Statistics Netherlands (Centraal Bureau voor de Statistiek; CBS). This project seeks an optimal solution accounting for scientific, technical, and ethical/legal challenges. We describe these challenges, our progress towards addressing them in a practical use case, and a simulation experiment.

Keywords:

Health Information Systems, Data Science, Machine Learning

Introduction

A growing amount of personal health data are being collected by a variety of entities, such as healthcare providers, insurance companies, and wearable device manufacturers. Use of personal health data such as health status, current and prior medications, lifestyle and behavior offers unprecedented opportunities to augment our understanding of human health and disease. This contributes to improved diagnostic accuracy and efficiency [1,2], and facilitates the transition to preventive [3,4] and precision medicine [5–7]. Moreover, the analysis of health data can help governments pursue effective health policies while minimizing healthcare costs. Such innovation arises from the secondary use of health data for research.

However, a major barrier to research lies in the difficulty of accessing and analyzing health data that are dispersed in both their form (e.g. medical records, consumer activity, and social media), representation (structured, semi-structured, and/or unstructured), and stewardship (who is responsible for data collection and governance?). While many methods to represent and exchange healthcare data have been developed [8], there has been a lack of focus on legal-ethical concerns such as data ownership and data stewardship as well as issues relating to privacy, security, and confidentiality [9]. Such considerations are particularly crucial when use and analysis of health data involve multiple legal entities, different data standards, a lack of detailed provenance, and unclear access authorization procedures.

Another significant challenge lies in the analysis of personal health data from multiple sources. The simplest case is where data are horizontally partitioned, such that data about different sets of individuals are located in different sites. Analyzing these distributed data is relatively well understood and reduces to combining a set of models from each site. A more challenging case is where data are vertically partitioned: different attributes about a particular individual are distributed over a set of data sources. While in the case of horizontally partitioned data analytical results are combined afterwards, this is not possible in the vertically partitioned case since none of the data providers can execute the complete analysis independently of the other providers. This is particularly challenging either when there is a legal impediment to link records across data providers with a unique identifier or when this unique identifier is unavailable. Addressing this challenge effectively requires a great level of technical sophistication to simultaneously address legal and/or privacy constraints.

Instead of centralizing the data for the analysis, one could use distributed learning methods, which operate over vertically partitioned data. In such a scenario, data-processing algorithms are sent to each site, and can only return the results of an analysis rather than any of the original data. One such infrastructure is the Personal Health Train (PHT) [10,11], which sends applications (the trains) containing algorithms to the data sources (the stations). The station can inspect whether the train is allowed to execute the application on (a subset of) the available data. The PHT empowers data subjects with more control (who can access the data?) and transparency (what are the trains requesting?). Hence, the PHT facilitates authorized algorithmic processing in a secure manner at multiple data sites without requiring a transfer of (original) data to a centralized location. Moreover, the PHT implements privacy-by-design in the following ways: 1) it can restrict which data elements are available to an application, 2) it can restrict the results of the analysis to only processed data, rather than original data, and 3) no data party can see the data of other parties in the network.

Here, we describe an implementation of the PHT that uses a Trusted Secure Environment (TSE) to analyze vertically

partitioned data that are prepared in line with the FAIR principles (Findable, Accessible, Interoperable, Reusable) [12]. By describing data using the FAIR principles, the infrastructure becomes ambivalent to certain syntactic data structures (e.g. OHDSI, CDISC-ODM or HL7 v2/v3/FHIR), as the applications, executed at the data source, should be able to interpret different types of data structures. To test the feasibility of this infrastructure, we combine data from two independent data providers to investigate how Type 2 Diabetes Mellitus (T2DM) status affects healthcare cost. The first dataset comes from the Maastricht Study[1], an observational prospective population-based cohort study focusing on the etiology of T2DM, and the second comes from the official statistics office in the Netherlands: Statistics Netherlands[1] (Centraal Bureau voor de Statistiek; CBS). We present preliminary results involving simulated data and discuss the challenges and feasibility of such an infrastructure to be scalable and secure.

Methods

In this section, we describe the development of our proposed infrastructure from a scientific, technical, and legal perspective to support the workflow. Following is the description of our simulation experiment to test the usability of our infrastructure.

Development Workflow

The PHT architecture has been previously used to analyze horizontally partitioned datasets [13–16]. Here, we extend this work to include vertically partitioned data. While several studies discuss exchanging and analyzing vertically partitioned data [17,18], these are largely theoretical and overlook practical challenges, e.g. legal and ethical considerations, incompatible data management standards, scalability of the infrastructure, lack of financial support to sustain such efforts, and the technical requirements of learning from vertically partitioned data. To tackle these challenges, our team has established three interlocking work packages that target: i) the scientific questions in the medical domain; ii) the ethical, legal, and societal issues; and iii) the technical aspect. These packages are highly intertwined to ensure the development of practical solutions.

Scientific Perspective

To develop infrastructure that is useful to scientific researchers, we have identified key research questions that the infrastructure should help answer. Answering these research questions should require the combination of sensitive (non-public) data from multiple providers. To combine data from multiple providers, a substantive set of individuals should be shared by the providers and at least some attributes of these individuals are present in both datasets to enable linking of the data records (and not necessarily by some specific individual identifier).

ELSI Perspective

The Ethical, Legal, and Societal Issues (ELSI) team deals with two types of challenges: i) privacy concerns that arise from the special nature of personal health data[3]; and ii) the legal challenges that arise from working with multiple data providers with each a distinct governance framework. Combining data from multiple parties is a relatively new phenomenon, and often not foreseen when establishing the legal framework when the data are collected. Therefore, one of the major challenges has been to facilitate this study whilst adhering to the original legal framework and defined purpose. In doing so, the ELSI team has examined the reach of the original legal basis (i.e. informed consent) and purpose for which each data provider obtained the personal data, and is further analyzing the legal basis and purpose for which secondary processing can occur. Options that are being considered include but are not limited to the route of compatible processing and the route of scientific research in the public interest. Additionally, there are a number of limitations from the data providers themselves regarding accessing, sharing, and linking data. In addition, for this challenge, a legal framework has to be formulated in order to establish collaborations between the data providers, among themselves and with the research team. Constructing this legal framework and finding the proper legal basis for the researchers' team is a valuable contribution from the ELSI team.

Technical Perspective

Following the PHT architecture[2], we use the concepts of (FAIR data) stations[3], rails (infrastructure) and (applications) trains. The minimal requirement of a FAIR data station is to enable execution of applications, where data providers decide whether to execute the application. These FAIR data stations are based on Semantic Web technologies such as the Resource Description Framework (RDF) [19], to convert the source data[4], and make the converted data FAIR.

Application (train) developers (i.e., researchers) can create the application trains using Docker containers [20], which are lightweight virtual machines. The Docker container carries all required software packages to execute the application on board. These applications can for instance query data available in the data station, perform data cleaning/formatting, and execute machine learning or statistical analysis [15]. Only the results of these (analytical) applications are sent back to the application developers.

To implement the proposed infrastructure, we created three stations. Two FAIR data stations are at the Maastricht Study and at CBS. A third station was configured as a "Trusted Secure Environment" (TSE), containing no data by itself, however, acting as a trusted and independent entity. Additionally, we created two application trains. The first application train extracts the data from two data stations, pseudonymizes the personal identifiers, encrypts the dataset, and sends the data to the TSE station. The second application train decrypts the data and analyzes the data at the TSE. For every execution, both application trains are configured for proper encryption and security measures.

Experiment Design

Prior to feeding our infrastructure with real data, we conducted a simulation experiment with two scenarios where researchers combine data from two independent providers using a TSE station. We monitor time to obtain the analytical results for each scenario. Scenario 1 consists of two providers, A and B, each having the same (small) number of individuals; Scenario 2 consists of providers A and B, but provider B has a much larger set of individuals, including all Provider A's individuals. For these scenarios, we use data from a publicly available dataset which contains attributes that could be interpreted as sex, body mass index (BMI), number of children, smoking status, region, and health insurance reimbursement of participants [21]. Additionally, we generated artificial personal identifiers including date of birth, zip code, house number, and sex for linking purpose [22]. In practice, combining multiple datasets might be prone to record-linking errors. We will discuss this in more detail in the Discussion section. Please find this synthetic

[1] Statistics Netherlands is a Dutch governmental institution that gathers statistical information about the Netherlands: https://www.cbs.nl/en-gb
[2] PHT architecture: https://bitbucket.org/jvsoest/pytaskmanager.git
[3] FAIR stations: http://github.com/maastroclinic/DataFAIRifier
[4] Convert CSV file to RDF file: https://github.com/sunchang0124/FAIRHealth/

dataset in Figshare[5]. This dataset is vertically split over the two providers: both have artificial personal identifiers (date of birth, zip code, house number, and sex). Only Provider A has BMI, number of children, and smoking status, while only Provider B has living region and health insurance reimbursement. In scenario 1, both providers have 1338 patients. In scenario 2, Provider A still has 1338 patients while Provider B hosts 64,400 patients. Since, Provider A in the second scenario only hosts a small subset of Provider B, a single record of Provider A might match with several records from Provider B. Even though this scenario is often encountered in practice, few solutions are available to address this linking challenge for vertically partitioned data [23].

For our experiment, we developed application trains using Docker 18.03.1. Pseudonymization, encryption, verification, and record linkage were implemented in Python 2.7. The infrastructure was tested with a 2.5GHz PC with 16GB RAM and 500 GB hard disk.

Result

In this section, we detail the contributions of each of the three work packages. Next, we discuss the outcome of the experiment.

Figures 1 and 2 provide an illustration of the infrastructure. In Figure 1, an overview of the operational framework for two providers, A and B, and a trusted secure environment, TSE, is presented. In Figure 2, we present the technical and legal requirements of the FAIR data stations. Researchers request permission to access and process data from the data provider. Once permission is granted, application trains to pseudonymize and encrypt the data are sent and executed in the data stations. Next, the encrypted data are sent to the TSE, followed by the data analysis application (from the researchers).

In the FAIR data stations (Figure 2), personal identifiers are pseudonymized by one-way hashing and salting techniques. One-way hashing turns any format of data into a fixed-length "fingerprint" that cannot be reversed. Salt, as a random string, is appended to data before hashing, to eliminate the risk of malicious decryption. We used Secure Hash Algorithm 2 (SHA-512) as the one-way hashing function and random salts are shared by two data providers to make personal identifiers pseudonymized on both sites. This results in a unique code per record, allowing linking the same records from all data providers. Every time data providers grant researchers permission to process/analyze the data, the personal identifiers get pseudonymized using different salts. The salt needs to be created and agreed upon by all data providers. Additionally, to safeguard secure transfer, processed data are encrypted, prior to sending them to TSE. The same as with the salt, encryption keys are re-generated every time.

The procedure then continues as follows: when the encrypted data are sent to the TSE, a notification is generated by the data stations to confirm the successful execution and departure of the train. After all encrypted data arrive at the TSE station, the researchers trigger analysis at the TSE with a set of keys and an application that includes code for the analysis. There is one private key per data station to decrypt the dataset, and one verification key to test the dataset integrity. The data station can only encrypt using the public key but cannot decrypt. The TSE station maintains the private key to decrypt for this specific data provider. After getting verified and decrypted data from both providers, the data can be linked and merged by pseudonymized personal identifiers. As the salted hashes performed at the data station are unknown to the TSE, it is not able to reverse or decrypt sensitive data such as personal identifiers. Thus, in addition to pseudonymization and encryption, the privacy of information is further protected as no data provider has direct access to the TSE. After executing the analytical algorithms on the merged dataset, the TSE checks whether the results reveal any personal identifiable information. Only the validated results such as figures and/or tables that do not contain any personal identifiable information are returned to the researchers. Finally, all (received and created) data in the TSE are destroyed.

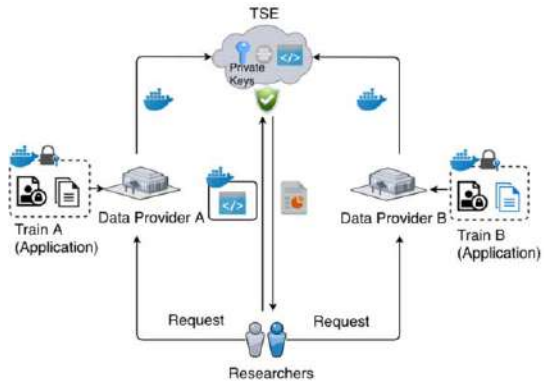


Figure 1 - Conceptual overview of the proposed infrastructure. Data access is regulated by the data provider hosting the stations. If access is granted, the data providers encrypt the data and send these to the TSE. The TSE executes the researchers' application and allows aggregated results to be returned to the researcher.

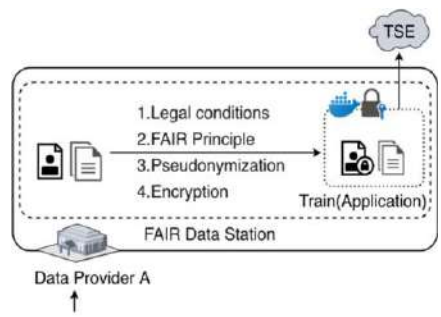


Figure 2 - Overview of data stations and application trains. Within each station, data are prepared, i.e., legal conditions are checked, FAIR principles implemented, personal identifiers pseudonymized, and encrypted. The application train enters the data station with algorithms and leaves with results or processed data.

Simulation Experiment

We used our proposed infrastructure to analyze synthetic data (discussed in the Methods section) that was vertically partitioned to form two datasets, each with a different data provider. Figure 3 shows one such result: a plot of BMI and health insurance reimbursement over one calendar year. While simple, the simulation experiment provides evidence for the feasibility of the infrastructure to execute an analysis, in this case, retrieval of a relation between two attributes in separate datasets in a secure and privacy-preserving manner.

[5] Find our synthetic datasets: https://doi.org/10.6084/m9.figshare.7379810.v2.

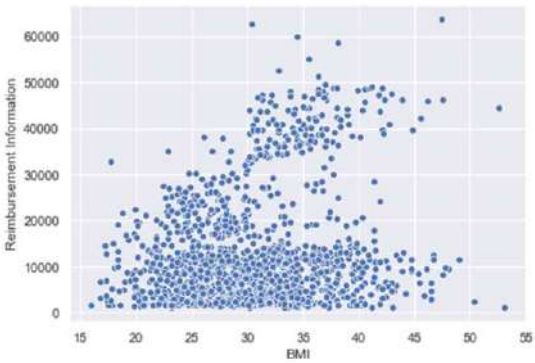


Figure 3. Plot of body mass index (BMI) versus health insurance reimbursement in the past year (dollars) from the analysis of a synthetic and vertically partitioned dataset using the proposed infrastructure.

We conducted an experiment with two scenarios. In the first scenario, where both providers host 1338 individuals, pseudonymization took 0.4 - 0.5 seconds and encryption took 0.1 - 0.2 seconds for each data station. At the TSE, verification and decryption spent merely 0.1 - 0.2 seconds, while record linkage took around 7.2 seconds. For the second scenario, where provider A hosts 1338 individuals and Provider B 64,400 individuals, pseudonymization for Provider B took 7.3 seconds and 2.5 seconds to encrypt. The total time cost at the TSE increased to about 15 seconds. From our experiment, we found that pseudonymization (at data stations) and record linkage (at the TSE) consumed approximately 80% of the running time. Future work will focus on operational performance measures, and among others, the size of provider datasets and number of attributes considered in linking.

Discussion

We have described and demonstrated a distributed learning infrastructure using artificial and vertically partitioned data involving two providers and a trusted secure environment. This is a preliminary, but promising result.

Our long-term goal is to deploy the infrastructure to analyze actual data from two independent organizations - Statistics Netherlands and the Maastricht Study. Thus far, we have requested data for 3451 consenting participants from the Maastricht Study, which is characterized by extensive phenotyping and provides information on the etiology, pathophysiology, complications, and comorbidities of T2DM. All participants are aged between 40 and 75 years and live in the southern part of The Netherlands. We requested those attributes which were complete and consented. Attributes include socio-demographic factors, lifestyle factors, the status of T2DM, physical function, mental functions, BMI, and cardiovascular disease history. From CBS, we requested regional population data of health insurance reimbursement available at Statistics Netherlands. As of November 2018, all application trains have been developed. We are in the stage of approving and building data stations for the Maastricht Study and CBS. A joint controllership agreement between the two organizations is established to enable the TSE. We are preparing analytic algorithms that will 1) answer scientific questions regarding the associations between T2DM status and healthcare costs, and 2) to evaluate the performance and security of our infrastructure.

Applying the infrastructure to real-world situations will present several challenges. Although we have only explored a two-data-provider scenario, we anticipate that it can be extended to more than two providers. While the performance of this system will depend on the size of the data and the algorithms used for encryption, merging, and analysis, we believe that the biggest bottleneck is in creating consortium agreements and deploying the infrastructure in individual facilities. As such, there is a need to further develop a "PHT" deployment kit that enables stakeholders to consider all the issues and options and make informed decisions in the most efficient manner. A second challenge in our implementation lies in the possibility of errors caused by linking vertically partitioned datasets. The accuracy of matching across these will decrease owing to missing data, typographical errors, differences in pseudonymization procedures, and different formats of identifying information. In addition, to match a fraction of records from multiple large datasets, the data providers could limit the size of their data by sending only a selection to TSE. This selection can be discovered and defined by sending exploratory or individual selection algorithms first. For instance, in our case, instead of sending the information of the entire Dutch population to the TSE, only a subset of the Dutch population which meets the criteria of the Maastricht Study sample is sent to the TSE. However, note that this selection might also leak information about the individuals in the data of (one or more) data providers. We intend to explore the impact of such aspects in future studies. A third challenge is how to manage and transport the keys securely among different parties. The TSE requires decryption and verification keys to decrypt the data and run the analysis algorithms. This approach must be agreed on by all parties from both technical and ethical-legal perspectives.

Conclusions

To analyze vertically partitioned data, we extended a Personal Health Train (PHT) infrastructure to send data analysis algorithms to multiple data stations and return only the results instead of the original data to the researchers.

This infrastructure was developed in a coordinated manner across multiple scientific, technical, ethical, legal, and societal aspects involving several units and organizations. This coordination across interests is essential to explore viable solutions for data sharing and reuse, as envisioned by the proponents of the FAIR principles. In particular, the idea of bringing the algorithm to the data, rather than obtaining consent to receive a copy of the data, offers an entirely new paradigm that has not been considered by most organizations. Having a new paradigm will require stakeholders to take the time and effort to thoroughly evaluate this in terms of their legal and technical requirements. However, as our experiment shows, it offers a more scalable and secure solution to analyze vertically partitioned data in a secure and privacy-preserving manner. Additional operational and security enhancements are still needed before the infrastructure is suited to deal with real (sensitive) data. Future work will explore the quality of scientific discovery (accuracy of outcome), the security, scalability, sustainability, and performance of computation. While no solution will be perfect for all situations, we believe that this adaptation of the PHT model will find utility in situations involving sensitive data with a multitude of stakeholders.

Acknowledgements

This project is funded by the Dutch National Research Agenda (NWA; project number: 400.17.605) and the provincial funding for the Limburg Meet (LIME) project. We thank our partners Statistics Netherlands (CBS) and the Maastricht Study (DMS) for contributing knowledge and support.

References

[1] G. Dougherty, Digital Image Processing for Medical Applications, (n.d.) 485.

[2] H. Elshazly, A.T. Azar, A. El-korany, and A.E. Hassanien, Hybrid system for lymphatic diseases diagnosis, in 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, Mysore, 2013: pp. 343–347. doi:10.1109/ICACCI.2013.6637195.

[3] E.A. Clarke, What is Preventive Medicine?, *Can Fam Physician*. **20** (1974) 65–68.

[4] I. Barbier-Feraud, J.B. Malafosse, P. Bouexel, C. Commaille-Chapus, A. Gimalac, G. Jeannerod, M. Léo, B. Leroy, B. Nordlinger, M. Paoli, I. Assistance, P. Prados, and J.-Y. Robin, Big data and prevention from prediction to demonstration, (n.d.) 80.

[5] R.B. Stricker, and L. Johnson, Lyme disease: the promise of Big Data, companion diagnostics and precision medicine, *Infect Drug Resist*. **9** (2016) 215–219. doi:10.2147/IDR.S114770.

[6] J.S. Beckmann, and D. Lew, Reconciling evidence-based medicine and precision medicine in the era of big data: challenges and opportunities, *Genome Medicine*. **8** (2016). doi:10.1186/s13073-016-0388-7.

[7] S. Dolley, Big Data's Role in Precision Public Health, *Front Public Health*. **6** (2018). doi:10.3389/fpubh.2018.00068.

[8] E.H. Shortliffe, and J.J. Cimino, eds., Biomedical Informatics: Computer Applications in Health Care and Biomedicine, 4th ed., Springer-Verlag, London, 2014. //www.springer.com/us/book/9781447144731 [accessed November 8, 2018].

[9] M.J. Bietz, C.S. Bloss, S. Calvert, J.G. Godino, J. Gregory, M.P. Claffey, J. Sheehan, and K. Patrick, Opportunities and challenges in the use of personal health data for health research, *Journal of the American Medical Informatics Association*. **23** (2016) e42–e48. doi:10.1093/jamia/ocv118.

[10] van S. Johan, S. Chang, M. Ole, P. Marco, van den B. Bob, M. Alexander, van O. Claudia, T. David, D. Andre, and D. Michel, Using the Personal Health Train for Automated and Privacy-Preserving Analytics on Vertically Partitioned Data, *Studies in Health Technology and Informatics*. (2018) 581–585. doi:10.3233/978-1-61499-852-5-581.

[11] Personal Health Train, *Dutch Techcentre for Life Sciences*. (n.d.). https://www.dtls.nl/fair-data/personal-health-train/ [accessed October 22, 2018].

[12] M.D. Wilkinson, M. Dumontier, Ij.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data*. **3** (2016) 160018.

[13] J.P.A. van Soest, A.L.A.J. Dekker, E. Roelofs, and G. Nalbantov, Application of Machine Learning for Multicenter Learning, in: I. El Naqa, R. Li, and M.J. Murphy (Eds.), Machine Learning in Radiation Oncology, Springer International Publishing, Cham, 2015: pp. 71–97. doi:10.1007/978-3-319-18305-3_6.

[14] A. Damiani, M. Vallati, R. Gatta, N. Dinapoli, A. Jochems, T. Deist, J. van Soest, A. Dekker, and V. Valentini, Distributed Learning to Protect Privacy in Multi-centric Clinical Studies, in: J.H. Holmes, R. Bellazzi, L. Sacchi, and N. Peek (Eds.), Artificial Intelligence in Medicine, Springer International Publishing, 2015: pp. 65–75. http://link.springer.com/chapter/10.1007/978-3-319-19551-3_8 [accessed June 25, 2015].

[15] T.M. Deist, A. Jochems, J. van Soest, G. Nalbantov, C. Oberije, S. Walsh, M. Eble, P. Bulens, P. Coucke, W. Dries, A. Dekker, and P. Lambin, Infrastructure and distributed learning methodology for privacy-preserving multi-centric rapid learning health care: euroCAT, *Clinical and Translational Radiation Oncology*. **4** (2017) 24–31. doi:10.1016/j.ctro.2016.12.004.

[16] A. Jochems, T.M. Deist, J. van Soest, M. Eble, P. Bulens, P. Coucke, W. Dries, P. Lambin, and A. Dekker, Distributed learning: Developing a predictive model based on data from multiple hospitals without data leaving the hospital – A real life proof of concept, *Radiotherapy and Oncology*. **121** (2016) 459–467. doi:10.1016/j.radonc.2016.10.002.

[17] J. Vaidya, A Survey of Privacy-Preserving Methods Across Vertically Partitioned Data, in: C.C. Aggarwal, and P.S. Yu (Eds.), Privacy-Preserving Data Mining, Springer US, Boston, MA, 2008: pp. 337–358. doi:10.1007/978-0-387-70992-5_14.

[18] T.Z. Gál, G. Kovács, and Z.T. Kardkovács, Survey on privacy preserving data mining techniques in health care databases, *Acta Universitatis Sapientiae, Informatica*. **6** (2014) 33–55. doi:10.2478/ausi-2014-0017.

[19] RDF 1.1 Concepts and Abstract Syntax, (n.d.). https://www.w3.org/TR/rdf11-concepts/ [accessed March 22, 2019].

[20] What is a Container, *Docker*. (2018). https://www.docker.com/resources/what-container [accessed October 22, 2018].

[21] B. Lantz, Machine learning with R: learn how to use R to apply powerful machine learning methods and gain an insight into real-world applications, Packt Publ, Birmingham, 2013.

[22] Welcome to Faker's documentation! — Faker 1.0.2 documentation, https://faker.readthedocs.io/en/master/ [accessed March 22, 2019].

[23] S. Hardy, W. Henecka, H. Ivey-Law, R. Nock, G. Patrini, G. Smith, and B. Thorne, Private federated learning on vertically partitioned data via entity resolution and additively homomorphic encryption, *ArXiv:1711.10677 [Cs]*. (2017). http://arxiv.org/abs/1711.10677 [accessed September 18, 2018].

Address for Correspondence

Chang Sun, Universiteitssingel 60, 6229ER Maastricht, chang.sun@maastrichtuniversity.nl, +31 6 36373387.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190247

Validating Auto-Suggested Changes for SNOMED CT in Non-Lattice Subgraphs Using Relational Machine Learning

Qi Sun[a,b], Guo-Qiang Zhang[b,d,e], Wei Zhu[b,c], Licong Cui[d]

[a] *Department of Computer Science, University of Kentucky, Lexington, KY, USA*
[b] *Institute for Biomedical Informatics, University of Kentucky, Lexington, KY, USA*
[c] *Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH, USA*
[d] *School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX, USA*
[e] *McGovern School of Medicine, University of Texas Health Science Center at Houston, Houston, TX, USA*

Abstract

An attractive feature of non-lattice-based ontology auditing methods is its ability to not only identify potential quality issues, but also automatically generate the corresponding fixes. However, exhaustive manual evaluation of the validity of suggested changes remains a challenge shared with virtually all auditing methods. To address this challenge, we explore machine learning techniques as an aid to systematically evaluate the strength of auto-suggested relational changes in the context of existing knowledge embedded in an ontology. We introduce a hybrid convolutional neural network and multilayer perception (CNN-MLP) classifier using a combination of graph, concept features and concept embeddings. We use lattice subgraphs to generate a curated, loosely-coupled training set of positive and negative instances to train the classifier. Our result shows that machine learning techniques have the potential to alleviate the manual effort required for validating and confirming changes generated by non-lattice-based auditing methods for SNOMED CT.

Keywords:

Non-lattice-based auditing, SNOMED CT, neural networks

Introduction

Non-lattice-based ontology auditing methods are based on the principle of Formal Concept Analysis. They have shown effectiveness in suggesting missing hierarchical relations (or IS-A relations) and concepts in SNOMED CT (SNCT) [1-3] as well as other teminology systems. A non-lattice-based approach consists of the following steps: extracting non-lattice subgraphs with concept pairs voilating the lattice property; detecting defects in the extracted non-lattice subgraphs; suggesting relational changes automatically; and reviewing and validating suggested changes by an ontology curator or a domain expert. Figure 1 shows a non-lattice subgraph in the September 2017 release of SNOMED CT (US edition) [3]. Examination of this subgraph reveals a missing hierarchical relation between nodes 4 and 5: "*Transient neonatal hyperglycemia*" IS-A "*Acute hyperglycemia.*"

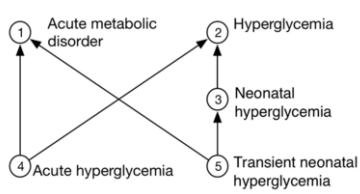


Figure 1 - An example of non-lattice subgraph [3].

However, human review of change remediations requires significant manual effort. The main motivation of this paper is to explore machine learning (ML) techniques as an aid to systematically evaluate the strength of auto-suggested relational changes using existing knowledge embedded in an ontology. Specifically, change predictions resulting from the IS-A relation prediction classifier can be compared with change suggestions using non-lattice subgraphs by assessing their agreement. When such an approach is used to audit the IS-A relations, it can bring benefits in two ways: the agreement between the ML prediction and non-lattice-based prediction can provide independent confirmation of the suggested IS-A changes, while disagreements may point to areas for improvements for both types of approaches. Furthermore, change suggestions generated by ML techniques may provide ranked list of changes based on numeric "strengths." If an ML technique can reach a high level of performance (using e.g. precision and recall measures), it may serve as an independent auditing approach by itself independent of non-lattices.

ML techniques have been widely used in knowledge graphs [4]. To arrive at an optimal subset of features, CNN has been used to automatically learn features [5]. Also pretrained word embeddings [6] have proven to be useful in neural network models for relation extraction. Recent study further improved performance by incorperating additional hand-crafted features [7]. Inspired by this, we propose a hybrid CNN-MLP classifier for IS-A relation prediction by exploring various knowledge in SNCT. Besides of concept embeddings, we analyze lexical and graph structural information in the entire directed acyclic graph (DAG) of SNCT, such as semantic meaning inheriting concept ancestors. Pesquita et al. [8] summerized two main semantic similarity approaches in an ontology graph for comparing concepts: node-based and edge-based. In this work, we aggregate node-based and edge-based similarity approaches as graph level features. Despite concept embeddings can present semantic meaning of concepts, we still handcraft concept level features as task specific features. Non-lattice-based auditing methods suggest IS-A relational changes to correct defects in non-lattice subgraphs and form lattice subgraphs [1-3]. Accordingly, constructing training set from lattice subgraphs equips us with the same objective of non-lattice-based methods.

In this paper, we introduce an ML approach to validate auto-suggested insertion changes from non-lattice-based auditing methods. Combining various knowledge in lattice subgraphs and DAG of SNCT, we implement IS-A relation prediction and auto-insertion change suggestion as two subtasks in the workflow of validation. We apply CNN to explore the concept pair features in concept embeddings and combine the output of CNN with handcrafted graph, concept features via MLP to

predict IS-A relations of concept pairs. Also evaluation metrics for subtasks are provided. Evaluation of our validation method is implemented on a reference set.

Methods

Our method consists of two main steps. In the first step, learning from various knowledge embedded in lattice subgraphs of SNCT, a hybrid CNN-MLP classifier is designed to predict IS-A relations of concept pairs. In the second step, based on the idea of majority voting [10], we validate the given insertion changes with predicted insertion changes generated by the classifier.

Relation Prediction

Figure 2 is an overview of our relation prediction workflow. First, we generate a training set using lattice subgraphs. Second, we pretrain concept embedding and develop graph and concept level features. Finally, we predict IS-A relations of concept pairs with a hybrid CNN-MLP classifier.

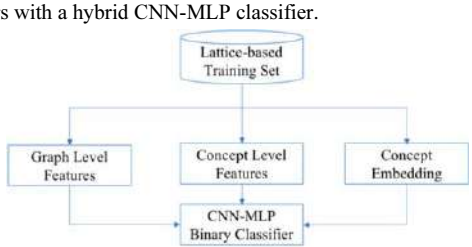


Figure 2-An overview of the IS-A relation prediction workflow

Lattice-based Training Set

A desirable property of the IS-A relationship is that it conforms to the lattice property [9]. This is also a by-product of non-lattice-based auditing methods: non-lattice subgraphs often become lattice-conforming by correcting hierarchical defects (e.g., inserting IS-A relations). Using this as working principle, we construct the training set for our CNN-MLP classifier using relations embedded in lattice subgraphs. Lattice subgraphs also allow us to naturally select a balanced set of positive samples and negative samples. Extracting all lattice subgraphs in SNCT, we construct a lattice-based training set with the following steps.

1. Select a subset of concepts with label length, which is the number of words in a label, no longer than α. We denote this set of concepts as $Con(\alpha)$.
2. Among all lattice subgraphs, choose those with size, which is the number of concepts in a subgraph, no more than β, denoted as $Lattice(\beta)$, and find those that contain any concept in $Con(\alpha)$.
3. Form a positive training set by extracting IS-A edges and their transitive closure from step 2. Form a negative training set by extracting non-edge node pairs in step 2.

Concept Embedding

Learning from current ontologies, the embedding techniques [6] allow us to represent concepts and capture latent semantic properties of concepts. In SNCT, each concept contains an associated human readable label, and concepts occuring in the same relation could describe each other. We take relations represented with labels as sentences to learn word embeddings. After perfoming a few basic text preprocessing steps, such as removing punctuations and digits, converting to lowercase and stemming, we use a skip gram word2vec model [6] to produce word embeddings with 200 dimensions. To generate concept embeddings, a sequence of words in concept labels are transformed to a set of vectors by looking up pretrained word embeddings. The out-of-vocabulory (OOV) word is assigned a vector with zeros. Then the word embeddings are summed into a single embedding as a concept embedding.

Feature Development

The IS-A relation comes with a source concept C_1 and a destination concept C_2. We develop features based on two types of observations. One is that from global graph level of entire IS-A DAG in SNCT, a concept at a lower level inherits semantics from its ancestors and is more specific in terms of the biomedical meaning [11]. The other is that measured from local level of concept pairs, if there are overlaps in noun phrases of C_1 and C_2, the noun phrase in C_1 usually contains more words than C_2 as C_1 becomes more specific. The more overlaping words the two concepts have, the more likely they should form an IS-A relation.

Graph level features. The work of Zhang et al. [11] and Cui et al. [3] provides several heuristics to generate graph level features. Considering concept semantics inherited from its ancestors and path information, we design three graph level features for situations of a concept pair to be in an IS-A relation. All graph level features are explored from the entire active IS-A graph of SNCT.

1. Graph level dissimilarity. The semantic of a concept could be specified through a path to its descendants, that is, a concept inherits all semantics from its ancestors. A pair of concepts, inheriting semantics from different paths, may have some common ancestors. Non common ancestors of C_1 distinguished C_1 from C_2, then the graph-level dissimilarity can be characterized by
$$GraphDissim(C_1,C_2)=1\text{-}CA(C_1,C_2)/A(C_1)$$
where $CA(C_1,C_2)$ is the common ancestor set of C_1 and C_2, and $A(C_1)$ is the ancestor set of C_1.
2. Graph level similarity. The semantic of a concept could be simplified with a concept label. While a given concept inherits semantics from its ancestors, the concept labels of ancestors could be used to enrich the label of the concept. For an IS-A relation (say C_1 IS-A C_2), because C_1 is more specific than C_2, the enriched label of C_2 maybe a part of enriched label of C_1, so we define graph level similarity as
$$GraphSim(C_1,C_2)=CEL(C_1,C_2)/EL(C_2)$$
where $CEL(C_1, C_2)$ is the common enriched label set of C_1 and C_2, and $EL(C_2)$ is the enriched label set of C_2.
3. Path comparison. If C_1 IS-A C_2, we assume that C_2 has a shorter path to the root R than C_1 does. We define a flag feature indicating whether the shortest path from C_2 to R is shorter than that of C_1 to R.

Concept level features. This category of features concerns about properties of labels in concept pairs.

1. Chunk comparison. We observe that if there are overlaps between noun phrases in C_1 and C_2, the noun phrase in C_1 usually has more words than that in C_2. We design two flag features from all the extracted noun phrases in both C_1 and C_2 to indicate word number comparison of noun phrases in C_1 and C_2 separately.
2. Concept level overlap. This feature considers overlaps of pairwise concepts. We define it as:
$$ConceptOLP(C_1,C_2)=W(C_1)\cap W(C_2)/W(C_2)$$
where $W(C_1)\cap W(C_2)$ is the overlapped word set of C_1 and C_2, and $W(C_2)$ is the word set of C_2. High value lesser than one implies that the meaning of C_1 approaches the meaning of C_2.
3. Concept level similarity. This feature concerns about the consine similarity of disjoint words in C_1 and C_2, and

disjoint words are represented with the summation of word embeddings.

Neural Network Classification

We design a hybrid CNN-MLP binary classifier to incorporate concept embedding and handcrafted features (see Figure 3).

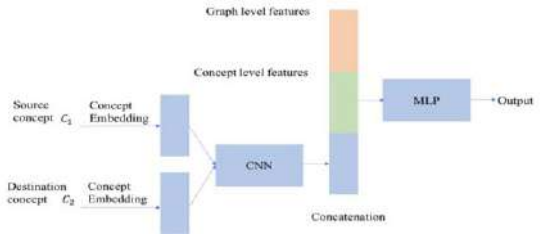


Figure 3 - Architecture of hybrid CNN-MLP binary classifier

We convolute the embeddings of concept pairs with 100 filters of window size 2 to get local features and use ReLU activation function. In order to fomulate a balanced feature vector between embeding features and handcrafted features, we transfer 100 dimension vector into 10 dimension vector by fully connected layer in CNN, which then is concatenated with graph level and concept level features. The concatenated feature vector is fed into a one layer MLP with active function of ReLU for the hidden layer and sigmoid for the output layer. In order to evaluate the performance of features, we feed MLP with different feature combinations, thus the number of neurons of MLP hidden layer is decided by the average of neurons in MLP input and output layers. Time complexity of neural network depends on the architechture, especially the number of neurons, of the network.

Validation of Auto-Suggested Insertion Changes

The objective of our work is using an ML approach to validate auto-suggested insertion changes. The validation workflow involves four stages: test set generation, relation prediction, auto-insertion change suggestion, and majority voting based validation. In stage 1, non-lattice-based auditing method [3] provides a set of non-lattice subgraphs containing auto-suggested insertion changes. IS-A relations and their transitive closure (positive instances), and non-edge concept pairs (negative instances) are extracted from the given non-lattice subgraphs to form a test set. In stage 2, the IS-A relations of extracted positive instances and negative instances are predicted by our hybrid CNN-MLP binary classifier. In stage 3, we retrieve top N predictions ranked by IS-A relation probability outputs. The predicted false positives are the set of auto-suggested insertion changes provided by our classifier. In stage 4, inspired by majority voting [10], we confirm the auto-suggested changes with the agreement between our classifier and the given auditing method. In this way, relational ML techniques alleviate the manual validation effort by confirming and validating changes automatically.

In the validation workflow, our classifier is used in two subtasks: predicting IS-A relation of concept pairs in stage 2, and automatically suggesting insertion changes in non-lattice subgraphs in stage 3. To evaluate performance of subtasks and determine configuration of the classifier for the validation task, we specify a test set, a reference set and evaluation metrics next.

Test Set and Reference Set

To evaluate our method, we use relations in a random sample of 200 non-lattice subgraphs from [3] as a test set, which contains 1,545 IS-A and transitive closure of IS-A edges (positive instances), 3,019 non-edge node pairs (negative instances). A total of 223 insertion changes were auto-suggested for the 200 non-lattice subgraphs [3]. Two domain experts confirmed 185 suggested insertions (a precision of 82.96%), which serve as a reference set for evaluating the performance of this work.

Evaluation of Relation Prediction

To thoroughly analyze the relation prediction part along, e.g., comparing training set parameters (α,β) and different categories of features, we predict 1,545 positive instances and 3,019 negative instances in the test set. For evalution, we report the precision-recall curve.

Evaluation of Auto-Suggested Insertion Changes

Motivated by precision and recall metrics, Zhang et al. introduced insertion recall, insertion precision and insertion G-measure [12] to evaluate an ontology quality assurance (OQA) method against validated changes. They considered an OQA method M as a group of subgraphs. Each xsubgraph may potentially capture missing IS-A relations as edges. A subgraph s is a graph consisting of a set of IS-A relations. E is a reference set of validated missing IS-A relations. Then the insertion recall, precision and G-measure are defined as:

$$R_{insert}=\frac{|\{r\in E/\exists s\in M, r\in s\}|}{|E|}$$

$$P_{insert}=\frac{|\{s\in M/\exists r\in s, r\in E\}|}{|M|}$$

$$G_{insert}=\sqrt{P_{insert}\times R_{insert}}$$

Our classifier outputs probabilities which can be used to rank instances in the test set from most probable IS-A relation to the least. In order to demonstrate the ability to validate auto-suggested changes, we treat our method as an OQA method and take top N predictions as predicted IS-A relations. False positives predicted by our classifier are auto-suggested insertion changes. The reference set contains 185 validated missing IS-A edges and their transitive closure. We use R_{insert}, P_{insert}, G_{insert} [12] to evaluate auto-suggested insertion changes obtained via our method. The higher the G_{insert}, the greater the agreement between our method and domain expert's validation, that is, more likely our method has the ability to alleviate the manual effort required for validating auto-suggested changes.

Results

Dataset and Implementaion

We used the September 2017 version of SNOMED CT (US edition) in this work. We constructed a lattice-based training set [1] by varying label length α and lattice size β. The statistics of training set are shown in Table 1. While relation prediction performance was evaluated with the test set, the reference set was used to evaluate auto-insertion change suggestion. We implemented the model using Keras [13]. Declaring binary cross entropy as the loss function, we ran 10 epochs for all the training examples. Noun phrase chunk detection and preprocessing for learning embeddings were implemented with NLTK [14].

Table 1 - Statistics of Training Set

	# of positives	# of negatives
$\alpha=5,\beta=5$	151,553	243,197
$\alpha=5,\beta=10$	515,954	1,005,074
$\alpha=10,\beta=5$	197,630	319,286
$\alpha=10,\beta=10$	693,776	1,351,872

Relation Prediction

Experiments were performed to evaluate IS-A relation prediction by permuting training set construction parameters and various feature categories. The experiments were performed on a MacBook Pro running MacOS Sierra, with 16 GB RAM and 2.7 GHz Intel Core i7 processor.

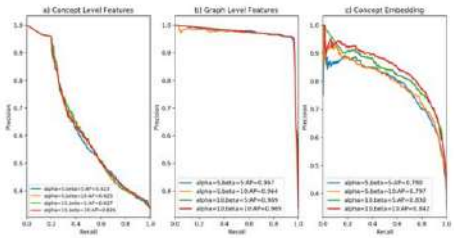


Figure 4 - The performance of relation prediction varying feature categories and training set construction parameters

Comparing Figure 4a), 4b) and 4c), the best prediction performance was produced with graph level features. When recall of IS-A relations was increased, the precision was decreased with concept level features. Though concept embeddings and concept level features both describe sementics of concept pairs, concept embeddings worked better than handcrafted concept level features.

Figure 4a) shows that longer concept label expresses concept level features more effectively, slightly leading to prediction performance improvement. Figure 4b) shows that graph level features are independent with label length α and lattice size β which is coincident with development of graph features, especially while increasing recall. As for the feature category of concept embeddings, prediction performance varied with different settings of training set parameters, which may be explained by Table 2 by OOV word rate over test set while looking up pretrained word embeddings. More OOV words in test set resulted in decrease of relation prediction performance.

Table 2 - Out-of-vocabulary word rate over test set

	$\alpha = 5$ $\beta = 5$	$\alpha = 5$ $\beta = 10$	$\alpha = 10$ $\beta = 5$	$\alpha = 10$ $\beta = 10$
OOV rate	4.4%	3.7%	3.0%	2.6%

Figure 5 shows that combining graph level features, concept level features and concept embeddings as the input of MLP had led to performance improvement. The best average precision was 0.972, achieved with $\alpha=5$, $\beta=5$. Overall, our experiments showed that exploring various features was effective for IS-A relation prediction.

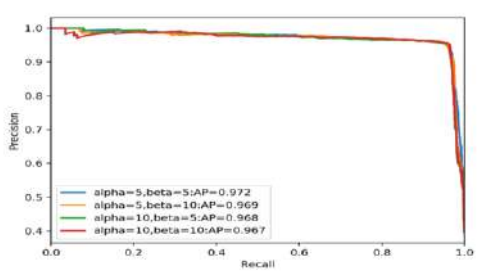


Figure 5-The performance of relation prediction with feature combination varying training set construction parameters

Auto-Suggested Insertion Changes

In this set of experiments, we evaluated our auto-suggested insertion changes for non-lattice subgraphs. A set of 185 validated missing IS-A relations was taken as the reference set. Among top N IS-A relation predictions, we evaluated auto-suggested insertion changes by R_{insert}, P_{insert}, G_{insert} values as defined in the subsection of auto-suggested insertion changes evaluation. Informed by the previous experiments, we configurated this set of experiments with feature combination and setting $\alpha=5$, $\beta=5$.

The test set contained 1,545 positive instances and 3,019 negative instances. Ranked by probability outputs of our classifier, we kept top N predictions as IS-A relations (predicted positive instances). Predicted false positives constituted insertion change suggestions in the non-lattice subgraphs. The top 1,500 predictions contained 129 IS-A relations already identified in the reference set of 185 validated missing IS-A relations. As we retrieve more predicted IS-A relations from the ranked list (i.e., as N increases), we obtain more validated missing IS-A relations. Note that both insertion recall (R_{insert}) and insertion precision (P_{insert}) increase as N increases. We stopped retrieving predictions at N=2,000 in case more irrelevant IS-A relations out of the referece set would be included.

Table 3 - Evaluation of auto-suggested insertion changes at top N predictions

Top N	R_{insert}(%)	P_{insert}(%)	G_{insert}(%)
1500	69.57	59.80	64.50
1600	80.98	69.35	74.94
1700	85.32	72.36	78.57
1800	86.41	72.86	79.35
1900	89.67	75.38	82.22
2000	90.21	75.88	82.74

Compared with the reference set, Table 4 shows examples of auto-suggested insertion changes from top 2,000 predictions in and out of the reference set. For those out of the reference set, they may reveal additional insertation changes needed to make non-lattice subgraphs conforming to the lattice property.

Table 4-Examples of auto-suggested insertion changes compared with the reference set

In the reference set

Source Concept	Destination concept
Thoracic spondylosis	Spondylosis
Sclerema neonatorum	Neonatal dermatosis
Acute empyema of sphenoidal sinus	Sphenoidal sinusitis
Oculocutaneous albinism	Congenital anomaly of eye
Degloving injury of genitalia	Degloving injury of perineum

Out of the reference set

Source concept	Destination concept
Echography scan B-mode for fetal growth rate	Ultrasound scan of fetus
Dilatation of anastomosis of bile duct	Biliary dilatation procedure
Feeding problems in newborn	Feeding problem
Physiological mobilization of the shoulder	Procedure on shoulder region
Carcinoma in situ of lower labial mucosa	Tumor of lower labial mucosa

The evaluation results and examples of auto-suggested insertion changes indicated that our method is effective to predict IS-A relations of concept pairs. Verified by the reference set, our method automatically suggested insertion changes for non-lattice subgraphs.

Validation of Auto-Suggested Insertion Changes

The performances of two subtasks in the workflow of validation, relation prediction and auto-insertion change suggestion, have been demonstrated to be promising by two sets of experiments. Configured with feature combination and $\alpha=5$, $\beta=5$ for the classifier, we evaluated 223 auto-suggested insertion changes in the test set. We compared insertion

changes resulting from our classifier with the given 223 insertion changes, and those changes with agreement were validated. The number of insertion changes validated by our classifier is shown in Table 5. Since 185 insertion changes were validated by domain experts, we also showed the number of expert-validated insertion changes among our classifier's validation. Validated by our method, the auto-suggested insertion change precision improved from 82.96% to 86.46%. We can further alleviate manual effort by narrowing down the number of insertion changes with the improvement of precision.

Table 5 – Validation evaluation at top N predictions compared with the reference set

Top *N*	# of Classifier Validated	# of Expert Validated	Precision (%)
1500	150	129	86.00
1600	174	149	85.63
1700	183	157	85.79
1800	185	159	85.95
1900	191	165	86.39
2000	192	166	86.46

Discussion

In this paper, we introduced a relational ML-based validation approach, CNN-MLP, to alleviate manual effort required for validating change suggestions in ontology auditing work. In addition to validating auto-suggested insertion changes, our experiments indicated that CNN-MLP classifier may be effectively used in tasks of concept pair relation prediction and ontology auditing. As shown in Table 4, five auto-suggested insertion changes were out of the reference set; although they were not used to validate change suggestions, they may provide additonal candidates for change redemiations.

Evaluated with the reference set, the performances of auto-suggested insertion change and auto-suggested insertion change validation look promising. However, note we have not addressed the task of auto-deletion change suggestion and deletion change validation due to the lack of the reference set containing validated deletion changes. Labeled data (or reference set) is required in the framework of supervised ML approach. SNCT is updated by domain experts twice a year, and it may be possible to treat insertion and deletion changes in different versions of SNCT as labeled data for ML based auditing or validation.

Conclusions

We have presented an ML approach to validate insertion changes generated by non-lattice-based auditing methods. In validation workflow, we introduced a hybrid CNN-MLP classifier to predict IS-A relations of concept pairs, and automatically suggest insertion changes for non-lattice subgraphs. Experiments on IS-A relation prediction achieves an average precision of 0.972, and auto-suggested insertion changes achieves F_{insert} of 82.43% with top 2,000 predictions, which indicate the potential of our classifier to validate IS-A insertion changes. Validated with insertion changes resulting from our classifier, the precision of a given set of auto-suggested IS-A insertion changes is improved from 82.96% to 86.46%.

By cumulating SNCT deletion changes as a surrogate reference set for ML approaches [12], our future work will focus on exploring features to predict potentially incorrect IS-A relaitons for validating auto-suggested deletion changes in non-lattice subgraphs.

Acknowledgements

This work was supported by the U.S. National Science Foundation under grant 1816805, and by the U.S. National Institutes of Health under grant R21CA231904.

References

[1] G.-Q. Zhang, W. Zhu, M. Sun, S. Tao, O. Bodenreider, and L. Cui, MaPLE: A MapReduce Pipeline for Lattice-based Evaluation and Its Application to SNOMED CT, *Proc IEEE Int Conf Big Data*. **2014** (2014) 754–759.

[2] L. Cui, W. Zhu, S. Tao, J.T. Case, O. Bodenreider, and G.-Q. Zhang, Mining non-lattice subgraphs for detecting missing hierarchical relations and concepts in SNOMED CT, *J. Am. Med. Inform. Assoc.* **24** (2017) 788–798.

[3] L. Cui, O. Bodenreider, J. Shi, and G.-Q. Zhang, Auditing SNOMED CT hierarchical relations based on lexical features of concepts in non-lattice subgraphs, *J. Biomed. Inform.* **78** (2018) 177–184.

[4] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich, A Review of Relational Machine Learning for Knowledge Graphs, *Proc. IEEE*. **104** (2016) 11–33.

[5] D. Zeng, K. Liu, S. Lai, G. Zhou, and J. Zhao, Relation Classification via Convolutional Deep Neural Network, *Proc. of COLING* (2014), Technical Papers.

[6] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, and J. Dean, Distributed Representations of Words and Phrases and their Compositionality, *Advances in Neural Information Processing Systems* **26** (2013) 3111–3119.

[7] E. Park, X. Han, T.L. Berg, and A.C. Berg, Combining multiple sources of knowledge in deep CNNs for action recognition, *IEEE Winter Conference on Applications of Computer Vision* (2016) 1–8.

[8] C. Pesquita, D. Faria, A.O. Falcão, P. Lord, and F.M. Couto, Semantic similarity in biomedical ontologies, *PLoS Comput. Biol.* **5** (2009).

[9] P. Zweigenbaum, B. Bachimont, J. Bouaud, J. Charlet, and J.F. Boisvieux, Issues in the structuring and acquisition of an ontology for medical language understanding, *Methods Inf. Med.* **34** (1995) 15–24.

[10] M. van Erp, L. Vuurpijl, and L. Schomaker, An overview and comparison of voting methods for pattern recognition, *Proceedings Eighth International Workshop on Frontiers in Handwriting Recognition* (2002) 195–200.

[11] S.-B. Zhang, and J.-H. Lai, An integrated information-based similarity measurement of gene ontology terms, *Computer Science and Information Systems*. **12** (2015) 1235–1253.

[12] G.-Q. Zhang, Y. Huang, and L. Cui, Can SNOMED CT Changes Be Used as a Surrogate Standard for Evaluating the Performance of Its Auditing Methods? *AMIA Annu. Symp. Proc.* **2017** (2017) 1903–1912.

[13] Keras documentation, (2016). http://keras.io [accessed on March 27, 2019]

[14] Toolkit-NLTK 3.4 documentation, https://www.nltk.org [accessed on March 27, 2019]

Address for correspondence

Licong Cui (Licong.Cui@uth.tmc.edu)

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190248

Detecting Systemic Data Quality Issues in Electronic Health Records

Casey N. Ta[a], Chunhua Weng[a]

[a] *Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY, USA*

Abstract

Secondary analysis of electronic health records for clinical research faces significant challenges due to known data quality issues in health data observationally collected for clinical care and the data biases caused by standard healthcare processes. In this manuscript, we contribute methodology for data quality assessment by plotting domain-level (conditions (diagnoses), drugs, and procedures) aggregate statistics and concept-level temporal frequencies (i.e., annual prevalence rates of clinical concepts). We detect common temporal patterns in concept frequencies by normalizing and clustering annual concept frequencies using K-means clustering. We apply these methods to the Columbia University Irving Medical Center Observational Medical Outcomes Partnership database. The resulting domain-aggregate and cluster plots show a variety of patterns. We review the patterns found in the condition domain and investigate the processes that shape them. We find that these patterns suggest data quality issues influenced by system-wide factors that affect individual concept frequencies.

Keywords:

Electronic Health Records; Cluster Analysis; Data Accuracy

Introduction

Secondary reuse of observational health data for research is increasing in importance and popularity as electronic health data become more available and analytic methods become more powerful. However, analysis of electronic health records (EHR) continues to face many social and technical challenges, including data inaccuracy, incompleteness, and biases implicit in the healthcare recording process [1–5]. Since observational clinical data are sourced from disparate systems not designed for research purposes, researchers consuming these data must clearly understand the nuances of their data. Due to the complexity of interactions, ongoing development, and technical challenges of these systems, events that impact data integration can occur without awareness from data managers. Further, experts knowledgeable of the intricacies of the data may not be available to researchers, especially in the context of shared data in collaborative research networks, such as the Observational Health Data Sciences and Informatics (OHDSI) [6].

EHR data quality is receiving increased attention for its effects on secondary analysis for clinical research and population health informatics [7–10]. Temporal trend analysis is a common method of assessing data quality. Kahn's harmonized data quality framework includes temporal plausibility metrics, such as comparing data value density over time against expected values [7]. Hall's database guidelines recommend counting frequency of records and field occurrences over time to identify blocks of missing data or changes in data volume [8]. Brown et al demonstrate visually exploring trends of intrasite and intersite frequency to detect data anomalies [9]. ACHILLES performs temporal analysis of concepts in Observational Medical Outcomes Partnership (OMOP) databases to assess month-to-month stability of counts and flags changes that exceed thresholds [6]. These methods enable detection of data quality issues in individual variables but do not detect systemic issues that may affect multiple variables.

The Columbia University Irving Medical Center (CUIMC) Clinical Data Warehouse (CDW) formatted using the OMOP Common Data Model (CDM) (herein referred to as the OMOP database) contains observational data on nearly 60,000 concepts. Individual analysis of 60,000 concepts would be prohibitively time consuming. Although population prevalence of clinical concepts can change with time, we hypothesize that common patterns of temporal trends in EHRs reveal systemic factors that influence observational data and data quality.

In this manuscript, we present methods of detecting systemic, population-level data quality issues in electronic health records by analyzing domain-level aggregate statistics and identifying common temporal patterns across concepts. We apply these methods to the CUIMC OMOP database and infer and investigate system wide processes that affect the data quality.

Methods

Data Source

This study received institutional review board approval with waiver for obtaining informed consent. The CUIMC CDW contains records from multiple sites and was converted to OMOP CDM v5.1 in March 2018. Diagnoses ("conditions" in OMOP nomenclature), drugs, procedures, and visit types were collected from CUIMC's OMOP database from the *condition_occurrence*, *drug_exposure*, *procedure_occurrence*, and *visit_occurrence* tables, respectively. We analyze conditions, drugs, and procedures to provide multiple views of patient data and visit types to indicate clinical service capacity.

Conceptual Framework

Figure 1 shows our conceptual framework for analysis. There are four entities: patients, concepts, domains, and time. Patient health is characterized in the OMOP database by individual concepts, which are categorized by domains (e.g., conditions, drugs, and procedures). For each concept, we count the number of patients in each year and plot the counts over time. Data captured in EHRs are affected by population health status, health service capacity, medical coding protocols, and operation of clinical databases. These parameters change over time, as reflected in the counts and frequency of concepts.

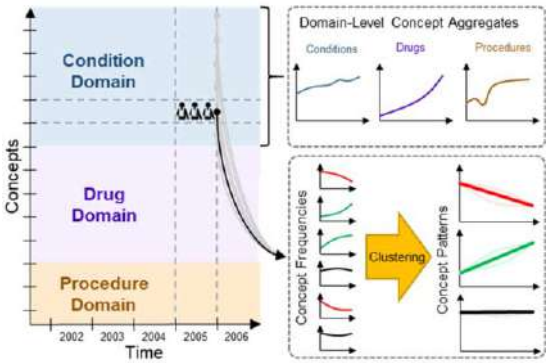


Figure 1– Conceptual framework for analysis.

Concept Prevalence

We assume that population health is consistent with only minor changes annually and gradual changes over many years [11]. Correspondingly, we assume the annual frequency of most clinical concepts to be relatively stable over time if the health system is stable. Observed fluctuations in concept frequencies may have causes intrinsic to the concept (e.g., outbreaks or new interventions) or extrinsic to the concept (e.g., expansion of clinical services). We measured yearly counts and frequencies of concepts from the OMOP database. We defined counts as the number of unique patients with the concept per calendar year. We defined frequency as the count divided by number of patients in that year. We estimated the number of patients per year from the number of unique patient identifiers associated with at least one condition, drug, or procedure per year.

Domain-Level Concept Aggregates

Given a stable healthcare system, we hypothesize the domain-level aggregated metrics are highly stable over time. Fluctuations in these metrics may reflect changes to healthcare services or the EHR recording process. To identify global trends in each domain, we analyzed the yearly patient count and yearly aggregated statistics across concepts in condition, drug, and procedure domains using the domain of each concept as defined in the OMOP standard vocabulary to aggregate concepts. We analyzed four aggregated statistics across each domain per year: 1) the total patient counts across concepts; 2) the count-per-capita (total patient count divided by number of patients), indicating the average number of conditions, drugs, or procedures per patient; 3) the number of unique concepts; and 4) the mean frequency across concepts.

Clustering of Concept Patterns

Analyzing thousands of individual concept frequencies would be overwhelming and difficult to differentiate between changes in the population prevalence versus changes affecting the recorded prevalence in EHRs. To identify common behavior patterns with similar trends among concepts, we performed K-means clustering [12] within each domain. Each concept's annual frequencies were normalized to unit magnitude and treated as vectors. We used concept frequency instead of count to normalize against population changes. We normalized frequencies to unit magnitude to match trends with similar proportional changes. Annual frequencies from calendar years 1985 to 2017 were included; we excluded 2018 since data was not yet available for the entire year. We excluded concepts with maximum frequency below 0.01% of the population to reduce noise from infrequent concepts. We performed K-means clustering on the normalized frequency vectors. We manually selected K by clustering each domain with different levels of K (4, 6, 8, 10, 12, 15, 20, and 25) and selecting the smallest K before intracluster variance ceased to improve substantially.

Post-hoc Analyses

To determine if the clustered concept patterns reveal systemic factors (e.g., changes in health service capacity or medical coding protocols) that affect concept frequency in the database, we manually reviewed the domain- and concept-level patterns, arriving at hypotheses on the causes of several patterns. We performed post-hoc analyses to validate these hypotheses.

Results

As of March 2018, the CUIMC OMOP database contains EHR records on 5,368,414 patients, covering 59,583 concepts, including 18,399 conditions, 18,691 drugs, and 22,787 procedures spanning October 1985 to March 2018.

Domain-Level Concept Aggregates

Figure 2 shows plots of the domain-level count-per-capita and total counts of a) conditions, b) drugs, c) procedures, and d) people per year. Patient counts are relatively flat between 1986-2001, except for a slight depression from 1992-1995, and grows steadily from 2001-2010 and rapidly after 2010, with a spike in 2014. The condition domain total count and count-per-capita increase nearly linearly over time. The condition count-per-capita is slightly unstable before 2000 but steadily increases after 2000, peaking in 2015 and dropping in 2016-2017. The total count of conditions spikes in 2014. Drug data is nearly non-existent before 2001 but steadily increases until it peaks in 2013. The drug count-per-capita drops substantially in 2014 and fluctuates through 2017. The procedure domain is the most unstable. Procedures display notable growth in 1989, a large spike in 1996-1997, rapid increase from 2001-2005 and steady

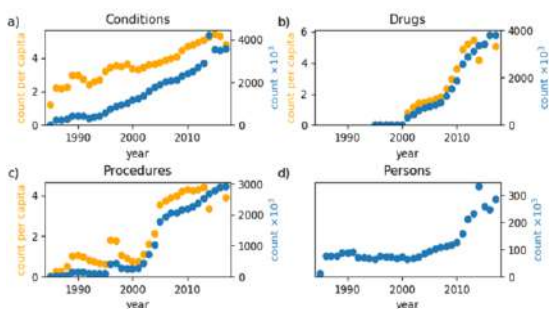


Figure 2– Total count (blue) and count-per-capita (orange) of a) conditions, b) drugs, c) procedures and d) people per year.

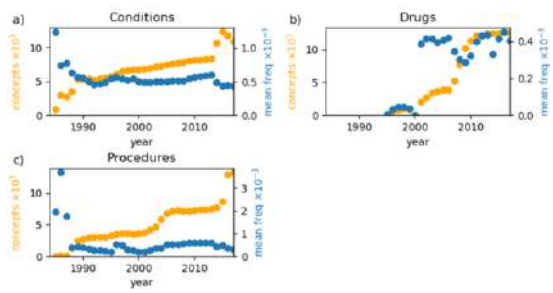


Figure 3– Count of unique concepts (orange) and the mean frequency of concepts (blue) per year for a) conditions, b) drugs, and c) procedures.

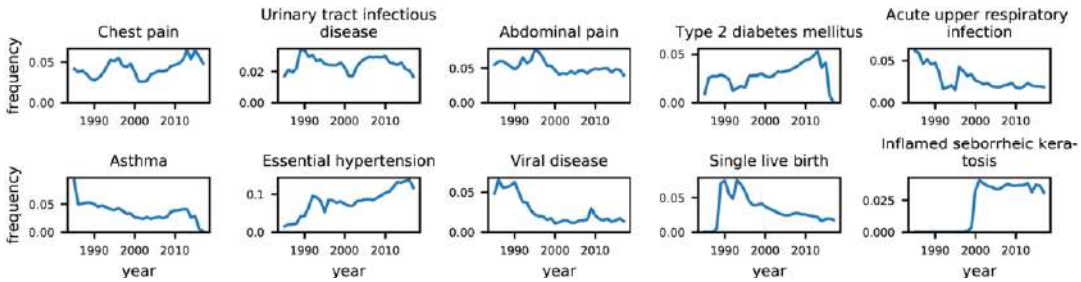


Figure 4– Example annual concept trends for the ten most prevalent conditions.

increase thereafter. Similar to drug count-per-capita, procedure count-per-capita drops in 2014 and fluctuates through 2017.

Figure 3 shows the annual number of distinct concepts and the mean frequency of patients across concepts for each domain. The condition domain stably grows from 1989-2013 but fluctuates thereafter. The drug domain is relatively stable from 2001-2006 but fluctuates thereafter. In the procedure domain, the number of concepts stably grows from 1989 to 2001, grows quickly to 2005, slowly grows through 2014, and quickly grows again. The mean frequency of procedure concepts has similar periods of behavior with an additional sharp increase in 1996.

Concept-Level Analyses

Due to space limitations, we primarily present and discuss the results for concept-level analyses in the condition domain and exclude drug and procedure results. Figure 4 plots annual frequencies for the ten conditions with the highest total counts across all years, showing multiple remarkable behaviors. Type 2 diabetes mellitus and asthma both begin to drop in 2014 and are nearly zero by 2017. Type 2 diabetes mellitus and acute upper respiratory infection both exhibit abrupt depressions from 1992-1995. Inflamed seborrheic keratosis is nearly zero until 1999 but rapidly increases and plateaus after 2001. Chest pain and abdominal pain both increase in frequency from roughly 1990-1995 and decrease from 1995-2000.

K-means clustering was performed for conditions (K=20). Figure 5 shows plots of cluster centroids (i.e., the temporal pattern) for condition concepts. The clusters are sorted in descending order of the number of concepts belonging to each cluster. The two largest condition clusters have nearly zero frequency until a rapid increase in 2014. Clusters 4, 5, 6, 11, 12, and 19 drop in frequency starting around 2014. Cluster 12 is the only cluster with constant frequency over a large time period but only contains 3.47% of condition concepts. Clusters 3 and 16 increase almost linearly with time. Other clusters display patterns that are more dynamic. Clusters 7, 8, 10, 13, 14, 17, and 18 exhibit peaks at different times. Condition cluster 11 decreases from 1989-1999 and then increases until 2014.

Post-hoc Analyses of Trends

To demonstrate that the above methods reveal systemic data quality issues, we manually reviewed the domain- and concept-level patterns, hypothesized the causes of several patterns, and performed post-hoc analyses to test these hypotheses.

Condition clusters 0 and 1 behave like step functions with a rapid increase in 2014 while many other clusters begin to decrease around the same time. The number of condition concepts (Fig. 2a) also suddenly increases in 2014. We hypothesized that these patterns were caused when the health systems converted from using one medical coding system to another, e.g., from ICD9CM to ICD10CM. To investigate this, we analyzed the composition of medical coding systems (i.e., OMOP source vocabularies) contributing to these clusters. Table 1 shows the source vocabulary composition of condition clusters 0 and 1 compared against the aggregate of all other trends. Condition clusters 0 and 1 are composed of 93.9% and 92.0% ICD10CM source concepts, respectively. The remaining clusters are composed of 83.2% ICD9CM concepts.

The domain-level count of condition concepts displays an abrupt depression from 1992-1995 (Fig. 1a). Condition clusters 9, 15, and 19 also display similar depressions from 1992-1995, but this behavior is not noticeable in other clusters. Since the behavior was isolated to a few trends, we hypothesized that this depression was caused by changes affecting specific service types (e.g., temporary loss of data or reduction of services) which cover the conditions in these clusters.

To investigate whether changes in health services occurred, we analyzed the number of patient visits per year stratified by visit type (e.g., inpatient, outpatient, etc.) from the OMOP *visit-occurrence* table. Figure 6 shows plots of the annual counts for the 10 most frequent visit types, which account for 97.3% of

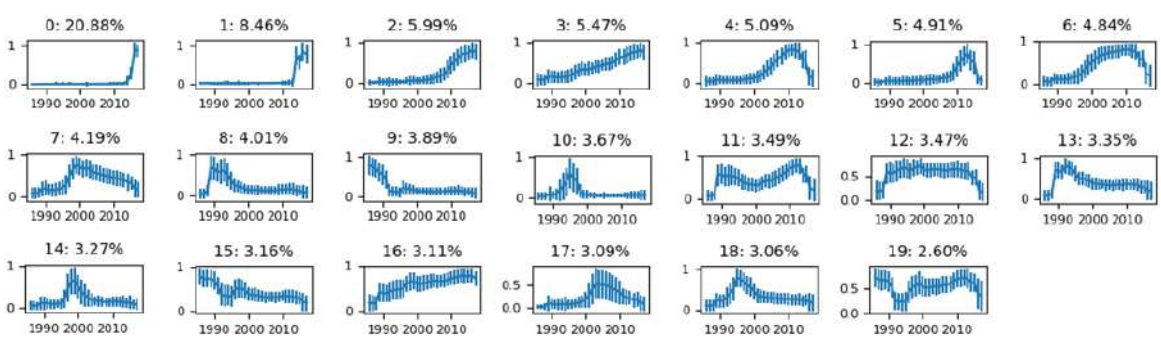


Figure 5– K-means clusters for conditions. Plots show cluster centroids (the cluster trend over time) with standard deviation across concepts as error bars (intracluster vaiance). Subplot titles show cluster labels and the percent of concepts belonging to each cluster.

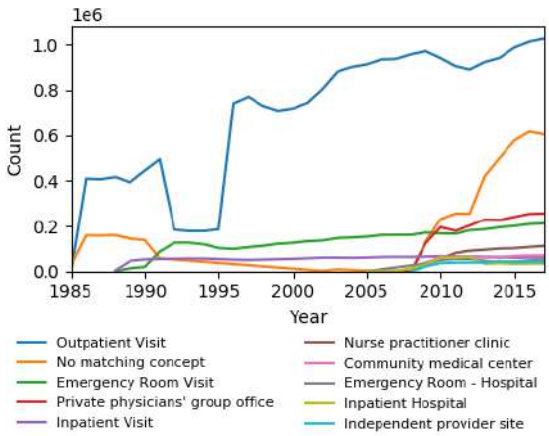


Figure 6– Annual number of visits for the 10 most common visit concepts.

Table 1– Source vocabulary composition of condition clusters.

Vocabulary ID	Cluster 0	Cluster 1	Remaining clusters
SNOMED	0.2%	1.6%	0.0%
ICD10CM	93.9%	92.0%	16.8%
ICD9CM	6.0%	6.4%	83.2%

visits. We excluded 17 less frequent visit types, including cardiology clinic, general surgery clinic, community clinic, etc., from the figure for visual clarity. Outpatient visit is the most frequent type and exhibits a large depression from 1992-1995. The other visit types do not display similar depressions. Inpatient visits appear in 1989 and are very stable throughout the years. Emergency room visits appear in 1989 with unstable growth until 1995 and stable growth thereafter. 11.7% of visits did not have a specified visit type (No matching concept).

Discussion

Concept Prevalence

Reviewing plots of annual frequencies of individual concepts, we found that the large majority of concepts exhibited a variety of unstable behaviors. Figure 1 shows a sample of plots, including the ten most prevalent conditions. We show the most frequent conditions because we assumed they would generally behave more stable than less frequent conditions. If frequencies of clinical terms in the EHR are representative of general population prevalence, then the behavior of terms in the EHR should reflect the behavior of the clinical concepts in the population. We should expect relative stability from the conditions in Figure 1, but from this small sample, it is immediately apparent that concepts exhibit a wide variety of behaviors in the EHR. Most of these fluctuations are too dramatic to be reflective of real health changes in the population. For example, Type 2 diabetes mellitus and asthma fall to nearly 0% frequency in 2015 while inflamed seborrheic keratosis was nearly 0% until 1999 but plateaus from 2000 onward. These patterns suggest underlying changes in the usage of medical coding systems (e.g., changing from ICD9CM to ICD10CM). Type 2 diabetes mellitus and acute upper respiratory infection both have abruptly depressed frequencies from 1992-1995. Chest pain and abdominal pain have similarly shaped growth and decline patterns in the 1990s.

Domain-Level Analysis

The domain-level aggregate plots show high-level changes observable in the EHR. Figure 1 shows the annual number of patients, total patient count across concepts, and the count of concepts per capita for each domain. Some of the patterns observed in individual concepts also exist at the domain-level. From 1992-1995, the number of patients abruptly drops relative to surrounding years, corresponding to decreased condition counts. This suggests that the dips in type 2 diabetes mellitus and acute upper respiratory infection were caused by system level events, potentially affecting many conditions. Drug counts were practically non-existent before 2001 and almost linearly increased after 2001, indicating that drug data was not available before 2001. Changes in the annual number of unique concepts and the mean patient frequency across concepts (Fig. 2) may suggest changes in the usage of medical coding systems. The increased number of concepts in the condition domain from 2014-2015, in the drug domain from 2007-2011, and in the procedure domain from 2002-2005 and 2015-2017 all suggest that changes occurred in the medical coding system.

Concept-Level Analysis

As discussed above, many concept frequency curves display fluctuations too dramatic to be caused by changes in population prevalence. We applied K-means clustering to find groups of concepts that have similar behavioral patterns. Clusters contain hundreds to thousands of concepts, so their average frequency curves are minimally influenced by true changes in individual concepts' population prevalences. These cluster patterns may suggest changes in the medical system that influence the clinical recording process, including service capacity changes (e.g., expansion of medical system) or events that affect patient data records (e.g., deployment of new clinical databases). In Figure 5, various patterns among condition concepts can be easily identified along with the proportion of concepts that follow the pattern. A small number of concepts remain stable and constant over time (e.g., condition cluster 12, 3.47% of conditions), and the stability only lasts for a period (1989-2013). The majority of concepts are stable only during certain time windows, and these windows vary by cluster.

Several clusters have sharp increases that correspond with decreases in other clusters. Clusters 0 and 1 increase while clusters 4, 5, 6, 11, 12, and 19 drop around 2014. We hypothesized that these patterns were caused by changes in the medical coding systems in 2014. To test this hypothesis, we analyzed the composition of source vocabularies among these clusters to determine if these clusters are primarily represented by different vocabularies (Table 1). Indeed, clusters 0 and 1 are composed of over 90% ICD10CM concepts, while all other clusters are cumulatively composed of 83.2% ICD9CM concepts, supporting the hypothesis above. Similar patterns were observed in the procedure domain associated with changes from ICD9Proc to ICD10PCS in 2016 (data not shown).

Condition clusters 9, 15, and 19 exhibit a marked drop in frequency between 1992-1995, corresponding with dips in domain-level counts (Fig. 2a). Since this behavior was not observed among most other clusters, we hypothesized that this depression was isolated to specific sites or services. To test this, we analyzed annual visit type occurrences (Fig. 5). Outpatient visits display a similar depression between 1992-1995 while the other visit types behave stably during this time, suggesting that clusters 9, 15, and 19 are primarily composed of conditions observed in outpatient settings. Inpatient visit and emergency room visit types both begin in 1989, corresponding with the opening of the Milstein Hospital Building of Presbyterian

Hospital. A flurry of other visit types, including private physicians' group office, nurse practitioner clinic, and community medical center, begin to appear in 2009, suggesting that the network of clinical data sources connected to the clinical data warehouse was expanding to more sites.

Limitations

There are a number of limitations to our methods. While clustering the annual concept frequencies reveals patterns of behavior among the concepts, the clusters on their own do not provide meaningful insight for the causes of these patterns. For some patterns, we were able to generate hypotheses and corroborate the causes by investigating other information sources (e.g., visit types and source vocabularies) to reveal systemic forces behind these patterns. However, many patterns are unexplained, and we cannot infer causality for the hypothesized causes of patterns as we did not apply statistical analyses. Even without causal explanations, these methods still allow researchers to identify time frames of data instability to exclude from their analyses or interpret with caution. These methods do not produce quantitative, scalar, and invariant results, which are desired characteristics of data quality metrics to allow comparison of data quality statistics across data sets and time [13]. The domain- and concept-level analyses were performed at annual intervals, which may be too coarse to detect short data aberrations. However, the method should perform well with shorter intervals.

The selected number of clusters for K-means clustering was arbitrary. The choice of K influences the resulting clusters and, thus, the interpretation of results. We attempted to use the elbow method [14] to guide our choice, but the intercluster sum of squared differences had broad-shouldered changes over K in all domains (data not shown). This indicates the presence of many small clusters or many concepts may not follow common patterns, thus, there may not exist an optimal choice for K.

We determined the annual number of patients based on observations in the condition, drug, and procedure tables and subsequently used this patient count in calculating concept frequencies of those conditions, drugs, and procedures. This introduces bias into the calculation of concept frequency. Ideally, the patient population should be identified externally from the variables undergoing analysis. The OMOP CDM contains *observation_periods* which should identify the start and end dates when the patient was observed in the database. However, in our database, 66.6% of patients have missing start dates, which may indicate an error in the extract-transform-load process, which converts data from the clinical data warehouse to the OMOP CDM. Also, we assume that each patient has a single unique identifier. This may not be true if patients have duplicate registrations within the CDW.

Developing this analysis on an OMOP database confers both advantages and disadvantages. By analyzing the transformed database, we lose some ability to investigate data provenance and sources of data quality issues. However, developing this analysis on a widely adopted CDM allows other institutions to immediately benefit and learn from these results. Although other data models will not be able to replicate this analysis directly, the methods are simple to implement in other systems.

Conclusions

We contribute methods for analyzing EHR data quality aspects (e.g., temporal plausibility, consistency, etc.) for secondary analysis using domain-level aggregate statistics and clustering the annual concept frequency trends to find common temporal patterns across concepts. These patterns may indicate EHR data quality issues caused by operational or system-wide factors that affect multiple concepts simultaneously.

Acknowledgements

This study was supported in part by grants OT3TR002027 from NCATS and R01LM009886-08A1 from NLM.

References

[1] G. Hripcsak, and D.J. Albers, Next-generation phenotyping of electronic health records, *J Am Med Inform Assoc* **20** (2013), 117–121.
[2] R. Platt, and T. Lieu, Data enclaves for sharing information derived from clinical and administrative data, *JAMA* (2018), E1–E2.
[3] T. Botsis, G. Hartvigsen, F. Chen, and C. Weng, Secondary use of EHR: data quality issues and informatics opportunities, *Summit on Translat Bioinforma* **2010** (2010), 1–5.
[4] H. Kharrazi, C. Wang, and D. Scharfstein, Prospective EHR-based clinical trials: the challenge of missing data, *J Gen Intern Med* **29** (2014), 976–978.
[5] H. Kharrazi et al., Comparing population-based risk-stratification model performance using demographic, diagnosis and medication data extracted from outpatient electronic health records versus administrative claims, *Med Care* **55** (2017), 789–796.
[6] G. Hripcsak et al., Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers, *Stud Health Technol Inform* **216** (2015), 574–578.
[7] M.G. Kahn et al., A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data, *EGEMS (Wash DC)* **4** (2016), 1244.
[8] G.C. Hall et al., Guidelines for good database selection and use in pharmacoepidemiology research, *Pharmacoepidemiology and Drug Safety* **21** (2012), 1–10.
[9] J. Brown, M. Kahn, and S. Toh, Data quality assessment for comparative effectiveness research in distributed data networks, *Med Care* **51** (2013), S22–S29.
[10] H. Kharrazi et al., A proposed national research and development agenda for population health informatics: summary recommendations from a national expert workshop, *J Am Med Inform Assoc* **24** (2017), 2–12.
[11] S.H. van Oostrom et al., Time trends in prevalence of chronic diseases and multimorbidity not only due to aging: data from general practices and health surveys, *PLoS One* **11** (2016), e0160264.
[12] J.A. Hartigan, and M.A. Wong, Algorithm AS 136: a K-means clustering algorithm, *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **28** (1979), 100–108.
[13] S.G. Johnson, S. Speedie, G. Simon, V. Kumar, and B.L. Westra, Application of an Ontology for Characterizing Data Quality for a Secondary Use of EHR Data, *Appl Clin Inform* **07** (2016), 69–88.
[14] R. Tibshirani, G. Walther, and T. Hastie, Estimating the number of clusters in a data set via the gap statistic, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63** (2001), 411–423.

Address for correspondence

Chunhua Weng, chunhua@columbia.edu, 622 West 168th Street, Presbyterian Hospital Floor 20, New York, NY 10032

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190249

An Exploratory Study on Pseudo-Data Generation in Prescription and Adverse Drug Reaction Extraction

Carson Tao[a], Kahyun Lee[b], Michele Filannino[b], and Özlem Uzuner[b]

[a]The State University of New York at Albany, Albany, NY, USA, [b]George Mason University, Fairfax, VA, USA

Abstract

Prescription information and adverse drug reactions (ADR) are two components of detailed medication instructions that can benefit many aspects of clinical research. Automatic extraction of this information from free-text narratives via Information Extraction (IE) can open it up to downstream uses. IE is commonly tackled by supervised Natural Language Processing (NLP) systems which rely on annotated training data. However, training data generation is manual, time-consuming, and labor-intensive. It is desirable to develop automatic methods for augmenting manually labeled data. We propose pseudo-data generation as one such automatic method. Pseudo-data are synthetic data generated by combining elements of existing labeled data. We propose and evaluate two sets of pseudo-data generation methods: knowledge-driven methods based on gazetteers and data-driven methods based on deep learning. We use the resulting pseudo-data to improve medication and ADR extraction. Data-driven pseudo-data are suitable for concept categories with high semantic regularities and short textual spans. Knowledge-driven pseudo-data are effective for concept categories with longer textual spans, assuming the knowledge base offers good coverage of these concepts. Combining the knowledge- and data-driven pseudo-data achieves significant performance improvement on medication names and ADRs over baselines limited to the use of available labeled data.

Keywords:

Natural Language Processing, Machine Learning, Information Storage and Retrieval

Introduction

Electronic health records (EHR) contain significant amounts of medical information. However, most of this information is presented in unstructured narrative free-text and remain unavailable to computerized systems that rely on structured representations for access and retrieval [1,2]. This unavailable information includes medication and prescription information, which is often described in narrative portions of electronic health records (EHRs) [3] and can complement the information represented in prescriptions themselves [4].

Natural language processing (NLP) methods such as information extraction (IE) can open up this information to downstream uses [2,3], such as comparative effectiveness [5] cohort selection [6,7], medication reconciliation [8], and pharmacovigilance and pharmacogenomics [9]. Extraction of medication information from narratives enhances patient safety by making this more complete medication information available to use for clinical decision support [8]. Extraction of ADRs supports goals highlighted by the U.S. Food and Drug Administration (FDA) [10] for comparing the ADRs present in labels from different manufacturers for the same drug, and for performing post-marketing safety analysis by identifying new ADRs not currently present in the labels [11].

Over the last few decades, there have been many efforts for extracting information from clinical narratives [4,11]. An important factor in the development of these systems is the availability of sufficient training data [12]. However, preparation of training data is time-consuming and labor-intensive; as a result, such data sets are scarce. Even when such data are available, some concept types in the data tend to be sparsely-represented [13]. NLP systems either relatively underperform on concept categories that are more sparse or they complement their supervised machine-learning models with hand-build rules [14,15]. An alternative to these approaches requires augmenting labeled data with pseudo-data. Pseudo-data are synthetic data that are derived from samples in labeled data, e.g., by combining elements of existing samples. Despite its promise for IE, to the best of our knowledge, pseudo-data generation has received limited attention in the clinical domain. Instead, researchers have focused on modifying the architectures of their systems to improve recognition of concepts, potentially over-fitting their systems to the data and to the task [4]. We aim to address this gap in the literature.

Our proposed approach develops and compares knowledge- and data-driven pseudo-data generation methods for IE from narratives of EHRs [4] and from drug labels [11]. We use the resulting pseudo-data to enhance NLP methods for extracting medication and ADR information. Despite focus on medication names and ADRs, we also study concepts that relate to medications and ADRs. Our IE methods extract concepts in a sentence; therefore, improving identification of a key concept in a sentence can have a spillover effect and help extraction of the rest of the concepts in the sentence, even when those concepts themselves are not directly enhanced with pseudo-data. We, therefore, report results on the following set of concept types: medication names, dosages, frequencies, modes, durations, and reasons; ADRs, animal species, drug classes, factors, negated phrases, and severities. Our results show that pseudo-data helps improve performance on medication names and ADRs over methods that are limited to the use of available labeled data.

Related Work

Extraction of ADR and medication information is commonly solved using rules and/or machine learning classifiers. Rule-based systems have proven to be useful [4,14] for this purpose but come with limitations on performance, scalability, and generalizability [16]. They suffer particularly on concept types that show much syntactic variation [13]. In response to these limitations of rule-based systems, most high-performing

methods utilize machine learning classifiers [13,15-17]. The primary challenge facing these systems is data sparsity [13]. As a result, machine learning solutions often require support from rules [16] and have given promising results. The current state of the art in medication extraction is one such hybrid system that combines a machine learning classifier with post-processing rules and achieves 89.6% F-measure on medication names in narrative EHRs [15]. In comparison, the current state of the art in ADR extraction is a machine learning system that achieves 82% F-measure [17] on ADRs. Given their reliance on labeled data, performance on both tasks can be improved by more data. Given the high cost of generating manual gold standard data, we explore an automatic method, pseudo-data generation, for augmenting available labeled data. To assess generalizability, we incorporate pseudo-data into two tasks tackled on two different kinds of corpora: medication extraction from EHRs and ADR extraction from drug labels.

Pseudo-data generation is a common way of resolving imbalanced datasets. Despite its potential benefit for addressing data sparsity, pseudo-data generation remains relatively unexplored in the clinical domain. Outside of the clinical domain, Chawla et al. [18] proposed an oversampling method, called synthetic minority over-sampling technique (SMOTE), that generates pseudo-samples for minority classes by joining existing samples with their nearest neighbors in terms of feature vectors. SMOTE takes the size of the largest class as a parameter to produce a fully balanced training set for all classes. Within the clinical domain, pseudo-data generation has been tackled only in very few studies: e.g., Keretna et al. [19] formulated an extended segment representation technique to improve biomedical IE, with emphasis on ambiguous named entities.

In our study, we evaluate the efficacy of pseudo-data in clinical IE on medications and ADRs. We propose two methods for pseudo-data generation. The first method utilizes a combination of random duplication and external gazetteers. Being one of the simplest oversampling methods, random duplication on its own is rarely effective [18]. Literature [20] fails to show significant improvement on minority class recognition when data sizes of the minority class are increased through repetition of existing data. We expect that this finding results from lack of any "new" information. Therefore, we incorporate information from external gazetteers and generate new samples from them so that we can address the sparsity of unique samples. The second method takes advantage of word embeddings. Pennington et al. [21] have shown that semantically similar words cluster together in the word embedding space. Given this, we generate pseudo-data by replacing tokens (from clinically-relevant concepts) in the existing training data with their nearest neighbors in terms of embeddings. After generating pseudo-data, we use them to enrich our training data. We hypothesize that pseudo-data can help improve IE performance.

Data

We base our studies on two established NLP data sets for medication and ADR extraction. Our first corpus consists of 991 discharge summaries from Partners Healthcare and was the basis for the 2009 i2b2 Medication Challenge [4] which focused on the extraction of medication and prescription information from narratives of EHRs. This dataset contains 145 annotated summaries and 846 unannotated summaries in the training set. The test set contains 252 annotated summaries. Our second data set consists of 2,309 drug labels from the FDA and were the basis for the 2017 TAC ADR Challenge [11], which focused on the extraction of ADRs, related drug classes, ADR severities, and other information from narratives. This dataset contains 101 annotated labels and 2,208 unannotated labels in the training set. The test set contains 99 annotated labels. In our study, we build systems based on the training data. We then run an end-to-end system against the test set. See **Table 1** for the per-category statistics for each dataset.

Table 1 - Number of records, tokens, and phrase-level concepts per category (training and test set)

2009 i2b2 Medication Challenge								
Dataset	**Record**	**Token**	**Medication**	**Dosage**	**Mode**	**Frequency**	**Duration**	**Reason**
Train (L)	145	250,436	7,988	4,132	3,052	3,881	508	1,516
Train (U)	846	925,703	N/A	N/A	N/A	N/A	N/A	N/A
Test (L)	252	284,311	8,440	4,371	3,299	3,925	499	1,325
2017 TAC ADR Challenge								
Dataset	**Record**	**Token**	**ADR**	**Animal**	**Drug**	**Factor**	**Negation**	**Severity**
Train (L)	101	222,429	12,323	44	244	600	87	760
Train (U)	2,208	3,576,500	N/A	N/A	N/A	N/A	N/A	N/A
Test (L)	99	224,035	12,693	86	164	562	173	947

(L) indicates labeled data. (U) indicates unlabeled data.

Methods

We study pseudo-data generation on two types of concepts, medication names and ADRs, which lie at the heart of any prescription and ADR extraction task. Given our focus on pseudo-data generation, we first introduce a baseline IE system that can extract these concepts from labeled data. We then present two pseudo-data generation methods and measure the contribution of pseudo-data to the gold standard data and to extraction performance.

Creating a Baseline IE System

Our baseline IE system [16] uses conditional random fields (CRFs) [22] trained with: (1) Normalized tokens: lower-cased tokens with punctuation or special characters removed. We replace all numbers with a generic placeholder (e.g., 300 ml → DDD ml, 3rd degree → D degree). (2) Part-of-speech (POS) tags. (3) Temporal properties: A set of seven binary features that indicate whether a token represents a signal for temporal expressions such as time, temporal period, and temporal reference. (4) Real-valued word embeddings trained on the unannotated portions of our datasets, as indicated by prior work [23]. When including these features into the training process, we use window size of ±2 around the target word. This window size is selected as the optimal parameter from cross-validation experiments which checked window sizes from ± 1 to ± 5 in our previous study [16].

We evaluated this system on both ADR and medication extraction using only gold standard labeled data (see **Table 2** for results [16]). We then enhanced it with pseudo-data as follows: we collected sentences from the annotated training data that contained at least one ADR or medication name. We generated pseudo-data based on these sentences using two different methods: Knowledge-driven and data-driven.

Generating Knowledge-Driven Pseudo-data

This method takes advantage of external gazetteers. Using ADR extraction as an example for describing the methodology, we select and access a database (i.e., Vigibase from VigiAccess.org) that can serve as a lookup table for ADRs [23]. In our study, most ADR concepts found in drug labels use normalized terms and controlled vocabularies (i.e., the FDA

publishes regulations governing the content and format of ADR information [10]). With the help of Vigibase, we manually create a gazetteer that contains 18,310 ADRs commonly found in the drug labels. We then randomly select one sentence from the ones previously collected from our gold standard and replace the ADR in that sentence with a different one from the gazetteer. If we have more instances in the gazetteer than the total number of sentences that contains ADRs in the gold standard, then we continue this step until all instances in the gazetteer are injected to the sentences in the gold standard, creating at least one pseudo-data sentence per original that contains an ADR. For example, the original sentence "the most common adverse reactions (>=20%), regardless of causality, were stomatitis, infections, and asthenia" becomes "the most common adverse reactions (>=20%), regardless of causality, were dizziness, pyrexia, and fatigue". We use the same method for pseudo-data generation for medication names, creating a gazetteer from Drugs.com which contains 5,394 commonly used medication names.

Generating Data-Driven Pseudo-data

This method takes advantage of patterns observed in the data. In particular, we study similarity of concepts using word embeddings. We utilize word embeddings trained on MIMIC-III (a large database of about 2 million clinical notes for about 46 thousand patients [24]). For each sentence containing an ADR/medication, we replace each token within the concept with the top three closest tokens represented by word embeddings (i.e., replacing each ADR/medication token with its first, second, and third nearest neighbor, respectively) to create pseudo-concepts. For multi-token concepts, we create multiple pseudo-concepts by combinatorically putting together the nearest neighbors of each of their constituent tokens. We measure word-to-word distances using Euclidean distance, as suggested by the literature [21] as a proxy for semantic similarity. This measure of similarity between tokens is a likelihood rather than a certainty; however, it has been widely-accepted and used in NLP with reliable results [21]. The purpose of the data-driven pseudo-data is to enrich the training data with more tokens that resemble ADRs/medication names [25-27]. We do not aim to replace each word with its synonym. Rather, we aim to replace each concept with a concept that is semantically similar or related [25-27], and is of the same type, i.e., replace a drug with a drug in order to improve learning contexts in which drugs occur.

Unlike knowledge-driven pseudo-data that is guaranteed to contain pseudo-concepts of the correct type, concepts that are related in the word embedding space may not all be concepts of the same type. In order to identify and utilize the most useful pseudo-concepts from data-driven pseudo-data, we use confidence thresholds and filter examples with lower confidence. For this, we start with training a supervised model using labeled data. We apply this model to predict concepts on the data-driven pseudo-data. We obtain confidence rates for each of the predicted labels and utilize sentences from pseudo-data that contain at least one ADR concept (or medication concept) that scores higher than the confidence threshold. We repeat this process until no ADRs/medications in the pseudo-data are introduced to the labeled data.

The literature suggests picking a confidence threshold of 0.90 or 0.95 [28] for filtering. In these cases, the system only utilizes high-confidence concepts from the pseudo-data. Such high thresholds would minimize noise that can come from incorrect pseudo-concepts. However, setting a high confidence threshold also potentially excludes some correctly predicted pseudo-concepts from the training model. Given sufficient samples of training data, the optimal threshold can be tuned via cross-validation. In our case, we observe that a threshold of 0.85 would filter out some potentially useful pseudo-concepts. Therefore, we set the threshold to 0.8 in order to be conservative without being overzealous. **Figure 1** shows some example pseudo-concepts.

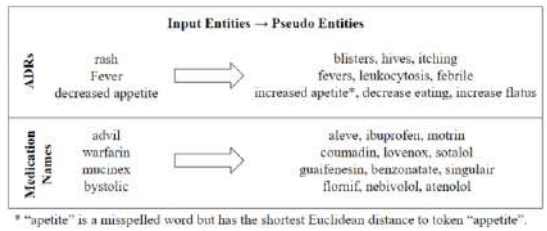


Figure 1 - Examples of data-driven pseudo-concepts

Generating Hybrid Pseudo-Data

Knowledge- and data-driven methods can generate different pseudo-data. Therefore, as a final step, we combine these two datasets to create a hybrid pseudo-dataset. **Figure 2** shows the workflow we designed to generate the hybrid pseudo-data.

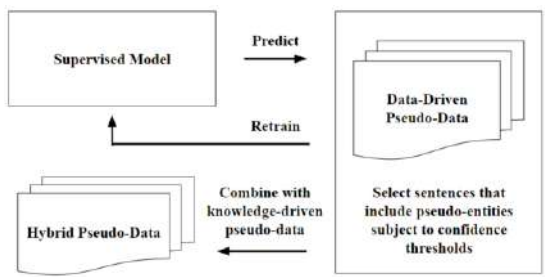


Figure 2 - Generating hybrid pseudo-data

Results and Discussion

We tune the parameters of our models with five-fold cross-validation on the training set. Then we train our IE system on the complete training set and evaluate on the test set for each task. Feature sets and window sizes remain identical to those described in Baseline IE System section.

Table 2 shows phrase-level F1-measure with and without pseudo-data on the test set for both IE tasks. When extracting ADRs and medication names, with respect to the baseline using no pseudo-data, the model performs significantly better when enhanced with either knowledge-driven or data-driven pseudo-data. The performance gain also spills over to other concept categories, with significant performance improvement on the severity class.

In medication name extraction, the model with data-driven pseudo-data performs significantly better than the model with knowledge-driven pseudo-data. However, there is no significant advantage of data-driven pseudo-data in ADR extraction when compared to knowledge-driven pseudo-data. Manual analysis of the system output and pseudo-data showed that medication names are mostly single token nouns in the MIMIC III vector set. As a result, these concepts can be used interchangeably without breaking the surrounding context and can benefit IE performance. In contrast, in spite of 53.60% of the ADRs being single-token concepts in the test set, many of

them have longer textual spans. Given that our data-driven methods utilize word embeddings at the token level, even after filtering through confidence thresholds, these pseudo-concepts may not benefit IE.

Our results and manual analyses suggest that data-driven pseudo-data are suitable for concept categories with short textual spans, whereas the knowledge-driven pseudo-data are effective for concept categories with longer textual spans, assuming that the gazetteer offers a good coverage of the concept types in focus. Hybrid pseudo-data puts the strengths of the two together and shows statistically-significant performance improvement on the extraction of both ADRs and medication names. The biggest gains come from improved performance on concepts that appear only once in the data. For example, on the ADR extraction task, 10.92% of ADR concepts fall under this category. Furthermore, 53.95% of these ADRs appear in the VigiAccess gazetteer. We also checked the ADRs that were not extracted by the model with knowledge-driven pseudo-data but by the model with hybrid pseudo-data. The model with hybrid pseudo-data performed particularly well on ADRs that are modified by adjectives or prepositional phrases. For instance, unlike the model with hybrid pseudo-data, the knowledge-driven model failed to recognize "peripheral sensory neuropathy" and "residual neuropathy" when "sensory" and "residual" are modifying the ADR term ("peripheral neuropathy").

Table 2 - Phrase-level F1-measure with and without pseudo-data on the test set for both IE tasks

Pseudo-Dataset	ADR	Animal	Drug	Factor	Negation	Severity
No	0.727	**0.738**	**0.398**	**0.540**	**0.308**	0.502
Gazetteer	0.744	**0.756**	**0.408**	**0.537**	**0.315**	0.546
Optimal Euclidean	0.745	**0.737**	0.349	0.488	**0.313**	0.515
Optimal Hybrid	**0.765**†	**0.737**	**0.381**	**0.538**	**0.308**	**0.559**†
Pseudo-Dataset	**Medication**	**Dosage**	**Mode**	**Frequency**	**Duration**	**Reason**
No	0.888	**0.926**	**0.936**	**0.941**	**0.558**	**0.426**
Gazetteer	0.891	**0.928**	**0.938**	**0.942**	**0.564**	**0.430**
Optimal Euclidean	0.899	**0.926**	**0.935**	**0.940**	**0.559**	**0.436**
Optimal Hybrid	**0.907**†	**0.929**	**0.942**	**0.940**	**0.561**	**0.432**

* Bold indicates the highest score in each categor. Numbers followed by † indicate statistical significance with respect to other results in the same column. We use a Z-score of 1.96 ($\alpha = 0.05$).

The model with hybrid pseudo-data also shows statistically significant gains on severity. However, it fails to show any performance gains against the baseline (differences are not statistically significant) on the rest of the minority classes. These concept categories have limited training samples therefore the classifier faces difficulties when generalizing patterns. For example, there are only 44 samples (5 unique samples) for Animal class in the training data. Therefore, the model fails to recognize other animal entities in the test data (e.g., "cynomolgus", "minipig", "ferret").

Error Analysis and Limitation

We analyzed the system's output to pinpoint incorrect predictions with respect to the gold standard. In medication extraction, we found the model still cannot distinguish some generic medication names (e.g., "his medication", "other antibiotics", "this medication", "trial drug"). We found 70 instances related to "medication(s)", 53 instances related to "antibiotic(s)", and 14 instances related to "home" in the remaining errors. Also, the system cannot recognize some ambiguous medication names. For example, the extraction of "No Doz" and "One A Day" is problematic because the first one starts with a negated token, whereas the second one looks like a medication frequency. In ADR extraction, generic ADRs are also a common theme in erroneous predictions (e.g., "gi adverse reactions", "autoimmune disorders", "interstitial lung disease"). Furthermore, some ADR concepts contain (or are surrounded by) multiple generic concepts and are syntactically complex, (e.g., "irreversible vision damage with vision damage from SABRIL", "structure infections with moderate/severe pre-existing renal impairment", "drug reaction with eosinophilia and systemic symptoms"). What is more, some concepts contain negated tokens and mislead the system. For example, the model failed to extract "negative thoughts" and "non-melanoma skin cancer" due to the included negation signals, "negative" and "non".

Other errors arise because of misleading lexical context. For example, "alcoholic beverage" and "vomiting" are both predicted as medication names when appearing respectively in "she took 600 ml alcoholic beverage for the day" and "vomiting x 5 hours". Similarly, the system does not recognize animal concept "rodent" in "rodent models inhibition of IL-12/IL-23p40" because it is presented with a protein product. Finally, our system still misses concepts that have longer textual spans or lack temporal signals (e.g., "disruption of the body ability to reduce core body temperature", "progression of pre chronic myelomonocytic leukemia with nras mutation", "after your cardiac catheterization", "possibly lifelong").

In addition, our method faces its biggest challenges when differentiating certain concepts that look like ADRs but are not caused by a reaction to a drug (and are therefore not ADRs). These include: (1) Concepts that look like ADRs but actually refer to the medical problems of patients. For example, in "fever and persistent constipation upon admission", where the underlined concepts are not ADRs. This is the most prominent type of challenge. (2) Hypothetical or negated ADRs that do not actually occur. For example, "evaluate if neuropathy is suspected" and "if muscle signs and symptoms persist".

Our future work will focus on these challenges and explore more comprehensive knowledge-bases, such as the Unified Medical Language System (UMLS) [29] as well as phrase-level [30] or sentence-level embeddings [31] that can generate more reliable pseudo-data for the training models.

Conclusion

In this paper, we evaluated the contribution of pseudo-data to clinical IE, using ADR and medication extraction as test bed applications. We developed two methods for generating pseudo-data. The first method is knowledge-driven and uses gazetteers that contain clinically-relevant concepts, whereas the second method is data-driven and takes advantage of word embeddings. We found that the data-driven method is suitable for concept categories with short textual spans, whereas the knowledge-driven methods are effective for concept categories with longer textual spans, assuming the gazetteers offer good coverage of the concepts within each category. Hybrid pseudo-data takes advantage of strengths of both, improving performance on the extraction of two important concept categories in clinical IE: ADRs and medication names.

Acknowledgements

This work was supported by Philips Research North America, PI: Peter Szolovits. The content is solely the responsibility of

the authors and does not necessarily represent the official views of the sponsors.

References

[1] Mønsted T, Reddy MC, Bansler JP. The use of narratives in medical work: A field study of physician-patient consultations. In ECSCW 2011: Proceedings of the 12th European Conference on Computer Supported Cooperative Work, 24-28 September 2011, Aarhus Denmark 2011 (pp. 81-100). Springer, London.

[2] Uzuner Ö, Stubbs A. Practical applications for natural language processing in clinical research: The 2014 i2b2/UTHealth shared tasks. Journal of biomedical informatics. 2015 Dec;58(Suppl):S1.

[3] Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, Chute CG. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. Journal of the American Medical Informatics Association. 2010 Sep 1;17(5):507-13.

[4] Uzuner Ö, Solti I, Cadag E. Extracting medication information from clinical text. Journal of the American Medical Informatics Association. 2010 Sep 1;17(5):514-8.

[5] Capurro D, van Eaton E, Black R, Tarczy-Hornoch P. Availability of structured and unstructured clinical data for comparative effectiveness research and quality improvement: a multisite assessment. EGEMS. 2014;2(1).

[6] Klassen P, Xia F, Vanderwende L, Yetisgen M. Annotating clinical events in text snippets for phenotype detection. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014) 2014.

[7] Mehrabi S, Krishnan A, Roch AM, Schmidt H, Li D, Kesterson J, Beesley C, Dexter P, Schmidt M, Palakal M, Liu H. Identification of patients with family history of pancreatic cancer-Investigation of an NLP System Portability. Studies in health technology and informatics. 2015;216:604.

[8] Cornish PL, Knowles SR, Marchesano R, Tam V, Shadowitz S, Juurlink DN, Etchells EE. Unintended medication discrepancies at the time of hospital admission. Archives of internal medicine. 2005 Feb 28;165(4):424-9.

[9] Wilke RA, Xu H, Denny JC, Roden DM, Krauss RM, McCarty CA, Davis RL, Skaar T, Lamba J, Savova G. The emerging role of electronic medical records in pharmacogenomics. Clinical Pharmacology & Therapeutics. 2011 Mar;89(3):379-86.

[10] Drug OD. Guidance for Industry. Center for Drug Evaluation and Research (CDER). 2011 Oct;1000.

[11] Roberts K, Demner-Fushman D, Tonning JM. Overview of the TAC 2017 Adverse Reaction Extraction from Drug Labels Track. In Proceedings of the 2017 Text Analysis Conference. 2017 Nov 13-14. National Institute of Standards and Technology.

[12] Friedman C, Johnson SB. Natural language and text processing in biomedicine. In Biomedical Informatics 2006 (pp. 312-343). Springer, New York, NY.

[13] Tao C. Clinical information extraction from unstructured free-texts. State University of New York at Albany. 2018.

[14] Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR, Denny JC. MedEx: a medication information extraction system for clinical narratives. Journal of the American Medical Informatics Association. 2010 Jan 1;17(1):19-24.

[15] Patrick J, Li M. A cascade approach to extracting medication events. In Proceedings of the Australasian Language Technology Association Workshop 2009 (pp. 99-103).

[16] Tao C, Filannino M, Uzuner Ö. Prescription extraction using CRFs and word embeddings. Journal of biomedical informatics. 2017 Aug 1;72:60-6.

[17] Xu J, Lee HJ, Ji Z, Wang J, Wei Q, Xu H. UTH_CCB System for Adverse Drug Reaction Extraction from Drug Labels at TAC-ADR 2017. In Proceedings of the 2017 Text Analysis Conference. 2017 Nov 13-14. National Institute of Standards and Technology.

[18] Chawla NV, Bowyer KW, Hall LO, et al. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research. 2002 Jun 1;16:321-57.

[19] Keretna S, Lim CP, Creighton D, et al. Enhancing medical named entity recognition with an extended segment representation technique. Computer methods and programs in biomedicine. 2015 Apr 1;119(2):88-100.

[20] Japkowicz N. The class imbalance problem: Significance and strategies. In Proceedings of the International Conference on Artificial Intelligence 2000 Jun 26.

[21] Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) 2014 (pp. 1532-1543).

[22] Okazaki N. Crfsuite: a fast implementation of conditional random fields (CRFs). 2017.

[23] Tao C, Lee K, Filannino M, Buchan K, Uzuner Ö. Extracting and Normalizing ADRs from Drug Labels. In Proceedings of the 2017 Text Analysis Conference. 2017 Nov 13-14. National Institute of Standards and Technology.

[24] Johnson AE, Pollard TJ, Shen L, Li-wei HL, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG. MIMIC-III, a freely accessible critical care database. Scientific data. 2016 May 24;3:160035.

[25] Henry S, Cuffy C, McInnes BT. Vector representations of multi-word terms for semantic relatedness. Journal of biomedical informatics. 2018 Jan 1;77:111-9.

[26] McInnes BT, Pedersen T. Evaluating semantic similarity and relatedness over the semantic grouping of clinical term pairs. Journal of biomedical informatics. 2015 Apr 1;54:329-36.

[27] Henry S, McQuilkin A, McInnes BT. Association measures for estimating semantic similarity and relatedness between biomedical concepts. Artificial intelligence in medicine. 2018 Sep 7.

[28] Abney, S. Semi-supervised learning for computational linguistics. 2007. CRC Press.

[29] Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic acids research. 2004 Jan 1;32(suppl_1):D267-70.

[30] Yu M, Dredze M. Learning composition models for phrase embeddings. Transactions of the Association for Computational Linguistics. 2015 Dec;3:227-42.

[31] Conneau A, Kruszewski G, Lample G, Barrault L, Baroni M. What you can cram into a single vector: Probing sentence embeddings for linguistic properties. arXiv preprint arXiv:1805.01070. 2018 May 3.

Addresses for Correspondence

Carson Tao, mtao@albany.edu

UAB 431, 1400 Washington Avenue, Albany, NY, USA, 12222

Özlem Uzuner, ouzuner@gmu.edu

Nguyen Engineering Building, Rm. 5359, 4400 University Drive, Fairfax, VA, USA, 22030

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190250

Identifying Diabetes in Clinical Notes in Hebrew: A Novel Text Classification Approach Based on Word Embedding

Maxim Topaz [a,b], Ludmila Murga [c], Chagai Grossman [d], Daniella Daliyot [d], Shlomit Jacobson [d], Noa Rozendorn [d], Eyal Zimlichman [d], Nadav Furie [d]

[a] *School of Nursing & Data Science Institute, Columbia University, New York City, NY, USA*
[b] *Visiting Nurse Service of New York, New York City, NY, USA*
[c] *Cheryl Spencer Department of Nursing, University of Haifa, Haifa, Israel*
[d] *Sheba Medical Center, Tel Hashomer, Israel.*

Abstract

NimbleMiner is a word embedding-based, language-agnostic natural language processing system for clinical text classification. Previously, NimbleMiner was applied in English and this study applied NimbleMiner on a large sample of inpatient clinical notes in Hebrew to identify instances of diabetes mellitus. The study data included 521,278 clinical notes (one admission and one discharge note per patient) for 268,664 hospital admissions to medical-surgical units of a large hospital in Israel. NimbleMiner achieved overall good performance (F-score =.94) when tested on a gold standard human annotated dataset of 800 clinical notes. We found 15% more patients with diabetes mentioned in the clinical notes compared with diagnoses data. Our findings about underreporting of diabetes in the coded diagnoses data highlight the urgent need for tools and algorithms that will help busy providers identify a range of useful information, like having a diabetes.

Keywords:
Natural language processing, Text classification, Diabetes.

Introduction

With the wide international adoption of electronic health records and other information technology in healthcare, there is an exponential growth in health data. It is estimated that as much as 50 to 80% of all healthcare data are captured as unstructured, mostly free text narrative data. Examples of health narrative data include daily inpatient progress notes that are completed by physicians and nurses, primary care clinics follow-up notes, radiology notes, surgical notes, etc. [1].

With large potential benefits that are projected from using these data, these exponentially growing narrative information sources pose several significant challenges. Busy clinicians are now required to go through an increasing number of narrative notes about a patient to understand their medical history and diagnose their conditions correctly [2]. Health researchers are dealing with an increasing number of clinical notes they need to review in order to identify important aspects of care and patient characteristics [2].

Some of these challenges can be solved through natural language processing (NLP) with a variety of techniques aimed at extracting meaning from narrative data [1]. NLP can be broadly divided into linguistic approaches and statistically based approaches. Linguistic NLP systems tend to use complex language characteristics, for example large vocabularies of pre-defined terms and expressions, to identify a concept of interest in clinical texts. For example, past studies extracted wound characteristics from outpatient notes [3] and identified patients with poor treatment adherence in inpatient notes [4] using pre-build vocabularies of clinical terms for these domains. Statistically-based NLP approaches - sometimes also called text classification- are often based on probabilistic estimations of a presence of a specific phenomenon in the clinical texts. For example, text classification was used to classify clinical notes in terms of information completeness and adequacy [5] or to identify depression and depression symptoms among patients [6].

Although significant advancements in clinical NLP were made over the last decades, most of the published NLP studies and existing NLP systems focus on processing information in English. In general, a vast majority of languages around the world are considered low-resource languages in terms of NLP [7]. More tools and approaches are urgently needed to overcome the resource barrier so advances in NLP can deliver more widespread benefits to health providers, researchers and patients [8].

This paper presents a novel open-source NLP system called NimbleMiner. NimbleMiner was previously tested in text classification tasks in English. For example, we successfully applied NimbleMiner to identify fall-related information in clinical notes, while achieving higher system performance in significantly less time compared to a previous rule-based systems [9]. By design, NimbleMiner implements a language-agnostic NLP approach and we estimated that our system can be applied to languages other than English. To evaluate this assumption, this study applied NimbleMiner to process clinical narratives in Hebrew.

Similar to many other languages, Hebrew has very scarce resources to implement NLP [8]. In healthcare, we found only one published effort that reported building Hebrew NLP healthcare pipeline in early 2010 [10] and since then, no other reports were identified. The aim of this study was to apply NimbleMiner on a large sample of inpatient clinical notes in Hebrew to identify instances of diabetes mellitus (both insulin dependent type I diabetes and insulin resistant type II diabetes). We worked with clinicians in the largest hospital in Israel (Sheba Medical Center) to develop the NLP pipeline and evaluate its performance.

Methods

Our methods are summarized in Figure 1. We first manually reviewed a random sample of admission and discharge narrative notes for 400 patients, identified diabetes cases, and created a gold standard of human annotated notes with the help of clinical experts. We then trained and tested our NLP system—called NimbleMiner—and classification algorithms. Finally, we assessed our automated approach by comparing system-generated classification against the gold standard. We also compared diabetes cases extracted from narrative data versus coded diagnoses for pateints included in the gold standard.

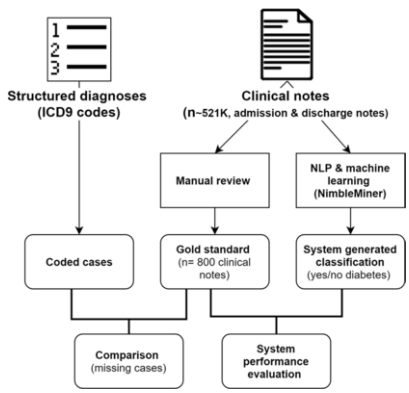


Figure 1– *Study Methods Overview*

Dataset

We extracted a large sample of clinical notes of all the patients admitted to the Sheba Medical Center over two years 2016-17. Sheba Medical Center located in the Tel Aviv District, is the largest hospital in Israel with about 1,700 beds, over 1,400 physicians, and 2,600 nurses. Only patients admitted to medical-surgical units were included in this study. We excluded any pediatric patients or women who were admitted for pregnancy/delivery reasons. Over the study time period, there were 268,664 admissions for 127,851 unique patients. Patients' average age was 65.1.

In this project, we used a subset of admission and discharge narrative clinical notes for all the patients in the sample. There were 521,278 clinical notes (one admission and one discharge note per patient's admission) after excluding missing data. Admission notes consisted of narrative description of patient's state at the hospital admission, including previous medical history, patient's diagnoses, medications, reason for admission, patient's signs and symptoms, relevant social history, etc. Discharge notes included information similar to the admission note, with addition of hospital treatment course and discharge recommendations. On average, clinical notes had roughly 2,000 characters, with admission notes being slightly longer than discharge summaries (about 200 charters longer on average). It is important to note that the written Hebrew alphabet does not include vowels, thus writing is more concise compared to other languages. This study received an Institutional Review Board approval from the University of Haifa, Israel.

Domain Definitions

We used clinical literature [11] and our team's expertise to define the domain of interest. In this project, we focused on identifying clinical notes where diabetes is very likely to be described. To accomplish that, we defined several categories of terms that will suggest high likelihood of diabetes presence, as follows:

- **Diagnosis of diabetes**: any mention of patient's having diabetes in the text (e.g., terms like "diabetes mellitus", abbreviations like "DM", etc.).
- **Mention of one or more diabetic medications** in the text: including – Insulin derivatives (e.g., Lantus, Apidra, Novorapid, Actrapid), Insulin pump, Glucomin, Glucophage, Methformin, Januet, Jardiance, Sulfanyl urea, Gluben, Forxiga, DPP4, Januvia, GLP1, Victoza, Trajenta, etc.).
- **Indication of hemoglobin A1c** >6.5% or **fasting glucose levels** > 125 mg/dL in the text.
- **Diabetes complications** in the text (e.g., diabetic foot, diabetic neuropathy, diabetic retinopathy, etc.).

NLP System Description

NimbleMiner is an open source NLP system developed by our team [9]. User manual and download options can be accessed at: http://github.com/mtopaz/NimbleMiner. Other research or clinical teams can use the system under the GNU General Public License v3.0. NimbleMiner includes several methodological stages of clinical note processing that are briefly described below and presented in Figure 2.

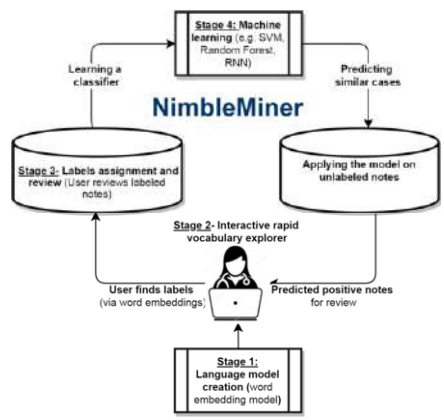


Figure 2– *NimbleMiner System Process Stages*

Stage 1- Language model creation: The user selects a large corpus of clinical notes and defines language model characteristics. We use a word embedding model (specifically skip-gram model) for language model generation and users can change model settings based on their preferences. **Stage 2- Interactive rapid vocabulary explorer**: The user enters a query term of interest, and the system returns a list of similar terms it identified as relevant. The list of suggested similar terms is based on the cosine term similarity metric extracted from the word embedding model. The user selects and saves the relevant terms by clicking on them. Negated or other irrelevant terms and expressions that are not selected by the user are saved by the system for further tasks, such as negation detection. Figure 3 describes the steps of the vocabulary explorer stage. **Stage 3- Labels assignment and review**: The system uses the stage 2 discovered similar terms to assign labels to clinical notes (while excluding notes with negations and other irrelevant terms). Assigning a positive label means that a concept of interest is present in the clinical note. When needed, the user reviews and updates lists of similar terms and negated similar terms. The user reviews the clinical notes with assigned labels for accuracy. This weakly supervised rapid labeling approach

is based on a positive labels learning framework validated in previous research [12,13]. **Stage 4- Machine learning:** The user choses a machine learning algorithm to be applied to create a predictive model (e.g., Support vector machines, recurrent neural network, random forest). The model is then applied to predict which clinical notes might have the concept of interest. The user reviews the predicted notes and can go through stages 2-4 again to add new labels.

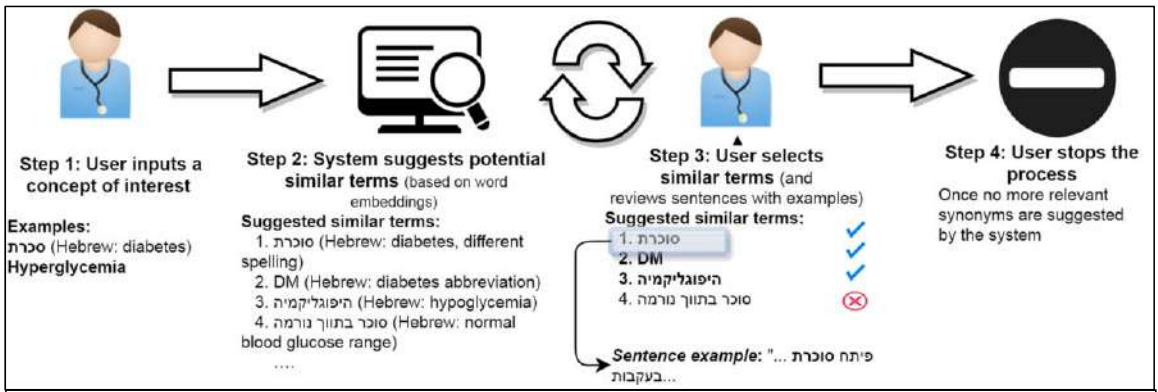


Figure 3– *NimbleMiner Stage 2 -Interactive Rapid Vocabulary Explorer Process Overview*

Legend*: Interactive rapid vocabulary explorer: The user enters a query term of interest (step 1), and the system returns a list of similar terms it identified as relevant (step 2). The list of suggested similar terms is based on the cosine term similarity metric extracted from the word embedding model. The user selects and saves the relevant terms while negated or other irrelevant terms that were not selected by the user are also saved in the system for further tasks (step 3). The system iteratively identifies new potential similar terms and presents them to the user for review (steps 2-3) and the process continues until there no new similar terms to suggest (step 4)*

NLP System Settings

Narrative data were pre-processed by removing punctuation and lowercasing all English words (Hebrew letters do not have upper case). All numeric symbols were kept in the text. Additionally, we converted frequently co-occurring words in the clinical notes into phrases with lengths of up to four words (4-grams) [14]. This is a common process in NLP where sets of co-occurring words are combined into phrases. For example, "pt has diabetes" might be a common 3-gram. We used a phrase2vec algorithm with default settings to implement this in NimbleMiner.

NimbleMiner has a user interface that is implemented in R statistical package. We used a skip-gram model implementation called word2vec and phrase2vec to create a word embedding model in R [14]. Parameters of the word embedding model were held constant based suggestions in other studies of word embedding [15]. Specifically, we used a model with window width size = 10, vector dimension = 100, minimum word count = 5, negative sample size = 5, and sub-sampling = 1e-3. We used all the available clinical documents (~ 521K clinical notes) to train the model.

For each similar term entered by the user, the system presented 50 potentially similar terms based on the cosine term similarity. Our previous experiments [9] showed that the random forest algorithm outperforms other approaches (e.g., J48 Decision trees, Support Vector Machines), hence we used this algorithm in the machine learning stage of this study. Random Forest algorithm was used with default settings (number of iterations = 100, minimum number of instances = 1, minimum variance for split = 1e-3, depth = unlimited).

To identify negation terms for medical domain in Hebrew, we started from translating a commonly used vocabulary of negation terms in English, called NegEx [16]. Next, we used our clinical and Hebrew linguistic knowledge to identify additional negation terms. We then used NimbleMiner's vocabulary explorer to identify other potential negations and variations in negations spelling. Our final list of medical negations in Hebrew includes 118 terms, which might appear before or/and after a diabetes term (e.g., "שולל"- denies or "ללא"- has no). n-grams (irrelevant terms not selected by the user) were used to detect family and past history contexts.

NLP Implementation

Similar terms were explored by a team of 2 nurses and 2 physicians who used NimbleMiner. First, two nurses and one physician used NimbleMiner to identify and review potentially relevant terms independently (NimbleMiner stage 2) and then the lists of terms were combined into one list. Another physician then reviewed the combined list and finalized any disagreements or additional terms that appeared in one list but not in the other.

System Evaluation- Gold Standard Creation

We extracted a random sample of admission and discharge notes for 400 patients in our database (800 clinical notes total). Each note was reviewed by 3 clinicians on our team (2 nurses and 1 doctors) for presence of diabetes. Clinicians were asked to use their clinical judgment and indicate if each specific note has a mention of diabetes. There was moderate inter-rater agreement (Cohen's Kappa inter-rater agreement = .69). A senior internal medicine physician then reviewed each of the cases and adjudicated whether diabetes was present or absent in cases of disagreement. This corpus of documents was not used for algorithm training but only as a gold standard for algorithms evaluation.

We applied our NLP algorithms on the gold standard dataset to predict presence of diabetes in each of the documents. We calculated precision (defined as the number of true positives out of the total number of predicted positives), recall (defined as the number of true positives out of actual number of positives) and F-score (F1, weighted harmonic mean of the precision and recall) to evaluate and compare the performance of our algorithms.

Comparison with Structured Diagnosis Data

In this project, we only had access to coded diagnoses of patients included in the gold standard set. Based on expert judgment, we labeled each note as either positive or negative for presence of diabetes based on information identified in the clinical notes. We then compared our results (per hospital admission and discharge) with structured data on patient diagnoses. Diagnoses data included primary diagnosis at admission, a list of secondary diagnoses, and discharge diagnoses (that might be different from admission diagnoses because they were updated during hospitalization). All diagnoses data were stored as International Calcification of Diseases version 9 (ICD 9) codes [17] and we used all codes from category 250.XX (diabetes mellitus) and V58.67 (Long term, current insulin use).

Results

Overall, 1,427 terms and expressions were discovered for diabetes using NimbleMiner. There was a moderate inter-rater agreement (Cohen's Kappa inter-rater agreement = .69) on the relevant terms between the three clinicians. Another physician reviewed all the terms, forming the final terms corpus. Table 1 presents the number of terms identified for each diabetes category. Since the system is using cosine similarity as a measure of term relatedness, our approach identified terms in Hebrew, English, misspellings in both languages, abbreviations and other lexical variants specific to Hebrew (for example specific Hebrew suffixes indicative of male or female).

Table 1– Categories and Examples of Diabetes Expressions

Domain	Number of terms	Examples
Diagnosis of diabetes	928	"dm 2", "סכרת type 2" (diabetes type 2), "ברקע גבולית סכרת" (diabetic background)
Diabetic medications	358	"באינסולין מטופלת" (uses insulin), "lantus במינון", "ב glucophage"
Hemoglobin A1c >6.5% or fasting glucose levels > 125 mg/dL	49	"hba1c>6.5", "מעל סוכר 180" (blood sugar above 180)
Diabetes complications	92	"diabetic foot", "נפרופתיה סוכרתית" (diabetic nephropathy)
Total	**1,427**	

NLP System Performance Evaluation

We first evaluated NimbleMiner's performance using just the labeling process (see NimbleMiner stage 3, weakly supervised labeling) and then compared these results with random forest algorithm performance. Table 2 shows that NimbleMiner's labeling alone showed better results than further machine learning or codes alone.

Table 2– Codes and NLP performance on the gold standard

	Diagnoses codes	NimbleMiner labeling	Random forest
Precision	.78	**.91**	.87
Recall	.85	**.96**	.86
F-score	.82	**.94**	.87

Comparison with Structured Diagnosis Data

Overall, we found 176 patients with diabetes in the coded diagnoses data while our experts found 159 patients with diabetes based on clinical notes review (800 clinical notes included in the gold standard). This difference is explained by the fact that diabetes for these patients (n=17) was coded, but was not mentioned in the text. On the other hand, we found that our experts identified an additional 15% (n= 26) more patients with diabetes described in the text compared to the diagnoses data.

Discussion

This study is one of the first to implement NLP in medical domain in Hebrew. In general, NimbleMiner's approach proved useful for clinical narratives in Hebrew. First, our team of clinicians have successfully interacted with NimbleMiner to identify a large vocabulary of terms and expressions describing diabetes in clinical notes. This list included misspellings, expressions in both Hebrew and English, specific Hebrew lexical variants, etc. Compiling such a vocabulary is often a challenge in NLP projects and it is usually conducted thorough literature review, reading large amounts of clinical notes, etc. [8]. Our approach enables clinicians to identify a comprehensive vocabulary of terms specific to their domain within a short period of time, while being assisted by machine learning components.

This comprehensive vocabulary of terms is then applied to label clinical notes. During labeling, negated terms are removed from positively labeled cases. We also remove irrelevant terms that were reviewed by the user during the vocabulary exploration process (stage 2). For example, similar terms related to presence of diabetes, like "dm 2" or are sometimes included in a larger irrelevant term, like "family history of dm 2", "son has dm 2", or "screened for dm2". These irrelevant terms are excluded, which helps our system to conduct high quality labeling. Our results support the current literature suggesting that machine learning with human-in-the-loop is an effective approach in healthcare.[18,19] Human experts can interact with machine learning to reduce NLP task complexity and improve machine learning speed.

Machine learning is then performed using the set of positively labeled clinical notes and similar in size sample of notes labeled as "unknown" to train the algorithm. In this project, our labeling approach outperformed random forest algorithm. One possible explanation is that for domains with relatively straightforward words and expressions, like presence of a diagnosed disease, rule-based approaches like ours work better than machine learning. On the other hand, machine learning can work better for domains where more ambiguity is present. Further research is needed to validate these claims. Other studies of NLP in the diabetes domain in English achieved similar NLP system performance.[20,21]

Importantly, our results indicate that about 15% of patients with diabetes mentioned in the clinical notes, did not have a diabetes-related diagnoses in the coded data. This is a serious concern for various reasons. First, not having a diabetes diagnosis might affect patient safety. For example, diabetic patients might be given a contra-indicated medication or provided with an inappropriate diet. Additionally, there are several significant implications for the organization in terms of reimbursement for patient care. When diagnoses are underreported to the health services payer—whether it is an insurance or ministry of health—medical institutions can potentially receive less reimbursement for the care they

provided. Third, underreported major diagnoses—like diabetes—pose significant limitations for any research that is conducted with clinical data.

Limitations: This study has several important limitations. First, we have only experimented with one data mining algorithm and more work is needed to examine more algorithms, such as Support Vector Machines. In addition, the use of NimbleMiner might be restricted to classifying documents based on phrases with high positive predictive value, such as diabetes terms and expressions, and further work is needed to explore the generalizability of our approach.

Conclusions

Our pioneering results indicate the feasibility of an NLP approach that can span across different languages, like English and Hebrew. Our finding about underreporting of diabetes in the coded diagnoses data highlight the urgent need in tools and algorithms that will help busy providers identify a range of useful information, like having diabetes. Identifying this information can also help health organizations receive optimal reimbursement for their services while health researchers will glean a more accurate picture of the patient population for research.

Acknowledgements

Funding for this work was provided through Sheba Innovation Fund.

References

[1] D. Demner-Fushman, W.W. Chapman, and C.J. McDonald, What can natural language processing do for clinical decision support?, *J. Biomed. Inform.* **42** (2009) 760–72. doi:10.1016/j.jbi.2009.08.007.

[2] N. Griffon, J. Charlet, and S.J. Darmoni, Managing Free Text for Secondary Use of Health Data, *Yearb. Med. Inform.* **23** (2014) 167–169. doi:10.15265/IY-2014-0037.

[3] M. Topaz, K. Lai, D. Dowding, V.J. Lei, A. Zisberg, K.H. Bowles, and L. Zhou, Automated identification of wound information in clinical notes of patients with heart diseases: Developing and validating a natural language processing application, *Int. J. Nurs. Stud.* **64** (2016). doi:10.1016/j.ijnurstu.2016.09.013.

[4] M. Topaz, K. Radhakrishnan, S. Blackley, V. Lei, K. Lai, and L. Zhou, Studying Associations Between Heart Failure Self-Management and Rehospitalizations Using Natural Language Processing., *West. J. Nurs. Res.* (2016). doi:10.1177/0193945916668493.

[5] W. Al Assad, M. Topaz, J. Tu, and L. Zhou, The application of machine learning to evaluate the adequacy of information in radiology orders, in: 2017 IEEE Int. Conf. Bioinforma. Biomed., IEEE, 2017: pp. 305–310. doi:10.1109/BIBM.2017.8217668.

[6] L. Zhou, A.W. Baughman, V.J. Lei, K.H. Lai, A.S. Navathe, F. Chang, M. Sordo, M. Topaz, F. Zhong, M. Murrali, S. Navathe, and R.A. Rocha, Identifying Patients with Depression Using Free-text Clinical Documents, 2015. doi:10.3233/978-1-61499-564-7-629.

[7] S. Abney, and S. Bird, The Human Language Project: Building a Universal Corpus of the World's Languages, *Proc. 48th Annu. Meet. Assoc. Comput. Linguist.* (2010) 88–97. https://aclanthology.coli.uni-saarland.de/papers/P10-1010/p10-1010 (accessed November 21, 2018).

[8] A. Névéol, H. Dalianis, S. Velupillai, G. Savova, and P. Zweigenbaum, Clinical Natural Language Processing in languages other than English: opportunities and challenges, *J. Biomed. Semantics*. **9** (2018) 12. doi:10.1186/s13326-018-0179-8.

[9] M. Topaz, L. Murga, K.M. Gaddis, M. V. McDonald, O. Bar-Bachar, Y. Goldberg, and K. Bowles, Mining fall-related information in clinical notes: comparison of rule-based and novel word embedding-based machine learning approaches, *J. Biomed. Inform.* (2019) 103103. doi:10.1016/J.JBI.2019.103103.

[10] R. Cohen, Y. Goldberg, and M. Elhadad, Transliterated pairs acquisition in medical Hebrew, in: Mach. Transl. Morphol. Lang. Work., 2011.

[11] American Diabetes Association, Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2018., *Diabetes Care*. **41** (2018) S13–S27. doi:10.2337/dc18-S002.

[12] Y. Halpern, S. Horng, Y. Choi, and D. Sontag, Electronic medical record phenotyping using the anchor and learn framework., *J. Am. Med. Inform. Assoc.* **23** (2016) 731–40. doi:10.1093/jamia/ocw011.

[13] C. Elkan, and K. Noto, Learning Classifiers from Only Positive and Unlabeled Data, *KDD* . (2008).

[14] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, and J. Dean, Distributed Representations of Words and Phrases and their Compositionality, (2013) 3111–3119. http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality (accessed October 2, 2017).

[15] B. Chiu, G. Crichton, A. Korhonen, and S. Pyysalo, How to Train good Word Embeddings for Biomedical NLP, in: Proc. 15th Work. Biomed. Nat. Lang. Process., 2016: pp. 166–174. doi:10.18653/v1/W16-2922.

[16] W.W. Chapman, W. Bridewell, P. Hanbury, G.F. Cooper, and B.G. Buchanan, A simple algorithm for identifying negated findings and diseases in discharge summaries, *J. Biomed. Inform.* **34** (2001) 301–310. doi:10.1006/jbin.2001.1029.

[17] WHO, WHO | International Classification of Diseases (ICD), (2014). http://www.who.int/classifications/icd/en/ (accessed February 23, 2015).

[18] A. Holzinger, Interactive machine learning for health informatics: when do we need the human-in-the-loop?, *Brain Informatics*. **3** (2016) 119–131. doi:10.1007/s40708-016-0042-6.

[19] A. Holzinger, From Machine Learning to Explainable AI, in: 2018 World Symp. Digit. Intell. Syst. Mach., IEEE, 2018: pp. 55–66. doi:10.1109/DISA.2018.8490530.

[20] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, Machine Learning and Data Mining Methods in Diabetes Research, *Comput. Struct. Biotechnol. J.* **15** (2017) 104–116. doi:10.1016/J.CSBJ.2016.12.005.

[21] S.M. Meystre, and P.J. Haug, Comparing natural language processing tools to extract medical problems from narrative text., *AMIA ... Annu. Symp. Proceedings. AMIA Symp.* **2005** (2005) 525–9. http://www.ncbi.nlm.nih.gov/pubmed/16779095 (accessed March 19, 2019).

Address for correspondence

Max Topaz, mt3315@cumc.columbia.edu.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190251

The Value of Aggregated High-Resolution Intraoperative Data for Predicting Post-Surgical Infectious Complications at Two Independent Sites

Roshan Tourani[a], Dennis H. Murphree[b], Genevieve Melton-Meaux[a,c], Elizabeth Wick[d], Daryl J. Kor[e], Gyorgy J. Simon[a,f]

[a] *Institute for Health Informatics, University of Minnesota, Twin Cities, MN*
[b] *Department of Health Sciences Research, Mayo Clinic, MN*
[c] *Department of Surgery, University of Minnesota, Twin Cities, MN*
[d] *Department of Surgery, University of California San Francisco, CA*
[e] *Department of Anesthesia, Mayo Clinic, MN*
[f] *Department of Medicine, University of Minnesota, MN*

Abstract

Surgical procedures carry the risk of postoperative infectious complications, which can be severe, expensive, and morbid. A growing body of evidence indicates that high-resolution intraoperative data can be predictive of these complications. However, these studies are often contradictory in their findings as well as difficult to replicate, suggesting that these predictive models may be capturing institutional artifacts. In this work, data and models from two independent institutions, Mayo Clinic and University of Minnesota-affiliated Fairview Health Services, were directly compared using a common set of definitions for the variables and outcomes. We built perioperative risk models for seven infectious post-surgical complications at each site to assess the value of intraoperative variables. Models were internally validated. We found that including intraoperative variables significantly improved the models' predictive performance at both sites for five out of seven complications. We also found that significant intraoperative variables were similar between the two sites for four of the seven complications. Our results suggest that intraoperative variables can be related to the underlying physiology for some infectious complications.

Keywords:

Postoperative complications, Machine learning.

Introduction

Surgical procedures carry the risk of postoperative infectious complications, which can be severe, expensive, and can put patients' lives at risk [1-5]. With the ubiquity of decision support tools integrated into electronic health record (EHR) systems, perioperative decision support capabilities represent a promising direction for reducing postoperative risk or tailoring care including interventions aimed at individualized anticipation and management of the complication risk.

Initial surgical planning is primarily based on a preoperative assessment. While clearly essential, this assessment is insufficient to capture dynamic risks or changes in patient status, which may occur in the operating room. Preoperative assessment is, by definition, based on preoperative data and the planned procedure. Surgeries can deviate from the original plan for a variety of reasons or may involve unexpected physiologic changes such as bleeding, rendering the original risk estimates less valid. While the original estimates reflect the general risk of the patient and the planned surgery, they reflect neither new findings, changes in patient status, nor an altered course of surgery.

A growing body of published literature indicates that high-resolution intraoperative data can be predictive of complications [6-9]. However, the findings in these studies often contradict each other, are difficult to replicate at other sites, and are often based on very limited use of intraoperative data. For example, several studies have reported intraoperative hypothermia increases the risk of post-surgical infectious complications [6-8]. On the contrary, other studies found that there were no differences in the minimum, maximum, or ending temperatures between patients with surgical site infection (SSI) and those without – potentially due to an institutional effect: better temperature control at certain institutions. While temperature was unassociated with the surgical site infection, other intraoperative factors, most prominently blood loss, showed a significant association [9]. Besides institutional differences in clinical practice, differences in findings may also arise due to differing analytical approaches. Studies differ in the way they aggregate intraoperative variables: some studies use moments and extrema (minimum and maximum), while others use the ending measurement (the last measurement or the average of the last few measurements). Finally, differences in findings can also stem from discrepancies in definitions of the variables and outcomes.

Several high-quality registries form the backbone of surgical quality improvement research aimed at understanding and reducing postoperative complications. The National Surgical Quality Improvement Project (NSQIP) registry [10, 11] uses high-quality manually curated data [1, 12] and stands out as the gold-standard for surgical quality improvement and surgical outcomes research. NSQIP offers standardized definitions for exposure/risk variables and outcomes, helping to ensure consistency across studies. Unfortunately, it lacks detailed information regarding intraoperative risk factors, including physiologic data, laboratory data, medications, or other treatments, limiting its use for our purpose.

We conducted this study at two independent sites, Mayo Clinic and University of Minnesota-affiliated Fairview Health Services. These are two large Midwestern academic health systems with tertiary centers providing a wide range of surgical services. Both sites are members of the NSQIP registry; so, we were able to use outcome data from NSQIP patients at each site, serving as a high quality gold-standard. We used preoperative and intraoperative EHR data collected from the

clinical data repositories of the respective sites and standardized the data elements across the two sites to facilitate direct comparison of the models.

The study aims to answer the following questions: (i) In the context of perioperative decision support, does the use of intraoperative data improve the performance of 30-day postoperative risk models? (ii) Do significant intraoperative variables in the risk models pertain to the same physiological concepts across the two sites? We proceed with the assumption that when significant intraoperative variables differ despite having been defined identically, the models may capture institutional differences; when the significant variables coincide, they are more likely to relate to the physiological processes underlying the postoperative complications.

Methods

Setting and cohort definition. We consider two independent Midwestern health care systems: Mayo Clinic (MC) and Fairview Health Services (FHS). We include all patients from MC and FHS between 2010 and 2017 who are part of the NSQIP sample. For these patients we collected all available information about their NSQIP index surgery and a 30-day history before the index surgery from the respective institutions' EHR repositories. The NSQIP registry collects complications within a 30-day postoperative window. If a patient had another surgery in the 30-day postoperative window, we used the index surgery and measured the 30-day postoperative window for the outcome from the index surgery. For each complication, patients with the same pre-existing complication at the time of surgery were excluded.

Independent variables. We primarily rely on known risk factors of infection [13-15]. Independent variables were divided into three groups: demographic, preoperative, and high-resolution intraoperative. We limit ourselves to basic demographic information, such as age, sex, and body mass index (BMI) that are generally available. Preoperative variables are historic diagnoses including the problem list, procedures, and medications, as well as the preoperative indication for surgery. We use laboratory results and vitals to establish a preoperative baseline. Data from the preoperative assessment were preferentially used; if the data is not available, we use measurements from no more than 30-days before surgery. Diagnosis codes are rolled up into complications using the Clinical Classification Software [16].

Aggregating intraoperative variables. The intraoperative variables include orders, medications, and high-resolution vitals and labs. The stream of high-resolution variables needs to be divided based on the three stages of anesthesia. During the first (approximately) 15 minutes, called the *induction* phase, the vitals drop and deviate heavily from the normal. The last 15 minutes is called the *emergence* phase, where vitals are expected to return to close to normal but also involves significant changes as the patient is transitioned from anesthesia—often with full ventilation—to no anesthesia. The operation takes place between these two phases, in the so-called *maintenance* phase, where the vitals are expected to remain stable, although different from the preoperative baseline. In this work, we focus on the maintenance phase and use the mean value of labs and vitals during this phase.

Outcomes. We consider seven infectious outcomes: sepsis, septic shock, urinary tract infection (UTI), pneumonia (PNA) and the three kinds of SSI (superficial, deep tissue, and organ space SSI). We extract the outcome information from the NSQIP registry.

Modeling. Each outcome was modeled independently using logistic regression. For each outcome, two models are constructed. The "*pre*" model only uses demographics and preoperative data (30-day history of complications, baseline labs and vitals), while the "*pre+intra*" model uses aggregated intraoperative measurements on top of the preoperative and demographic data.

All models are logistic regression models. Missing labs and vitals were imputed using the middle of the normal range for the measurement. This assumes that the measurement is missing because it was deemed unnecessary to measure. The initially high number of independent variables was reduced by causal variable screening [19]. We used the PC-Simple algorithm [20] with a maximal condition set size of three [21]. This algorithm discards all variables that are independent of the outcome given at most three other variables. The rationale is that these variables do not affect the outcome directly, they only affect the outcome through other variables. Subsequent backwards elimination was applied to the remaining set of independent variables with a significance level of .05. The R statistical computing environment was used for all modeling. The PC-Simple algorithm is available in the pcalg R package [20].

Evaluation. Consistent with the intended use in perioperative decision support, we evaluated the models based on their predictive performance using concordance as the metric. Concordance is the probability that between two randomly selected patients, among which one has the complication in question while the other does not, the one with the complication has the higher predicted risk. Concordance is equivalent to the commonly reported area under the receiver-operating-characteristic curve (AUC).

Internal validation and effect of the intraoperative variables. Bootstrap estimation with 200 replications was used to estimate the concordance of the models. For each complication, two models were built on the same bootstrap replication: one with and one without the intraoperative variables. We measured the effect of the intraoperative variables as the mean difference in concordance between the models with and without intraoperative variables across the 200 bootstrap replications. The statistical significance of the difference was determined through a paired t-test.

External comparison. If the intraoperative variables capture a valid physiological phenomenon, then the variables selected at the two sites relate to the same quantity. We manually examined the models with the intraoperative data across the two sites to determine whether the selected variables coincide or at least relate to the same physiological concept. For example, oxygen saturation and ventilator settings relate to the same physiological concept (oxygen in the blood). When the selected variables relate to different concepts, there is a risk that the model simply captures an institutional artifact.

Results

Cohort Description

Table 1 contains a description of the analytic cohort at the two health systems. It contains demographic information, outcomes, history of the outcomes, history of relevant complications, and baseline (preoperative or at most 30 days before the index surgery when preoperative measurement is unavailable) lab results and vitals. Due to the high number of lab results and vital signs, we only report those that were significant in at least one of the models. Categorical variables are reported as count and percentage in parenthesis; continuous variables are reported as median and interquartile range in parenthesis.

Table 1 – Cohort Description.

	Mayo Clinic	Fairview
Number of patients	38,045	9,044
Demographics		
Age	61 (49, 71)	55 (40, 66)
Gender (male)	18,769 (48.7%)	4053 (45%)
Outcomes		
SSI Superficial	586 (1.5%)	107 (1.2%)
SSI Deep Tissue	186 (0.5%)	45 (0.5%)
SSI Organ Space	417 (1.1%)	119 (1.3%)
UTI	282 (0.7%)	147 (1.6%)
PNA	453 (1.2%)	111 (1.2%)
Sepsis	494 (1.3%)	88 (0.98%)
Septic Shock	186 (0.5%)	40 (0.44%)
History of Complications		
SSI	1,232 (3.2%)	91 (1.01%)
UTI	1,271 (3.3%)	127 (1.4%)
PNA	777 (2.0%)	73 (0.81%)
Bacteremia	133 (0.3%)	23 (0.25%)
Infections	3,953 (10.4%)	165 (1.8%)
Opport. Inf.	432 (1.1%)	45 (0.50%)
Malnutrition	792 (2.1%)	60 (0.66%)
Cancer	13,547 (35.6%)	1413 (16%)
Metastatic Disease	4,359 (11.5%)	112 (1.2%)
Transplant	624 (1.6%)	224 (2.5%)
Diabetes (T1&2)	5,215 (13.7%)	462 (5.1%)
COPD	2,372 (6.2%)	94 (1.04%)
Baseline labs and vitals		
BMI	28.4 (24.8, 32.9)	27.9 (24.6, 33.6)
Pulse	70.0 (62.6, 80.0)	75.0 (66.3, 85.0)
Respiration	9.5 (8.1, 10.8)	16 (16, 18)
Bilirubin	0.45 (0.45, 0.45)	0.67 (0.4, 1)
BUN	26 (18, 35)	14 (10, 20)
RBC	3.79 (3.13, 4.36)	4.27 (3.77, 4.70)
WBC	7.85 (4.45, 11.5)	8.50 (6.23, 12.1)
MCV	88.8 (87.2, 94.1)	89.4 (86.0, 93.0)
RDW	14.8 (13.5, 16.6)	13.9 (13.0, 15.3)
Hematocrit	33.1 (27.6, 39.7)	38.2 (33.8, 41.6)
PH	7.38 (7.32, 7.42)	7.38 (7.32, 7.42)
CO_2	40.5 (36.5, 43.5)	25.5 (24.0, 27.5)
Ca	4.71 (4.55, 4.92)	4.59 (4.30, 4.78)
K	4.15 (3.90, 4.50)	3.9 (3.7, 4.2)
Na	138 (137.8, 138.4)	139 (137, 141)

(1) Abbreviations: UTI: urinary tract infection; PNA: pneumonia; SSI: surgical site infection; COPD: chronic obstructive pulmonary disease; BMI: body mass index; BUN: blood urea nitrogen; RBC: red blood cell [count]; WBC: white blood cell [count]; MCV: mean corpuscular volume; RDW: red blood cell distribution width; Ca: calcium ion concentration; K: potassium ion concentration; Na: sodium ion concentration. (2) Only variables that were significant in at least one model are presented.

Model performance and effect of the intraoperative variables

Table 2 presents the predictive performance of the models for each of the outcomes. For each outcome, the table contains three rows: the first one shows the performance as measured by AUC of the "pre" model (that uses only demographics and preoperative data), the second one shows the performance of the "pre+intra" model, which uses the intraoperative data in addition to the demographics and preoperative data, and the third row shows the difference in performance between the "pre" and "pre+intra" models. For the "pre" and "pre+intra" models, the standard deviation of the performance (obtained from 200 bootstraps) is shown in parenthesis; for the difference, the p-value of the paired t-test is shown in the parenthesis using the scientific notation (e.g. 2e-8 is $2*10^{-8}$). All differences are statistically significant at 0.05 confidence level, but superficial and deep tissue SSI are not significant on FHS side after Bonferroni correction.

Table 2 – AUC performance of the "pre", the "pre+intra" models, and the difference in performance.

Outcome	Model	Mayo Clinic	Fairview
PNA	Pre	.877 (±.021)	.693 (±.044)
	Pre+intra	.881 (±.022)	.734 (±.044)
	Difference	***.004 (1e-8)***	***.041 (1e-16)***
Sepsis	Pre	.736 (±.024)	.695 (±.045)
	Pre+intra	.755 (±.021)	.711 (±.046)
	Difference	***.019 (1e-16)***	***.016 (8e-13)***
Septic Shock	Pre	.827 (±.031)	.715(±.078)
	Pre+intra	.834 (±.030)	.755(±.079)
	Difference	***.007 (1e-13)***	***.040 (1e-14)***
SSI Superficial	Pre	.667 (±.018)	.560 (±.044)
	Pre+intra	.688 (±.017)	.563 (±.042)
	Difference	***.021 (1e-16)***	*.003 (0.022)*
SSI Deep Tissue	Pre	.660 (±.041)	.568 (±.069)
	Pre+intra	.678 (±.035)	.575 (±.064)
	Difference	***.018 (1e-7)***	*.007 (0.3)*
SSI Organ Space	Pre	.732 (±.026)	.657 (±.037)
	Pre+intra	.750 (±.022)	.699 (±.039)
	Difference	***.021 (1e-16)***	***.042 (1e-16)***
UTI	Pre	.665 (±.022)	.717 (±.031)
	Pre+intra	.672 (±.021)	.727 (±.032)
	Difference	***.007 (1e-13)***	***.010 (5e-14)***

External Comparison of the Significant Variables

Table 4 displays the number of bootstrap iterations in which the variable was selected for each complication at each site. We only list variables that were selected in at least 100 of the 200 bootstrap iterations for at least one outcome at one site.

Table 3 offers a summary of the information in Table 4. It lists the most important intraoperative variables that are common between the two sites for each outcome.

Table 3 – Selected variables

	Significant intraoperative variables	
Outcome	**Mayo Clinic**	**Fairview**
PNA	antibiotic use, lactate, CO_2, FIO2	Lactate, MCV, pulse, SpO_2
Sepsis	Glucose, isoflurane expired, PEEP, pulse	PO2, RBC, pulse, hematocrit
Septic Shock	Antibiotic use, glucose, WBC, pulse, PEEP	Lactate, FIO2
SSI Superficial	CVP, PEEP, pulse	PEEP
SSI Deep Tissue	Antibiotic use, steroid use, PEEP	PO2
SSI Organ Space	Antibiotic use, bicarb, Ca, glucose, CVP, isoflurane expired	Hemoglobin, PH arterial, PO2, pulse
UTI	Antibiotic use, steroid use	Ca, PO2, CVP, PEEP

Abbreviations: FIO2: Fraction of Inspired Oxygen; PEEP: Positive End Expiratory Pressure; PO2: Partial pressure of oxygen; WBC: white blood cell [count]; CVP: central venous pressure; Ca: calcium ion in the blood; MCV: mean corpuscular volume; SpO_2: blood oxygen saturation; RBC: red blood cell [count].

Discussion

We view EHR-integrated perioperative clinical decision support as a key avenue towards reducing postoperative complications or towards better anticipating and managing them. Central to such systems are accurate risk models with the ability to produce risk estimates on demand based on contemporaneous data, including intraoperative data. In this work, we sought to answer the following two questions: (i) Do intraoperative data elements improve our ability to predict postoperative complications? and (ii) Are these models specific to institutions or do they describe possible physiological phenomena that are institution-independently predictive of postoperative complications?

Regarding the first question, we found that the inclusion of intraoperative variables led to statistically significant improvements in the risk models' predictive ability for five of the seven infectious complications at both sites. These complications were pneumonia, sepsis, septic shock, organ space SSI, and urinary tract infection. At the Mayo Clinic site, the inclusion of the intraoperative variables led to statistically significant (albeit not necessarily clinically relevant) improvements for all complications, likely due to the much larger sample size.

The lack of statistically significant improvement for the other two SSI types, superficial and deep tissue, is unsurprising. These diagnoses have more ambiguity and our previous work on detecting SSI retrospectively also suggests that predicting superficial and deep tissue SSI is difficult [17, 18].

Regarding the second question, we found that some of the intraoperative variables overlap across the two sites for all four of the complications in which intraoperative variables improved the predictive performance. In the cases of pneumonia, sepsis, and organ-space SSI, some of the variables (lactate, pulse) matched exactly, while others referred to a shared concept. For example, FIO2 and SpO2 are related to the sufficiency of oxygen in the bloodstream: the first one is the ventilator setting (to supply the oxygen into the bloodstream) while the second one is the direct measurement of oxygen saturation. For sepsis, partial pressure from oxygen and PEEP have an analogous relationship; and for organ space SSI, bicarbonate and PH both measure the acidity of blood. The fact that many of the significant variables are common across the two sites suggests that these variables can be related to the physiological process(es) underlying the outcomes.

Conclusion

Intraoperative variables were found to statistically significantly improve the performance of 30-day post-operative risk models for five of the seven infectious complications at both sites. There was considerable overlap in the significant intraoperative variables across the two sites and the overlapping variables were related to lactate, acidity, and blood oxygen.

Acknowledgements

This work was supported in part by NIGMS award R01 GM 120079, AHRQ award R01 HS024532, and the NCATS University of Minnesota CTSA UL1 TR002494. The views expressed in this manuscript are those of the authors and do not necessarily reflect the views of the funding agencies.

References

[1] McCoy CC, Englum BR, Keenan JE, Vaslef SN, Shapiro ML, Scarborough JE. Impact of specific postoperative complications on the outcomes of emergency generalsurgery patients. J Trauma Acute Care Surg. 2015 May;78(5):912-8.

[2] Vonlanthen R, Slankamenac K, Breitenstein S, Puhan MA, Muller MK, Hahnloser D, Hauri D, Graf R, Clavien PA. The impact of complications on costs of major surgical procedures: a cost analysis of 1200 patients. Ann Surg. 2011 Dec;254(6):907-13.

[3] Khan NA1, Quan H, Bugar JM, Lemaire JB, Brant R, Ghali WA. Association of postoperative complications with hospital costs and length of stay in a tertiary care center. J Gen Intern Med. 2006 Feb;21(2):177-80.

[4] Brown SR1, Mathew R, Keding A, Marshall HC, Brown JM, Jayne DG. The impact of postoperative complications on long-term quality of life after curative colorectal cancer surgery. Ann Surg. 2014 May;259(5):916-23.

[5] Artinyan A, Orcutt ST, Anaya DA, Richardson P, Chen GJ, Berger DH. Infectious postoperative complications decrease long-term survival in patients undergoing curative surgery for colorectal cancer: a study of 12,075 patients. Ann Surg. 2015 Mar;261(3):497-505.

[6] Baucom RB, Phillips SE, Ehrenfeld JM, Muldoon RL, Poulose BK, Herline AJ, Wise PE, Gieger TM. Association of Perioperative Hypothermia During Colectomy With Surgical Site Infection. J Am Med Assoc, 2015.

[7] Leaper D, Ousey K. Evidence update on prevention of surgical site infection. Curr Opin Infect Dis. 2015 Apr;28(2):158-63.

[8] Seamon MJ, Wobb J, Gaughan JP, Kulp H, Kamel I, Dempsey DT. The effects of intraoperative hypothermia on surgical site infection: an analysis of 524 trauma laparotomies. Ann Surg. 2012 Apr;255(4):789-95.

[9] Melton GB, Vogel JD, Swenson BR, Remzi FH, Rothenberger DA, Wick EC. Continuous Intraoperative Temperature Measurement and Surgical Site Infection Risk. Ann Surg. 2013, 258:606–613.

[10] Henderson. WG, Daley J. Design and statistical methodology of the National Surgical Quality Improvement Program: why is it what it is? Am J Surg 2009; 198:S19.

[11] Shiloach M, Frencher SK, Jr., Steeger JE, Rowell KS, Bartzokis K, Tomeh MG, Richards KE, Ko CY, Hall BL. Toward robust information: data quality and inter-rater reliability in the American College of Surgeons National Surgical Quality Improvement Program. J Am Coll Surg. 2010 Jan;210(1):6-16.

[12] Hall BL, Hamilton BH, Richards K, Bilimoria KY, Cohen ME, Ko CY. Does surgical quality improve in the American College of Surgeons National Surgical Quality Improvement Program: an evaluation of all participating hospitals. Ann Surg. 2009 Sep;250(3):363-76.

[13] Phillips J, O'Grady H, Baker E. Prevention of Surgical Site Infections. Surgery 2014; 32(9).

[14] Korol E, Johnston K, Waser N, Sifakis F, Jafri HS, Lo M, Kyaw MH. A Systematic Review of Risk Factors Associated with Surgical Site Infections among Surgical Patients. Plos One 2013; 8(12).

[15] Pessaux P, Msika S, Atalla D, Hay JM, Flamant Y. Risk factors for postoperative infectious complications in noncolorectal abdominal surgery: a multivariate analysis based on a prospective multicenter study of 4718 patients. Arch Surg. 2003 Mar;138(3):314-24.

[16] Elixhauser A, Steiner C, Palmer L. Clinical Classifications Software (CCS), 2015. U.S. Agency for Healthcare

Research and Quality. Available: http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp.
[17] Hu Z, Melton GB, Moeller ND, Arsoniadis EG, Wang Y, Kwaan MR, Jensen EH, Simon GJ. Accelerating Chart Review Using Automated Methods on Electronic Health Record Data for Postoperative Complications. AMIA Annual Symposium Proceedings 2016, 1822.
[18] Z Hu, GJ Simon, EG Arsoniadis, Y Wang, MR Kwaan, GB Melton. Automated detection of postoperative surgical site infections using supervised methods with electronic health record data. Studies in health technology and informatics 2015; 216, 706.
[19] Ma S, Statnikov A. Methods for computational causal discovery in biomedicine. Behaviormetrika. 2017 Jan 1; 44 (1): 165-91.
[20] Kalisch M, Maechler M, Colombo D, Maathuis MH, Buehlmann P. Causal Inference Using Graphical Models with the R Package pcalg. Journal of Statistical Software, 2012; 47(11), 1-26. URL http://www.jstatsoft.org/v47/i11/.
[21] Aliferis CF, Tsamardinos I, Statnikov A. HITON: a novel Markov Blanket algorithm for optimal variable selection. In AMIA Annual Symposium Proceedings 2003 (Vol. 2003, p. 21). American Medical Informatics Association.

Address for correspondence

Gyorgy J. Simon
Institute for Health Informatics,
Department of Medicine,
University of Minnesota
simo0342@umn.edu

Table 4 – Frequency at which each intraoperative variable was selected (out of 200 bootstrap iteration)

	PNA		SSI Superf.		SSI Deep T.		SSI Organ		Sepsis		Septic Shock		UTI	
	MC	**FV**	**MC**	**FV**	**MC**	**FV**	**MC**	**FV**	**MC**	**FV**	**MC**	**FV**	**MC**	**FV**
Calcium Ion	7	17	31	10	62	9	124	66	93	38	7	41	2	94
CO_2	120	18	1	20	19	3	4	24	80	5	1	2	0	0
Glucose	8	1	16	0	19	11	160	0	105	0	115	36	66	0
Lactate	158	144	5	21	1	1	43	27	29	23	28	163	7	1
MCV	5	153	27	6	48	64	41	16	10	41	25	5	0	2
PO2	42	90	6	17	24	63	11	192	12	95	40	12	4	110
CVP	11	8	136	12	43	40	122	34	30	8	46	5	31	137
Isoflurane Exp.	1	0	71	1	4	1	127	61	190	0	0	0	0	1
FIO2	115	6	1	13	12	0	53	1	1	5	6	120	9	8
PEEP	16	49	200	107	141	2	136	53	156	4	86	7	75	129
Pulse	31	88	173	22	28	13	167	63	197	142	152	0	4	9
Steroid	22	32	5	36	115	50	3	8	1	8	7	0	106	6
Abx	162	6	72	4	136	0	147	8	88	0	190	0	178	13

Notes: (1) Only variables selected in at least 100 iterations are displayed. (2) Abbreviations; MC: Mayo Clinic; FV: Fairview Health Services; MCV: mean corpuscular volume; CVP: central venous pressure; FIO2: Fraction of Inspired Oxygen; PEEP: Positive End Expiratory Pressure; PO2: Partial pressure of oxygen; Abx: Antibiotic use.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190252

Development and Validation of a Controlled Vocabulary: An OWL Representation of Organizational Structures of Trauma Centers and Trauma Systems

Joseph Utecht[a], Jane Ball[b], Stephen M. Bowman[c], Jimm Dodd[b], John Judkins[d], Robert T. Maxson[e], Rosemary Nabaweesi[f], Rohit Pradhan[c], Nels D. Sanddal[b], Robert J. Winchell[g], Mathias Brochhausen[a]

[a] Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA,
[b] American College of Surgeons, Chicago, Illinois, USA
[c] Department of Health Policy and Management, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA
[d] Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[e] Department of Surgery, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA
[f] Department of Pediatrics, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA
[g] Department of Surgery, Weill Cornell Medical College, New York, New York, USA

Abstract

In trauma care and trauma care research there exists an implementation gap regarding a consistent controlled vocabulary to describe organizational aspects of trauma centers and trauma systems. This paper describes the development and evaluation of a controlled vocabulary for trauma care organizations. We give a detailed description of the involvement of domain experts in the domain analysis workflow and the authoring of definitions and additional term descriptions. Finally, the paper details the evaluation methodology to assess the initial version of the controlled vocabulary. The results of the evaluation show that our development process yields terms most of which find approval from domain experts not involved in the development. In addition, our evaluation tools resulted in valuable domain expert input to optimize the controlled vocabulary.

Keywords:

Trauma Centers; Vocabulary, Controlled; Biomedical Ontologies

Introduction

Since the 1970s, the American College of Surgeons (ACS) has worked to establish and refine criteria that set the standards for trauma centers. Due to the inherent difficulty in obtaining reliable outcome data, these standards have focused largely on measures of structure and process. The most recent standards were published in *Resources to Optimal Care of the Injured Patient (2014)* [1]. In 2008, the ACS established the Trauma Quality Improvement Program (TQIP) to enable trauma centers to measure and compare their risk-adjusted patient outcomes to similar organizations [2,3]. Though the TQIP program continues to grow, at this time the relationship between structure and process measures and clinical outcomes remains elusive. Further, beyond the standards themselves, the ways in which the particular attributes of trauma centers, and at a larger scale, of regional trauma systems, contribute to optimal patient outcomes have not yet been identified or measured.

While the trauma center standards have paved the way for a consistent and reusable terminology, there still exists an implementation gap in providing and using a well-structured controlled vocabulary to describe the structure of trauma centers and trauma systems. Having such a resource will facilitate transferring knowledge and experience from one trauma center or trauma system to another. In addition, controlled vocabulary also will facilitate comparison of organizational structures and procedures both nationally and internationally.

CAFÉ (Comparative Assessment Framework for Environments of Trauma Care) is an NIH-funded project that aims to collect detailed information on the particular organizational attributes of trauma systems and trauma centers. Ultimately, the goal is to link that data to clinical outcome data to identify those organizational attributes of trauma centers and trauma systems that are of high impact on patient outcomes. The project creates a web-based service that allows an interested individual to enter data about their trauma center or trauma system and to conduct an anonymous self-assessment of the organizational structures of their trauma center or trauma system. The first step to enable such a comparison is to provide a common controlled vocabulary covering all relevant aspects of trauma center and trauma system management. Preliminary CAFÉ project work that leveraged natural language processing to create lists of relevant terms from published abstracts pertinent to the domain was described in a previous publication [4]. In this paper we report the workflow of creating the initial version of the vocabulary in close collaboration. In addition, we describe our evaluation methodology, describe the results of the evaluation, and draw conclusions on whether the development methodology of the controlled vocabulary has been successful. For the purpose of the CAFÉ project, the controlled vocabulary needs to exist as a Web Ontology Language (OWL2) to be used with other semantic web technologies. The general outline of the project plan was published elsewhere [4].

Methods

The first step necessary for developing the CAFÉ system was the domain analysis of organizational structures of trauma centers and trauma systems. A domain analysis for the purpose of developing a controlled vocabulary consists of identifying the relevant terms to represent the domain of interest and providing definitions for those terms. For the CAFÉ project, that meant collecting relevant terms to represent and analyze the organizational structure of trauma centers and trauma systems. Since our aim is to create an ontology that follows the OBO Foundry principles, it was prudent to use the OBO

Foundry suggested form of definition, the *genus-differentia* (http://www.obofoundry.org/principles/fp-006-textual-definitions.html) form. In the OBO Foundry community, definitions of that form are frequently provided using the IAO:definition annotation property (http://purl.obolibrary.org/obo/IAO_0000115). In addition to the *genus-differentia definition* being an OBO Foundry requirement, these definitions are also particularly useful when creating a taxonomy for the controlled vocabulary since these definitions always refer to the parent term of the term being defined.

Earlier work on OBO Foundry ontologies has informed the project leads for ontology development that domain experts and potential users often find the *genus-differentia definition* a poor representation of how they view the domain and communicate about its phenomena. Hence, we decided to follow the advice of Hogan et al. [5] to provide definitions that are easily accessible for domain experts using an annotation property different from the IAO:definition. To allow that, we created the OOSTT user-centered description annotation property (http://purl.obolibrary.org/obo/OOSTT_00000030).

In addition, we also created the annotation property "CAFE application label" for cases where the label that was created to follow naming conventions in ontology development would be a suboptimal label to display to the user of the system (http://purl.obolibrary.org/obo/OOSTT_00000043).

The effect of this consideration on our domain analysis is that we are collecting one RDFS:label, one IAO:definition and, in cases where the *genus-differentia definition* is deemed hard to parse for domain experts, one OOSTT user-centered description.

After the CAFÉ project was funded by the National Institute of General Medical Science (NIGMS) in spring 2015 (R01GM111324), a weekly (later bi-weekly) expert call was started to identify relevant terms to assess the organizational structure of trauma centers and trauma systems and, in future work, its relationship to patient safety and outcomes. The committee consisted of two trauma surgeons, both bearing roles in the American College of Surgeons (ACS) Committee on Trauma (COT); two representatives of the COT with experience in trauma center verification and trauma system consultation; two researchers with experience in analyzing data on trauma care, including data from trauma centers and trauma systems; and one expert in organization analysis. To accomplish the domain analysis part of the project, a multi-step process was created during which the members of our expert group reviewed material from the ACS COT and added material based on their experience to identify the relevant terms and provide user-centered descriptions for those. The goal of this process was the implementation of our controlled vocabulary in the OWL language.

Step 1: *Domain set of terms.* In the first step, the expert group agreed on a set of terms necessary to cover the domain of organizational structures of trauma centers and trauma systems. This step used the American College of Surgeons' *Resources for Optimal Care of the Injured Patient, Sixth Edition* from 2014 as a starting point [1]. The experience of the domain experts regarding the practical implementation of trauma center verification, trauma system consultation, and quality improvement on trauma, in general, informed the final decision on which terms to include.

Step 2: *Preliminary descriptions provided.* Once the initial set of terms was agreed upon by the domain expert group, preliminary descriptions of each term were provided by the domain experts. The goal was to capture a description that the domain experts were able to agree upon regardless of whether the description met the requirements of an OBO Foundry definition.

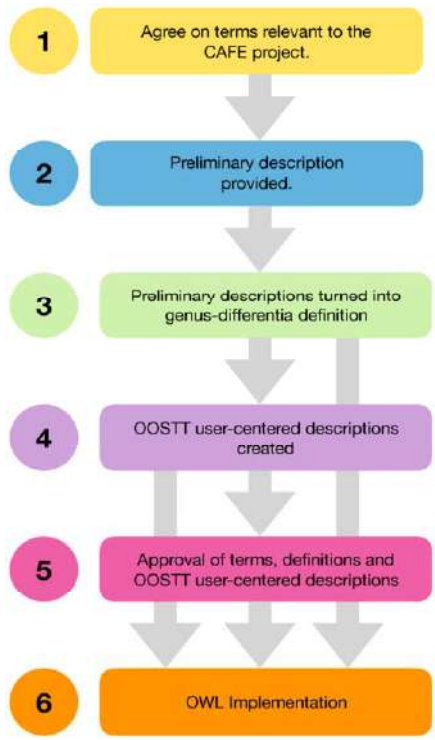


Figure 1 – Development workflow of the controlled vocabulary on organizational structures of trauma centers and trauma systems.

Step 3: *Preliminary descriptions turned into genus-differentia definitions.* Based on the descriptions, the CAFÉ OWL implementation team suggested *genus-differentia definitions*. In doing so, the OWL implementation teams thought about definitions and descriptions in a set-theoretical way. We assumed that for each definition/description there is a set of entities for which that definition/description is true. In creating a *genus-differentia definition* for a description it is crucial that the former is true for the same set of entities as the latter. In other words, that the p *genus-differentia definition* has the same extension as the initial description authored by the domain experts. In many cases, this process involved iterative communication between the domain experts and the ontology developers to a) achieve a clear understanding of the intended meaning and b) ensure congruency between the preliminary description and the *genus-differentia definition*.

Step 4: *OOSTT user-centered description created.* Based on whether the domain experts felt that the *genus-differentia definition* was useful for the target audience or not, OOSTT user-centered definitions were created where usefulness for the target audience was in doubt. To the extent the delimitation of the CAFÉ domain expert group agreed that the *genus-differentia definitions* are useful to the target audience was one of the parameters that we plan to assess with our evaluation approach. For the terms that have both a OOSTT user-centered description and a *genus-differentia definition* the same kind of congruency as described in Step 3 was required.

Step 5: *Approval of terms, definitions and OOSTT user-centered descriptions by domain experts.* Our aim was to submit all final terms, their definitions, and their OOSTT user-centered descriptions to another round of expert approval. While this strategy worked for many terms, restrictions created by our timeline limited another approval round for all terms.

Step 6: *OWL implementation.* Finally, all terms were included in the OWL file step-by-step during the rollout of the OOSTT ontology, which started in 2015 and was described in more detail in a previous publication [4]

To help determine the quality and acceptance by our target audience of the terms and definitions created, we conducted a survey with feedback. A new surveying tool was created for the task. Our goal for this new tool was to show a small number of random terms and definitions to members of our target audience while minimizing inconvenience to them. This tool also needed to ensure a balanced number of reviews for each term and easily allow us to load the up-to-date version of the 216 terms and definitions. The term survey tool (https://github.com/cafe-trauma/term_survey) was built over two days in Python using the Django Framework. To ensure current terms and definitions could easily be loaded, the tool was built with the functionality to parse an OWL file for OWL classes from which the RDFS:label would be used for the "term" and the URI of an annotation would be passed to query for the "definition". As mentioned above, we used a custom annotation property for "User Centered Definitions." After the terms and definitions were loaded, an administrator set a welcome message, a logo, and a number of terms for each respondent to review (Figure 2).

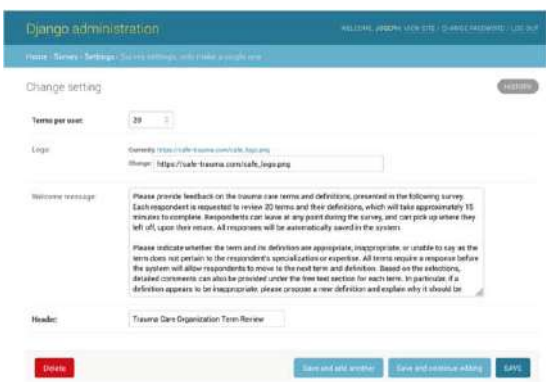


Figure 2 - Term Tool Configuration

At this point the survey was ready. To distribute it to respondents we emailed a link to the landing page of the survey, where the respondent was greeted by the welcome message and could begin reviewing terms (Figure 3). We distributed the survey to 75 members of the ACS COT. The terms for each respondent to review were not preselected. Each time a respondent attempted to review a new term the tool would query for the current list of terms with the lowest number of reviews and return a random term from this list, ensuring an equal coverage of reviews for our terms. On the term review page, a respondent was presented with a term, its definition, radio buttons to indicate the acceptability of a definition, and space to provide feedback (Figure 4). After clicking to move to the next term, a respondent's feedback was immediately saved to the server and a new randomly selected term was presented. The term review page also had a progress bar to indicate the number of terms remaining for review. After giving their initial feedback, a cookie was created in the respondent's browser that tied them to their session of terms in case they closed the browser before finishing. Since we saved feedback after each term we still had feedback even if the respondent did not finish the requested number of term reviews. After reviewing the assigned 20 terms and definitions, the respondent was shown a summary screen with all of their feedback and given the choice to edit any of their feedback or to clear their session and review more terms if so inclined (Figure 5).

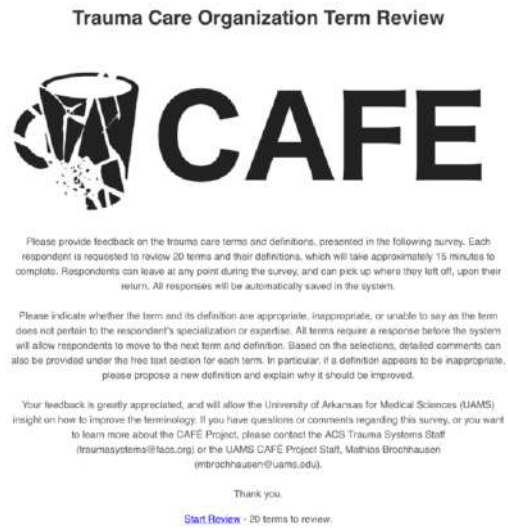


Figure 3 - Welcome Page

Figure 4 - Term Review


Thank you for your participation.

Term	Good	Proposal	Comment
trauma program role	✔		
trauma program manager role	✔		
trauma medical director role	✔		
orthopedic surgery residency program	✔		
board eligible trauma surgeon role	✔		
record of participation of continuing trauma care education	✖	This could be clearer.	
trauma team activation	✔		
trauma center designation	✔		
jurisdictional lead agency	✔		
successful completion of trauma surgery fellowship information	✖	The term seems too specific	
continuous anesthesiology coverage policy	✔		
injury			
radiology liaison role			
prehospital protocol	✔		
tqip coordinator obligee role	✔		
trauma program manager role	✔		
emergency medical services agency association	✔		
prehospital care provider role	✔		
trauma nursing core course certificate	✔		
board eligible orthopedic surgeon role	✔		

Perform another term review?

Figure 5 – Survey Completion Summary

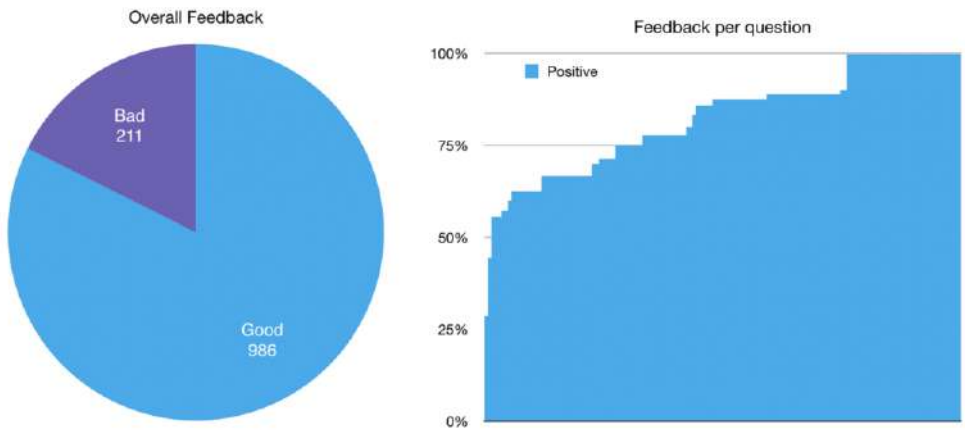


Figure 6 - Survey Feedback

Results

Using our newly built survey tool, we received a large amount of good feedback about our terms. The survey was distributed to 75 people in the trauma care domain. We received at least one term review from 74 people. In total, our 142 terms and definitions received 1,197 reviews. That response rate allowed us to have, on average, eight different people review each term and definition. Of the 1,197 individual reviews 986 of them considered the term and definition suitable (Figure 6). The reviewers largely agreed with each other when reviewing the same terms both positively and negatively. Our highest rated terms, such as "trauma program leadership," "hospital governing body role," and "trauma program" received uniformly good feedback, while the lowest rated terms, "trauma nursing evaluator obligee role," "trauma quality improvement and patient safety program lead role," and "emergency medical services provider association" were negatively reviewed by multiple people. Overall, 103 or 72% of our total terms received positive feedback from over 75% of reviewers (Figure 6).

The responses and comments were provided to the ontology developers who reviewed the comments and aimed to correct the issues pointed out by the domain experts. For the lowest rated terms, this included a complete revision of the definition, as in the case of "emergency medical services provider association," which was updated based on reviewer feedback[1]. The most reported problem with "trauma quality improvement and patient safety program lead role" was that the label was singular, while the definition implied plural. The definition was changed to better reflect the label[2]. The label "trauma nursing evaluator obligee role" was critized for the use of uncommon language, namely "obligee role." Since that class did not have a CAFÉ application label, we added the CAFÉ application label "trauma nursing evaluator"[3].

Table 1 – Highest rated terms in our controlled vocabulary and their initial definition/OOSTT user-centered description.

Term label	OOSTT user-centered description or definition
trauma program leadership	The individuals who provide oversight, direction, coordination, and management for the facility's trauma program.
hospital governing body role	A group of individuals appointed to a hospital's board of directors who provide governance, direction, and oversight of the overall operation of a hospital.
trauma program	The organizational unit of a healthcare facility that is designated to provide and coordinate care for injured patients.

Table 2 – Lowest rated terms in our controlled vocabulary and their initial definition/OOSTT user-centered description.

Term label	OOSTT user-centered description or definition
trauma nursing evaluator obligee role	The individual with authority and responsibility for evaluating the nursing care provided to trauma patients.
trauma quality improvement and patient safety program lead role	The role of an authorized group of healthcare providers that ensures trauma quality improvement and patient safety within the trauma program.
emergency medical services provider association	An association of EMS agencies that any agency may join voluntarily to share information or collaborate on issues of state or national importance.

[1] https://github.com/OOSTT/OOSTT/commit/38c6666824d57da695d38710241495673085b942

[2] https://github.com/OOSTT/OOSTT/commit/73f72efc598a813018daaa9397d54657e0ac2107

[3] https://github.com/OOSTT/OOSTT/commit/3bcbbfa990e86533068055946e4af21c9833bf71

Discussion

The results presented in this paper provide insight into two main areas: a) domain analysis and definition authoring in a clinical domain and b) providing a novel resource for trauma care and trauma research to facilitate comparison between trauma centers and trauma systems by creating a controlled vocabulary.

The current evaluation of the initial version of the domain analysis yielded primarily positive results. Not only did participants rate most terms as "good," they also provided insightful and highly actionable comments to make the terms and definitions more useful or correct. The curators of the controlled vocabulary have already incorporated multiple improvements based on reviewer comments and are currently working to finalize that process. At this point in time, the survey has only been distributed to trauma care experts affiliated with the American College of Surgeons. We are planning to distribute the survey to clinical staff in trauma management in order to gather their input on these terms.

Larger, international evaluations of the vocabulary are planned for 2019. Distributing the survey to experts from diverse healthcare systems will provide interesting insights into which terms are useful and which definitions are shared across healthcare systems.

While efforts to improve terminologies to advance trauma care and trauma research are not uncommon [6,7,8], literature searches conducted on PubMed reveal an astonishing lack of effort to create terminologies or controlled vocabularies relevant to trauma care in general and its organization. Oliver and Walter [9] point out the problems arising from the lack of a common terminology in trauma research related to preventable death. Searching PubMed for ["trauma system" terminology], ["trauma center" terminology], ["trauma system" vocabulary], ["trauma center" vocabulary], yields 38 distinct papers, only one of which specifically deals with the problem of providing a terminology/vocabulary for organizational structures in trauma systems and trauma centers. Notably, that paper is also the output of the CAFÉ project [4]. So, while there is an established field of work on terminologies and classifications for specific types of trauma and a corpus of research on outcome measures, there is not much effort to provide a unified language to describe the organizational structure of trauma centers and trauma systems.

Table 3 – Number of publications retrieved from PubMed using a set of queries related to trauma systems, trauma centers and terminology/vocabulary.

Query	Number of retrieved publications
["trauma system" terminology]	2
["trauma center" terminology]	34
["trauma system" vocabulary]	1
["trauma center" vocabulary]	1

Conclusions

While some of the terms still required additional refinement, a vast majority of the terms was deemed "good" by a large group of domain experts working in collaboration with the American College of Surgeons. Based on these results we conclude that our methodologies for domain analysis and creation of definition and user-centered descriptions have successfully created an initial version of a controlled vocabulary for the ACS community and the trauma care community in the U.S.

We also believe that the number of targeted and useful comments we received from the domain experts participating in our survey highlights the fact that dividing review tasks into small portions and distributing them over a larger number of participants is a valuable approach to encourage expert feedback.

Acknowledgements

The research presented in this paper is funded by the National Institute of General Medical Sciences of the National Institutes of Health under award number R01GM111324.

Sarah Bost participated in the technical editing of this paper.

References

[1] Committee on Trauma American College of Surgeons, Resources for Optimal Care of the Injured Patient, American College of Surgeons, Chicago, Illinois, 2014.

[2] S. Shafi, A.B. Nathens, H.G. Cryer, et al., The trauma quality improvement program of the American college of surgeons committee on trauma, *J Am Coll Surg* **209** (2009) 521-530.

[3] M.R. Hemmila, A.B. Nathens, S. Shafi, et al., The trauma quality improvement program: Pilot study and initial demonstration of feasibility, *J Trauma* **68** (2010) 253-262.

[4] J. Utecht, J. Judkins, J.N. Otte, et al., OOSTT: A resource for analyzing the organizational structures of trauma centers and trauma systems, In *CEUR Workshop Proceedings*, Annecy, France, 2016, 1747.

[5] W.R. Hogan, M.M. Wagner, M. Brochhausen, et al., The Apollo structured vocabulary: An OWL2 ontology of phenomena in infectious disease epidemiology and population biology for use in epidemic simulation, *J Biomed Semantics* **7** (2016) 50.

[6] A. Weir, P. Brukner, E. Delahunt, et al., Doha agreement meeting on terminology and definitions in groin pain in athletes, *Br J Sports Med* **49** (2015) 768-774.

[7] D. Patel, Eye injuries: Improving our practice, *Community Eye Health* **28** (2015) 41-43.

[8] H.W. Mueller-Wohlfahrt, L. Haensel, K. Mithoefer, et al., Terminology and classification of muscle injuries in sport: The Munich consensus statement, *Br J Sports Med* **47** (2012) 342-350.

[9] G. Oliver and D. Walter, A call for consensus on methodology and terminology to improve comparability in the study of preventable prehospital trauma deaths: A systematic literature review, *Academic Emergency Medicine*, **23** (2016) 503-510.

Address for Correspondence

Mathias Brochhausen,
mbrochhausen@uams.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190253

Normalizing Dietary Supplement Product Names Using the RxNorm Model

Jake Vasilakes[a,b], Yadan Fan[a], Rubina Rizvi[a,b], Anusha Bompelli[a], Olivier Bodenreider[c], Rui Zhang[a,b]

[a] *Institute for Health Informatics, University of Minnesota, Minneapolis, MN, USA*
[b] *Department of Pharmaceutical Care & Health Systems, University of Minnesota, Minneapolis, MN, USA*
[c] *Lister Hill National Center for Biomedical Communications, National Library of Medicine, National Institutions of Health, USA*

Abstract

The use of dietary supplements (DSs) is increasing in the U.S. As such, it is crucial for consumers, clinicians, and researchers to be able to find information about DS products. However, labeling regulations allow great variability in DS product names, which makes searching for this information difficult. Following the RxNorm drug name normalization model, we developed a rule-based natural language processing system to normalize DS product names using pattern templates. We evaluated the system on product names extracted from the Dietary Supplement Label Database. Our system generated 136 unique templates and obtained a coverage of 72%, a 32% increase over the existing RxNorm model. Manual review showed that our system achieved a normalization accuracy of 0.86. We found that the normalization of DS product names is feasible, but more work is required to improve the generalizability of the system.

Keywords:

Dietary supplements; RxNorm; Natural Language Processing

Introduction

Dietary supplements (DSs) are defined as "products taken by mouth that contain a dietary ingredient that includes vitamins, minerals, amino acids, and herbs/botanicals, as well as other substances that can be used to supplement the diet" [1]. They comprise one of the fastest growing categories of complementary and alternative medicines [2]. According to the National Health and Nutrition Examination Survey (NHANES), the age adjusted consumption of DSs has steadily increased, both in male (28% to 44%) and female (38% to 53%) populations [3], especially among adults aged ≥60 years where 70% have reported using one or more DS [4]. Increasing usage of DSs has led to substantial market growth resulting in wide availability of dietary supplement products.

The regulations covering DSs are much less stringent than those covering commonly consumed foods and clinical drugs [1], even though DS adverse events and DS-drug interactions are common [5, 6] and potentially severe [7]. DS products and dietary ingredients are regulated by the U.S. Food and Drug Administration (FDA) under the Dietary Supplements Health and Education Act (DSHEA). As part of this, the FDA developed guidelines to help ensure that DSs sold in the United States (produced both domestically and abroad) are properly labeled. However, according to the FDA, "those guidance documents only represent the agency's current perspective and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited" [8, 9]. Thus, it is not required to obtain approval of a label in order to import or distribute a DS, and failing to comply with the guidelines does not entail any legally enforceable consequences.

To make matters worse, DS product names express ingredient and brand information in a large variety of ways. Product names often include additional components such as ingredient qualifiers (e.g. "leaf", "dried", "extract"), dose information (e.g. "capsules", "10mg"), and flavors. This, along with loose labeling guidelines, have resulted in DS product names that lack a consistent structure, which hinders critical tasks such as cross-platform communicability and the reuse of DS knowledge.

The situation is very different for clinical drugs. In addition to stricter regulations regarding drug naming, the U.S. National Library of Medicine develops RxNorm, a normalized naming system for generic and branded drugs [10]. It supports semantic interoperability between sixteen drug terminologies and pharmacy knowledge bases. RxNorm normalizes drug names using a set of 15 term types corresponding to drug entities [11]. Term types are codes which indicate the level of specificity of a given drug name or qualifier. For example, the drug name "Fluoxetine" is assigned the term type *IN* (ingredient) and the qualifier "Oral solution" is assigned *DF* (dose form). Some RxNorm term types are the combination of two or more atomic term types. For example, the *IN* and *DF* term types combine into *SCDF* (Semantic Clinical Drug Form), such as in "Fluoxetine Oral Solution".

As a means of normalizing drug names, RxNorm plays an essential role in decision support, quality assurance, healthcare research, reimbursement, and mandatory reporting [12]. Similarly, normalizing DS product names is an important step. By providing a reliable way to refer to DS products, it would facilitate DS pharmacovigilance and knowledge discovery such as in [13, 14]. However, Y. Wang et al. showed that existing normalization resources such as RxNorm and UMLS cover only a fraction of DS terms, indicating a need for DS-specific resources [15]. Sharma and Sarkar developed such a resource to extract DS mentions from adverse event reports and clinical notes, but their system is restricted to ingredients and do not consider related concepts such as dose form or strength, which are crucial in the RxNorm model [16, 17]. A recent study by L. Wang et al. showed promising results applying and extending the RxNorm model to Chinese clinical drugs [18]. In a similar vein, this study evaluates the feasibility of applying an RxNorm-like normalization approach to DS product names. We developed a rule-based natural language processing (NLP) system which is able to find various components of the product names and assign them to term types, which can be used for normalization. Our system leverages three existing terminologies to develop the NLP patterns: The Therapeutic Goods Administration (TGA) [19], RxNorm, and iDISK - an integrated knowledge base of DSs and related terms [20]. We evaluated the generalizability of the system by reporting its coverage and accuracy on a set of product names extracted from the Dietary Supplement Label Database (DSLD) [21].

Methods

This study is comprised of three phases: data extraction and preprocessing, NLP pattern development, and evaluation. Figure 1 illustrates the overall process and each phase is detailed below.

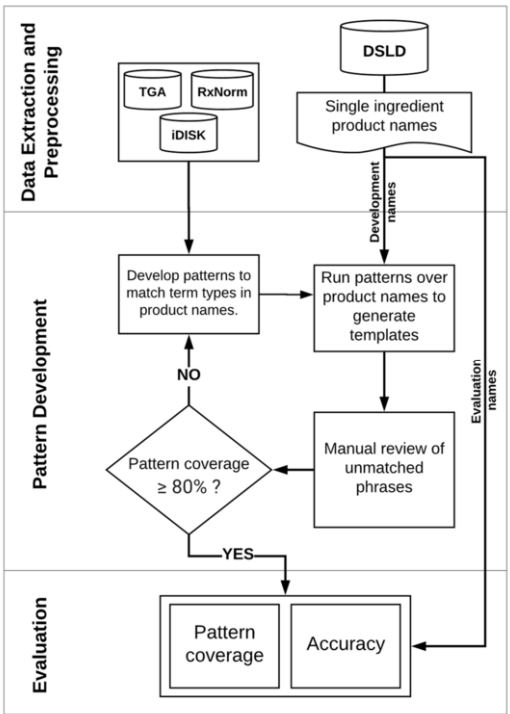


Figure 1– The study design.

Data Extraction and Preprocessing

We extracted 12,383 product names from DSLD using a web scraper. We restricted the extracted names to those listed as containing a single dietary ingredient in order to reduce the amount of variability in the product names. This set of names was then split into a development set (9,906, 80% of the original data) and evaluation set (2,477, 20% of the original data). In order to ensure both sets were representative of the full data set, the split was stratified on the LanguaL (http://www.langual.org) product type assigned to the DS product by DSLD.

NLP System Development

The development set of product names was used to build the NLP system. This system was built using an iterative process comprised of four stages:

1. We developed a set of term types corresponding to components of the product names, detailed in Table 1. We also developed regular expression patterns to match these term types in the product names. These patterns used keyword lists obtained from TGA and RxNorm to match components such as dosages and dose forms, plant preparations (e.g. dried leaf), etc, as detailed in Table 1. We used the ingredient name thesaurus from the iDISK knowledge base to match ingredient names as well as a regular expression for certain vitamins. Where necessary, we manually augmented these keywords lists with lexical variants such as abbreviations and plural forms (e.g. "cap" and "capsules" in addition to "capsule"). Brand names were matched using a combination of a rule based method and manually curated list of brand names extracted from the development set. We removed common stop words from the product names in addition to defining a stop word term type (*STOP*) in order to designate which words should not be included in the normalized product names.

Table 1– Term types used in the product name normalization system.

Term Type (Abbreviation)	Description	Example	Pattern Source
Animal Source (ANM)	The part of an animal from which the ingredient is derived.	Bone Marrow	TGA
Brand Name (BN)	Manufacturer's name.	GNC	Annotation, rules
Certification (CERT)	Official certifications claimed by the product.	USP certified	TGA
Claim or Use (USE)	A description of the purported use of a dietary supplement.	Sleep aid	Annotation
Dose Form (DF)	The phsyical from of the product.	Capsule	TGA, RxNorm
Dose Form Group (DFG)	A grouping of dose forms related by route of administration.	Topical	TGA, RxNorm
Flavor (FLV)	The flavor of a supplement.	Strawberry	Annotation
Ingredient (IN)	Name of the dietary supplement ingredient.	Gingko Biloba	iDISK, rules
Plant Source (PLNT)	The part of a plant from which the ingredient is derived.	Leaf	TGA
Demographic or Population (POP)	The group of persons for whom the product is intended.	Children's	TGA
Preparation (PREP)	A descriptor of how an ingredient is prepared.	Dried	TGA
Stop Word (STOP)	Uninformative words that are to be excluded from the normalized form.	With, Natural	Annotation
Strength (STR)	The quantity of the ingredient in a product.	100 mg	TGA
Time of Use (TIM)	When the product is intended to be used.	Night time	TGA

2. For each name in the development set, we searched for each pattern in turn. Thus the output of this step is an ordered list of term type codes each corresponding to a matched span in the product name string. The ordering of this list matches as closely as possible the RxNorm term types. For example, running the patterns on the product name "Herb Pharm Elderberry" returns the list of matched term types *BN IN*, where *BN* (brand name) matches "Herb Pharm" and *IN* (ingredient) matches "Elderberry". We call each unique list of term types a template. Note that the *STOP* term type is not included in the final templates. Templates correspond to the higher-level RxNorm term types such as *SCDF*. Ambiguous contexts were handled either by the regular expressions themselves (e.g. "mg" for milligrams must be preceded by a number to avoid confusion with magnesium), or by the order in which the

patterns were searched. Regarding the latter case, brand names were search first, followed by ingredients, as these have the most potential for overlap with other term types.

3. We computed the coverage of the patterns on the development set. This included the number of fully matched (i.e. all parts of the name were matched to one or more patterns), partially matched (i.e. some substring of the name was matched), and unmatched product names. Our target full-match coverage on the development set was 80%. If our system did not reach this target, we reviewed the partially matched and unmatched names (step 4 below) and proceeded with the next round of development. If it met or exceeded 80% we moved on to evaluation.
4. At each iteration of the pattern development cycle while the full-match coverage was below 80%, Two health informaticians (RR and AB) manually reviewed 20% of the partially matched and unmatched product names. The results of this review were used to modify existing patterns and create new patterns to improve the coverage of the system.

Evaluation

Evaluation proceeded after our pattern matching system obtained the target 80% full-match coverage on the development set. At this point we ran the pattern matching system on the 2,477 held out evaluation product names and computed the coverage on the evaluation set. Each fully matched product name corresponds to a template which is output by our system. For each template that is also present in RxNorm, we report the frequency with which it occurred in the development and evaluation sets.

Additionally, we evaluated the accuracy of our system on the evaluation set in two ways:

1. We measured the accuracy of the term type patterns on the evaluation set. Each *matched span - term type* pair in each product name in the evaluation set was annotated according to its correctness. We assigned a 1 if the words within the span belonged to the corresponding term type, or a 0 otherwise. We then computed the accuracy for each term type using these annotations.
2. We measured the accuracy over the product names in the evaluation set. This was computed by averaging the accuracies of the product names, where the accuracy of a given product name *n* is the mean of the labels assigned to each token in the name in step 1, computed by

$$accuracy(n) = \frac{1}{|T(n)|} \sum_{t \in T(n)} \ell(t)$$

Where $T(n)$ is a function that returns the tokens in the product name n and $\ell(t)$ is a function that returns the label (1 or 0) for token t.

In the case of partially matched or unmatched product names, each unmatched token is implicitly assigned a 0. This allows us to compute the accuracy of partially matched and unmatched product names and thus an accuracy value for the entire evaluation set.

Results

Running the pattern matching system over the development set produced 129 unique templates using all 13 term types after removing *STOP*. Running the system on the evaluation set produced 62 unique templates, 7 of which were not seen in the development set, for a total of 136 templates. The *TIME* term type was not present in any full matches on the evaluation set.

The 5 most frequent templates across the development and evaluation sets are shown in Table 2.

8 of the 129 development templates and 5 of the 62 evaluation templates matched existing RxNorm term types. The frequencies of these templates in the development and evaluation sets are given in Table 3. In both the development and evaluation sets the *BN IN STR* template (*SBDC* in RxNorm) accounted for about one third (33%) of the fully matched product names. The most frequent of these templates are also the first, third, and fourth most frequent templates overall in both the development and evaluation sets, shown in Table 2. Note that the second most frequent template, *BN IN*, does not have a corresponding RxNorm term type.

Table 2– The 5 most common templates and their product name coverage across the development and evaluation sets along with examples for each.

Frequency ranked templates	Example product name
BN IN STR (32.0%)	Bronson Laboratories Vitamin E 200 IU
BN IN (21.3%)	NutraBio Melatonin
BN IN DF (3.4%)	TERRAVITA Potassium Citrate Powder
BN IN STR DF (3.0%)	Optimum Nutrition Tribulus 625 MG Caps
BN IN PLNT (1.9%)	Nature's Answer Hawthorn Berry

Table 3– Frequencies of templates generated on the development and evaluation sets that match RxNorm term types, computed using the fully matched product names. We do not include the following RxNorm term types: Precise Ingredient (PIN), Multiple Ingredients (MIN), Generic Pack (GPCK), Brand Name Pack (BPCK) as they are not applicable to this study.

RxNorm Term Type	Corresponding Template	Dev Frequency	Eval Frequency
Ingredient (IN)	IN	1 (0.01%)	1 (0.04%)
Semantic Clinical Drug Component (SCDC)	IN STR	1 (0.01%)	0
Semantic Clinical Drug Form (SCDF)	IN DF	2 (0.02%)	0
Semantic Clinical Dose Form Group (SCDG)	IN DFG	0	0
Semantic Clinical Drug (SCD)	IN STR DF	3 (0.03%)	0
Brand Name (BN)	BN	209 (2.11%)	10 (0.40%)
Semantic Branded Drug Component (SBDC)	BN IN STR	3353 (33.85%)	812 (32.78%)
Semantic Branded Drug Form (SBDF)	BN IN DF	370 (3.74%)	80 (3.23%)
Semantic Branded Dose Form Group (SBDG)	BN DFG	0	0
Semantic Branded Drug (SBD)	BN IN STR DF	325 (3.28%)	85 (3.43%)
Total		**4264 (43.04%)**	**988 (39.89%)**

The coverage of the final NLP system on the evaluation set, after obtaining 80% full match coverage on the development set, was 71.9% full match, 27.6% partial match, and 0.5%

unmatched. Thus only 11 (0.5%) evaluation product names were completely unmatched by our system. Compared to the coverage of RxNorm term types (39.89%), our system improves full-match coverage by 32% on the evaluation set.

Table 4 shows the average accuracy of the pattern matching system on the evaluation product names. The average of the fully matched names is 0.30 greater than the partially matched names. This is expected due to the fact that each unmatched token in the partially matched names is treated as incorrect. Still, because the majority of the names in the evaluation set were fully matched, the average accuracy (0.86) is closer to the fully matched accuracy.

Table 4– Overall accuracy on the evaluation set, reported for all evaluation names, only those which were fully matched, and only those that were partially matched.

Match Type	Accuracy
Full + Partial + None	0.86
Full match only	0.95
Partial match only	0.65

The accuracy of each term type, computed over the fully matched and partially matched evaluation set names, is given in Table 4. We report both the average accuracy of each term type over all the evaluation names (given by the bars) as well as the accuracy on the fully matched and partially matches names separately (given by the triangles and Xs, respectively).

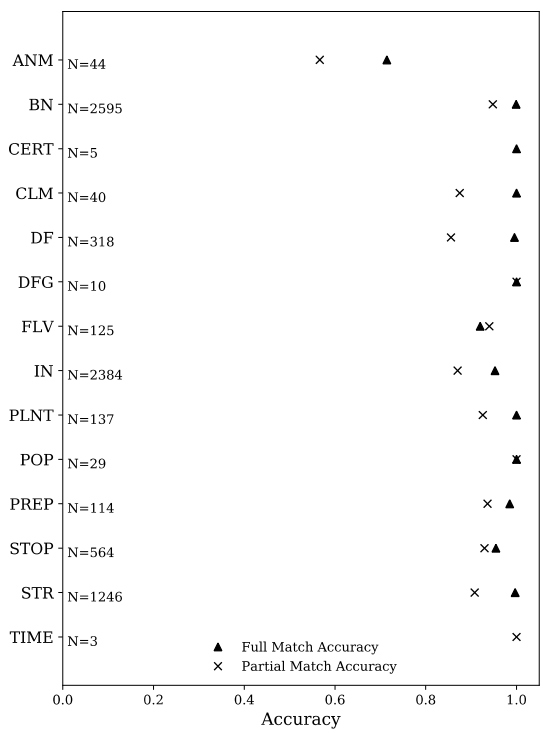


Figure 2– Accuracy of each term type on the evaluation set. The bars indicate the average accuracy of the term types over the full and partially matched product names in the evaluation set. The points indicate the accuracy on the fully and partial matched names, respectively.

A number of term types achieved a perfect 1.00 accuracy, and the most common term types, *BN* and *IN*, achieved accuracies above 0.90. *ANM* (animal source) obtained the lowest accuracy, with 0.61.

Discussion

As shown in Table 3, only about 40% of DS product names fit existing RxNorm term types. This suggests that RxNorm is not well suited to the space of DS products. Table 2 and Table 3 show that brand names (*BN*) play a significant role in DS product labeling, with most or all of the most frequent templates containing *BN*. Indeed, further investigation revealed that 91% of the patterns generated on the development and evaluation sets contain *BN.* This is not surprising, given the different ways in which drugs and DS products are marketed. Drugs are carefully prescribed and regulated, with brand name and generic drugs being for the most part interchangeable, meaning that the *IN* term type is most useful for clinicians, patients, and regulators. On the other hand, there are many competing DS products containing similar ingredients so DS manufacturers emphasize product branding to appeal to consumers.

Nevertheless, the accuracy of our system indicates promising potential for normalizing DS product names. Even when treating unmatched words as misses, our system was able to achieve an accuracy of 0.86 on the evaluation set, which improves to 0.95 on fully-matched product names only (Table 4).

On the other hand, the generalizability of our system is limited by its coverage. Our system was only able to fully match 71.9% of the evaluation set names, a difference of 8% from the development set coverage (80%). Still, a majority of the remaining names were partially matched (27.6%) and our system was unable to find a match for only 11 (0.5%) of the evaluation names.

Reviewing examples of unmatched words in the development and evaluation sets revealed that many were unseen brand names. For example, our system missed the brand name “Cellucor COR-Performance” because it did not occur in the development set. In the future, the use of machine learning methods could improve the coverage and accuracy of the system on brand names, which are too numerous and varied to be manually curated. Many other unmatched words were uninformative buzzwords such as “High Intensity Training Program”. We found that most errors for the *IN* term type were due to the inclusion of phrases that belong to *PLNT* or *PREP*, such as in “Peppermint *Leaf”* and “Green Tea *Extract”*. These occurred because our system searched the *IN* patterns before *PREP* and *PLNT* and the iDISK ingredient thesaurus often includes these phrases in ingredient names. Also, many brand names contain implicit information regarding claims or ingredients which our system could not match. For example, “PomGuard” in “Jarrow Formulas PomGuard” suggests the inclusion of pomegranate as in ingredient.

This study has the following limitations: First, we only include single ingredient products from DSLD. Single ingredient products comprise 22% of all product names extracted from DSLD, so the generalizability of our method to multi-ingredinet products remains to be investigated in future work. Second, because of the above limitation, we assume the presence of at most one ingredient in each product name. Still, some product names listed as single ingredient in DSLD contain more than one ingredient mention, e.g. “Physician's Preference *Royal Garlic* with *Hawthorn* and *Cayenne*”. It would be straightforward to modify our system to search for multiple ingredient mentions, which would increase coverage. Third, our system has limited ability to disambiguate context.

Therefore, only one meaning was chosen for any polysemous keywords, e.g. keywords that occurred in more than one TGA list. Important future work would be to employ more advanced NLP and machine learning methods to disambiguate context in product names. This could vastly improve the accuracy of term types such as ANM, which contains keywords (e.g. “Heart” and “Liver”) that are often confused with claims or uses.

Conclusions

In this study we developed and evaluated an NLP system to apply an RxNorm-like normalization approach to dietary supplement product names. As has been done for drugs, normalization is important to facilitate interoperability and the search for information about DSs. We found that the existing RxNorm drug normalization templates do not generalize to dietary supplements and that it is necessary to extend the RxNorm model to sufficiently cover DS product names. The normalization system outlined here obtains a substantial increase (32%) in coverage on DS product names over RxNorm as well as an accuracy of 0.86. Nevertheless, there is great variability in supplement product names and more work is required to improve the performance of our system.

Acknowledgements

This research was supported by National Center for Complementary & Integrative Health Award (#R01AT009457) (PI: Zhang). The content is solely the responsibility of the authors and does not represent the official views of the National Center for Complementary & Integrative Health. This work was also supported by the Intramural Research Program of the NIH.

References

[1] "FDA 101: Dietary Supplements." https://www.fda.gov/ForConsumers/ConsumerUpdates/ucm050803.htm (accessed 2017).

[2] M. B. Royne, A. K. Fox, G. D. Deitz, and T. Gibson, "The effects of health consciousness and familiarity with DTCA on perceptions of dietary supplements," *Journal of Consumer Affairs,* vol. 48, no. 3, pp. 515-534, 2014.

[3] R. L. Bailey *et al.*, "Dietary supplement use in the United States, 2003-2006," (in eng), *J Nutr,* vol. 141, no. 2, pp. 261-6, Feb 2011, doi: 10.3945/jn.110.133025.

[4] J. J. Gahche, R. L. Bailey, N. Potischman, and J. T. Dwyer, "Dietary Supplement Use Was Very High among Older Adults in the United States in 2011-2014," (in eng), *J Nutr,* Aug 2017, doi: 10.3945/jn.117.255984.

[5] J. J. Knapik, D. W. Trone, K. G. Austin, R. A. Steelman, E. K. Farina, and H. R. Lieberman, "Prevalence, Adverse Events, and Factors Associated with Dietary Supplement and Nutritional Supplement Use by US Navy and Marine Corps Personnel," (in eng), *J Acad Nutr Diet,* vol. 116, no. 9, pp. 1423-1442, 09 2016, doi: 10.1016/j.jand.2016.02.015.

[6] I. Levy, S. Attias, E. Ben-Arye, L. Goldstein, and E. Schiff, "Adverse events associated with interactions with dietary and herbal supplements among inpatients," (in eng), *Br J Clin Pharmacol,* vol. 83, no. 4, pp. 836-845, 04 2017, doi: 10.1111/bcp.13158.

[7] A. I. Geller *et al.*, "Emergency Department Visits for Adverse Events Related to Dietary Supplements," *N Engl J Med,* vol. 373, no. 16, pp. 1531-40, Oct 15 2015, doi: 10.1056/NEJMsa1504267.

[8] "Dietary Supplement Labeling Guide." https://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/DietarySupplements/ucm2006823.htm (accessed 2018).

[9] "Guidance for Industry: Statement of Identity, Nutrition Labeling and Ingredient Labeling of Dietary Supplements; Small Entity Compliance Guide." https://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/DietarySupplements/ucm073168.htm (accessed 2018).

[10] "RxNorm overview." https://www.nlm.nih.gov/research/umls/rxnorm/overview.html# (accessed April, 2018, 2018).

[11] "RxNorm Technical Documentation," vol. 2018, 01/02/2018,

[12] B. o. D. o. t. A. M. I. Association, "Standards for medical identifiers, codes, and messages needed to create an efficient computer-stored medical record," *J. Am. Med. Informatics,* vol. 1, pp. 1-7, 1994.

[13] J. Vasilakes, R. Rizvi, J. Zhang, T. Adam, and R. Zhang, "Detecting Signals of Dietary Supplement Adverse Events from the CFSAN Adverse Event Reporting System (CAERS)," presented at the In press at the 2019 AMIA Informatics Summit.

[14] S. A. Jordan, S. A. Jack, and R. P. Pless, "Safety Surveillance of Dietary Supplements: Importance, Challenges, and New Horizons," (in eng), *Clin Pharmacol Ther,* vol. 104, no. 3, pp. 432-434, Sep 2018, doi: 10.1002/cpt.1101.

[15] Y. Wang, T. Adam, and R. Zhang, "Term Coverage of Dietary Supplements Ingredients in Product Labels," *AMIA Annual Symposium,* pp. 2053-2061, 2016.

[16] V. Sharma and I. N. Sarkar, "Identifying natural health product and dietary supplement information within adverse event reporting systems," (in eng), *Pac Symp Biocomput,* vol. 23, pp. 268-279, 2018.

[17] V. Sharma and I. N. Sarkar, "Identifying Supplement Use Within Clinical Notes: An Applicationof Natural Language Processing," (in eng), *AMIA Jt Summits Transl Sci Proc,* vol. 2017, pp. 196-205, 2018.

[18] L. Wang *et al.*, "Toward a normalized clinical drug knowledge base in China-applying the RxNorm model to Chinese clinical drugs," (in eng), *J Am Med Inform Assoc,* vol. 25, no. 7, pp. 809-818, Jul 2018, doi: 10.1093/jamia/ocy020.

[19] "Therapeutic Goods Administration Code Tables." https://www.ebs.tga.gov.au/ (accessed Nov 12, 2018).

[20] R. Rizvi, J. Vasilakes, T. Adam, G. Melton, J. Bishop, and R. Zhang, "Towards Building and Integrated Dietary Supplements Knowledge Base (iDISK) from Online Resources," *Journal of the American Medical Informatics Association (submitted)*.

[21] "Dietary Supplement Label Database." http://www.dsld.nlm.nih.gov/dsld/ (accessed Nov 12, 2018).

Address for Correspondence

Rui Zhang, PhD. Assistant Professor, Institute for Health Informatics and College of Pharmacy, University of Minnesota, MN, USA. Email: zhan1386@umn.edu.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190254

Identifying Suicidal Adolescents from Mental Health Records Using Natural Language Processing

Sumithra Velupillai [a,b], Sophie Epstein [a,c], André Bittar [a], Thomas Stephenson[c], Rina Dutta [*,a,c], Johnny Downs[*,a,c],

[a] *Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK,*
[b] *School of Electrical Engineering and Computer Science, KTH, Stockholm, Sweden*
[c] *South London and Maudsley NHS Foundation Trust, London, UK*

[*] *contributed equally to this work*

Abstract

Suicidal ideation is a risk factor for self-harm, completed suicide and can be indicative of mental health issues. Adolescents are a particularly vulnerable group, but few studies have examined suicidal behaviour prevalence in large cohorts. Electronic Health Records (EHRs) are a rich source of secondary health care data that could be used to estimate prevalence. Most EHR documentation related to suicide risk is written in free text, thus requiring Natural Language Processing (NLP) approaches. We adapted and evaluated a simple lexicon- and rule-based NLP approach to identify suicidal adolescents from a large EHR database. We developed a comprehensive manually annotated EHR reference standard and assessed NLP performance at both document and patient level on data from 200 patients (~5000 documents). We achieved promising results (>80% f1 score at both document and patient level). Simple NLP approaches can be successfully used to identify patients who exhibit suicidal risk behaviour, and our proposed approach could be useful for other populations and settings.

Keywords:

Natural Language Processing; Suicide; Electronic Health Records

Introduction

Suicidal ideation (SI) occurs in approximately 10% of adolescents in the general population [1,2]. It is a risk factor for self-harm and completed suicide, and is associated with a number of mental disorders. Amongst young people attending Child and Adolescent Mental Health Services (CAMHS), who are often receiving therapeutic support for mental health disorders, the prevalence is higher. SI is an important risk indicator that is used in clinical practice by CAMHS professionals. It is asked about in routine clinical assessments, particularly in situations where patients report low mood, anxiety, emotional regulation difficulties, trauma or distress following difficult life experiences.

Information about suicide risk including suicidal thoughts can be documented in structured risk assessment forms [3], or in free-text in Electronic Health Records (EHRs). Whilst admissions to general hospitals following suicide attempts would be routinely recorded with structured diagnostic codes, routine assessments of suicidal behaviour risk in clinical practice are predominantly recorded as free-text [4–6]. To capture and extract suicide-related information from text within EHRs, Natural Language Processing (NLP) approaches are essential. Common to many NLP solutions for extracting particular pieces of information from EHR text are: to 1) identify concepts of relevance (e.g. the term *suicide*) and 2) classify contextual attributes that semantically alter the meaning of the identified mention. Negation detection is particularly important for clinical constructs that are routinely documented as part of clinical assessments (e.g. *denies suicidal ideation*). Current state-of-the-art NLP methodologies that model such problems in the clinical domain rely on machine learning approaches or symbolic approaches (or a combination of both) [7,8].

However, extracting *patient*-level risks or classifications from (a collection of) documents presents some challenges. Each patient EHR might contain several documents, and each document might contain several relevant (and irrelevant) mentions of clinically important information. Particularly for complex clinical constructs such as suicidality, generating sufficiently large and high-quality datasets that could be used to develop machine-learning based models is costly due to the time and human effort required for manual data annotation. Inferring patient-level labels from individual mentions can be done by post-processing identified mentions using currently available clinical evidence, as was done in a recent study for asserting asthma status [9], or by other heuristic approaches.

Haerian et al. (2012) found that NLP approaches combined with ICD-9 codes had higher precision/positive predictive value (PPV) than using ICD-9 codes alone when evaluated on data from the New York Presbyterian Hospital/Columbia University Medical Center [4]. Similarly, Anderson et al. (2015) found that only 3% of patients who had an indication of suicidal ideation written in text had a corresponding ICD-9 code in a study performed on a large distributed health network of primary care organizations [5]. In a study to identify suicidal behaviour amongst pregnant women, similar findings are reported: using diagnostic codes alone reduced sensitivity considerably [10]. Rule- and machine-learning based approaches to extract suicide ideation and suicide attempt mentions from psychiatric EHR text have been developed with promising results [6], but were evaluated on the mention level, not the patient level.

We have previously developed a step-wise rule-based NLP approach that 1) identifies suicide-related mentions and filters out negated instances using lexicons (for concepts as well as negation terms) and rules based on syntactic information (constituency-based parse tree of sentences), 2) determines a document-level label based on heuristics (counting the number of positive and negative instances and assigning a document-level label based on a majority rule), and 3) determines a patient-level label based on document labels (patient positive for suicidal behaviour = the patient had one (or more) documents labelled as positive for suicidality). This approach was evaluated on a cohort of adolescents diagnosed with autism

spectrum disorder (ASD), resulting in precision, recall, and f1 scores all > 0.85 at the document and patient level [11].

In the study presented here, we extend this approach. We had two main aims: 1) to generate a manually annotated reference standard of an adolescent cohort that was inclusive of all mental health conditions and 2) to apply and modify an existing NLP approach for mention-level extraction of specified clinical constructs [12] that only relied on target (relevant clinical concepts) and modifier (terms that alter the contextual meaning of the target, e.g. negation) lexicons and a sentence tokenizer; thus, minimizing the need for time-costly pre-processing steps such as syntactic parsing, and test its performance on the new cohort. Our focus was on finding *asserted* suicidality risk, filtering out negated cases. Further, we applied previously published target and modifier lexicons (in this case, negation terms), to assess their applicability on a new clinical use-case and dataset. Finally, we developed and evaluated these lexicons at both document and patient level. To our knowledge, this is the first study that focuses on a general adolescent patient cohort and that systematically analyses the applicability of published lexical resources on the problem of retrieving asserted suicide risk from EHR text.

The capacity to extract data on suicidality from free-text EHRs has many implications: it provides an estimate of the prevalence of this problem in a clinical CAMHS population; it generates a cohort of high-risk patients in which to study risk factors for self-harm and suicide; and it contributes to an emerging field of research using text from routinely collected health data and data-driven methods to help understand suicide risk at scale.

Methods

Data Source and Clinical Cohort

We used data from the Clinical Records Interactive Search (CRIS) database [13,14], which contains anonymized EHRs from the South London and Maudsley (SLaM) NHS Foundation Trust and has ethical approval for research use (Oxford REC C, reference 08/H0606/71+5) under an extensive governance model. The full cohort consisted of all patients aged 11-17 and in contact with CAMHS in SLaM between April 1st 2009 and March 31st 2016. All documents for these patients in this time period were extracted (documents include event notes, correspondence letters, and free-text sections from different types of semi-structured assessments), yielding a total of 1,601,422 documents (derived from 23,455 patients)[1]. No structured data was included, and access to patient outcomes was not covered by the ethical approval for this project.

Reference Standard Development

EHR Corpus

All documents for a sample of 200 patients were extracted for manual review. This subset was split into a training set (100 patients, 2883 documents), and a test set (100 patients, 2602 documents). The sample was randomly extracted from the patients who had a number of EHR documents within the 1st and 3rd quartiles (12,146 patients, 342,037 documents in total, cut-off points: minimum of 10, maximum of 61 per patient).

Annotation

A review of similar studies was performed to guide the definition of suicide-related information, and suicidal behaviour. In summary, any mentions that indicated the desire to kill oneself, end one's life, or wanting to be dead/to die (including having tried to kill oneself) were included. Mentions related to self-harm without suicidal intent or no explicit mention of suicidal intent were excluded. No assumptions on implicit information about intent (e.g. overdoses) were made and were thus not annotated. All documents for each patient were manually reviewed and annotated by a child psychiatrist (SE) for a document-level label (non-relevant, suicidal, non-suicidal, or uncertain). Each document labelled as suicidal, non-suicidal or uncertain was also annotated with mention-level annotations (positive, negated, uncertain or unrelated). The definition of a mention-level annotation was deliberately not constrained to specific textual units (e.g. word or paragraph), instead, any text segment that was indicative of the mention-level labels could be marked by the annotator to allow for expressivity and clinically relevant segments. Guidelines were iteratively developed to refine inclusion and exclusion criteria for each label. A small subset (100 documents, randomly extracted with a 2:1 ratio of not positive:positive) was annotated by a second clinical annotator (TS) to measure inter-annotator agreement (IAA). We used the Extensible Human Oracle Suite of Tools (eHost)[2] annotation tool.

Document- and Patient-Level Label Heuristics

Because the purpose of the clinical use-case for which this approach was developed was to generate a cohort of patients who had *ever* exhibited suicidal tendencies, priority was placed on detecting *affirmed* suicide-related information. Thus, if there was *at least one* affirmed mention of suicidality in a document, the document label was set to suicidal. Similarly, a patient-level label of suicidal was set if there was at least one document labelled as positive for suicidality (Figure 1).

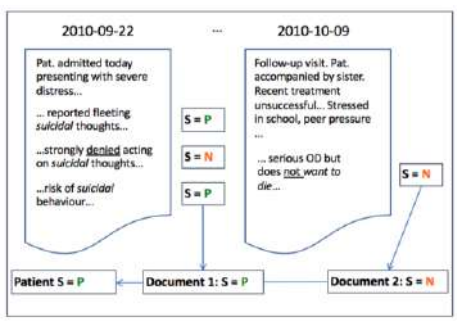


Figure 1 - Example illustrating document and patient-level label heuristics. A patient may have several EHR documents. This example illustrates two fictitious documents from two fictitious dates, where there are multiple suicide-related mentions in the first document (left), 2 positive (P) for suicidality (S) and one negated (N), which leads to a positive document-level label, and only one mention in the second document (right) which is negated, leading to a negated document-level label. Because one document was labelled as positive, the patient-level label is positive (Patient S = P)

Natural Language Processing Approach

We extended our previous approach by relying solely on pyConTextNLP [12] (version 0.6.0.0) that only requires a sentence tokenizer for pre-processing, and on two lexicons: one that defines relevant concepts/mentions (*target terms*, e.g. *suicidal*) and one that defines modifiers (e.g. negations) and so-called termination terms. Termination terms define if the scope of a target modifier should be triggered within a sentence. For instance, in a sentence like *'The patient **denies** wanting to die **but** has had suicidal thoughts.'*, the scope of the negation term

[1] As of 12 April 2018

[2] http://blulab.chpc.utah.edu/content/ehost-extensible-human-oracle-suite-tools

'denies' covers *'wanting to die'* and not '*suicidal'* due to the termination term ***but*** pyConTextNLP allows for matching with character-based regular expressions. We used spaCy[3] (version 2.0.9) for sentence tokenization. The work was performed in a Python 2.7.6 environment. The development and analysis of target and modifier term lexicons was the main focus, and was done in two main phases. For target terms, we reviewed previous similar studies [6,11,15] and experimented empirically with these to identify a good coverage lexicon. As a baseline, we used a minimal target lexicon only containing the term *suicide* and its variants (*suicid**) to assess the impact of the addition of other target terms. For modifier terms, we used three previously published off-the-shelf lexicons: 1) negation terms used for detecting suicidality in adolescents with ASD ("AMIA2017") [11], 2) negation terms used to detect suicidal ideation ("SREP2018") [6], and 3) a collection of negation terms developed for a variety of clinical use-cases (not specifically suicidality) and languages ("MEDINFO2013") [16]. The off-the-shelf lexicon that yielded best results on the development set was then further iteratively adapted based on findings from manual error analyses. To assess the applicability of the lexicons on the development set, priority was given to optimizing performance on detecting *affirmed* suicidality at both document and patient level. The test set was held aside throughout the development phase.

Table 1 – Reference standard document label distributions

document label	development set	test set
non-relevant	2570	2393
suicidal	136	54
non-suicidal	46	33
uncertain	131	122
total	*2883*	*2602*

Table 2 – Reference standard patient label distributions

patient label	development set	test set
non-relevant	52	66
suicidal	37	26
non-suicidal	11	8
total	*100*	*100*

Evaluation

We measured inter-annotator agreement on a subset with Cohen's κ and accuracy. NLP performance was measured against the annotated reference standard development and test sets with precision (PPV), recall (sensitivity) and f1-score at a document- and patient-level, focusing on results for *affirmed* suicidality. Finally, qualitative error analysis on results was performed to inform future improvements.

Results

Reference Standard and Inter-Annotator Agreement

The overall distribution of the resulting annotations on a document and patient level is presented in Tables 1 and 2. Non-relevant documents are in the majority of both the development and test sets (89% and 92%, respectively). A 4.7% of the documents in the development set were positive for suicidality, and 2.1% in the test set. On a patient level, the distribution of patients labelled as suicidal was: 37% (development set) versus 26% (test set). Inter-annotator agreement on the subset of 100 documents was very high: κ 0.96, 98% accuracy.

Natural Language Processing Results

Adaptations of the target term lexicon involved adding terms like *end his/her life*, *kill him/herself*. Overall performance with these additions resulted in 15-17% f1-score improvement on document-level classification (baseline f1-score results using only *suicid** were 64-72% on the development set (56-68% on the test set). On a patient level, baseline results ranged between 70-79% on the development set (71-79% on the test set), compared to 83-86% using the extended target lexicons on the development set (82-83% on the test set).

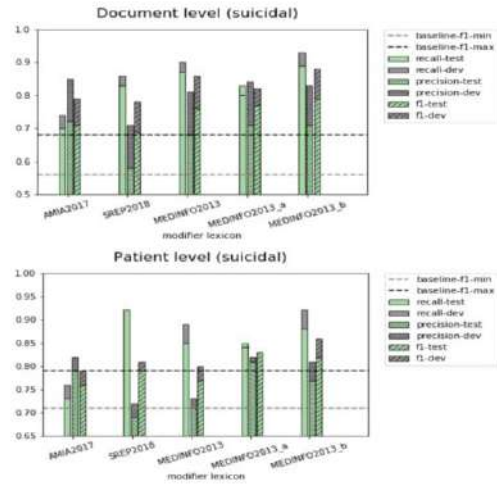


Figure 2 - Results (precision, recall and f1-score) on development (dev, grey) and test (green) sets for document (top) and patient (bottom) level classification (positive for suicidality). Each value on the x-axis represents a modifier lexicon configuration: 1) previous studies: AMIA2017 [11]; SREP2018 [6] and MEDINFO2013 [16] or 2) modifications to the MEDINFO2013 lexicon based on iterative error analysis on the dev set (MEDINFO2013_a and _b). The dashed lines represent f1-scores using the baseline target lexicon (suicid only) on the test set: grey line shows lowest results (min), black line highest results (max)*

Results for the different configurations of modifier lexicons are shown in Figure 2 (document-level on top, patient-level bottom). The precision results on the document-level ranged from 71% (using the SREP2018 lexicon) to 85% (using the AMIA2017 lexicon); recall results ranged from 74% (AMIA2017 lexicon) to 90% (MEDINFO2013 lexicon); f1-score results from 78% (SREP2018) to 86% (MEDINFO2013) without any further adaptation on the development set. The MEDINFO2013 lexicon resulted in best overall off-the-shelf f1-score results and was used for additional adaptation. Errors produced by the MEDINFO2013 lexicon were analysed and iteratively revised based on the development set by adding missing negation terms and revising conjunction terms. We present results for two adapted lexicons: MEDINFO2013_a and _b. In both lexicons, the following negation terms were added: *doesn't*, *doesnt* (sic), *never*, *none*. The main difference between the two adaptations is the list of conjunction/termination terms. In MEDINFO2013_b, 24 conjunction terms, such as *unless* and *whereas,* were added. Furthermore, some conjunction terms that were included in the original MEDINFO2013 lexicon were excluded in both adapted versions, like *patient*, *recent*, *today*.

[3] https://spacy.io/

Document-level results on the blind test set were overall lower compared to the development set (58-72% precision, 70-87% recall and 69-75% f1-score) when using the off-the-shelf modifier lexicons. The adapted lexicons also resulted in lower overall performance compared to the development set, except for recall using MEDINFO2013_a, which improved from 80% to 83%, and MEDINFO2013_b which improved from 88% to 89% (but resulted in lower f1-score).

On a patient level, results for the off-the-shelf lexicons ranged from 72-82% precision, 76-92% recall and 79-81% f1-score on the development set. The adapted lexicons resulted in more balanced, and, on average, higher precision and recall results (81-82%, 84-92% respectively) and slightly higher f1-scores (83-86%). On the blind test set, best f1-score results were obtained with the two adapted lexicons (82% for MEDINFO_a and 83% for MEDINFO_b).

On the test set, the lowest baseline f1-score document-level results were produced by the AMIA2017 lexicon (56%), the highest with the adapted lexicons MEDINFO2013_a and _b (68%). However, on patient-level results the highest baseline result was 79% (MEDINFO2013_b) but results were similar for MEDINFO2013 and MEDINFO2013_a (78%), the lowest results were produced with the AMIA2017 lexicon (71%).

Discussion

We have developed a new comprehensive manually annotated reference standard of EHRs from an adolescent mental health patient cohort. Most documents do not contain any suicide-related information (~90%), only 2-5% documents contain affirmed suicide-related information. On a patient level prevalence is higher (26-37%). This is slightly higher than was found in our previous work [11], but the definition of what constituted affirmed suicide-related information was not identical. Inter-annotator agreement on a small subset was high, indicating that the task is well-defined. Previous studies have shown that using diagnostic codes alone to identify risk of suicidal behaviour is insufficient [4,5,10]. We performed a preliminary assessment of this on our development data and found similar results: only 6 of 37 patients (16.2%) would have been identified as at risk of suicidal behaviour if using information from structured risk assessment forms only.

Our adaptation of pyConTextNLP resulted in similar or higher overall results as compared with our previous NLP approach [11] when applied on the development set, with the added benefit of more efficient processing time (approx. 10 times faster) mainly due to not requiring time-costly pre-processing steps such as syntactic analysis. This shows that a lexical, surface-level approach is sufficient to deal with syntactic phenomena such as negation in these contexts. Off-the-shelf modifier lexicons unsurprisingly yielded varying results – adapting these lexicons based on findings from iterative error analysis on the development set resulted in improvements also on the blind test set.

For our clinical use-case, accurate patient-level classifications are the priority, not document-level. In general, high recall was easier to achieve than high precision. This is probably due to the NLP approach failing to correctly classify certain types of negations (e.g. semi-structured forms, general advice paragraphs) and third party mentions (e.g. relatives), which generate false positives and lower precision, without effecting recall. Precision is of greater importance for this use-case, as false positives are a bigger concern than false negatives.

Extending the target lexicon with additional terms improved document-level results more than patient-level results. This indicates that the term *suicide* and its variants really is the key term overall on patient-level assertions (i.e., it is sufficient for one document to have one affirmed mention of *suicid**, and this term is by far the most common suicide-related term overall). However, if it is important for the end use-case to e.g. identify *when* the first mention of affirmed suicide risk is documented, optimal performance on the document level would be essential. To our knowledge, our study is the first of its kind to extensively analyse the relation between document- and patient-level assertions using this particular type of NLP approach. Our findings could be valuable for further studies in this area, e.g., distributions of suicide-related information in affirmed, as well as negated or other, semantic modifications overall in EHR notes, to inform post-processing heuristics or other data-driven representations of this clinical construct. All lexicons, guidelines and scripts are made available online[4].

Our study has some limitations. The majority of the documents were only annotated by one clinician. An extension of our work would involve double-annotating a larger subset for further inter-annotator agreement analysis. Because the agreement was high on our subsample, we would expect to get high agreement also on a larger sample. We have only focused on applying negation detection on this data and with this NLP approach. We plan to further study the uncertain and unrelated annotations and develop our lexicons to capture these semantic modifiers as well as filtering out third party mentions and references to the past. The off-the-shelf lexicons were not developed specifically for the NLP system we used in this study and in particular were not designed with the same approach for finding the scope of modifiers by defining conjunction terms. We have not studied the individual effects of these terms, which would be an interesting future area to study. In addition, we plan to extend the analysis of lexical terms with other external resources such as the UMLS and SNOMED-CT. Clinical text is often written in non-standard grammatical structures (e.g. long sentences, abbreviations, lack of subjects and/or predicates) which motivates further analysis in the pre-processing steps (sentence and word tokenization) as well as document sections. Furthermore, we have prioritized results on extracting asserted suicide-related information. It might be the case that improving results on the negation detection part would lead to improved results on the assertion detection part, which is something we will study further. Finally, lexicon- and rule-based approaches always suffer from poor generalizability in the sense that spelling variants and new terms need to be added manually. We plan to extend our comparison with other approaches such as Metamap, and also to apply more data-driven approaches for finding relevant terms e.g. by developing word- and sentence embedding models on larger datasets to automatically generate extended representations. However, because of the low prevalence problem on the document level, this approach will need careful study design to ensure that sufficient data samples are used. One approach, similar to Downs et al. [11], would be to filter documents using specific suicide-related target terms and use these documents for generating semantic representations. The applicability of this approach on EHR data from other institutions would also need further investigation.

By adapting and developing a simple NLP tool to identify text relating to affirmed mentions of suicidality, we have been able to identify a cohort of patients who are at higher risk of future adverse outcomes including self-harm. This cohort could be studied for a number of important clinical outcomes. Initially, we are planning a study which will explore socio-demographic,

[4] https://github.com/KCL-Health-NLP/camhs_pycontext_adaptation

clinical and educational risk factors for hospital presentations with self-harm in this high-risk population. This could provide additional information to clinicians on which to base assessments of risk for self-harm and could therefore have an important impact on clinical management.

Conclusions

We have developed a comprehensive manually annotated EHR corpus with rich suicide-related information from a general adolescent cohort. Our adaptation of a rule-based NLP approach and systematic analysis of existing lexicons applied to this data is a contribution to this understudied field for several reasons: we show that simple NLP approaches can have promising results; we confirm previous published results that most suicide-related information is written in free-text; we illustrate the challenges in defining patient-level labels based on mention- and document-level extraction for this low prevalence problem. Our adapted approach could be used in different populations and settings to define cohorts for similar use-cases. We look forward to seeing further developments of NLP approaches for this important problem.

Acknowledgements

Funding support: SV: Swedish Research Council (2015-00359), Marie Skłodowska Curie Actions, Cofund, Project INCA 600398. SE and the project: an MQ Data Science award and The Psychiatry Research Trust (grants held by RD). RD: a Clinician Scientist Fellowship (project e-HOST-IT) from the Health Foundation in partnership with the Academy of Medical Sciences, also funds AB. JD: Medical Research Council (MRC) Clinical Research Training Fellowship (MR/L017105/1) and Psychiatry Research Trust Peggy Pollak Research Fellowship in Developmental Psychiatry.

References

[1] M.K. Nock, J.G. Green, I. Hwang, K.A. McLaughlin, N.A. Sampson, A.M. Zaslavsky, and R.C. Kessler, Prevalence, correlates, and treatment of lifetime suicidal behavior among adolescents: results from the National Comorbidity Survey Replication Adolescents Sup, *JAMA Psychiatry* **70** (2013) 300–310.

[2] B. Mars, J. Heron, E.D. Klonsky, P. Moran, R.C. O'Connor, K. Tilling, P. Wilkinson, and D. Gunnell, What distinguishes adolescents with suicidal thoughts from those who have attempted suicide? A population-based birth cohort study, *J Child Psychol Psychiatry* (2018).

[3] C.J. Hawley, B. Littlechild, T. Sivakumaran, H. Sender, T.M. Gale, and K.J. Wilson, Structure and content of risk assessment proformas in mental healthcare, *J Ment Health* **15** (2006) 437–448.

[4] K. Haerian, H. Salmasian, and C. Friedman, Methods for identifying suicide or suicidal ideation in EHRs, in: *AMIA Proc*, 2012: pp. 1244–1253.

[5] H.D. Anderson, W.D. Pace, E. Brandt, R.D. Nielsen, R.R. Allen, A.M. Libby, D.R. West, and R.J. Valuck, Monitoring suicidal patients in primary care using electronic health records, *J Am Board Fam Med* **28** (2015) 65–71.

[6] A.C. Fernandes, R. Dutta, S. Velupillai, J. Sanyal, R. Stewart, and D. Chandran, Identifying Suicide Ideation and Suicidal Attempts in a Psychiatric Clinical Research Database using Natural Language Processing, *Sci Rep* **8** (2018) 7426.

[7] Y. Wang, L. Wang, M. Rastegar-Mojarad, S. Moon, F. Shen, N. Afzal, S. Liu, Y. Zeng, S. Mehrabi, S. Sohn, and H. Liu, Clinical information extraction applications: A literature review, *J Biomed Inform* **77** (2018) 34–49.

[8] S. Wu, T. Miller, J. Masanz, M. Coarr, S. Halgrim, D. Carrell, and C. Clark, Negation's not solved: generalizability versus optimizability in clinical natural language processing, *PloS One* **9** (2014) e112774.

[9] H. Kaur, S. Sohn, C.-I. Wi, E. Ryu, M.A. Park, K. Bachman, H. Kita, I. Croghan, J.A. Castro-Rodriguez, G.A. Voge, H. Liu, and Y.J. Juhn, Automated chart review utilizing natural language processing algorithm for asthma predictive index, *BMC Pulm Med* **18** (2018) 34.

[10] Q.-Y. Zhong, E.W. Karlson, B. Gelaye, S. Finan, P. Avillach, J.W. Smoller, T. Cai, and M.A. Williams, Screening pregnant women for suicidal behavior in electronic medical records: diagnostic codes vs. clinical notes processed by natural language processing, *BMC Med Inform Decis Mak* **18** (2018).

[11] J. Downs, S. Velupillai, G. George, R. Holden, M. Kikoler, H. Dean, A. Fernandes, and R. Dutta, Detection of Suicidality in Adolescents with Autism Spectrum Disorders: Developing a Natural Language Processing Approach for Use in Electronic Health Records, *AMIA Proc* **2017** (2017) 641–649.

[12] B.E. Chapman, S. Lee, H.P. Kang, and W.W. Chapman, Document-Level Classification of CT Pulmonary Angiography Reports Based on an Extension of the ConText Algorithm, *J Biomed Inf* **44** (2011).

[13] G. Perera, M. Broadbent, F. Callard, C.-K. Chang, J. Downs, R. Dutta, A. Fernandes, R.D. Hayes, M. Henderson, R. Jackson, A. Jewell, G. Kadra, R. Little, M. Pritchard, H. S., A. Tulloch, and R. Stewart, Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) Case Register: current status and recent enhancement of an Electronic Mental Health Record-derived data resource, *BMJ Open* **6** (2016).

[14] J. Downs, R. Gilbert, R.D. Hayes, M. Hotopf, and T. Ford, Linking health and education data to plan and evaluate services for children, *Arch Dis Child* **102** (2017) 599.

[15] C. Polling, A. Tulloch, S. Banerjee, S. Cross, R. Dutta, D.M. Wood, P.I. Dargan, and M. Hotopf, Using routine clinical and administrative data to produce a dataset of attendances at Emergency Departments following self-harm., *BMC Emerg Med* **15** (2015) 15.

[16] W. Chapman, D. Hillert, S. Velupillai, M. Kvist, M. Skeppstedt, B. Chapman, M. Conway, M. Tharp, D. Mowery, and L. Deleger, Extending the NegEx Lexicon for Multiple Languages, *Stud Health Technol Inform* **192** (2013) 677–681.

Address for correspondence

Sumithra Velupillai, sumithra.velupillai@kcl.ac.uk

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190255

Annotating Temporal Relations to Determine the Onset of Psychosis Symptoms

Natalia Viani[a], Joyce Kam[a], Lucia Yin[a], Somain Verma[a], Robert Stewart[a,b], Rashmi Patel[a,b], Sumithra Velupillai[a,c]

[a] Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK,
[b] South London and Maudsley NHS Foundation Trust, London, UK,
[c] School of Electrical Engineering and Computer Science, KTH, Stockholm, Sweden

Abstract

For patients with a diagnosis of schizophrenia, determining symptom onset is crucial for timely and successful intervention. In mental health records, information about early symptoms is often documented only in free text, and thus needs to be extracted to support clinical research. To achieve this, natural language processing (NLP) methods can be used. Development and evaluation of NLP systems requires manually annotated corpora. We present a corpus of mental health records annotated with temporal relations for psychosis symptoms. We propose a methodology for document selection and manual annotation to detect symptom onset information, and develop an annotated corpus. To assess the utility of the created corpus, we propose a pilot NLP system. To the best of our knowledge, this is the first temporally-annotated corpus tailored to a specific clinical use-case.

Keywords:
Schizophrenia; Electronic Health Records; Natural Language Processing

Introduction

For patients with a diagnosis of schizophrenia, duration of untreated psychosis (DUP) is the period of time between the onset of first symptoms and the initiation of adequate treatment [1]. As shown in previous studies, prolonged DUP is associated with poor intervention outcomes, both in the first years of treatment and in the long-term [2,3]. Therefore, to enhance the management of symptoms and improve social functioning, timely treatment is crucial. For determining symptom onset and ultimately reducing DUP, the information collected in clinical practices could be successfully re-used.

With the rapid adoption of electronic health records (EHRs), clinical data are increasingly available in electronic format, allowing for large-scale retrospective research. However, especially in the field of mental health, clinically relevant information (e.g., symptoms, diagnoses, medication) is often documented in unstructured form (free text), for instance through letters and progress notes. To allow analyzing the information enclosed in such clinical text, natural language processing (NLP) techniques are becoming increasingly popular [4]. In the case of determining symptom and treatment onset from clinical notes, NLP methods are needed to both extract clinical concepts (events) and anchor them on a timeline. To this end, two types of temporal information have to be identified: time expressions such as dates and times (TIMEXes), and temporal links (relations) between these and the available events ({hallucinations} in {2002}).

In recent years, a few clinical corpora have been annotated for temporal relations (TLINKs), and used for developing NLP systems tailored to this task. The 2012 i2b2 Temporal Relations Challenge focused on temporal relations in narratives from an intensive care unit [5]: a total of 310 discharge summaries were annotated with events, temporal expressions, and 8 types of temporal relations (e.g., "before", "overlap"). Styler IV et al. developed a corpus (THYME) of 1,254 cancer patient records, which were annotated with clinical and temporal information [6]. This corpus was then used in the 2015 and 2016 Clinical TempEval challenges (440 and 591 documents, respectively), which focused on determining two types of TLINKs [7,8]: relations between events and the document creation time (DCT), and relations between an event or a TIMEX and a narrative container. Successful NLP systems developed on these corpora have mainly relied on supervised machine learning algorithms, using lexical, morphological and syntactical features. A few systems also included heuristics and rule-based components.

Despite the recent advances in temporal relation extraction, developing temporal NLP systems in different clinical domains remains a challenge, due to the inherent complexity of the task – each patient can have several EHRs with clinically relevant information, and in each document every clinically relevant event can in principle be linked to every TIMEX. In this paper, we address the problem of determining temporal relations in mental health records, with a focus on symptom onset identification for schizophrenia patients. To the best of our knowledge, this is the first study on temporal relation extraction that was driven by a specific psychiatric clinical use-case. The NLP task was defined from a clinical perspective, with the final goal to extract relevant information at the patient level. As a crucial initial step to reach this goal, we needed to not only identify the clinically relevant events to be extracted and anchored in time, but also the set of documents that were likely to contain this information.

We have three main aims in this study. First, we propose a methodology for selecting the most relevant documents for the considered use-case. Then, we develop a manual annotation process to temporally anchor all the relevant symptoms, thus enabling the extraction of symptom onset and other information of interest. Finally, we propose a preliminary NLP system to assess the utility of the created corpus.

Methods

Dataset

In this study, we used mental health records from the Clinical Record Interactive Search (CRIS) database [9]. This research

repository contains anonymized patient data (structured and unstructured), derived from the EHR system used at the South London and Maudsley National Health Service (NHS) Foundation Trust (SLaM). Textual documents typically consist of either notes related to specific events or attachments of different types (e.g., assessments, discharge letters). In the system, there are no structured elements indicating whether a document represents a first assessment, which would be helpful to identify relevant content for our use-case.

For selecting the records that would most likely include the information of interest, we focused on documents from early intervention services for people with first episode psychosis (FEP). We considered documents written within a 3-month window by six intervention teams from the team's acceptance date in April 2018, on the assumption that these "early" documents would include the initial assessment and the richest description of a patient's clinical history. The steps followed for document extraction are shown in Figure 1. We focused on longer documents, in which clinicians typically document the presenting history and mental state examination, and excluded questionnaires and forms consisting of short lines. To identify how symptom onset information was typically documented, 70 documents were double-annotated for relevant paragraphs such as: "difficulties were noted for the first time when the patient was 7 years old, as he was displaying aggressive behaviour."

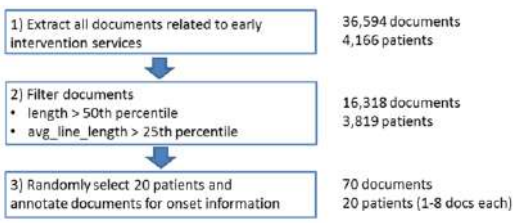


Figure 1 – First Document Extraction.

We analysed the annotated documents in terms of clinical and temporal content, automatically identifying symptoms (i.e., events) and time expressions. For extracting symptoms, we used a previously-developed keyword list [10] containing 598 psychosis-specific symptom terms. To identify TIMEXes, we used SUTime [11], a rule-based tool for temporal expression extraction which we have adapted to the mental health domain [12]. As a result of these analyses, we selected the final set of documents to be used for annotation. In general, documents do not follow a standard format. In some cases, there is a semi-structured format, with section headings (e.g., history of presenting complaint, clinical history, mental state examination), but there is a large variability. It is important to note that, while some events are reported with a specific date (e.g., "presented on 1st Jan. 2014 with hallucinations"), others are not clearly linked to a temporal point (e.g., references to "the past", ongoing symptomology).

Temporal Relation Annotation

Documents were pre-annotated with symptoms and time expressions, using the same tools as for corpus selection. Annotators were asked to try to link each pre-annotated event to a TIMEX, if such relation could be inferred from the text. In addition, they had to assign to each event a polarity value, which could be either "positive" or "negative" (e.g., "denied hallucinations"). This distinction is important, as negated symptoms would not likely indicate onset information, and should be represented differently on a patient's clinical timeline. Figure 2 shows an annotation example.

Figure 2 – Annotation Example.

The corpus was divided in batches, each including documents belonging to 9-10 patients (Table 1). Annotations were carried out by three medical students, and all documents were double-annotated. To guide the annotation task, two NLP researchers created specific guidelines, which were enriched with relevant example cases. To create the final version of the corpus, all annotated documents were adjudicated, resolving disagreements and performing corrections when needed.

To investigate if symptoms were linked to time points prior to the clinic visit, we also analysed, for each patient, how far back in time (in terms of days) a symptom referred to. In this analysis, we did not consider symptoms with negative polarity, as these are not likely to represent onset information.

Automated Temporal Relation Extraction

We used the annotated corpus to develop two NLP modules: TLINK extraction and polarity classification. The dataset was randomly split into training, development, and test sets. The training set was used for system development and manual rule engineering, with validation on the development set. The test set was set aside for final evaluation (in future studies).

In the temporal relation module, we addressed TLINK extraction (i.e., determining if the event can be linked to a time expression in the document) and TIMEX assignment (i.e., finding the normalized value of the TIMEX linked to the event) simultaneously. To perform both tasks, we developed a rule-based system relying on a number of features, such as the section in which the event is found and the presence of anchor dates in the text (admission, discharge, clinic dates). More specifically, section labels were identified by using a set of keywords (e.g., "history", "examination"), while anchor dates were extracted with regular expressions. System development was carried out on a subset of the training set, iteratively adding/refining rules. As a result, we developed ten rules to be applied following an order of relevance. For example, if only one TIMEX (representing a date like YYYY-MM-DD) is found in the same sentence including the symptom, a link is created. As another example, if a symptom is mentioned in a section named "mental state examination on admission" and an admission date is available, a link is created.

To assign a polarity value to events, we used ConText [13], a rule-based algorithm which relies on modifiers (e.g., "no", "denies") to determine whether a concept is negated. These modifiers are looked for in a window of words surrounding the event. In this paper, we used 11 "negation" modifiers representing the terms that were found in our corpus.

Evaluation

To evaluate the quality of the developed corpus, we computed inter-annotator (IAA) agreement. For polarity, we calculated accuracy on symptoms: an agreement is obtained when both annotators marked the same value. For the TLINK task, we defined an "adapted" accuracy, considering for each symptom two cases of agreement: i) both annotators identified 0 links, or ii) both annotators identified a link to the same time expressions (in terms of normalized value). All other combinations were regarded as disagreements. Automated extraction systems were evaluated with the same metrics.

Results

Corpus Selection

Figure 3 shows the results of the analysis conducted for corpus selection (on a subset of 70 documents). For each document, we computed the number of automatically extracted symptoms/TIMEXes (orange lines) and of manually annotated onset paragraphs (blue lines). Results (normalized counts) indicate that documents with many clinical/temporal elements are more likely to contain information on symptom onset.

Starting from this observation, we filtered the initial corpus by adding the following criteria: Symptom_count > 0 and Timex_count > 5. The number of symptoms was computed by using a list of 26 keywords developed by two psychiatrists, and the number of TIMEXes was found with the adapted SUTime. This filtering step led to a final set of 9,779 unique documents for 3,433 patients. From this corpus, we extracted 645 documents for 239 randomly selected patients (an average of 2.7 documents per patient) grouped into 24 batches.

Corpus Annotation

Table 1 reports the number of patients, documents, events, and TIMEXes in our corpus.

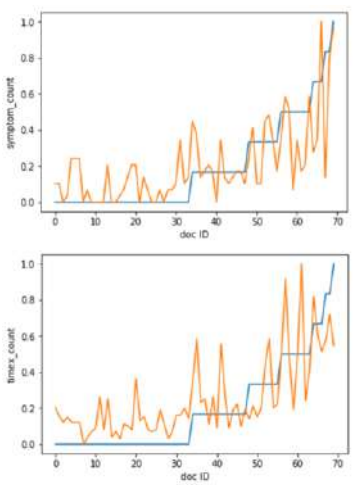


Figure 3 – Relation between Symptom/TIMEX Counts (orange lines) and Manually Annotated Onset Paragraphs (blue lines).

There are 2,590 symptoms (on average, 4 per document). The 5 most frequent (73% of events) are the following (raw counts between brackets): hallucinations (736), delusions (430), delusional (398), paranoia (179), and thought disorder (159). Each symptom was manually annotated for a Polarity value and an optional TLINK. Table 2 reports the IAA, while Table 3 shows the prevalence of Polarity values and temporal relations (for TLINKs, "Yes" represents the existence of a link). These counts were computed on the adjudicated dataset.

Table 1 – Patient, Document, Event and TIMEX Counts

	Total	Train	Dev
Patients (batches)	239 (24)	140 (14)	49 (5)
Documents	645	361	133
Events	2,590	1,465	515
TIMEXes	24,135	13,502	5,061

Table 2 – Inter-Annotator Agreement (IAA) per Annotated Item

Item	IAA (average)	IAA (range/batch)
TLINK	0.73	0.60 - 0.84
Polarity	0.95	0.81 - 1

Table 3 – Annotation Results for TLINK and Polarity

Item	Value	Total	Train	Dev
TLINK	Yes	1,661 (64.1%)	945	302
	No	929 (35.9%)	520	213
Polarity	Pos	1,900 (73.4%)	1,110	368
	Neg	690 (26.6%)	355	147

Starting from the adjudicated annotations, for each patient we computed the difference between the maximum and the minimum dates associated to any "positive" symptom (diff). It is important to note that, for those symptoms that were not explicitely linked to a date, this difference could not be computed. As a result, we were able to compute diff values for 206 patients (Figure 4). Out of these, 41 (20%) had a diff value longer than one year, while 71 (34%) had a zero diff value.

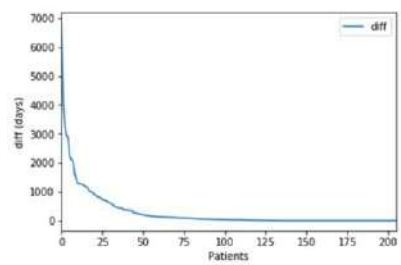


Figure 4 – Difference (in days) between Minimum and Maximum Symptom Dates for 206 Patients

Automated Temporal Relation Extraction

Table 4 shows preliminary results for the two developed NLP modules, using the same metrics as for IAA. For comparison purposes, we also report two baseline results: for the Polarity attribute we classified each event as "positive", while for TLINKs we did not link any event to any specific TIMEX.

Table 4– Performance of NLP modules

Item	Model	Train	Dev
TLINK	baseline	0.47	0.54
	Rule-based	0.67	0.58
Polarity	baseline	0.76	0.72
	ConText	0.93	0.95

Discussion

We have developed a corpus of mental health records for patients with schizophrenia who have been admitted to early onset intervention services, annotated with temporal relations to capture the onset of psychosis symptoms. To the best of our knowledge, this is the first temporally-annotated corpus that was developed for a specific clinical use-case besides clinical timeline reconstruction. In particular, our use-case is related to the analysis of symptom onset and to the calculation of DUP on a large patient cohort. To address this long-term goal, dataset selection was crucial: we applied symptom/TIMEX-based filtering steps to the available CRIS data, and selected multiple documents referred to each patient. Starting from a mention-level annotation task, we aim at proposing a framework that could be also relevant for information extraction on a patient-level. The guidelines and the keywords used in the annotation process, as well as the code for NLP development, are available at: https://github.com/medesto/.

Besides this underlying perspective, our corpus differs from related datasets (i.e., i2b2 2012 and THYME) in two main ways. First, to allow capturing onset information which can be reported across different sentences/paragraphs, we did not require linked entities to be close to each other: each event could be linked to any time expression written in the document. This is a first step towards reconstructing timelines across multiple documents, a problem that remains understudied in the clinical domain. Raghavan et al., for example, proposed a system for cross-document alignment of event sequences [14]. Second, to simplify the annotation task and still obtain useful data, we asked annotators to associate each event to only one time expression (the most relevant one), thus considering only one type of TLINK. Given these differences, we proposed the use of an adapted accuracy to measure the IAA on temporal links, with a final value of 0.73. This represents a promising result, especially considering the inherent difficulty of the task. In some cases, for example, annotators found it difficult to decide whether a symptom should be linked to a given date: even if a temporal link could be reasonably inferred, the relation was not clearly stated in the text. As another cause of disagreements, for a few symptoms that were clearly related to the visit date, this date was not explicitly written in the text: in these cases, the "most likely" date was often chosen. As for the Polarity attribute, we obtained a particularly high IAA, with an accuracy of 0.95. As expected, classifying a symptom as positive or negative was easier than contextualizing it from the temporal point of view.

In the adjudicated corpus, 1661 symptoms (64%) were linked to a specific date. It is interesting to note that 541 (33%) of these symptoms were negated, and therefore do not play a role for symptom onset extraction. As a matter of fact, 197 negated events (36%) were found in "examination" sections, thus representing results of patient visits. We aimed at capturing these temporal links as they could be important for general timeline reconstruction, however they might not be directly relevant to our long-term goal. To assess the utility of our data for symptom onset extraction, we analyzed the temporal gap between the first and the last symptom dates available for each patient. Our assumption was that symptoms going far back in time could represent the actual onset of psychosis. To verify this, we reviewed the documented "early symptoms" for 41 patients having a diff value of more than one year. Out of these 41 instances, 17 corresponded to a clear onset date, while 15 were a close approximation to the onset date (which was specified in other parts of the texts). An example of the first type is given by: "he has been suffering from psychosis since he was 10 years old when he started experiencing hallucinations". The remaining 9 instances resulted either from erroneous dates written in the text (3), or from a long temporal gap between documents associated to the same patient (6). These results indicate that the proposed annotation schema could be useful for correctly capturing information on early symptom onset, as well as for retaining more general temporal information for timeline reconstruction. To further assess this point, we plan to analyze the information annotated for the 165 patients with a diff value lower than one year.

As regards NLP system development, preliminary results indicate that our gold data are consisent enough to allow for automated system development (Table 4). However, the performance of the TLINK module on the development set (0.58) was lower than that on the training set (0.67), showing that more effort should be put into developing a generalizable system. To address this, we plan to both improve the available rules and explore supervised machine learning methods. Given the complexity of our problem, human-in-the-loop approaches could be explored [15]. Moreover, to support real-world usability, it would be important to focus on explainable methods [16]. Once the NLP system is completed, we will run it on a large patient cohort, to quantify the number of patients for which an early symptom onset is documented in free text. In addition, we are interested in assessing which types of TIMEXes are most frequently associated to onset information.

Our study presents two main limitations. First, given the huge amount of textual information available in CRIS, the proposed corpus selection might not be ideal. To investigate this, we are currently annotating different types of documents, in particular those related to first referrals to SLaM (without focusing on early intervention services). Second, the way in which we modeled the problem could be potentially improved. To simplify the annotation task, we only focused on a limited set of symptom keywords; however, these keywords are not suitable to capture more complex linguistic variants. Moreover, addressing the extraction problem at a mention-level is not necessarily the best option. As future work, we will investigate other ways to model our problem, for example by following a question-answering annotation approach. One drawback of this approach could be the low prevalence of symptom onset descriptions in the texts and the need to review even larger sets of documents.

Conclusions

In this study we described a gold standard for temporal relation extraction in the mental health domain, with a focus on symptom onset and DUP extraction. We presented a method for corpus selection and an annotation schema, with promising IAA results. As a proof of concept, we proposed an early rule-based system for TLINK extraction. In the future, this system could be used to temporally anchor symptoms and treatments extracted from mental health records, thus enabling the calculation of DUP and other relevant concepts.

Acknowledgements

NV and SV are supported by the Swedish Research Council (2015-00359), Marie Skłodowska Curie Actions, Cofund,

Project INCA 600398. Ethical approval: Oxford REC C, reference 08/H0606/71+5. RS and RP are part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. RP has received

support from a Medical Research Council (MRC) Health Data Research UK Fellowship (MR/S003118/1) and a Starter Grant for Clinical Lecturers (SGL015/1020) supported by the Academy of Medical Sciences, The Wellcome Trust, MRC, British Heart Foundation, Arthritis Research UK, the Royal College of Physicians and Diabetes UK.

References

[1] M. Marshall, S. Lewis, A. Lockwood, R. Drake, P. Jones, and T. Croudace, Association between duration of untreated psychosis and outcome in cohorts of first-episode patients: a systematic review, *Arch. Gen. Psychiatry*. **62** (2005) 975–983.

[2] M. Hill, N. Crumlish, M. Clarke, P. Whitty, E. Owens, L. Renwick, S. Browne, E.A. Macklin, A. Kinsella, C. Larkin, J.L. Waddington, and E. O'Callaghan, Prospective relationship of duration of untreated psychosis to psychopathology and functional outcome over 12 years, *Schizophr. Res.* **141** (2012) 215–221.

[3] M. Penttilä, E. Jääskeläinen, N. Hirvonen, M. Isohanni, and J. Miettunen, Duration of untreated psychosis as predictor of long-term outcome in schizophrenia: systematic review and meta-analysis, *The British Journal of Psychiatry*. **205** (2014) 88–94.

[4] Y. Wang, L. Wang, M. Rastegar-Mojarad, S. Moon, F. Shen, N. Afzal, S. Liu, Y. Zeng, S. Mehrabi, S. Sohn, and H. Liu, Clinical information extraction applications: A literature review, *Journal of Biomedical Informatics*. **77** (2018) 34–49.

[5] W. Sun, A. Rumshisky, and O. Uzuner, Evaluating temporal relations in clinical text: 2012 i2b2 Challenge, *J Am Med Inform Assoc*. **20** (2013) 806–813.

[6] W.F. Styler IV, S. Bethard, S. Finan, M. Palmer, S. Pradhan, P.C. de Groen, B. Erickson, T. Miller, C. Lin, G. Savova, and J. Pustejovsky, Temporal annotation in the clinical domain, *Transactions of the Association for Computational Linguistics*. **2** (2014) 143–154.

[7] S. Bethard, L. Derczynski, G. Savova, J. Pustejovsky, and M. Verhagen, SemEval-2015 Task 6: Clinical TempEval, *Proceedings of SemEval 2015*. (2015) 806–814.

[8] S. Bethard, G. Savova, W.-T. Chen, L. Derczynski, J. Pustejovsky, and M. Verhagen, Semeval-2016 task 12: Clinical TempEval, *Proceedings of SemEval 2016*. (2016) 1052–1062.

[9] G. Perera, M. Broadbent, F. Callard, C.-K. Chang, J. Downs, R. Dutta, A. Fernandes, R.D. Hayes, M. Henderson, R. Jackson, A. Jewell, G. Kadra, R. Little, M. Pritchard, H. Shetty, A. Tulloch, and R. Stewart, Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) Case Register: current status and recent enhancement of an Electronic Mental Health Record-derived data resource, *BMJ Open*. **6** (2016) e008721.

[10] R. Jackson, R. Patel, S. Velupillai, G. Gkotsis, D. Hoyle, and R. Stewart, Knowledge discovery for Deep Phenotyping serious mental illness from Electronic Mental Health records, *F1000Res*. **7** (2018).

[11] A.X. Chang, and C.D. Manning, SUTIME: A Library for Recognizing and Normalizing Time Expressions, *Proceedings of LREC 2012*. (2012).

[12] N. Viani, L. Yin, J. Kam, A. Alawi, A. Bittar, R. Dutta, R. Patel, R. Stewart, and S. Velupillai, Time Expressions in Mental Health Records for Symptom Onset Extraction, *Proceedings of LOUHI 2018*. (2018) 183-192.

[13] B.E. Chapman, S. Lee, H.P. Kang, and W.W. Chapman, Document-level classification of CT pulmonary angiography reports based on an extension of the ConText algorithm, *Journal of Biomedical Informatics*. **44** (2011) 728–737.

[14] P. Raghavan, E. Fosler-Lussier, N. Elhadad, and A.M. Lai, Cross-narrative Temporal Ordering of Medical Events, *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics.* **1** (2014) 998–1008.

[15] A. Holzinger, Interactive machine learning for health informatics: when do we need the human-in-the-loop?, *Brain Inform*. **3** (2016) 119–131.

[16] A. Holzinger, From Machine Learning to Explainable AI, *2018 World Symposium on Digital Intelligence for Systems and Machines (DISA)*. (2018) 55–66.

Address for correspondence

Natalia Viani

Email: natalia.viani@kcl.ac.uk.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190256

Extracting Symptom Names and Disease-Symptom Relationships from Web Texts Using a Multi-Column Convolutional Neural Network

Shoya Wada[a], Ryu Iida[b], Kentaro Torisawa[b], Toshihiro Takeda[a] Shiro Manabe[a], and Yasushi Matsumura[a]

[a] Department of Medical Informatics, Osaka University Graduate School of Medicine, Suita, Osaka, Japan.
[b] National Institute of Information and Communications Technology, Seika, Kyoto, Japan.

Abstract

We propose a method to create large-scale Japanese medical dictionaries that include symptom names and information about the relationship between a disease and its symptoms using a large web archive that includes large amounts of text written by non-medical experts. Our goal is to develop a diagnosis support system that makes a diagnosis according to the natural language (NL) inputs provided by patients. To achieve this, two medical dictionaries need to be constructed: one that includes a wide variety of symptom names expressed in NL and another that includes information about the relationship between a disease and its symptoms. Dictionaries will then be used to predict the patient's disease via two developed methods that extract symptom names and disease–symptom relationships. Both methods retrieve sentences using WISDOM X and then apply neural classifiers to them. Our experimental results show that our methods achieved 93.8% and 88.3% in the F1-score, respectively.

Keywords:

Natural language processing, Pattern Recognition, Automated, Machine learning

Introduction

Recently, there have been rising expectations for the introduction of an automatic diagnosis support system (DSS) as advanced artificial intelligence techniques, including deep learning, have developed and electronic medical reports have become widely used in the medical field. To develop a DSS, first medical dictionaries that include a wide variety of symptom expressions and disease–symptom relationship information need to be created. To the best of the authors' knowledge, no such large-scale Japanese dictionaries exist.

This study aims to develop a DSS that can formulate diagnoses according to the NL inputs given by patients, such as short texts describing their health problems or symptoms. In this work, a method is proposed to create large-scale medical dictionaries that include NL expressions referring to symptoms and disease–symptom relationships. Hereafter, these NL expressions will be referred to as "symptom names."

To collect symptom expressions and relationships (shown in Table 1), various approaches using rule-based systems, NL techniques, and machine learning (ML) have been proposed. However, their target documents are relatively formal texts, such as medical textbooks, scientific papers, and electronic medical records. Unlike the texts written by medical experts, inputs given by patients contain a wide range of NL expressions about symptoms. For example, patients may state "tiredness" or "lack of energy" when referring to fatigue. In English, there are some available resources of patients' informal expressions such as the Patient-Free Term List [1] and Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE), a library of items for patient self-reporting of symptomatic adverse events (AEs) developed by the U.S. National Cancer Institute. PRO-CTCAE includes 78 symptomatic AE terms and is translated into Japanese [2], however, to obtain expressions more widely used in Japan, a Japanese medical dictionary that includes a comprehensive variety of NL symptom expressions needs to be created. In addition, to predict patients' diseases, another dictionary that includes information about the relationship between a disease and its symptoms needs to be generated. The disease–symptom relationships should cover a wide range of symptom names used by patients so that the system can predict diseases from the inputs given by the patients. This work aims to extract a broad range of symptom names and the causal relationship between diseases and their symptoms using a large web archive, which includes a large amount of text written by non-medical experts to cover symptom names as entered by the user and correlate the relationship to known diseases. To this end, the authors developed two neural methods that extract symptom names and information about disease–symptom relationships. Both methods initially retrieve sentences that include symptom candidate names using the open-domain question answering system WISDOM X [3], and then they apply two neural methods: one for extracting symptom names and one for extracting disease–symptom relationship. The experimental results show that the proposed symptom name extraction method achieved a 93.8% F1-score, and the disease–symptom relationship extraction method achieved an 88.3% F1-score. Moreover, the proposed methods can acquire symptom names that are not included in the training data and the existing Japanese medical term dictionaries; this finding exemplifies the possibility of developing this system further.

Definition of symptoms

In this work, all changes caused by the diseases of living organisms are regarded as symptoms and are referred to as "symptom names." They contain expressions of physical features observed by the patient and are hypothesized by the patient to be a sign of a disease. Symptoms are defined as such because the patients are expected to express general symptoms, including the expressed symptoms of subjective experiences (e.g., fatigue and paresthesia), the objectively observed findings (e.g., erythema and systolic murmur), and the disease names of the complications, to a DSS without distinction between these categories.

Methods

Materials

WISDOM X is a Japanese open-domain question answering system that uses four billion web pages as its information source (https://wisdom-nict.jp) [3]. It answers a wide range of questions, such as "What is AI used for?" and "What happens if global warming worsens?" This study used WISDOM X because it enabled the efficient extraction of sentences that include a symptom candidate. For example, when "What is caused by AMI?" was typed into WISDOM X, the program answered that "AMI causes life-threatening arrhythmia," in which "life-threatening arrhythmia" is regarded as a symptom candidate name (Figure 1).

A list of questions was created to be entered into WISDOM X; each of these questions consisted of a certain disease name and a question template (e.g., "what is caused by [DISEASE]"). First, 25,367 disease names were collected from the disease name master list of Japanese International Statistical Classification of Diseases and Related Health Problems Tenth Revision [4] and the doctor national examination question criteria [5]. This list was then expanded to 25,916 names using a contextually similar word database [6].

Figure 1 – An example of using WISDOM X.

Further, the authors manually generated 67 question templates (e.g., "What happens with [DISEASE]" and "What does [DISEASE] cause?") and obtained 1,736,372 questions by exhaustively combining disease names and question templates. All questions were entered WISDOM X and 3,894,833 sentences that contained both a disease name and a symptom candidate name were found. Out of the 25,916 disease names, 10,467 (40.4%) had at least one or more answers.

The symptom candidate names extracted as above contain expressions that are irrelevant to proper symptoms, and they were removed as follows. Specific noun classes were identified (10 classes) that contain expressions mostly relevant to symptoms in a noun semantic class dictionary created by applying the word semantic classification method [7] to a large-scale web document (containing approximately 600 million documents). For example, the classes mainly consist of suffixes related to illness, such as "-syndrome", "-ache" and "-emia". Then, expressions that include an entry in the selected 10 class dictionary were extracted from the above 3,894,833 sentences. As a result, 127,352 sentences were obtained. The number of unique disease names and unique symptom candidate names were 7,423 and 32,550, respectively.

Data construction for training and evaluation

In this work, data construction for training and evaluation was performed by the following two human judgment tasks. An example of the human judgment tasks is shown in Table 1.

Symptom Name Extraction (SNE)

Human annotators judged whether a given symptom candidate name is proper in a sentence or not. Even if a given expression is part of a longer symptom name, it was identified as proper (e.g., "loss" is regarded as SNE in the fourth example in Table 1).

Disease–Symptom Recognition (DSR)

Human annotators judged whether a given proper symptom name is caused by a given disease sentence. The disease–symptom relationship was judged based on the content of the sentence. In other words, the annotators did not care if the relationship is proper in the real world.

Table 1 – Annotation Examples
Examples were translated into English for readability. P and N denote proper and non-proper expressions (or relationships), respectively.

	EXAMPLES	SNE	DSR
1	This ($_{dis}$hay fever) has given me ($_{sym}$a runny nose).	P	P
2	($_{sym}$Method) for treatment of ($_{dis}$Aortic Regurgitation).	N	-
3	It causes such ($_{sym}$symptoms) as intracranial calcification, hepatosplenomegaly and liver disorder, jaundice, microcephaly, psychomotor development delay, convulsion, ($_{dis}$retinitis), and the like.	P	N
4	($_{dis}$Acute myocardial infarction) causes violent chest pain, cold sweat, and ($_{sym}$loss) of consciousness, ...	P	P

Seven annotators who are non-medical experts worked on these two tasks, and 50,000 sentences were sampled from the above 127,352 sentences used in the final dataset. Each sentence was triple-annotated for the SNE and DSR tasks. Then, the final label for each sentence was determined by a majority vote among the three annotators. As a result, 26,363 (52.7%) and 15,547 (31.1%) terms were labeled as proper symptom names and as disease–symptom relationships, respectively. The inter-rater agreement of Fleiss' kappa of SNE and DSR were 0.823 and 0.788, respectively, denoting substantial agreement.

The 50,000 samples were divided into training, validation, development and test sets, and they were used in the experiment's settings described later. A set was also created from the test (All) by excluding symptom candidate names contained within the training set, the validation set, and the existing Japanese medical term dictionaries [8, 9] to evaluate the system's ability to deal with newly discovered expressions (labelled as "Unknown" in Table 2).

Table 2 – Statistics of the datasets

	Total	Number of positive instances	
		SNE	DSR
Training	35,000	18,489	10,877
Validation	5,000	2,581	1,532
Development	5,000	2,624	1,608
Test (All)	5,000	2,669	1,530
Test (Unknown)	1,056	586	367

Neural network-based models for SNE and DSR

In the SNE and DSR methods, a multi-column convolutional neural network (MCNN) [10], which is a variant of a convolutional neural network [11], with several independent columns, was used. Each column has its own convolutional and pooling layers. The outputs of all the columns are combined in the last layer to provide a final prediction. MCNNs have been widely used in NL processing tasks, such as question answering [12]. MCNNs were also used in tasks such as causality recognition [13] and zero anaphora resolution [14], in which, target expressions and their surrounding contexts should be considered. In the study's tasks, such contextual clues are essential for DSR, it is expected that MCNNs will accurately recognize these relationships.

The architecture of the MCNN used and its inputs are shown in Figure 2. As an input to the MCNN, an input sentence was first divided into five consecutive word sequences: a symptom candidate name (SYM), a given disease name (DIS), the word sequence before the SYM (BFR), the word sequence between the SYM and the DIS (BTWN), and the word sequence after the DIS (AFTR). If the DIS precedes the SYM, the BFR is the word sequence before the DIS while the AFTR is the word sequence after the SYM. These five-word sequences are independently given to the corresponding columns (Column 1 to Column 5 in Figure 2). Additionally, the question used for retrieving the sentence (Q) was also used as the input for the other column (Column 6) to use the relationship between the Q and the corresponding answer. Hereafter, we refer to the combination of the DIS and the SYM as BASE and the combination of the BFR, BTWN, and AFTR as CONT[ext].

The divided word sequences lack the relative linear-position feature indicating whether the SYM precedes the DIS or not. To integrate the linear-position into the MCNN, one of the three symbols representing the relative position (i.e., L[eft] and R[ight] and S[ymptom]) was assigned to each word in the sentence. Each symbol was converted into randomly initialized 10-dimensional vectors and was concatenated with the original word embedding vector to use it as a column's input. Hereafter, we refer to the use of the position vectors as LOC[ation].

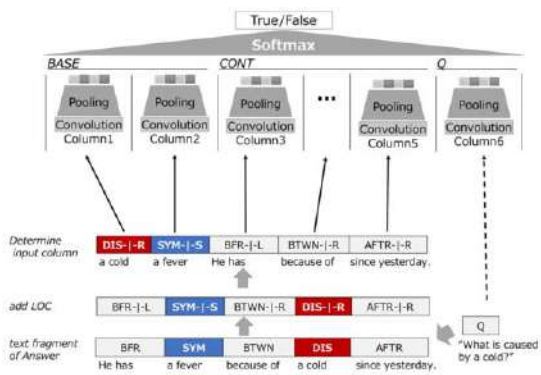


Figure2 – The MCNN architecture

Experimental settings

The performance of the proposed methods was evaluated using the datasets presented above. For the word embeddings used in the MCNN, 500-dimensional word embedding vectors were pre-trained for 1,658,487 words using the skip-gram model [15] on a set of all the sentences extracted from Wikipedia articles (jawiki-20150118; 35,975,219 sentences). Words that did not occur in the embedding vocabulary were treated as unknown words. Each of them was assigned a distinct random vector. The sentences in the datasets were tokenized by a Japanese morphological analyzer, MeCab [16].

For the convolutional layer of the MCNN, a combination of {100, 300} for the filter size and {(3, 4, 5), (4, 5, 6)} for the filter widths were tried. In the pooling layer, max-pooling is performed. The final layer has vectors coming from multiple feature maps in multiple columns. They are concatenated again and constitute a high-dimensional feature vector. The final layer applies a softmax function to produce the class probabilities of the SNE or the DSR labels: true and false. A mini-batch stochastic gradient descent (SGD) with the Adadelta update rule was used. To avoid overfitting, early stopping was administered using the validation dataset and the dropout; a dropout rate of 0.5 was applied to the final layer, and SGD with mini-batches of 100 and a learning decay rate of 0.95 were used. In training, the best epoch of the MCNN was selected by the F1-score on the validation set, and the optimal hyperparameters were selected using the development set.

To compare with our proposed method, we introduced two baseline methods. One extracts all symptom candidate names as proper ones and regards given pairs of a symptom and a disease name as proper disease–symptom relationships. This method was called the AllPositive method. The other is a traditional ML method of document classification based on support vector machines (SVM) with Latent Semantic Indexing features converted from a Bag of Words representation of context. This was called the SVM method. The McNemar test (α=0.05) was used to examine whether the classification performance of our method was significantly better than that of each of the baseline methods.

Results

The experimental results are shown in Table 3. On the evaluation of SNE, the results showed that the F1-score of our proposed method (Proposed (i.e. BASE+CONT+Q+LOC) in Table 3) was 93.8%, which was significantly better than those of the AllPositive and SVM (69.6% and 77.3%). We also conducted ablation test, that is, we evaluated performances of the Proposed methods that did not use one of Q, CONT, and LOC. However, as shown in Table 3, each of them didn't significantly contribute to performance improvement in both cases of evaluating test (All) and test (Unknown).

Of the DSR models, the Proposed method significantly outperformed the two baselines (AllPositive and SVM). In this case, the F1-score of the Proposed method was significantly better than the Proposed without CONT and that without LOC. That is, only the Q didn't contribute to performance improvement ($p = 0.071$).

The results using the test (Unknown) showed that the Proposed method significantly outperformed the AllPositive and SVM baselines in the cases of extracting unknown symptom names and recognizing unknown disease-symptom relationship, while F1-scores are relatively worse than those of the test (All).

Table 3 – Performance of Symptom Name Extraction (SNE) and Disease-Symptom Recognition (DSR)

		Development dataset				Test dataset							
						All				Unknown			
		Recall	Precision	F1-score	McNemar p-value	Recall	Precision	F1-score	McNemar p-value	Recall	Precision	F1-score	McNemar p-value
SNE	AllPositive	1.000	0.546	0.688	<0.001*	1.000	0.556	0.696	<0.001*	1.000	0.576	0.714	<0.001*
	SVM	0.761	0.782	0.771	<0.001*	0.761	0.785	0.773	<0.001*	0.765	0.821	0.792	<0.001*
	Proposed	0.955	0.935	0.945	-	0.942	0.934	0.938	-	0.867	0.922	0.894	-
	Proposed w/o Q	**0.956**	0.932	0.944	0.767	**0.945**	0.933	**0.939**	0.697	**0.877**	0.921	**0.899**	0.450
	Proposed w/o CONT	0.928	0.944	0.936	0.009*	0.921	0.947	0.934	0.306	0.843	0.925	0.882	0.177
	Proposed w/o LOC	0.939	**0.953**	**0.946**	0.323	0.921	**0.954**	0.937	0.930	0.838	**0.937**	0.885	0.265
DSR	AllPositive	1.000	0.348	0.487	<0.001*	1.000	0.332	0.469	<0.001*	1.000	0.370	0.516	<0.001*
	SVM	0.496	0.735	0.592	<0.001*	0.500	0.708	0.586	<0.001*	0.583	0.716	0.643	<0.001*
	Proposed	**0.901**	0.872	**0.886**	-	**0.904**	0.864	**0.883**	-	**0.880**	**0.868**	**0.874**	-
	Proposed w/o Q	0.879	**0.887**	0.883	0.869	0.871	**0.875**	0.873	0.071	0.845	**0.868**	0.856	0.108
	Proposed w/o CONT	0.825	0.800	0.812	<0.001*	0.834	0.791	0.812	<0.001*	0.793	0.764	0.778	<0.001*
	Proposed w/o LOC	0.875	0.843	0.859	<0.001*	0.871	0.825	0.847	<0.001*	0.823	0.823	0.823	<0.001*

*Note: w/o stands for without; the Proposed method consists of BASE, CONT, Q and LOC; and * indicates $p < 0.05$ compared with the Proposed method*

Discussion

In the SNE task, the performance of the Proposed method was not statistically significant compared with the model without Q, CONT, or LOC. It means that SNE should be performed using the clues that are extracted from the words of symptom candidate names.

On the other hand, in the DSR task, performance of the Proposed method was significantly improved by the CONT and LOC. This indicates that the accuracy is increased by utilizing the contexts surrounding the DIS and the SYM and by considering relative linear-position of DIS and SYM. However, the performance was not increased by the Q. The relationship between a question and the corresponding answer was used in hopes of improving the model's performance. It may be because the questions are indirectly related to DIS and SYM, and thus has no significant positive effect for performance improvement, while they are important to extract sentences that include both DIS and SYM.

It was also expected that the following two issues eventually led to high F1-score in DSR. The first one is related to the annotation guidelines of DSR. In the annotations of DSR, annotators read the context where the DIS and the SYM both appear and then judge their disease-symptom relationship. That is, if disease-symptom relationship is easily interpretable by its context, annotators tend to regard it as proper relationship. We expect that such contextual clues are relatively easily captured by MCNNs through the word embedding vectors and convolution-pooling operations. The second issue is related to the text format in which the DIS and the SYM are described. In our dataset, the DIS and the SYM co-occurred in the titles of kinds of medical case reports and thus disease and symptom names appeared involving with typical templates, such as "a case of DIS exhibiting SYM," "DIS accompanying SYM," and "DIS with SYM," making the recognition easy.

Finally, the ability of the prediction models to extract new symptom names and causal relationships between diseases and their symptoms was examined. Although the F1-score slightly decreased compared to when the model was used with the test (All) dataset, the prediction models maintained high accuracy in SNE and DSR. To explore the final goal, classifiers were developed that may permit the automatic discovery of new terms without added human effort. The results of the test (Unknown) dataset exemplify the ability to achieve this object on a web archive.

Limitations

There were some instances where the prediction models misjudged the information and provided false positives and false negatives; these events were sampled from DSR cases. False positives/negatives were mainly caused by annotation errors. They were generally caused by obfuscated sentences and ones that were translated using an automated translator. These sentences tended to be unnatural as Japanese, so it made human judgment difficult. Another issue was medical reliability. To evaluate the medical reliability with respect to DSR, we sampled 100 pairs of the disease and the symptom from the pairs that were automatically recognized by each method (AllPositive, SVM, and Proposed in Table 3) and then a medical expert (a physician) confirmed them whether each pair was proper or not. Note that the samplings and judgment by the expert were performed for both test (All) and test (Unknown) cases. The percentage of proper relationship judged by the expert for each method is shown in Table 4. The result was like the precisions of DSR in Table 3 and even in this case, the performance of the Proposed method drastically better than those of the remaining methods. The misjudged cases were classified into improper symptom names, apposition, complements, inclusion relation, reversed relation of causality, and medical errors. Compared to the SVM method, our methods seemed good at reading contexts. Only a few medical errors were located and discussed about the medical reliability. For example, "($_{dis}$unstable angina) complicated with ($_{sym}$severe carotid artery lesion)" and "($_{dis}$steroid-resistant nephrotic syndrome) with ($_{sym}$relapse after transplantation)", which were controversial among medical professionals. In this research, the annotators were instructed not to care if the relationship was proper in the real world. To avoid such controversial cases at the annotation phase, it might be useful to present multiple examples at once to easily find controversial cases.

Table 4 – Evaluation of DSR predictions

	Accuracy of each 100 samples	
	Test (All)	Test (Unknown)
AllPositive	38%	52%
SVM	75%	81%
Proposed	93%	86%

The symptom names that were extracted by our method also have noises due to the parallel structure of noun phrases (e.g., this system recognizes "fever•cough•snivel" as a new word)

and word segmentation errors by Japanese morphological analyzer (e.g. “A-22” and “stomachache” was concatenated into a single word as “A-22stomachache”). For example, the proposed symptom extraction method outputted 488 symptom names on the evaluation using the test (Unknown) set, but the 28 symptom names in the output included noises due to the parallel structure and the 111 symptom names also included unnecessary characters by word segmentation errors. To exclude the effect of such noises, we manually corrected and recounted; the number of unique symptom names was 524, of which 378 cases (72.1%) were not included in the training data and the existing medical term dictionaries. Certainly, the 378 newly discovered symptom names contain variant ones that were in the training data, but we think that extracting such variants is also important for a DSS because a wide variety of symptom names are a key to bridge between actual symptom expression given by patients and the disease that causes the symptom as explained in the Introduction.

In this study, a portion of phrases requiring modified symptom expressions (e.g., "a condition in which I felt languid", "loss of one’s eyesight" [No.4 in Table 1]) were collected. It is necessary to collect modified expressions related to nouns (phrases) from this study to use with the NL processing technique to complete such expressions.

One of the limitations of the DSR data collected by the proposed method is the insufficient medical reliability. This is a result of the information being derived from web documents, which are less reliable than medical books. Additionally, some of the content acquired from the titles of case reports existed in the datasets, so it is possible that rare events were included as a symptom name associated with a disease, such as “A Case Report of [DISEASE] With Unusual Presentation and Clinical Course”. Therefore, it is necessary to be cautious about symptom names with low relevance to the target disease. That is, the difficulty of merging the knowledge of experts and non-experts must be rectified. To solve this problem, it is necessary to develop new technologies and methodologies that enable only reliable data to be stored in the database, such as extracting symptom names from reliable medical books and identifying the relationships between a disease and its symptoms. In some cases, manual verification may be required to judge the synonymy and the semantic similarity of the candidates of the disease names and the symptom names collected in this study using a deep learning technique.

Conclusions

In this paper, the authors developed a method to extract symptom expressions and the causal relationships between diseases and their symptoms through deep learning from a large web archive, which included a large amount of text written by non-medical experts. The acquired symptom names contain terms that do not exist in medical term dictionaries; thus, various new expressions can be obtained using the proposed methods. However, since expressions requiring corrections were present, along with the problematic medical reliability of web texts, the new content needs to be reevaluated before it is used. For further improvement, the experiment could be more efficiently performed by focusing on the probability score of the output as it might help to combine the knowledge of experts and non-experts.

References

[1] MedDRA, Patient-Friendly Term List, https://www.meddra.org/patient-friendly-term-list.
[2] E. Basch, B. B. Reeve, S. A. Mitchell, S. B. Clauser, L. M. Minasian, A. C. Dueck, T. R. Mendoza, et al, Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE), *J Natl Cancer Inst* 106 (2014), dju244.
[3] J. Mizuno, J. M. Tanaka, K Ohtake, J. H. Oh, J. Kloetzer, C. Hashimoto, and K. Torisawa, WISDOM X, DISAANA and D-SUMM: Large-scale NLP Systems for Analyzing Textual Big Data, *COLING* (2016), 263–267.
[4] MEDIS, standard disease name master for ICD10, https://www2.medis.or.jp/stdcd/byomei/
[5] Ministry of Health Labour and Welfare, About medical doctor's edition national examination question standard, https://www.mhlw.go.jp/stf/shingi2/0000128981.html
[6] ALAGIN, a contextually similar word database Version 1.1.2, https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html
[7] J. Kazama, and K. Torisawa, Inducing gazetteers for named entity recognition by large-scale clustering of dependency relations, *Proceeding of ACL-08* (2008), 407–415.
[8] K. Sagara, Comejisyo V5-1, https://ja.osdn.net/projects/comedic/
[9] MedNLP, MANBYO_web_v5, http://mednlp.jp/DIC/
[10] D. Cireşan, U. Meier, and J. Schmidhuber, Multi-column Deep Neural Networks for Image Classification, *In Proceedings of Computer Vision and Pattern Recognition* (2012), 3642-3649.
[11] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, *In Proceedings of the IEEE* (1998), 2278–2324.
[12] L. Dong, F. Wei, M. Zhou, and K. Xu, Question Answering over Freebase with Multi-Column Convolutional Neural Networks, *In Proceedings of ACL* (2015), 260–269.
[13] C. Kruengkrai, K. Torisawa, C. Hashimoto, J. Kloetzer, and J. H. M. Tanaka, Improving Event Causality Recognition with Multiple Background Knowledge Sources Using Multi-Column Convolutional Neural Networks, *In Proceedings of AAAI* (2017), 3466–3473.
[14] R. Iida, K. Torisawa, J. H. Oh, C. Kruengkrai, and J. Kloetzer, Intra-Sentential Subject Zero Anaphora Resolution using Multi-Column Convolutional Neural Network, *In Proceedings of EMNLP* (2016), 1244–1254.
[15] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, Distributed representations of words and phrases and their compositionality, *In Proceedings of Advances in NIPS* (2013), 3111–3119.
[16] T. Kudo, MeCab 0.996, http://taku910.github.io/mecab/

Address for correspondence

Shoya Wada, MD, wada@hp-info.med.osaka-u.ac.jp

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190257

Map-Assisted Generation of Procedure and Intervention Encoding (Magpie): An Innovative Approach for ICD-10-PCS Coding

Kin Wah Fung[a], MD, MS, MA; Julia Xu[a], MD, PhD; Filip Ameye[b] MD, FRCS(Eng), FACS; Arturo Romero Gutiérrez[c], MD; Ariel Busquets[d], MD

[a]*National Library of Medicine, Bethesda, MD,*
[b]*National Institute for Health and Disability Insurance, Belgium,*
[c]*Ministry of Health, Social Services and Equality, Spain,*
[d]*Hospital Privado Universitario de Córdoba, Argentina*

Abstract

ICD-10-PCS coding is challenging because of the large number of codes, non-intuitive terms and paucity of the ICD-10-PCS index. We previously repurposed the richer ICD-9-CM procedure index for ICD-10-PCS coding. We have developed the MAGPIE tool based on the repurposed ICD-9-CM index with other lexical and mapping resources. MAGPIE helps the user to identify SNOMED CT and ICD-10-PCS codes for medical procedures. MAGPIE uses three innovative search approaches: cascading search (SNOMED CT to ICD-9-CM to ICD-10-PCS), hybrid lexical and map-assisted matching, and semantic filtering of ICD-10-PCS codes. Our evaluation showed that MAGPIE found the correct SNOMED CT code and ICD-10-PCS table in 70% and 85% of cases respectively, without any user intervention. MAGPIE is available online from the NLM website: magpie.nlm.nih.gov.

Keywords:

Controlled vocabulary, Operative Surgical Procedures, Clinical Coding

Introduction

In the U.S., ICD-9-CM procedure codes (also known as ICD-9-CM Volume 3, or ICD9V3 in short) had been used for over 30 years to encode hospital-based medical procedures and interventions for administrative and reimbursement purposes. In 2015, together with the replacement of ICD-9-CM Volumes 1 and 2 diagnosis codes by ICD-10-CM, ICD9V3 was replaced by ICD-10-PCS. Despite the similarity in name, ICD-10-PCS is not an evolutionary descendant of ICD9V3, but a brand-new procedure coding system [1-3]. While the tabular list of codes in ICD9V3 is a tree-shaped taxonomy similar to ICD-9-CM diagnosis codes or ICD-10-CM, ICD-10-PCS is built on a multi-axial structure. ICD-10-PCS codes are composed of seven characters. Each character is an axis of the classification that specifies some information about the procedure performed. Within a defined code range, an axis specifies the same type of information in that axis of classification. Within the Medical and Surgical Section (the first character is 0), which contains 87% if all ICD-10-PCS codes, the details of seven axes are shown in Figure 1.

Compared to ICD9V3, coding in ICD-10-PCS is more challenging because of three reasons. Firstly, there are 20 times more codes in ICD-10-PCS (78,705 codes in the 2018 version) compared to ICD9V3 (3,882 codes in the last updated version in 2013). Secondly, the terms used in ICD-10-PCS are not clinically intuitive. For example, in the clinical discourse, the three operations, extraction, removal and extirpation, are very close in meaning and can sometimes be used interchangeably. In ICD-10-PCS, those three operations are called 'root operations' and have very specific definitions within the coding system:

- Extraction - Pulling or stripping out or off all or a portion of a body part by the use of force
- Removal - Taking out or off a device from a body part
- Extirpation – Taking or cutting out solid matter from a body part

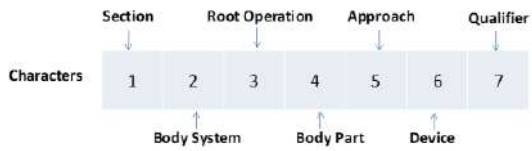


Figure 1— seven axes in ICD-10-PCS codes

This can be confusing to users of ICD-10-PCS, especially clinicians who may not be familiar with the definitions of root operation. For example, cataract removal is coded as extraction and not removal. Non-excisional wound debridment is coded as extraction, not excision or removal. Extraction of embolus (embolectomy) is coded as extirpation. Another area of potential confusion is the difference between resection and excision, which are mostly used interchangeably in clinical discourse. In ICD-10-PCS, excision is defined as "the cutting out or off, without replacement, a **portion** of a body part"; while resection is "the cutting out or off, without replacement, **all** of a body part". As a result, the removal of the appendix is coded as resection while partial nephrectomy as excision.

The third reason why ICD-10-PCS coding is more difficult is that the ICD-10-PCS index is not as rich as ICD9V3 index. The ICD9V3 index contains a lot of detailed terms that are commonly used in clinical records, including abbreviations and eponyms (e.g., Polya gastrectomy). As an example, Figure 2 and Figure 3 show the comparison of the entry for 'Gastrectomy' in the two indexes.

To overcome those challenges, the SNOMED CT to ICD-10-PCS Map Project Group is proposing an innovative solution called MAGPIE (Map-Assisted Generation of Procedure and Intervention Encoding). Our Project Group was formed under SNOMED International in 2015 and has been studying ways to map between SNOMED CT and ICD-10-PCS [4]. We have explored various ways of automatic mapping including lexical matching of the ICD-10-PCS index, ontological alignment between the SNOMED CT attributes and ICD-10-PCS axes, and the use of post-coordination to achieve logical equivalence

[5-9]. The latest study we carried out was repurposing the ICD9V3 index for ICD-10-PCS coding [10].

Gastrectomy (partial) (sleeve) (subtotal) NEC 43.89
with
anastomosis (to) NEC 43.89
duodenum 43.6
esophagus 43.5
gastrogastric 43.89
jejunum 43.7
esophagogastrostomy 43.5
gastroduodenostomy (bypass) 43.6
gastroenterostomy (bypass) 43.7
gastrogastrostomy (bypass) 43.89
gastrojejunostomy (bypass) 43.7
jejunal transposition 43.81
complete NEC 43.99
with intestinal interposition 43.91
distal 43.6
Hofmeister 43.7
laparoscopic, vertical (sleeve) 43.82
Polya 43.7
proximal 43.5
radical NEC 43.99
with intestinal interposition 43.91
sleeve
laparoscopic 43.82
total NEC 43.99
with intestinal interposition 43.91

Figure 2 – ICD9V3 index

Gastrectomy
Partial *see* Excision, Stomach 0DB6
Total *see* Resection, Stomach 0DT6
Vertical (sleeve) *see* Excision, Stomach 0DB6

Figure 3 – ICD-10-PCS index

We harvested the rich ICD9V3 index terms and matched them to SNOMED CT through the UMLS. We mapped the ICD9V3 codes to ICD-10-PCS using the General Equivalence Map (GEM) published by the Centers for Medicare and Medicaid (CMS) (Figure 4) [11]. We showed that the re-purposed ICD9V3 index out-performed the native ICD-10-PCS index in the retrieval of ICD-10-PCS codes based on common surgical procedure names.

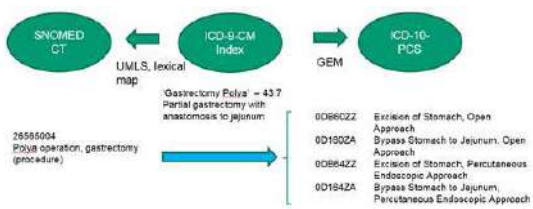


Figure 4– Repurposing the ICD9V3 index and mapping to SNOMED CT and ICD-10-PCS

Based on the repurposed ICD9V3 index and other lexical and mapping resources, we have developed the MAGPIE coding algorithm and tool that allow the user to search for SNOMED CT and ICD-10-PCS codes for a medical procedure or intervention.

Methods

Search Strategies

MAGPIE uses three innovative search strategies to help the user hone in on ICD-10-PCS codes in an interactive manner (Figure 5).

1. Cascading search

Starting from the search term that the user types in, MAGPIE looks sequentially for matches in SNOMED CT, ICD9V3, ICD-10-PCS tables, and ICD-10-PCS codes. The rationale for this approach is that SNOMED CT terms are closest to clinical parlance and meaning, so it is often possible to find a SNOMED CT term exactly matching the input search term. Identifying the correct SNOMED CT concept will help navigate the subsequent search for ICD9V3 codes. Going through ICD9V3 is necessary to take advantage of the repurposed ICD9V3 index and the GEM map. In the next step, MAGPIE will suggest candidate ICD-10-PCS tables that the user can pick to display for individual code selection. After the ICD-10-PCS code(s) are selected, MAGPIE will prompt the user for possible refinement of the SNOMED CT code in cases where the ICD-10-PCS codes selected could lead to a more specific SNOMED CT concept than the one chosen. At each step, MAGPIE will suggest

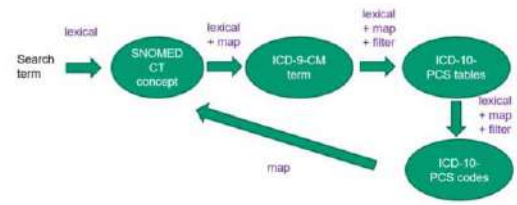


Figure 5 – Overall schema of MAGPIE search sequence and search methods

default codes, if available, based on lexical and map-assisted matching (see below), but the user can change the default if they see better matches. In this way, the user can guide MAGPIE interactively to arrive at the correct ICD-10-PCS codes.

2. Hybrid lexical and map-assisted matching

The first step of searching for a SNOMED CT code is based on lexical matching alone. (Figure 5) All subsequent steps use a combination of lexical and map-assisted matching. To support **lexical matching**, we have built synonym and entry term tables for SNOMED CT, ICD9V3 and ICD-10-PCS codes based on the UMLS, repurposed ICD9V3 index and ICD-10-PCS index. We use the open-source Apache Lucene information retrieval library functions as the search engine for lexical matching [12]. For each sequential search, we expand the search terms by including the synonyms of the selected code in the previous step. For example, the synonyms of the selected SNOMED CT concept are included in subsequent lexical searches for ICD9V3 and ICD-10-PCS codes. For **map-assisted matching**, we have built pairwise mapping tables between the codes from SNOMED CT, ICD9V3, and ICD-10-PCS based on the UMLS, repurposed ICD9V3 index and GEM. Based on the selected code at an earlier step, MAGPIE looks up the code-mapping tables to find target codes for the next step. For example, the selected SNOMED CT code will help to find ICD9V3 codes through the SNOMED CT to the ICD9V3 mapping table. Lexical matching almost always returns some codes, but the accuracy is variable. Map-assisted matching does not always return target codes but is generally more accurate. If both lexical and map-assisted matching return some codes, codes found by map-assisted matching will take precedence over lexical matching.

3. Semantic filtering

In the GEM map, one ICD9V3 code often maps to many ICD-10-PCS codes because of the different granularity between the two systems. For example, *65.29 Other local excision or destruction of ovary* maps to 21 ICD-10-PCS codes because of the combinations of the various options in laterality (left, right, bilateral) and approach (open, endoscopic, etc.). MAGPIE uses semantic filtering to narrow down the possible choices in the

suggestion of ICD-10-PCS tables and codes. (Figure 5) Semantic filtering is based on some keywords in the user's search string. For example, if the user types in 'left laparoscopic oophorectomy', MAGPIE will recognize the keywords 'left' and 'laparoscopic', and pre-select the ICD-10-PCS code *0UT14ZZ Resection of Left Ovary, Percutaneous Endoscopic Approach* while filtering out other choices pertaining to right ovary, bilateral ovaries, and open approach. To support semantic filtering, we have built keyword tables for body part and approach based on the ICD-10-PCS definition tables.

Evaluation

To evaluate the performance of MAGPIE, we use a list of the most commonly performed surgical procedures from a large health care institution that we have obtained in another project [13]. The procedure names have been manually mapped to ICD-10-PCS codes previously, and the maps are used as the reference standard to assess MAGPIE [10]. The procedure names are entered individually into MAGPIE and we evaluate the accuracy of the default SNOMED CT and ICD9V3 codes suggested by MAGPIE. We also evaluate the accuracy and completeness of the suggested ICD-10-PCS tables and ICD-10-PCS codes. If the correct ICD-10-PCS tables or codes are not found by the default selections, we would make necessary adjustments to the SNOMED CT and ICD9V3 code selections, just like a coder would normally do, to get to the correct codes.

Results

A testing version of MAGPIE was made available through the internet during the period of this study. Since the completion of the study, a stable version is now available from the NLM website: magpie.nlm.nih.gov.

MAGPIE users need to accept the UMLS user agreement since SNOMED CT is copyright protected. The users start searching by entering a procedure name, such as 'cesarean', and the autocomplete feature will show the matching terms based on our list of synonyms and entry terms. The users can either pick from the list or type in something not on the list. Upon submitting the search term, MAGPIE will suggest as default the SNOMED CT concept *Cesarean section (11466000)* and ICD9V3 code *Other cesarean section of unspecified type (74.99)* based on lexical and map-assisted matching (Figure 6).

In this case, both through lexical matching using the ICD-10-PCS index, and GEM mapping based on the selected ICD9V3 code, MAGPIE suggests the ICD-10-PCS table *10D Obstetrics | Pregnancy | Extraction*. The user picks the table (only one in this case) for further exploration. On opening the table, the codes found by the GEM map are pre-selected. (Figure 7) MAGPIE highlights the axes where there are multiple values in the GEM suggested codes (the Qualifier axis in this example) and prompts the user to narrow down the choices.

On picking the value 'Classical' for the Qualifier axis, the ICD-10-PCS code Extraction of Products of Conception, Classical, Open Approach (10D00Z0) is recorded by MAGPIE. MAGPIE further checks for potential refinement of the SNOMED CT concept based on the final ICD-10-PCS code. Based on the SNOMED CT to ICD-10-PCS map, the selected ICD-10-PCS code is related to three SNOMED CT concepts that are descendants of the concept Cesarean section (11466000). MAGPIE displays these concepts and asks whether the user wants to use one of these concepts instead. (Figure 8) On picking the refined SNOMED CT concept, MAGPIE displays the final result of the SNOMED CT and ICD-10-PCS codes selected.

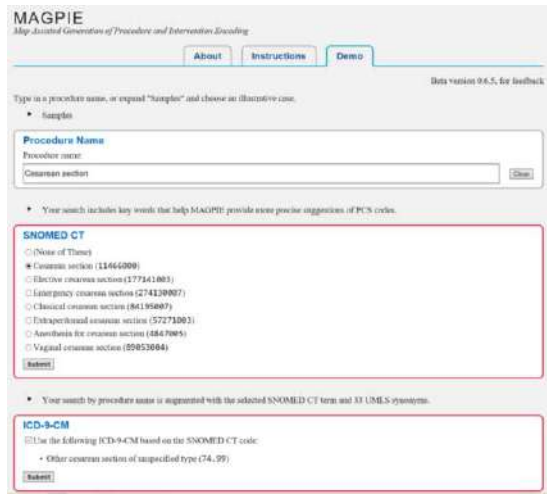


Figure 6 – MAGPIE screenshot 1 - default SNOMED CT and ICD9V3 codes for 'Cesarean section'

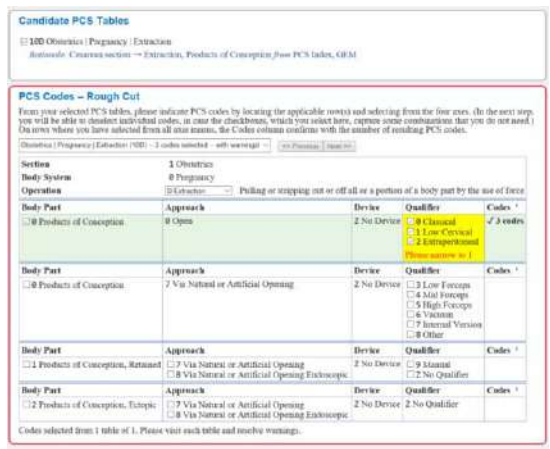


Figure 7 -- MAGPIE screenshot 2 - ICD-10-PCS tables and pre-selected codes for 'Cesarean section'

Evaluation

We used 88 common surgical procedure names to test MAGPIE. First, we evaluated the accuracy of the default SNOMED CT and ICD9V3 codes suggested by MAGPIE. The majority (70%) of the SNOMED CT codes were exact matches, while there were more close matches (48%) than exact matches for ICD9V3 codes (Table 1).

Figure 8– MAGPIE screenshot 3 - refinement of SNOMED CT codes and list of final codes.

Table 1 – Accuracy of default codes found by MAGPIE

Match category	SNOMED CT	ICD9V3
Exact match	62 (70%)	34 (39%)
Close match	26 (30%)	42 (48%)
Unrelated	0	12 (14%)
Total	88 (100%)	88 (100%)

Next, we assessed how well MAGPIE performed in finding the ICD-10-PCS tables and codes. Without changing the default SNOMED CT and ICD9V3 codes suggested by MAGPIE, the correct ICD-10-PCS table was found in 75 (85%) cases. The list of suggested tables was ranked by MAGPIE based on the number of corroborating sources. The sources included ICD-10-PCS index, ICD-10-PCS procedure name or GEM. The tables with the highest number of suggesting sources were ranked first. Among the 75 cases where the correct table was among the default MAGPIE suggestions, the correct table was ranked first in 55 (63%) cases. Overall, in 65 (74%) cases, the correct table was found among the top three (Table 2).

Table 2 – MAGPIE performance using the default selections

Result based on MAGPIE default choices	No. of cases (N=88)
Correct ICD-10-PCS table suggested	75 (85%)
- Ranked first	55 (63%)
- Ranked top 3	65 (74%)
Correct ICD-10-PCS codes suggested	25 (28%)

We further assessed the accuracy and completeness of the pre-selected ICD-10-PCS codes (suggested by GEM). Among the 75 cases where the correct table was found by MAGPIE, the pre-selected codes were exactly the same as in the reference standard in 25 (28%) cases. In the rest of the cases, the users need to adjust the pre-selected codes. Generally, the cases that needed user adjustment can be grouped into three categories. First, there were no pre-selected ICD-10-PCS codes because MAGPIE did not find any ICD9V3 codes, and so there were no GEM-suggested ICD-10-PCS codes. Second, the GEM suggestions were incomplete. One example was 'Ligation of Fallopian tube'. The default ICD9V3 code was *Other bilateral destruction or occlusion of fallopian tubes (66.39)*. GEM found only the bilateral codes, missing the codes for left and right Fallopian tubes in the reference standard. And in the last category, some of the GEM suggested codes were not in the reference standard. For example, in 'Hemorrhoidectomy, internal and external, simple', the GEM suggestions included codes for percutaneous approach, percutaneous endoscopic approach and use of extraluminal device, which were not present in the reference standard.

We further analyzed the 13 cases in which the correct ICD-10-PCS tables were not found by MAGPIE using the default selections. In some of the cases, there was more than one procedure involved. One example was 'Complete transurethral resection of the prostate including control of postoperative bleeding' which required two ICD-10-PCS codes for complete coding. MAGPIE was able to find the code for the 'resection' procedure, missing the 'control of bleeding' part. Some failure was caused by procedure names that were under-specified. One example was 'Shaving benign hyperkeratotic lesion, single'. In clinical discourse, this could be understood as referring to a procedure on the skin. However, since the skin was not explicitly stated in the search term, MAGPIE was not able to find the correct table. Had one used the search term 'Removal of skin lesion' MAGPIE would be able to find the correct codes. In some cases, the procedure could be coded in multiple ways in ICD-10-PCS and there was potential ambiguity. One example was 'Dilation and curettage of uterus'. If the operation was carried out in a non-pregnant woman, this would be coded as *0UDB7ZX Extraction of endometrium through natural or artificial opening*. However, if this was done for termination of pregnancy, it would be coded as *10A07ZZ Abortion of Products of Conception, Via Natural or Artificial Opening*.

Discussion

ICD-10-PCS is radically different from ICD9V3 and the traditional approach for ICD9V3 coding may not work equally well for ICD-10-PCS. Searching based on the ICD-10-PCS index will not work well because the ICD-10-PCS index is relatively lacking in useful clinical terms, compared to the ICD9V3 index. Searching based on the names of the ICD-10-PCS codes, which are generated from concatenating the values of the individual axes, will probably not work well either because some ICD-10-PCS terms (e.g., removal, extraction, extirpation, resection, excision) are not well-aligned with their meaning in clinical usage. We have shown in our previous study that recall and precision are both better when searching with the repurposed ICD9V3 index compared to the native ICD-10-PCS index. Based on the repurposed ICD9V3 index and other resources that the SNOMED CT to ICD-10-PCS Map Project Group has developed, we have built a tool that can be used for both ICD-10-PCS and SNOMED CT coding. SNOMED CT is an emerging international clinical terminology standard and is increasingly used in electronic health records. Using the MAGPIE tool, users can compare the two coding systems side-by-side and understand differences in their scope, granularity and organizing principles.

Since the goal of MAGPIE is to look simultaneously for both SNOMED CT and ICD-10-PCS codes, the cascading search approach is a natural choice. This is similar to the approach used in the I-MAGIC tool [14], which starts with a SNOMED CT concept then navigates to ICD-10-CM through a map. SNOMED CT is the preferred starting point because the SNOMED CT terms are closest to clinical parlance. SNOMED CT is also very comprehensive and has good coverage of both clinical diagnosis and procedures. The next link through ICD9V3 is necessary to make use of the repurposed ICD9V3 index and GEM. Since ICD9V3 is no longer in use, the ICD9V3 code is used only as a navigational pointer. In many cases, it is not necessary to find the perfectly matching ICD9V3 code to get to the correct ICD-10-PCS code, and an approximate ICD9V3 code match is sufficient, as shown in our results. While in only less than 40% of cases the default ICD9V3 codes are exact matches, overall 85% of the cases lead to the correct ICD-10-PCS tables by accepting the default values. The use of SNOMED CT and ICD9V3 codes as intermediate steps offers the opportunity for the user to 'correct course' if MAGPIE is going down a wrong path, which sometimes happens due to a vague or potentially ambiguous search term. One example is 'Construction of shunt', which can mean shunts for arteries, veins or the nervous system. The user can help to guide MAGPIE by choosing the correct SNOMED CT and ICD9V3 codes.

The mapping resources that MAGPIE depends on are used in either direction. Using maps in this way can sometimes be

problematic, especially when the source and target terms are not exactly equivalent. Sometimes maps may contain errors too. However, despite these limitations, our evaluation shows the performance of MAGPIE to be satisfactory.

One potential problem of the GEM is that one ICD9V3 code often ends up with multiple ICD-10-PCS codes because of the difference in the granularity of the two systems. In our previous study, three or four times more ICD-10-PCS codes were retrieved by the GEM compared to the reference standard. This is mostly the result of combinatorial explosion i.e., the total number of codes is the product of the number of options for each axis. To mitigate this problem, we use semantic filtering to prune the GEM-suggested ICD-10-PCS codes. If the user types in '*laparoscopic* excision of ovarian cyst', MAGPIE will exclude the options of 'open' or 'through natural or artificial orifice' from the approach axis. Similarly, if the user types in 'Total replacement of *left* hip', MAGPIE will exclude the values of 'right hip' or 'bilateral hips'.

One special feature of MAGPIE is that the final selection of the ICD-10-PCS code(s) can be used to suggest refinement for the initially selected SNOMED CT concept, in a way 'closing the loop' of the cascading search. In coding, one should always use the most specific code possible in a coding system for a clinical concept. However, coders tend to settle on the first code in the picklist that is applicable and do not look further for more specific codes. MAGPIE takes advantage of the information carried in the final ICD-10-PCS code to prompt the user to refine the initial SNOMED CT selection where appropriate. Therefore, in addition to helping with ICD-10-PCS coding, MAGPIE will also improve SNOMED CT coding.

We recognize the following limitations in our study. The list of common procedure names used in the evaluation was taken from one institution and might not be representative of other institutions. The reference standard for the evaluation was based on the judgment of two physicians with terminology expertise and was not independently validated. The assessment of the performance of MAGPIE was based on a single reviewer and a relatively small sample.

Conclusion

To overcome some of the difficulties in ICD-10-PCS coding, we have developed an innovative tool called MAGPIE (Map-Assisted Generation of Procedure and Intervention Encoding). MAGPIE makes use of the repurposed ICD-9-CM procedure index, which is much richer than the native ICD-10-PCS index. In addition, other lexical and mapping resources derived from the UMLS and GEM are also used. MAGPIE applies various innovative approaches to help the user hone in on the correct SNOMED CT and ICD-10-PCS codes. Our evaluation shows that MAGPIE finds the correct SNOMED CT code and ICD-10-PCS table in 70% and 85% of cases respectively, without any user intervention.

Acknowledgements

The authors thank Kaiser Permanente for providing the list of commonly performed procedure names. We also thank Phill Wolf for the development of the MAGPIE tool. This work is supported in part by the Intramural Research Program of the National Institutes of Health and the National Library of Medicine.

References

[1] R.F. Averill, R.L. Mullin, B.A. Steinbeck, N.I. Goldfield, and T.M. Grant, Development of the ICD-10 procedure coding system (ICD-10-PCS), *Top Health Inf Manage* **21** (2001), 54-88.

[2] US Centers for Medicare & Medicaid Services ICD-10-PCS web page, in https://www.cms.gov/Medicare/Coding/ICD10/2019-ICD-10-PCS.html.

[3] ICD-10-PCS Official Guidelines for Coding and Reporting, in https://www.cms.gov/Medicare/Coding/ICD10/2019-ICD-10-PCS.html.

[4] SNOMED CT to ICD-10-PCS Mapping Project Group, in https://confluence.ihtsdotools.org/display/SCTICD10PCS

[5] K. Fung, J. Xu, A. D'Havé, F. Ameye, and A. Gutiérrez, Leveraging lexical matching and ontological alignment to map SNOMED CT surgical procedures to ICD-10-PCS, *AMIA Annu Symp Proc 2016* (2016), 570-579.

[6] K.W. Fung, J. Xu, F. Ameye, A.R. Gutierrez, and A. D'Have, Achieving Logical Equivalence between SNOMED CT and ICD-10-PCS Surgical Procedures, *AMIA Annu Symp Proc* **2017** (2017), 724-733.

[7] K. Fung, J. Xu, A. D'Havé, F. Ameye, A. Gutiérrez, and A. Santos, Developing a Map from SNOMED CT Procedure Concepts to ICD-10-PCS, in: *SNOMED CT Expo 2015*, Montevideo, Uruguay.

[8] K. Fung, J. Xu, A. D'Havé, F. Ameye, and A. Gutiérrez, Mapping SNOMED CT procedure concepts to ICD-10-PCS, in: *SNOMED CT Expo 2016*, Wellington, New Zealand.

[9] K. Fung, J. Xu, A. D'Havé, F. Ameye, A. Gutiérrez, and A. Busquets, Demonstrating the Benefits of Mapping SNOMED CT to ICD-10-PCS through a Prototype Application for End User Implementation, in: *SNOMED CT Expo 2017*, Bratislava, Slovakia.

[10] K.W. Fung, J. Xu, F. Ameye, A.R. Gutierrez, and A. Busquets, Re-purposing the ICD-9 -CM Procedures Index for Coding in ICD-10-PCS and SNOMED CT, *AMIA Annu Symp Proc* **2018** (2018), 450-459.

[11] General Equivalence Maps 2018 ICD-10-PCS, Centers for Medicare & Medicaid Services, in https://www.cms.gov/Medicare/Coding/ICD10/2018-ICD-10-PCS-and-GEMs.html.

[12] Apache Lucene open source information retrieval software library.

[13] K.W. Fung, C. McDonald, and S. Srinivasan, The UMLS-CORE project: a study of the problem list terminologies used in large healthcare institutions, *J Am Med Inform Assoc* **17** (2010), 675-680.

[14] I-MAGIC demo tool, NLM, in: https://imagic.nlm.nih.gov/imagic/code/map.

Address for correspondence
Kin Wah Fung is the corresponding author. The preferred method of contact is through email: kfung@mail.nih.gov.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190258

Using Machine Learning to Integrate Socio-Behavioral Factors in Predicting Cardiovascular-Related Mortality Risk

Hanyin Wang[a], Yikuan Li[b], Hongyan Ning[a], John Wilkins[a], Donald Lloyd-Jones[a], Yuan Luo[a]

[a] Feinberg School of Medicine, Northwestern University, Chicago, IL, U.S.A
[b] Department of EECS, Northwestern University, Chicago, IL, U.S.A

Abstract

Cardiovascular disease is prevalent and associated with significant mortality rate. Robust lifetime risk stratification for cardiovascular disease is important for effective prevention, early diagnoses, targeted intervention, and improved prognosis. Health disparities, manifested as socio-behavioral factors, are believed to have multiple effects throughout life with great complexity. Multiple studies investigated lifetime cardiovascular-related mortality risk prediction focusing on subjects' pathophysiology and intervention profiles. In this study, we applied machine learning algorithms and focused on integrating socio-behavioral factors to pathophysiology and intervention profiles to predict cardiovascular-related mortality risk. Our results showed that multiple machine learning algorithms can predict risk with reasonable accuracy, using mixed types of features. Particularly, socio-behavioral factors contributed significantly to the improved accuracy of mortality risk prediction. Feature analysis identified the odds ratio of socio-behavioral factors for cardiovascular-related mortality and offered potential insights on how they impact subjects' long-term outcomes. Our results call for further investigation of this important topic.

Keywords:
Risk Factors, Cardiovascular Diseases, Machine Learning

Introduction

Cardiovascular Disease (CVD) is a major cause of disability and premature death throughout the world and substantially contributes to the escalating costs of health care, mortality, and morbidity [1]. According to the report from World Health Organization, an estimated 17.9 million people died from CVDs in 2016, representing 31% of all global deaths [2]. In the US, the mortality of cardiovascular disease is as high as 0.25 (95% CI 0.10-0.63) [3]. In order to reduce the physical, emotional, and economical burden caused by cardiovascular-related mortality, it is important to develop robust lifetime risk stratification algorithm that is accurate and that generalizes at a population level.

Previous studies revealed the association between cardiovascular disease and demographics, physiology, comorbidity, and medication features, such as body mass index [4], chronic kidney diseases [5], hypertension and obesity [6], anemia [7], etc. Social determinants and behavioral factors, such as education level, income level, smoking status, and diet habit, act upon populations at all times, and they have also been found to have wide-ranging health related effects across the life course of a population [8, 9]. For example, higher income is linked to better health outcomes, including lower prevalence of cardiovascular disease and lower mortality risk among populations with cardiovascular disease. With growing awareness of their importance, previous studies have explored relationships between social determinants and risks for developing cardiovascular diseases [10-13]. In addition to social determinants, behavioral factors, such as diet [14-16] and physical activity [17, 18], are investigated to be related to various mortality and cardiovascular-related outcomes. For example, poor diet, lack of physical activity, and cigarette smoking are all linked to premature deaths [19]. Unfortunately, a majority of existing studies focused on linking social and behavioral factors to risks for developing specific diseases, leaving systematic analysis on these factors' relations with mortality risks a largely uncharted territory. On the other hand, pooling the social determinants and behavioral factors together and utilizing more powerful approaches, such as machine learning [20-22], may shed more light on the landscape and yield a more robust mortality-risk prediction model.

In this study, we apply different machine learning models to predict cardiovascular-related mortality risks by integrating social behavioral factors with subjects' pathophysiology and intervention profiles as features. We aim to provide a comprehensive understanding of the socio-behavioral landscape and use the additional knowledge to improve cardiovascular-related mortality prediction. Our results demonstrate the impact of social and behavioral factors on cardiovascular health at population level and the feasibility towards more robust lifetime risk stratification by integrating them.

Methods

Study population

In this study, we used data from The Cardiovascular Disease Life Risk Pooling Project (LRPP) [23]. LRPP was designed as an individual-level pooled dataset from 20 US community-based cardiovascular disease cohorts that include demographics, physiological test results, medication status, socio-behavioral factors, and mortality indicators (for mortality of cardiovascular disease, coronary heart disease, and mortality resulting from all causes). From the overall sample size of 1,097,178 observations of 277,296 subjects, 40,711 completed cases with all predicting features available were selected for each mortality in this study. Shown in Table 1 is the data distribution after splitting training and test set into approximately 7:3 ratio.

Table 1– Data distribution

	Training set				Test set				All data			
	Total	Death	Non-event	% death	Total	Death	Non-event	% death	Total	Death	Non-event	%death
CVD	28,497	1,772	26,725	6.22%	12,214	846	11,368	6.93%	40,711	2,618	38,093	6.43%
CHD	28,497	1,772	26,725	6.22%	12,214	846	11,368	6.93%	40,711	2,618	38,093	6.43%
TOT	28,497	6,048	22,449	21.22%	12,214	2,703	9,511	22.14%	40,711	8,751	31,960	21.50%

CVD: mortality due to cardiovascular diseases; CHD: mortality due to coronery heart disease; TOT: all causes mortality

Statistical analysis

In total, 25 features were chosen from the entire LRPP dataset. They belong to the following domains: demographics, physiological test results, medication status, and socio-behavioral factors. Features were grouped into 3 different combinations, physiological test result (Phy), physiological test results combined with medication status (Phy+Med), and further including socio-behavioral factors (Phy+Med+Soc). While demographics were included in each group (Table 2).

Four supervised machine learning algorithms were applied for prediction of the three outcomes, including Naïve Bayes (NB), Logistic Regression (LR), Linear Support Vector Machine (SVM), and Random Forest (RF). Training and test set were split to approximately 7:3. Due to the imbalance of the data (incidence rate at approximately 6%), parameter *'class_weight'* was set to *'balanced'* for LR, SVM and RF. To reduce over-fitting and variance, five-fold cross validation was performed on training set. Best parameters for each model, tuned through cross validation, were then applied on test set to evaluate the predictive performance of each classifier. The area under the receiver operating characteristic curve (AUC) as well as predicted probability were acquired from each algorithm.

The AUCs of the three models are shown in Table 3, ranging from the lowest to the highest. Considering that the distribution of the predicted probability was unknown, permutation test [24] was applied to explore the difference between the probability of correct prediction of every two models (ie. tested between Phy and Phy + Med, Phy+Med and Phy+Med+Soc).

In RF, coefficient *'feature_importance'* was extracted to identify the importances of features. Generalized linear regression (logit link function) was employed to find the odds ratio (OR) to evaluate risk for mortality of each socio-behavioral factor. Data cleaning, permutation test, and generalized linear regression were performed by R 3.3.1. *'scikit learn'* package was used for machine learning algorithms.

Results

AUCs of different algorithms and models

Table 3 shows the AUCs yielded from each of the four machine learning algorithms as applied to predict the three outcomes, with the three combination of the features. P-value 1 was derived from the permutation test between the feature combination of Phy and Phy+Med, while P-value 2 shows the significance of difference between the feature combination of Phy+Med and Phy+Med+Soc. Result demonstrates that AUCs are in a range from 0.7263 to 0.8941 across all the models.

Table 2– Predictors features

Categories	Variable names	Description
Demographics	SEX	Gender
	AGE	Calculated age at exam, round to 0 .1
	BMI	Body mass index, kg/m^2
Physiological tests	SBP	Systolic blood pressure, mmHg
	DBP	Diastolic blood pressure, mmHg
	LDLCHL	LDL cholesterol, mg/dL
	HDLCHL	HDL cholesterol, mg/dL
	TOTCHL	Total cholesterol, mg/dL
	TRIGLY	Triglycerides, mg/dL
	GLU	Glucose level
	CVD_BASELINE	Prevalent of cardiovascular disease at enrollment
Medication statues	HXDIAB	Diabetic Medication Status
	RXCHL	Medication of lowering lipid
	RXHYP	Medication of anti-hypertension
Socio-behavioral factors	RACE	Race/ethnicity
	EDU	School education
	SMOKER	Current smoker
	FORSMOKER	Former smoker
	DRINKER	Current drinker
	ALCO_ML_DAY	Alcohol intake, mL /day
	DASH	Dietary Approaches to Stop Hypertension score
	aHEI	Alternate Healthy Eating Index score 2010
	aMed	Alternate Med diet score (all cohorts median)
	primary PA	Primary physical activity
	MVPA	Moderate to vigorous physical activity
Outcomes	CVD_DTH	Death caused by cardiovascular diseases
	CHD_DTH	Death caused by coronary heart disease
	TOT_DTH	All causes death

Table 3– AUCs of the three models and four algorithms

	Phy	PM	*P-val 1*	PMS	*P-val 2*
Mortality due to cardiovascular disease					
NB	0.7436	0.7579	*0.001*	0.8048	*0.010*
LR	0.8510	0.8554	*0.001*	0.8752	*0.001*
SVM	0.8510	0.8558	*0.298*	0.8754	*0.001*
RF	0.8461	0.8535	*0.001*	0.8941	*0.001*
Mortality due to coronary heart disease					
NB	0.7436	0.7579	*0.001*	0.8048	*0.010*
LR	0.8510	0.8554	*0.001*	0.8752	*0.001*
SVM	0.8510	0.8558	*0.298*	0.8754	*0.001*
RF	0.8461	0.8535	*0.001*	0.8941	*0.001*
Mortality result from all causes					
NB	0.7263	0.7349	*0.001*	0.7864	*0.001*
LR	0.8214	0.8247	*0.001*	0.8541	*0.001*
SVM	0.8208	0.8241	*0.001*	0.8538	*0.001*
RF	0.8325	0.8393	*0.001*	0.8921	*0.001*

PM: combination of Phy+Med; PMS: combination of Phy+Med+Soc

Considering NB algorithm as the baseline, overall result illustrates significant improvement in AUCs when applying LR, SVM, and RF algorithms. The Phy combination consisted of the 11 features from Demographics and Physiological tests categories. NB algorithm provided AUCs between 0.7263 and 0.7436 based on these features. RF showed a better performance within this combination that yielded AUCs of 0.8461 for prediction of mortality risk of CVD and CHD. The most promising algorithms for this model were LR and SVM which achieved AUCs of 0.8510 for prediction regarding CVD and CHD mortality. When adding the Medication status to the feature collection, we observed an elevation in all AUCs compared to the Phy model. Meanwhile, p-values ($\leq$0.001) generated from permutation test also confirmed the statistically significant improvement on the probability of correct prediction, except for SVM algorithm for CVD and CHD mortality risk prediction. Last but not least, much better performance was observed when integrating the Socio-behavioral factors into the feature combination. All AUCs yielded from RF algorithm became above 0.80, the highest AUC reached 0.8941. Comparing the AUCs to the previous model, this model with socio-behavioral factors demonstrated the best performance. Results from permutation test for difference between the correct prediction probability of Phy+Med and Phy+Med+Soc also helped verify the significant improvement.

Feature analysis

Shown in Figure 1 are the feature importances derived from RF for different feature combinations. In the combinations without socio-behavioral factors (Phy and Phy+Med), age served as the most important feature for classification. Blood pressure, BMI, triglycerides level, and the three kinds of cholesterol levels also played relatively important roles in the first two Phy and Phy+Med [25]. The pattern was different in the third feature combination (Phy+Med+Soc), race performed as the most important feature for all three outcomes, followed by age and SBP. Meanwhile, aHEI score also appeared to be an essential feature for classification. Other important factors shown in the previous two feature combinations still remained on top.

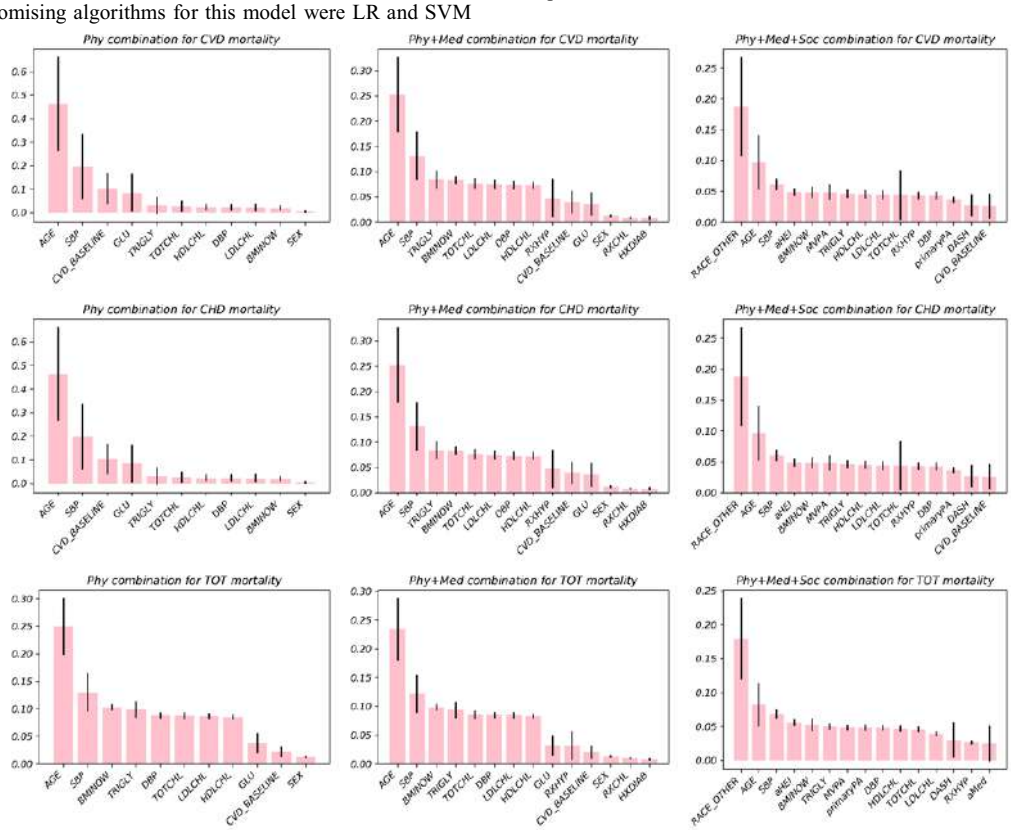


Figure 1– Feature importances for different feature combinations derived from Random Forest. For combination Phy+Med+Soc,only the first 15 highest ranked features were shown in the figure.

Discussions, Limitations and Future Work

With machine learning approaches, we were able to detect the hidden patterns within larg-scale data. Among the four algorithms applied in this study, LR, SVM and RF demonstrated reasonable accuracy. Among them, RF appeared to be the best for solving the current problem.

Significant improvements were observed from those machine learning configurations when adding socio-behavioral factors as well as medication status into predicting models. It should not be surprising to observe medication status (BP medication, lipids medication, and DM medication) as an important predictor, since only the subjects with such health disorders would take the medicine. Meanwhile, blood pressure disorders [26, 27], dyslipidemia [28, 29], and diabetes mellitus [30] were also found to be physiologically related with mortality caused by cardiovascular diseases. Eleven socio-behavioral factors were integrated to the previous models, covered the aspects of race, education, smoking, diet, and physical activity. To understand which sociobehavioral factors in particular have a higher risk for cardiovascular-related mortality, a further general linear regression (logit link function) was performed (Table 4). Comparing to Asian, subjects with race in African American and White were found to have odds ratio greater than 1 with p-values satisfying significant level, which indicates a higher mortality risk for African American and White population. Moreover, current smokers also experienced a higher risk than those who were not currently smoking. Compared with the educational level below, high school, higher educational level (high school or above) reduces the risk. Subjects with higher dietary scores (DASH and aHEI) were also found to be undertaking higher risk.

Previous studies with consideration of social factors were generally focused on social determinants [31] while ignored the importance of behavior. This study showed that socio-behavioral factors, such as smoking and diet, can also be essential predictors for mortality risk of cardiovascular diseases and is worthwhile to be included into predictive models. Similarly, socio-behavioral factors might also be worth considering for risk prediction of other health outcomes.

The LRPP study is a pooled project from a large variety of cohorts. Therefore, missing rate is relatively high for part of the features especially the socio-behavioral factors. This limited the size of training and test set and likely introduced biases. In future work, we will use modern imputation methods [32-34] to impute the missing data and scale our analysis on larger cohorts. Continuing from what we found in this study, we are planning to perform further analysis considering time sequence. With the dimension of time, we will generate more sophisticated models for prediction. At the same time, the approaches we demonstrated in this study could be applied to other CVD outcomes as well as other diseases.

Conclusions

In the present study, we applied multiple machine learning algorithms and integrated socio-behavioral factors with pathophysiology and medication profiles of subjects to predict cardiovascular-related mortality. All the machine learning algorithms, including Naïve Bayes, Logistic Regression, Linear Support Vector Machine, and Random Forest yielded reasonable accuracy. In addition to pathophysiological and interventional profiles, socio-behavioral factors were also included as predictive features. Integration of socio-behavioral factors brought statistically significant improvement of the all three mortality risk predictions, including the risk of mortality due to cardiovascular diseases, coronary heart disease, and mortality resulting from all causes. In the overall analysis, age, race, and blood pressure were found to be relatively more important for the prediction. While in the analysis for odds ratio of socio-behavioral factors, race, educational level, smoking, and diet were found to have significant impact on cardiovascular-related outcomes.

Given that socio-behavioral factors manifested long-term impacts on the health outcome disparities, we expect more exciting and promising developments of robust lifetime risk prediction algorithms with tighter integration of socio-behavioral factors to inform policy making and improve the population health and longevity.

Table 4 – Odds ratio and significance of socio-behavioral factors

Features	OR for CVD_DTH	P-value	OR for CHD_DTH	P-value	OR for TOT_DTH	P-value
RACE_ASIAN			(reference group)			
RACE_BLACK	1.7074	0.001***	1.7074	0.001***	2.6543	0.000***
RACE_HISPANIC	0.7668	0.173	0.7668	0.173	0.8067	0.059
RACE_OTHER	0.5646	0.476	0.5646	0.476	1.6610	0.399
RACE_WHITE	1.9269	0.000***	1.9269	0.000***	2.7296	0.000***
EDU < HIGH SCHOOL			(reference group)			
EDU > HIGH SCHOOL	0.6655	0.000***	0.6655	0.000***	0.5489	0.000***
EDU_HIGH SHCOOL	0.7089	0.000***	0.7089	0.000***	0.6733	0.000***
SMOKER_0			(reference group)			
SMOKER_1	1.6952	0.000***	1.6952	0.000***	2.6555	0.000***
FORSMOKER_0			(reference group)			
FORSMOKER_1	1.0318	0.549	1.0318	0.549	1.3092	0.000***
DRINKER_0			(reference group)			
DRINKER_1	0.8913	0.503	0.8913	0.503	1.1589	0.037*
ALCO_ML_DAY	1.0036	0.235	1.0036	0.235	1.0027	0.043
DASH	1.0651	0.000***	1.0651	0.000***	1.0657	0.000***
aHEI	0.9540	0.000***	0.9540	0.000***	0.9531	0.000***
aMed	1.0326	0.098	1.0326	0.098	1.0590	0.000***
primaryPA	0.9526	0.109	0.9526	0.109	0.8895	0.000***
MVPA	1.0416	0.164	1.0416	0.164	1.0130	0.495

** : <0.05; ** : <0.01; *** : <0.001*

Acknowledgements

This work was supported by NIH Grant 1R21LM012618. We thank Ying Liu at UT Dallas for helpful comment.

References

[1] S. Mendis, S. Davis, and B. Norrving, "Organizational update: the world health organization global status report on noncommunicable diseases 2014; one more landmark step in the combat against stroke and vascular disease," (in eng), no. 1524-4628 (Electronic).
[2] W. H. Orgnization, "Cardiovascular diseases (CVDs)," 2017.
[3] M. J. Benjamin Ej Fau - Blaha *et al.*, "Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association," (in eng), no. 1524-4539 (Electronic).
[4] S. S. Khan, H. Ning, J. T. Wilkins, and et al., "Association of body mass index with lifetime risk of cardiovascular disease and compression of morbidity," *JAMA Cardiology,* **3** (2018), 280-287.
[5] D. E. Weiner *et al.*, "Chronic kidney disease as a risk factor for cardiovascular disease and all-cause mortality: a pooled analysis of community-based studies," **15** (2004), 1307-1315.
[6] F. S. Ahmad, H. Ning, J. D. Rich, C. W. Yancy, D. M. Lloyd-Jones, and J. T. Wilkins, "Hypertension, Obesity, Diabetes, and Heart Failure–Free Survival: The Cardiovascular Disease Lifetime Risk Pooling Project," *JACC: Heart Failure,* **4** (2016), 911-919.
[7] M. J. Sarnak *et al.*, "Anemia as a risk factor for cardiovascular disease in The Atherosclerosis Risk in Communities (ARIC) study," **40** (2002), 27-33.
[8] B. E. Garrett, S. R. Dube, S. Babb, and T. McAfee, "Addressing the Social Determinants of Health to Reduce Tobacco-Related Disparities," (in eng), no. 1469-994X (Electronic).
[9] E. P. Havranek *et al.*, "Social Determinants of Risk and Outcomes for Cardiovascular Disease," **132** (2015), 873-898.
[10] N. E. Adler, M. M. Glymour, and J. Fielding, "Addressing Social Determinants of Health and Health Inequalities," (in eng), no. 1538-3598 (Electronic).
[11] E. Blas and A. S. Kurup, *Equity, social determinants and public health programmes*. World Health Organization, 2010.
[12] C. Kreatsoulas and S. S. J. C. J. o. C. Anand, "The impact of social determinants on cardiovascular disease," **26** (2010), 8C-13C.
[13] E. K. Chung *et al.*, "Screening for Social Determinants of Health Among Children and Families Living in Poverty: A Guide for Clinicians," (in eng), no. 1538-3199 (Electronic).
[14] J.-P. Després and B. J. N. r. r. Lamarche, "Effects of diet and physical activity on adiposity and body fat distribution: Implications for the prevention of cardiovascular disease," 6 (1993), 137-159.
[15] R. Estruch *et al.*, "Primary prevention of cardiovascular disease with a Mediterranean diet," 368 (2013), 1279-1290.
[16] J. H. O'Keefe Jr and L. Cordain, "Cardiovascular disease resulting from a diet and lifestyle at odds with our Paleolithic genome: how to become a 21st-century hunter-gatherer," in *Mayo Clinic Proceedings*, **79** (2004), 101-108: Elsevier.
[17] H. J. M. Kohl 3rd, s. i. sports, and exercise, "Physical activity and cardiovascular disease: evidence for a dose response," **6** (2001), S472.
[18] P. D. Thompson *et al.*, "Exercise and physical activity in the prevention and treatment of atherosclerotic cardiovascular disease: a statement from the Council on Clinical Cardiology (Subcommittee on Exercise, Rehabilitation, and Prevention) and the Council on Nutrition, Physical Activity, and Metabolism (Subcommittee on Physical Activity)," 107 (2003), 3109-3116.
[19] J. M. McGinnis, J. R. Williams-Russo P Fau - Knickman, and J. R. Knickman, "The case for more active policy attention to health promotion," (in eng), no. 0278-2715 (Print).
[20] H. O. Alanazi, A. H. Abdullah, and K. N. Qureshi, "A Critical Review for Developing Accurate and Dynamic Predictive Models Using Machine Learning Methods in Medicine and Health Care," (in eng), no. 1573-689X (Electronic).
[21] R. C. Deo, "Machine Learning in Medicine," (in eng), no. 1524-4539 (Electronic).
[22] T. Zheng *et al.*, "A machine learning-based framework to identify type 2 diabetes through electronic health records," (in eng), no. 1872-8243 (Electronic).
[23] J. T. Wilkins *et al.*, "Data Resource Profile: The Cardiovascular Disease Lifetime Risk Pooling Project," (in eng), no. 1464-3685 (Electronic).
[24] E. W. Noreen, *Computer-intensive methods for testing hypotheses: an introduction* (Computer). 1989.
[25] R. K. Upadhyay, "Emerging risk biomarkers in cardiovascular diseases and disorders," (in eng), *Journal of lipids,* (2015), 971453-971453.
[26] S. L. Stevens *et al.*, "Blood pressure variability and cardiovascular disease: systematic review and meta-analysis," (in eng), no. 1756-1833 (Electronic).
[27] X. Xie *et al.*, "Effects of intensive blood pressure lowering on cardiovascular and renal outcomes: updated systematic review and meta-analysis," (in eng), no. 1474-547X (Electronic).
[28] S. Karr, "Epidemiology and management of hyperlipidemia," (in eng), no. 1936-2692 (Electronic).
[29] S. Yusuf *et al.*, "Cholesterol Lowering in Intermediate-Risk Persons without Cardiovascular Disease," (in eng), no. 1533-4406 (Electronic).
[30] A. Rawshani *et al.*, "Mortality and Cardiovascular Disease in Type 1 and Type 2 Diabetes," (in eng), no. 1533-4406 (Electronic).
[31] S. N. Kasthurirathne, J. R. Vest, N. Menachemi, P. K. Halverson, and S. J. Grannis, "Assessing the capacity of social determinants of health data to augment predictive models identifying patients in need of wraparound social services," (in eng), no. 1527-974X (Electronic).
[32] Y. Luo, P. Szolovits, A. S. Dighe, and J. M. Baron, "Using Machine Learning to Predict Laboratory Test Results," *American Journal of Clinical Pathology,* **6** (2016), 778-788.
[33] S. Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in R," *Journal of statistical software,* 45 (2011).
[34] Y. Luo, P. Szolovits, A. S. Dighe, and J. M. Baron, "3D-MICE: integration of cross-sectional and longitudinal imputation for multi-analyte longitudinal clinical data," *Journal of the American Medical Informatics Association (JAMIA),* **6** (2017), 645-653.

Address for correspondence
yuan.luo@northwestern.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190259

Development of Deep Learning Algorithm for Detection of Colorectal Cancer in EHR Data

Yu-Hsiang Wang[a], Phung-Anh Nguyen[b,*], Md. Mohaimenul Islam[b], Yu-Chuan Li[b,c], Hsuan-Chia Yang[b]

[a] College of Medicine, Taipei Medical University, Taipei, Taiwan
[b] International Center for Health Information Technology, Taipei Medical University, Taipei, Taiwan
[c] College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan
[*] Corresponding author

Abstract

We aimed to develop a deep learning model for the prediction of the risk of advanced colorectal cancer in Taiwanese adults. We collected data of 58152 patients from the Taiwan National Health Insurance database from 1999 to 2013. All patients' comorbidities and medications history were included in the development of the convolution neural network (CNN) model. We also used 3-year medical data of all patients before the diagnosed colorectal cancer (CRC) as the dimensional time in the model. The area under the receiver operating characteristic curve (AUC), sensitivity, and specificity were computed to measure the performance of the model. The results showed the mean (SD) of AUC of the model was 0.922 (0.004). Moreover, the performance of the model observed the sensitivity of 0.837, specificity of 0.867, and 0.532 for PPV value. Our study utilized CNN to develop a prediction model for CRC, based on non-image and multi-dimensional medical records.

Keywords:
Colorectal Neoplasms; Algorithms; Electronic Health Records;

Introduction

Cancer has been the leading cause of death since 1986 [2] and colorectal cancer (CRC) has been ranked as the third cause of cancer-related death in Taiwan since 2008 [5]. The incidence rate of CRC was 43.6 per 100,000 population in 2015, the age-adjusted incidence rate of male was higher than female. In fact, it increased by 13% for male and 6% for a female since 2002. CRC is now a burden and challenge to Taiwan's health care [1]. Therefore, it is important to find the potential biomarkers that can contribute to early diagnosis of colorectal cancer.

Artificial intelligence (AI) and machine learning are taking all fields of science by storm including clinical medicine [7; 10]. Deep learning is a type of the highly flexible machine learning approach, has emerged a groundbreaking algorithm to enhance the performance of current machine learning techniques and to solve complex problems [7]. Several studies already showed that machine learning model can predict cancer that help physicians to reduce the number of false decisions and improve health outcomes.

In this study, we aimed to develop a deep learning model by applying all patients phenotypic information to detect the onset of colorectal cancer.

Methods

Data Sets

In this study, we used the reimbursement data from Bureau National Health Insurance (NHI) system in Taiwan that had registered all medical records since 1996 [8]. We obtained a randomization of two million population of the NHI claim data from January 01, 1999 to December 31, 2013. The study was approved by the institutional review board committee at Taipei Medical University and the need for informed consent was waived.

Study population

We initially identified all patients who were diagnosed with a primary diagnosis of colorectal cancer (CRC) (International Classification of Disease, Ninth Revision [ICD-9] codes 153, 154) between January 01, 1999, and December 31, 2013. The diagnosis accuracy of CRC was confirmed by both specific ICD-9 codes and inclusion in the Registry for Catastrophic Illness Patient (RCIP), a subpart of the NHI database [6]. Surgical pathological confirmation or undergoing treatment process of CRC is required for patients to be registered in the RCIP. However, the index date was defined as the date of diagnosis. Patients were excluded if they were younger than 20 years of age at the date of diagnosis CRC or if they did not have any outpatient claims at each calendar year of four years before the index date.

We then randomly selected 50,000 individuals without any cancer diagnosis during 15 years in the NHI claim database. All medical records of those were also collected. We defined the date of the last visit observed in the NHI database as the index date of patients with cancer-free. We then used the same conditions as CRC patients to cross-check their age and the history of medical data.

In this study, our purpose was to predict CRC in one year ahead; we, therefore, used the data of patients with and without cancer diagnosis from the fourth year to the first year before the index date and left out a year data to be the gap of developing CRC cancer **(Figure 1).**

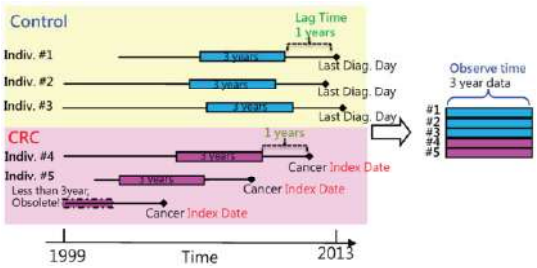


Figure 1. Observation Window and Lag Time

Feature definition

First, we checked the index date to observe the age and sex of overall individuals (i.e., with and without cancer diagnosis). We included the ICD9-CM 3-digit codes such as 001-999, V01-V91 in the features of the development of the model. The supplementary classification of external causes of injury and poisoning codes, E000-E999 were excluded in the feature list.

For the medications, the Taiwan medication codes were collected and mapped to WHO-ATC (Anatomical Therapeutic Chemical Classification System) codes. The ATC-4 (ie. ATC 5-digit) codes were used in the further development of the model. A total of 1931 features (ie. age, sex, 1099 ICD9-CM codes, and 830 ATC-4 codes) were used in the development of the CRC prediction model.

Time definition and data processing

We collected the data of all individuals in 1095 days (i.e., 3-year data). For each disease, the patient was considered as a diagnosis if at least three visits were observed with its ICD9 code during the 3-year period. For each medication, the patient was defined as a user if its ATC code was observed in the prescriptions, and the medication observation time was the number of days prescribed by physicians. For each patient, the 7-day data was summed up continuously, and the final data of 157 grids were derived from 1095-day data set **(Figure 2)**. The value of each feature of each patient was then normalized to scale up to the range in between 0 and 1 in order to make the similar severity of all individuals as the same cause. It expresses mathematically in (1) as: $x_{new} = {x - x_{min}}/{x_{max} - x_{min}}$ (1), where x = (x_1, …, x_n) and x_{new} is now the new normalized value. **Figure 3** shows an example of the normalized data.

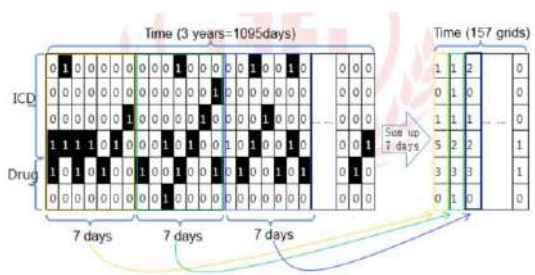


Figure 2. The processing of time data

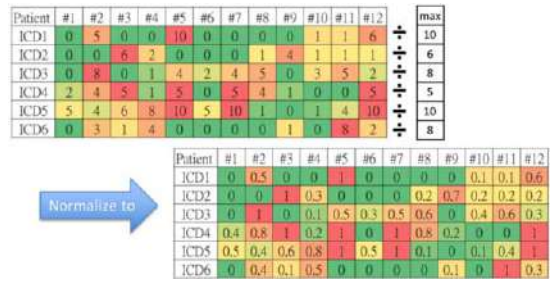


Figure 3. An example of normalized data

Development of the model

Deep learning is the process of training a neural network (a large mathematical function with millions of parameters) to perform a given task. **Figure 4** shows the convolution neural network (CNN) structure of the study. We created the function that computed the CRC severity from a large set of the matrix of all individuals. During the training process, the parameters of the neural network were initially set to random values. The total of 1929 features data including 1099 ICD9 codes and 830 ATC codes, were categorized into 19 groups (i.e., 18 groups for ICD9, for example, 001-139-infectious and parasitic disease group, and one ATC group) for all patients. We created two convolutional layers with 32 feature maps for each category. Two max-pooling layers with the filters of size 1x3 were applied to reduce the number of parameters, and to control overfitting. After that, we flattened the pooled feature maps with the addition of age and sex into another layer in order to process the artificial neural network. A hidden layer of 400 neurons was used in this fully-connected layer.

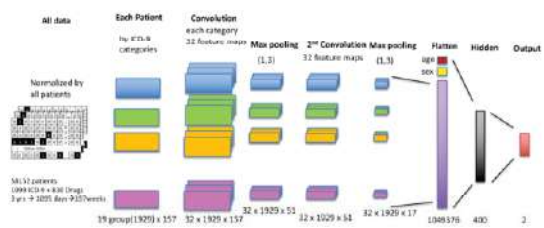


Figure 4. CNN structure of the study

The optimization algorithm used to train the network weights was a distributed stochastic gradient descent with a mini-batch size of 32. The model was optimized using Adadelta optimizer [13]. The input and hidden layers used a Rectified Linear Unit (ReLU) activation function (refs), while the output layer used the Softmax activation function. The dropout value of 0.1 used to prevent the overfitting and was applied to the input and hidden layer [9].

The average k-fold cross-validation accuracy, with a k-value of 5, was used as the metric to determine the best performance, optimizer, and loss functions of the model. We computed the area under the receiver operating curve (AUC) value, sensitivity, specificity, and positive predictive value (PPV) to observe the model performance.

The software was implemented using python v.3.7 and the model was created and trained with the Tensorflow framework v.1.9.

Results

The demographic characteristics of cancer and non-cancer patients are summarized in **Table 1**. Following the conditions, there were 58152 patients including in the study, of which 10185 diagnosed with CRC and 47967 patients were without a cancer diagnosis. The mean age (SD) of patients with CRC diagnosis was 62.7 (14.2) compared to patients without cancer of 47.6 (17.3). Most of the patients diagnosed with CRC were male (54.1%). In addition, the average number of diagnosis of a CRC patient was significantly higher compared to a non-cancer patient before the index date (e.g. 41.2 vs. 26.4 in the first observed year).

Table 1. Characteristics of patients in this study

	Case (n=10185)	Control (n=47967)
Age mean (sd)	62.7 (14.2)	47.6 (17.3)
Male sex (%)	5513 (54.1%)	23109 (48.2%)
Diagnosis count	1119580	3735982
Average in 1 year before index date*	41.2	26.4
Average 1~2 year before index date*	39.9	28.6
Average 2~3 year before index date*	37.7	27.4

*Note: * Number per patient per year*

Figure 5 shows the AUROC of the model with 5-fold cross-validation. The mean (SD) of AUC of the model was 0.922 (0.004). The performance of the model observed with the sensitivity of 0.837, specificity of 0.867, and 0.532 for PPV value.

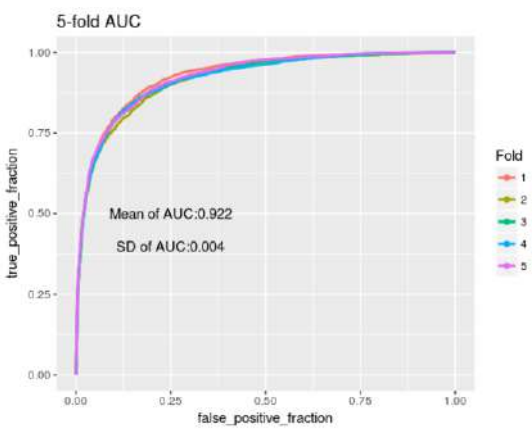


Figure 5. AUROC of the model with 5-fold cross-validation

Discussion

Our study findings demonstrate that the CNN model could appropriately predict CRC based on non-image and multi-dimensional medical records (ie. phenotype, diseases and medication). The results showed higher AUC (0.922) with very good sensitivity and specificity. This automated system for the detection of CRC offers several advantages, including consistency of interpretation (it could be a good screening tool for CRC for any person before deciding to take further some examination tests).

In this study, we tried to observe the features weight by using the odds ratio (ORs) and stepwise feature extraction (i.e., ANN model). We also selected top 10 features and most of the features were related to cardiovascular diseases. Comparing to other studies [3; 4; 12], our study showed higher performance (Area under receiver operating curve, sensitivity, and specificity). Since, our government has already launched a nationwide screening program of CRC since 2004 and a free fecal immunochemical test is offered biennially to individuals aged 50 to 75 [11]. According to our findings, the program may extend to cover all patients beyond this age range.

The main strength of this study lies in a use of a population-based cohort with a large and nationally representative sample, which increases its generalizability in Taiwan. However, there are several limitations in our study that need to be addressed. First, this is a retrospective cohort study, so selection bias and misclassification may be introduced. Second, inherent limitations of NHIRD hinder our ability to get some information related to the CRC, such as smoking habits, alcohol consumption, body mass index (BMI), family history of CRC, diet, and physical activity. However, we have included the comorbidities as surrogates for some risk factors of CRC, such as COPD and a stop-smoking clinic for smoking, alcohol-related illness for alcohol, and obesity for BMI. Finally, many chronic diseases such as diabetes mellitus (DM) and hypertension are not labeled every time on diagnosis, which may lessen the importance of those diseases.

Future direction

The ultimate goal of our study is to develop a web service for the public, helping them make predictions of risk of getting CRC one year before based on prior three-year-diagnosis and medication records. Therefore, this model provides a cheap and instant first-line screening for CRC which doesn't require medical examination in the first place. In the future, we would like to include more features in our study, including lifestyle, physical characteristics, smoking history, lab examination, and genetic information that help predict the risk more accurately.

Conclusions

In our study, we developed a deep learning model to predict CRC based on non-image and multi-dimensional medical records (diagnoses and medications). Our model showed higher performance for detecting CRC one year ahead. Indeed, previously published literatures did not show any time frame to predict CRC and traditional methods are not able take time frame consideration for predicting CRC. But, time sequence matters in the real world clinical setting. Therefore, our model can be considered in the real clinical settings.

Acknowledgements

This research was sponsored in part by Ministry of Science and Technology (MOST) 107-2634-F-038-002-, and Ministry of Science and Technology (MOST) 106-2634-F-038 -001 -CC2.
We would like to thank Mr. Chia-Wei Liang for his assistant in this study.

References

[1] T.C.R. Center, Report data, in.
[2] B.K. Chen and C.Y. Yang, Temporal trend analysis of avoidable mortality in Taiwan, 1971-2008: overall

progress, with areas for further medical or public health investment, *BMC Public Health* **13** (2013), 551.
[3] J. Grundner, H.U. Prokosch, M. Sturzl, R. Croner, J. Christoph, and D. Toddenroth, Predicting Clinical Outcomes in Colorectal Cancer Using Machine Learning, *Stud Health Technol Inform* **247** (2018), 101-105.
[4] M.-H. Hsieh, L.-M. Sun, C.-L. Lin, M.-J. Hsieh, K. Sun, C.-Y. Hsu, A.-K. Chou, and C.-H. Kao, Development of a Prediction Model for Colorectal Cancer among Patients with Type 2 Diabetes Mellitus Using a Deep Neural Network, *Journal of clinical medicine* **7** (2018), 277.
[5] T.N.H. Insurance, Top 10 cancer in Taiwan 2008-2014, in.
[6] U. Iqbal, P.A. Nguyen, S. Syed-Abdul, H.C. Yang, C.W. Huang, W.S. Jian, M.H. Hsu, Y. Yen, and Y.C. Li, Is long-term use of benzodiazepine a risk for cancer?, *Medicine (Baltimore)* **94** (2015), e483.
[7] C. Naylor, On the prospects for a (deep) learning health care system, *Jama* **320** (2018), 1099-1100.
[8] P.A. Nguyen, S. Syed-Abdul, U. Iqbal, M.-H. Hsu, C.-L. Huang, H.-C. Li, D.L. Clinciu, W.-S. Jian, and Y.-C.J. Li, A Probabilistic Model for Reducing Medication Errors, *PLOS ONE* **8** (2013), e82401.
[9] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, *The Journal of Machine Learning Research* **15** (2014), 1929-1958.
[10] B. Wahl, A. Cossy-Gantner, S. Germann, and N.R. Schwalbe, Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings?, *BMJ Global Health* **3** (2018).
[11] Y.W. Wang, H.H. Chen, M.S. Wu, and H.M. Chiu, Current status and future challenge of population-based organized colorectal cancer screening: Lesson from the first decade of Taiwanese program, *J Formos Med Assoc* **117** (2018), 358-364.
[12] H.-J. Yang, C.W. Cho, S.S. Kim, K.-S. Ahn, S.-K. Park, and D.I. Park, Application of deep learning to predict advanced neoplasia using big clinical data in colorectal cancer screening of asymptomatic adults, *Gastrointestinal endoscopy* (2018).
[13] M.D. Zeiler, ADADELTA: an adaptive learning rate method, *arXiv preprint arXiv:1212.5701* (2012).

Address for correspondence

Phung-Anh Nguyen, Ph.D.
International Center for Health Information Technology
College of Medical Science and Technology
Taipei Medical University
Address: 250-Wuxing Str., Xinyi Dist., Taipei 11031, Taiwan
Email: alexnthhp@tmu.edu.tw
Phone: +886 266382736 #1528

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190260

Construction of Disease Similarity Networks Using Concept Embedding and Ontology

Duo (Helen) Wei[a]*, Tian Kang[b], Harold Alan Pincus[c,d], Chunhua Weng[b,d]

[a] *Department of Computer Science, School of Business, Stockton University, Galloway, New Jersey, USA*
[b] *Department of Biomedical Informatics, Columbia University, New York, New York, USA*
[c] *Department of Psychiatry, Columbia University, New York, New York, USA*
[d] *Irving Institute for Clinical and Translational Research, Columbia University, New York, New York, USA*

Abstract

Discovering disease similarities are beneficial for the diagnosis and treatment of mental diseases. In this research, we proposed a data driven method, that is, integrating a variety of publicly available data resources including Unified Medical Language System (UMLS) Metathesaurus, Systematized Nomenclature of Medicine -- Clinical Terms (SNOMED CT) and cui2vec concept embedding to construct a mental disease similarity network. The resulting mental disease similarity network offered a new view for navigating and investigating disease relations; it also revealed popular mental disease in the literature in terms of the number of connections and similarities with other diseases. It shows that ***depressive disorder*** *is directly connected with nine other popular diseases and connects 52 other diseases in the network. The top three popular mental diseases are* ***depressive disorder****,* ***dysthymia (now known as persistent depressive disorder)****, and* ***neurosis****. Future research will focus on studying the clusters generated from the similarity network.*

Keywords:

Mental disorders, Disease similarity, Word2Vec, Concept embedding.

Introduction

Disease similarity study promises to enable better understanding of disease relations and to accelerate treatment discoveries based on disease relations.[1, 2]. It can also reveal the pathogenesis of common diseases [1] and help infer the mechanisms of complex diseases [2], which can yield insights into disease etiology and suggest treatment that can be appropriated from one disease to another [3]. Additionally, it provides a starting point for associating diseases to their genes as well as to help find potential targets for drugs [4].

We are motivated to study disease similarity using mental health as an example domain for methodology illustration. Mental disease/illnesses involve the change of people' emotion, thinking and behavior [5]. According to National Alliance on Mental Illness (NAMI), "…*One in five adults experiences a mental health condition every year. One in 17 lives with a serious mental illness such as schizophrenia or bipolar disorder. … Half of mental health conditions begin by age 14, and 75% of mental health conditions develop by age 24...*" Mental disease has become one of the most consequential health concerns in the US and in people across nations, races, and gender nowadays.

This research investigated mental disease similarity by leveraging word embedding, a language modeling technique that transform the vocabulary of an input corpus into low-dimensional vector representation. This leads to our research questions: Is it feasible to use existing literature and research repositories to explore the most popular mental disease topics? Would it be possible to create a mental disease similarity network so that relations of mental diseases could be identified, and similar mental diseases could be grouped together?

The objectives of this research are to investigate mental disease similarity and analyze similarity network. In particular, we intent to utilize the semantics of the mental disease concepts in the knowledge base and leverage data driven based method to analyze the closeness (similarities) among concepts extracted from the knowledge base.

Background

Related work for disease similarities

Various methods have been defined to measure or compute disease similarities. In [6], a method for measuring disease similarity has been proposed that integrates semantics and gene functional association, where two functions, one calculating functional similarity and the other calculating semantic similarity, are combined to measure disease similarity. In [2], a matrix-based method was introduced to measure disease similarity that incorporates the uniqueness of shared genes. For each disease pair, the uniqueness score was calculated, and disease similarity matrices were constructed using Online Mendelian Inheritance in Man (OMIM) and Disease Ontology annotation. An ontology-based disease similarity network was described in [7] for disease gene prediction based on semantic similarity measures on phenotype ontology database. In [4], an online system called DisSim was introduced for exploring similar diseases. The system implemented five state-of-the-art methods to measure the similarity between Disease Ontology (DO) terms and provided the significance of the similarity score. Some other disease similarities methods include International Classification of Disease (ICD) based disease similarity, probability-based disease similarity, and machine learning-based approaches.

Word embedding and concept embedding

Word embedding is an emerging technique that transforms a word from the literature (text corpus), to continuous decimal and low dimensional vectors based on the context of the word. There are two popular word embedding algorithms: word2vec and GloVe. Word2vec was first developed by a team of researchers led by Tomas Mikolov at Google [8]. The word2vec utilized two model architectures for computing continuous vector representations of words from very large data sets: the skip gram model and the continuous bag-of-words (CBOW) model, where the former uses the current word to predict the surrounding context words and the latter predicts the current word from surrounding context words [8]. The softmax

function was applied to calculate the probability of a word w given its context c:

$$p(w|c) = \frac{\exp(h^\top v'_w)}{\sum_{w_i \in V} \exp(h^\top v'_{w_i})}$$

Another embedding algorithm is a "count-based" model called GloVe proposed by Pennington [9] that tabulates how frequently words co-occur with one another in a given corpus that offered a way to train over larger scale of data.

As healthcare data come in a variety of forms, the aforementioned word embedding algorithms like word2vec and GloVe that were originally developed for text cannot be directly applied to many kinds of healthcare data. In this research, we leveraged medical concept embedding referred to as *cui2vec* developed by Harvard [10] which links word2vec to traditional count-based methods that are based on co-occurrence statistics. The pre-trained embedding *cui2vec* utilized an extremely large collection of multimodal medical data including insurance claims for 60 million Americans, 1.7 million full-text PubMed articles, and clinical notes from 20 million patients at Stanford [10]. The *cui2vec* contains embedding for 108,477 medical concepts.

Methods

Research pipeline

In the context of this paper, we assume the disease similarity is disease concept similarity without considering about biological, molecular, genetic information of the disease.

To calculate mental disease similarities, we need to retrieve as many mental disease concepts in computable format from the literature as possible. The Diagnostic and Statistical Manual of Mental Disorders (DSM), a widely used diagnostic manual that offers standard criteria for the classification of mental disorder, has included over 450 different definitions of mental diseases. For each disorder, there is a set of diagnostic criteria associated with specific symptoms that characterize the disorder. However, the definitions of the diseases on DSM are very subjective, the classifications are too vague, and the descriptive codes are not in computable format, which makes it impractical to be used for computational purpose.

The clinical terminologies or ontologies like The Systematized Nomenclature of Medicine Clinical Terms (SNOMED for short) can serve as a valuable resource to expand the search queries so that more mental disease concepts at varying levels of granularity can be captured. SNOMED is a standard clinical terminology system with more than 350,000 concepts released twice every year on January and July respectively. In this research, we retrieved SNOMED concepts under the Mental disorder hierarchy for three reasons: First, the total number of SNOMED mental disease concepts are approximately four times the size of the DSM codes and nine times the size of ICD listed concepts. Second, as we use *cui2vec* that utilized 500-dimension vectors (see Step 1) to represent a UMLS unique identifier (CUI), SNOMED ConceptIDs could be easily mapped to UMLS CUIs, so that we can apply pre-trained embedding to represent semantic similarities of mental disease concepts. Lastly, SNOMED is both a clinical vocabulary and an ontology. Due to SNOMED's broader concept coverage and richer relations, not only can we extract mental disease related concepts, but we can also capture the hierarchical relationships among them.

The major steps of our research pipeline are illustrated in Figure 1.

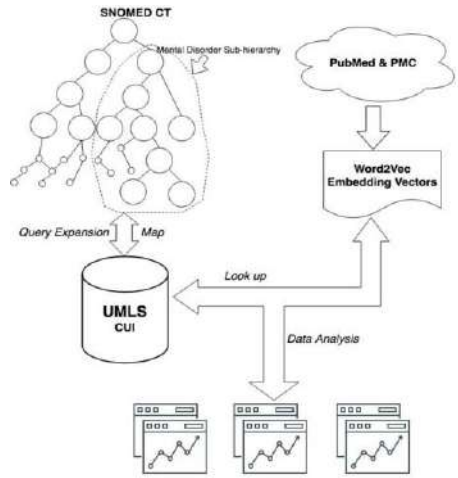


Figure 1. Illustrates the major steps for the research project.

Step 1. Identifying the mental disease concept hierarchies and mapping the extracted SNOMED concepts to UMLS CUI. In this step, we map SNOMED ConceptID, which is SCUI in MRCONSO of the UMLS Metathesaurus Rich Released Format to UMLS CUI.

Step 2. Matching the UMLS CUIs against the pre-trained *cui2vec* embedding vectors.

Step 3. Calculating pairwise disease similarities among the selected diseases based on cosine similarity.

$$sim\,(i,\,j) = cos\,(\theta) = \frac{vec(CUIi)\;\; vec(CUIj)}{\|vec(CUIi)\|\;\|vec(CUIj)\|}$$

$$= \frac{\sum_{k=1}^{n} veck(CUIi)\;\; veck(CUIj)}{\sqrt{\sum_{k=1}^{n} veck(CUIi)^2}\sqrt{\sum_{k=1}^{n} veck(CUIj)^2}}$$

Gensim, a free Python library, is used to calculate the cosine similarity with the results ranging from -1 to 1, where 1 means extremely similar and -1 means the opposite.

Step 4. Visualizing the disease similarity network based on the pairwise disease similarities and analyzing the similarity network. All networks are visualized using Cytoscape's Biolayout[11].

Mental disease similarity network analysis

We developed a network based on pairwise similarities among mental disease concepts. Let us assume that a mental disease concept is a vertex (a node) and an edge (a link) between two vertices represents a relation between the two concepts. In this context, relation refers to similarity. The value associated with each edge is the similarity value, where the higher value it is the more similar the two concepts are. The mental disease similarity network can be considered as graph G (denoted as G=(V,E)), consisting of a non-empty set of vertices V and a set of edges E.

As one of the research Methods in this study is to use literature and research repositories to find the most popular mental disease topics, we have defined the term "popular topics" (PT). Let us first define "Degree": The degree of a vertex *i* is a count of number of edges the vertex tie to other vertices, which can be denoted by D(*i*).

Now let us define the popular topics (PT). The term "popularity" is borrowed from the Latin term popularis, which originally meant "common." In a social network, the popularity of a person can be thought of as being liked or having influence on other people. According to Wikipedia, with respect

to interpersonal popularity, there are two primary divisions: sociometric popularity and perceived popularity. The former means how liked an individual is and the latter means how well known among their peers as being popular.

In the context of the mental disease similarity network, popular topics take into account both "sociometric" and perceived popularity. We use similarity to represent the sociometric popularity and use degree to represent the perceived popularity.

Popular topics in the graph are considered as all vertices in the disease similarity network whose similarity values with other vertices are no less than a threshold and the degrees of the vertices are also no less than a threshold. Here the "popularity" of a topic can be considered as a disease concept is that not only closely related to other concepts (high similarity value) but also possessing more connections to other concepts (high degree value). In a symbolic form, it can be represented as a set of vertices i as follows:

$$\{\forall i(\forall j\,(sim\,(i,\,j) > \theta) \rightarrow D\,(i) > \delta)\} \quad \text{where } i,\, j \in N$$

We use a popularity value to measure a popular topic. If a vertex i is a popular topic in the disease similarity network, then the popularity of a vertex i, denoted by *popularity*(i), can be defined as the aggregated similarity values of a vertex i with all vertices associated with the vertex i.

$$popularity(i) = \sum_{j=1}^{D(i)} sim(i, j)$$

Results

Using *cui2vec* to calculate disease similarities

We retrieved 1,756 concepts from the "Mental Disorder" subhierarchies of Clinical Finding hierarchy. Note that this research only concentrates on the concept level without considering about the synonyms or term level descriptions.

The retrieved ConceptIDs were mapped against the UMLS CUIs using the UMLS MRCONSO Rich Released Format. As each source SNOMED concept may be mapped to more than one CUIs (i.e. the SNOMED concept **Dissociative disorder** has been mapped to two UMLS CUIs C0020701 and C0020703), the Mental disorder concepts of SNOMED have been mapped to 1,768 UMLS CUIs.

After that, we matched the 1,768 UMLS CUIs with the *cui2vec* and we obtained 401 unique CUIs out of the 109,053 unique UMLS CUIs where each CUI is represented by 500 dimensions continuous decimal numbers. The Gensim package in Python is utilized to calculate similarities of a pair of disease concepts represented by CUI embedding vectors. For example, the CUI for **Hypomania** is C0241934 and the CUI for **Mania** is C0338831, and the similarity value is sim(*Vec*(C0241934), *Vec*(C0338831)) = 0.9901.

The results for similarities of all pairs of CUIs are summarized in Table 1. We categorized the disease pairs based on their similarity values. There are five major categories as illustrated in Table 1. Altogether, we obtained 80,200 pairwise similarity values out of the 401 unique CUIs based on cosine similarity measures (see **Method** section).

Mental disease similarity network

A mental disease similarity network was developed based on pairwise similarities (See **Method** section), where each disease concept was represented by a vertex (i.e. V = {**Depressive disorder**, **Hypomania**, **Mania**}) and an edge connecting between two vertices represented that there is a similarity value no less than certain threshold θ (i.e. θ = 0.9, E = {{**Hypomania**, **Mania**}, {**Severe major depression with psychotic features,**

Table 1– Similarity values and example pairs of mental diseases

Similarity value range	Number of pairs	Description	Example pairs
[0.9, 1)	488	Extremely similar	(hypomania, mania)
[0.8, 0.9)	1,432	Very similar	(identity disorder, moderate depression)
[0.5, 0.8)	9,915	Weakly similar	(asperger's disorder, psychoactive substance-induced withdrawal syndrome)
[0.0, 0.5)	59,820	Similar	(alzheimer's disease, developmental articulation disorder)
<0	8,545	Connect	(receptive language delay, acute depression)
			In total, there are 80,200 pairs

Mania}, ……}). Figure 2 presented a partial mental disease similarity network. In total, there are 252 vertices and 1,920 edges. In addition, we can easily observe that some disease concepts are naturally grouped together which offered a new view of among the same kind of disease.

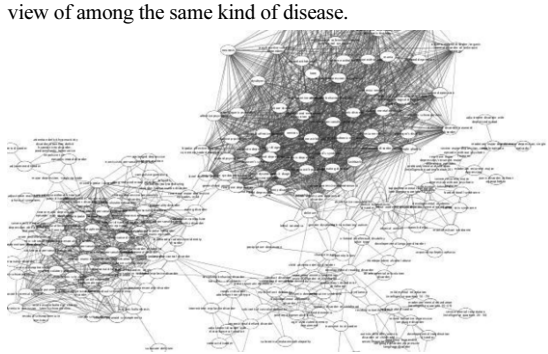


Figure 2. Partial mental disease similarity network θ = 0.8

Furthermore, we explored the popular topics in the literature (PT definition can be found in **Method** section). Table 2. showed top ten popular topics in the mental disease literature. As is seen in Table 2, the most popular mental disease topic is **Depressive disorder** (with popularity value = 121.7807), a synonym of **Depression**, which is a common but serious mood disorder causing a person persistent feeling of sadness and loss of interest in things that used to bring pleasure. **Depressive disorder** listed as the top one popular topic indicates that **Depressive disorder** possess most connections with other mental illnesses. This is due to the prevalence of depressive disorder and homogeneity of subthreshold form of depression [12]. The finding is also confirmed by WebMD that states "Depression can be triggered by other mental illnesses, but it can also lead to certain mental illnesses" and "Depression's link to 9 other mental illnesses." The UPMC released five major categories of mental illnesses: Anxiety disorder, Mood disorders, Schizophrenia and psychotic disorder, Dementia, and Eating disorder. As Table 2 is shown, the majority of the top ten PTs can be categorized into Mood disorders; **Psychoactive substance dependence** is associated with **Anxiety disorder**; **Neurosis** is characterized by both anxiety and depression.

Table 2. popular topics with $\delta = 50$ and $\theta = 0.8$

Disease Names	*popularity*(*i*)	CUI(s)
Depressive disorder	121.7807	C0025193/C0349217/ C0011581/C0344315/ C1269683
Dysthymia	88.4942	C0282126/C0011581/ C0013415/C0221508
Hypomania	46.0239	C0241934
Neurosis	45.5589	C0027932
Bipolar I disorder	45.284	C0853193
Mania	45.0888	C0338831
Severe depression	44.8137	C0588008
Reactive depression (situational)	44.5026	C0011579
Psychoactive substance dependence	44.3936	C0038580/C1510472
Agitated depression	44.3462	C0235136

Figure 3 illustrated an excerpt figure from Figure 2, which demonstrated the top ten most popular topics in mental disease similarity network along with the connections with other mental diseases. It shows that **depressive disorder** is directly connected with nine other popular diseases and in fact it connects 52 other diseases in the network.

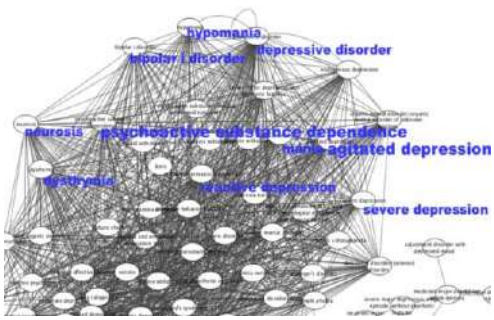


Figure 3. Excerpt mental disease similarity network for most popular topics mental diseases

Discussion

There are two major contributions in this research. First, we initially propose to integrate word embedding results and disease ontologies such as SNOMED-CT to generate a disease similarity network. Second, we defined the most popular topic based on the disease similarity network. The similarity network could serve as a starting point to navigate complex disease network and could be used to detect popular topics in literature.

Word2Vec to explore disease similarity

We create a mental disease similarity network integrating public data including Harvard *cui2vec*, UMLS *Metathesauru*, and SNOMED-CT. The proposed research pipeline can be generalized to build other disease related network.

The concept embedding captured the semantics of medical concepts in the literature and offered opportunity to study the underlying relations among disease concepts, which forms the foundation of the disease similarity network. Using the mental disease similarity network, we can observe that biologically and genetically related mental diseases are often grouped together, indicating that similar disease also often being studied together in the literature.

Our results have been reviewed by a psychiatrist, one of the co-author of the paper (HP). We obtained some interesting observations. For "Extremely similar" pairs, like **Hypomania** and **Mania** with similarity value of 0.9901, **Hypomania** is in fact a less severe version of **Mani**a. **Bipolar affective disorder, current episode depression** and **Depressive disorder** has similarity value of 0.9897 because bipolar patients who are in the midst of a depressive episode, by definition meet criteria for major depressive disorder. We can also observe that some mental disorders with very high similarity values are very similar conditions. For example, **Bulimia nervosa** and **Binge eating disorder**. Lastly, some share very similar symptoms or may co-occur. For example, **Panic disorder** and **Simple phobia (**with similarity value of 0.9663). Both conditions share same symptoms including intense fear, feelings of anxiety, and panic attacks. According to information found in the DSM-5, both conditions are classified as "anxiety disorders." However, panic disorder and phobias are considered separate conditions, each with a distinct set of diagnostic criteria.

Interestingly, some of Mental disease concepts pairs are negative related called "Connect", for example, **Receptive language delay** and **Acute depression** (with similarity value = -0.956). However, having negative similarity do not necessarily mean they are not related. They coexist but with negative similarity value may but not limited due to the following factors: (1) The two disease conditions may occur at different stage of one's life. For example, **Receptive language delay** and **Acute depression** with the former occurring in very young children. Another example, **Separation anxiety disorder of childhood** and **Schizoaffective disorder, bipolar type,** with the former only take place in children. One might speculate that separation anxiety during the childhood probably may cause schizophrenia in adulthood. (2) Two disease conditions are in different categories but can co-occur. For example, **Sexual relationship disorder** and **Acute depression**. (3) Two diseases have very different levels of severity. For example, **Adjustment disorder with mixed disturbance of emotions and conduct** and **Schizoaffective disorder, bipolar type** with the latter much severe than the former (4) Using variant terms to represent the same condition. For example, **Developmental receptive language disorder** and **Receptive language delay**. There are quite a few pairs with negative similarity values just because it makes no sense to put them together. For example, **Schizoaffective disorder, bipolar type** and **Facial tic disorder**.

Several mental diseases pairs with very high similarity values do not make sense to be together like **Nondependent alcohol abuse** and **X-linked intellectual disability atkin type** (similarity value = 0.968). It may imply that in the literature they are typically being studied together, which actually indicates a potential for future research studies. A number of mental diseases form a natural cluster implying that they potentially share common genes, proteins and chemicals. For instance, schizophrenia, bipolar disorder, autism, ADHD and depression share common gene *CACNA1C* and *CACNB2* [13].

The research pipeline can be used to assist pathologist or physicians to predict potential genetic relatedness of diseases, before rigorous experiment is conducted. It can also serve as a complementary method for existing disease similarity analysis.

Limitation and future work

Due to the granularity of the SNOMED concepts and the lack of occurrence of the CUI embedding, only less than a third of the SNOMED concepts can find the exact match in the *cui2vec* embedding, which limited our research scope.

The results of the disease similarity network highly depend on the quality of the concept embedding. The *cui2vec* embedding is based on the statistics of occurrences of medical concepts, and thus less frequently appeared concepts might not have their corresponding CUI embedding. Future research can explore

word embedding generated from other resources like electronic health records and derive the similarity network.

Besides, as the disease similarity network is developed based on concept embedding which is generated out of the existing literature, it is not applicable to use the disease similarity network for unknown/new disease discovery. Therefore, it is desirable to have more high quality and high medical concept coverage embedding publicly accessible.

Another limitation of the research is that due to many-to-many characteristics of the mapping process, which may affect the accuracy of the popular topic result. For example, each UMLS CUI may have multiple SCUIs, that is, each CUI can be mapped to more than one SNOMED concepts (i.e. the UMLS CUI C0025362 can be mapped to two SNOMED concepts: **Development academic disorder** and **Mental retardation**), the 401 unique CUIs in the vocab for Mental disorder can be mapped to 406 SNOMED concepts. In this research, we merged concept names if they share the same CUIs. For example, two SNOMED concepts **Delusional disorder** and **Paranoid disorder** share the same CUI C1456784. In the experiment, we merge them as one concept "**Delusional disorder / Paranoid disorder**". The merger of the concepts may affect the accuracy of the results.

As the development of disease similarity network, it is feasible to analyze disease clusters based on the similarity network. We can also interpret underlying meaning of the "centrality" of the cluster and the "connected components" among cluster. It is interesting to compare existing disease classification with the generated clusters. As many joint symptoms may appear for two different diseases, which may result in diagnostic mistakes. In the future, we may analyze the possibilities of using the similarity network to predict such situations.

At the same time, the disease similarity network reveals the potential to enrich existing ontologies as we observed a number of hierarchical relationships, such as **Bipolar affective disorder, current episode depression** is-a **Depressive disorder**.

Conclusions

This research integrates ontological method and machine learning method, concept embedding in particular, to develop a disease similarity network and observed several interesting findings regarding mental disease similarities. Popular topics are identified based on the similarity network. The top three popular topics are **depressive disorder**, **dysthymia (now known as persistent depressive disorder)**, and **neurosis**. Part of the findings are confirmed by the literatures or domain experts.

Acknowledgements

This project was funded by National Library of Medicine grant R01LM009886 (PI: Weng) and National Center for Advancing Clinical and Translational Science grant 5UL1TR001873-03 (PI: Reilly). This project was also funded by Stockton University Research and Professional Development grant and Stockton sabbatical subvention grant.

References

[1] Y. Hu, M. Zhou, H. Shi, H. Ju, Q. Jiang, and L. Cheng, "Measuring disease similarity and predicting disease-related ncRNAs by a novel method," *BMC medical genomics,* vol. 10, no. Suppl 5, pp. 71-71, 2017.

[2] M. B. Carson, C. Liu, Y. Lu, C. Jia, and H. Lu, "A disease similarity matrix based on the uniqueness of shared genes," *BMC medical genomics,* vol. 10, no. Suppl 1, pp. 26-26, 2017.

[3] S. Mathur and D. Dinakarpandian, "Finding disease similarity based on implicit semantic similarity," *Journal of Biomedical Informatics,* vol. 45, no. 2, pp. 363-371, 2012/04/01/ 2012.

[4] L. Cheng *et al.*, "DisSim: an online system for exploring significant similar diseases and exhibiting potential therapeutic drugs," *Scientific reports,* vol. 6, pp. 30024-30024, 2016.

[5] D. J. Stein, K. A. Phillips, D. Bolton, K. W. M. Fulford, J. Z. Sadler, and K. S. Kendler, "What is a mental/psychiatric disorder? From DSM-IV to DSM-V," *Psychological medicine,* vol. 40, no. 11, pp. 1759-1765, 2010.

[6] L. Cheng, J. Li, P. Ju, J. Peng, and Y. Wang, "SemFunSim: a new method for measuring disease similarity by integrating semantic and gene functional association," *PloS one,* vol. 9, no. 6, pp. e99415-e99415, 2014.

[7] D.-H. Le and V.-T. J. V. J. o. C. S. Dang, "Ontology-based disease similarity network for disease gene prediction," journal article vol. 3, no. 3, pp. 197-205, August 01 2016.

[8] T. Mikolov, Chen, K., Corrado, G.,Dean, J., "Efficient Estimation of Word Representations in Vector Space," arXiv:1301.3781 2013.

[9] J. Pennington, Socher, R., and Manning, C.D. . (2014). *GloVe: Global Vectors for Word Representation*. Available: https://nlp.stanford.edu/projects/glove/

[10] A. L. Beam, Kompa, B. , Fried, I. , Palmer, N. P., Shi, X. , Cai, T. , and Kohane, I. S. (2018). *Clinical Concept Embeddings Learned from Massive Sources of Medical Data*. Available: http://cui2vec.dbmi.hms.harvard.edu/

[11] P. Shannon *et al.*, "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome research,* vol. 13, no. 11, pp. 2498-2504, 2003.

[12] H. A. Pincus, W. W. Davis, and L. E. McQueen, "'Subthreshold' mental disorders: A review and synthesis of studies on minor depression and other 'brand names'," *British Journal of Psychiatry,* vol. 174, no. 4, pp. 288-296, 1999.

[13] C. Cross-Disorder Group of the Psychiatric Genomics, "Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis," *Lancet (London, England),* vol. 381, no. 9875, pp. 1371-1379, 2013.

Address for correspondence

Duo (Helen) Wei,

Associate Professor of Computer Science,

School of Business, Stockton University,

101 Vera King Farris Dr., Galloway, NJ 07751.

Email: duo.wei@stockton.edu

Phone: 001-609-626-3813

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190261

Process Mining in Primary Care: Avoiding Adverse Events Due to Hazardous Prescribing

Richard Williams[a], Darren M. Ashcroft[a], Benjamin Brown[a], Eric Rojas[b], Niels Peek[a], Owen Johnson[c]

[a] NIHR Greater Manchester Patient Safety Translational Research Centre, University of Manchester, Manchester, UK
[b] Internal Medicine Department. School of Medicine. Pontificia Universidad Católica de Chile
[c] NIHR Yorkshire and the Humber Patient Safety Translational Research Centre, University of Leeds, Leeds, UK

Abstract

Process mining helps healthcare professionals understand processes within healthcare. While often used in secondary care, there is little work in process mining using primary care data. Serious adverse events that result from hazardous prescribing are common and costly. For example, non-steroidal anti-inflammatory drugs (NSAIDs) and antiplatelets can cause gastro-intestinal bleeds (GiBs). Prescribing typically occurs during primary care; therefore we used this setting to attempt process mining. We extracted events (drug started, drug stopped, GiB) for understanding three prescribing pathways, and applied process mining. We found NSAIDs are often short-term prescriptions whereas antiplatelets are often long-term. This perhaps explains our finding that co-prescription of gastro-protection is more prevalent for antiplatelets than NSAIDs. We identified reasons why primary care data is harder to process mine and proposed solutions. Process mining primary care data is possible and likely useful for improving patient safety and reducing costs.

Keywords:
Data mining, Patient safety, Primary care.

Introduction

Process mining describes a collection of methods for extracting information about processes from event logs [1]. There are three distinct stages: detecting the underlying process from the event logs (process discovery); identifying deviations from what was expected (conformance checking); and generating suggestions for redesigning and improving the processes (enhancement) [1]. Process mining adds a temporal dimension to standard data mining methods. Originally applied to business processes, more recently has been applied to other domains including healthcare. In a recent literature review, we showed that while process mining within secondary and tertiary care has become more common, there is almost no work within primary care [2].

Patient safety is fundamental to healthcare systems. Within UK primary care this is true for medication prescribing where life-threatening errors appear in 1 in 550 prescriptions [3]. A recent economic analysis showed that: adverse drug reactions (ADRs) cost the NHS up to £1.6 billion a year; more than one third of ADR related hospital admissions are caused by non-steroidal anti-inflammatory drugs (NSAIDs), antiplatelets and anticoagulants; and half of the deaths associated with primary care ADRs are due to gastro-intestinal bleeds (GiBs) [4]. Studying the relationship between the prescribing practice of NSAIDs and antiplatelets with ADRs including GiBs is therefore important.

In the UK there are several large databases of coded primary care records available for research [5]. While the quality of coding may not be universally high [6], all practice-based prescribing in primary care is electronic so therefore this would be a suitable place to attempt to apply process mining.

While the epidemiology of hazardous prescribing in primary care has been extensively studied using large electronic health record databases, to date little is known about the typical processes that lead to such prescriptions. To design effective interventions for reducing hazardous prescribing, it is essential to get a better understanding of these processes. This could lead to better decision support systems for prescribers, and ultimately improve patient safety and reduce cost by reducing the number of ADRs.

Our objective was therefore to process mine UK primary care data to explore the relationship between the prescribing of NSAIDs, antiplatelets and the adverse outcome of gastro-intestinal bleeds.

Methods

A process model is a graphical representation of a process showing the events and how they interrelate via directed edges. Process discovery is the extraction of a process model from an event log via the application of an algorithm. There are many algorithms with various strengths and weaknesses. For example the α-algorithm is simple and therefore easy to undersand, but it does not deal well with noisy event logs which are typical of real world processes [1]. Heuristic Miner and, in particular, Fuzzy Miner are better able to deal with this noise [7]. Here we focus on process discovery to prove that process mining can be applied to primary care data, and use the Fuzzy Miner to best handle the messiness of routinely collected health data.

Anonymised patient data was obtained from the Salford Integrated Record (SIR); a data warehouse with contributions from 43 general practices in Salford, UK (population 0.25M). All coded data, including diagnoses and medications, for patients who have not opted out (1.5% opt outs) was available to extract from a SQL Server database. The earliest records are historic diagnoses from the 1940s, but the bulk of the data collection is from 2000 onwards. Approval was granted by the SIR governance board and all data was obtained pseudonymised (random identifier, no name, year of birth instead of age, geographic region instead of address).

A review by Spencer *et al.* [8] identified 56 prescribing safety indicators for use in primary care to improve patient safety. They each try to prevent a particular adverse outcome through safer prescribing, e.g., patients with chronic kidney disease should not be prescribed an NSAID because of the increased risk of acute renal failure. A subset of these indicators are

included in: electronic audit and feedback initiatives such as the national PINCER [9] rollout and the SMASH intervention [10]; and clinical decision support systems such as OptimizeRx [11]. We selected three prescribing safety indicators for further analysis that focus on NSAIDs and anticoagulants, and are designed for preventing GiBs in cohorts of patients at increased risk such as the elderly and those with a history of peptic ulceration. The indicators and the descriptions used in our analyses are provided in Table 1.

Table 1– Prescribing Safety Indicators for Preventing GiBs

Id	Short name	Description
I1	Age≥65 + NSAID	Patients aged 65 and over who are prescribed an NSAID should also be prescribed a gastro-protective medication (GPM).
I2	Pep + NSAID	Patients with a history of peptic ulceration who are prescribed an NSAID should also be prescribed a GPM.
I3	Pep + Antiplatelet	Patients with a history of peptic ulceration who are prescribed an antiplatelet should also be prescribed a GPM.

Prescription events are recorded automatically in a patient's record; however, the stopping of medication is not recorded. We have previously developed an algorithm to convert these prescription events into more meaningful events such as when a drug is started and stopped, and when a dose is changed [12]. This process is done by evaluating: the date of the prescription; the amount prescribed; and the rate at which it is consumed. For the medications of interest (NSAIDs, GPMs, and antiplatelets), we extracted the start and stop events.

For each indicator, we developed queries that would extract the patient data. First, we defined the cohort of patients from an initial event. For indicator I1, it was when a patient turned 65, and for I2 and I3, it was the first instance of a peptic ulceration. Next, we extracted the start and stop events for all medications of interest. Finally, we extracted other relevant events: GiB and peptic ulceration classified as either the first bleed, or a subsequent bleed; patient turned 65; and patient died. Clinical code sets were constructed for each event of interest [13].

The initial output of our medication algorithm gives the start and stop events for individual active ingredients so, for example, for two different NSAIDs we would have two different start events. However, we are only interested in whether any NSAID (or other medication) is started if the patient is not already taking an existing one. Similarly, when a drug is stopped it is only relevant if the patients are then not taking any other drug of the same type. An additional processing script was therefore required to produce the final event log. This additional data processing was done using JavaScript and nodejs [14].

The data was extracted on 2nd November 2018 and process mining was performed by the lead author. Process mining was performed using Fluxicon Disco (academic licence) [15] on a Dell XPS 15 laptop running Windows 10. All clinical code sets and processing code is at https://zenodo.org/record/1493640.

Results

The demographic information for the patient cohorts for each indicator are displayed in Table 2. The median duration time, interquartile range, and number of transitions between events are shown in Tables 3-5. For example, in Table 3, the event "Bleed" immediately followed the event "Age 65" in the event logs 390 times, with a median transition time of 49 (IQR [19,112]) months. The process mining diagrams extracted from Disco are shown in Figures 1-3. The numbers on the nodes in the diagrams represent the number of times each event occurred, while the edge numbers are the number of times the target event directly followed the source event.

Table 2– Patient Characteristics for Each Cohort and the Population of Salford. Values are n (%) unless otherwise specified.

Demographic	I1		I2, I3		All patients	
# of patients	38,936		3,477		270,412	
Age						
Mean (SD)	76	(8)	66	(16)	37	(23)
Sex						
Female	20,633	(53%)	1,238	(36%)	131,935	(49%)
Male	18,303	(47%)	2,239	(64%)	138,473	(51%)
Ethnicity						
White	16,291	(42%)	1,444	(42%)	96,696	(36%)
Other	643	(2%)	139	(4%)	24,124	(9%)
Unknown	22,002	(57%)	1,894	(54%)	149,592	(55%)
Deprivation [16] quintiles						
1st (highest)	15,023	(39%)	1,618	(47%)	115,593	(43%)
2nd	8,126	(21%)	694	(20%)	56,284	(21%)
3rd	7,536	(19%)	569	(16%)	44,770	(17%)
4th	3,533	(9%)	258	(7%)	19,591	(7%)
5th (lowest)	4,320	(11%)	269	(8%)	19,561	(7%)

Table 3– Median Duration in Months of Transitions between Key Events for Indicator I1. IQR in [square] brackets. Number of transitions in (round) brackets. NSAID – non-steroidal anti-inflammatory drug, GPM – gastro-protective medication.

Next event / Event	Bleed	NSAID (no GPM)	NSAID (GPM)	NSAID Stopped	GPM Started	GPM Stopped
Age 65	49 [19,112] (390)	33 [11,74] (10925)	17 [6,39] (1385)	9 [2,28] (1684)	48 [15,111] (9366)	12 [3,37] (1617)
Bleed	1 [0,5] (523)	14 [3,47] (48)	22 [7,46] (75)	1 [0,3] (54)	0 [0,3] (582)	10 [2,34] (276)
NSAID (no GPM)	6 [1,36] (34)			1 [1,3] (24106)	0 [0,0] (9478)	
NSAID (GPM)	2 [1,32] (11)			1 [1,3] (8733)		1 [1,5] (566)
NSAID Stopped	20 [5,57] (292)	11 [5,26] (14923)	10 [5,22] (4553)		18 [5,46] (6913)	0 [0,5] (7769)
GPM Started	5 [1,23] (483)		9 [3,27] (3966)	1 [1,2] (6548)		1 [1,5] (28868)
GPM Stopped	13 [3,38] (231)	11 [5,25] (5719)		1 [1,4] (2724)	6 [2,14] (21444)	

For indicator I1 there were 45,479 NSAID start events. Of these, 9,981 (22%) were for patients already prescribed a GPM. A further 9,478 (21%) then started GPM at a median duration of 0 days suggesting co-prescription. However, 24,106 (53%) NSAID start events were followed by an NSAID stop event at a median duration of 1 month (IQR [1, 3] months), suggesting a short term prescription without co-prescription of a GPM.

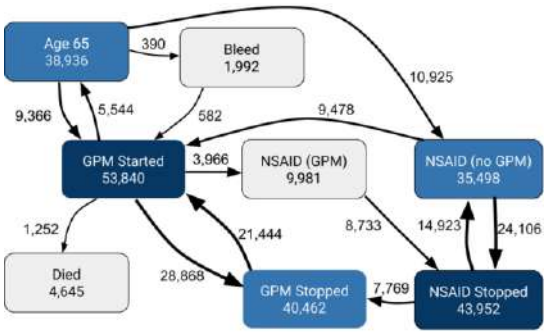


Figure 1– Process Diagram for Indicator I1. NSAID – non-steroidal anti-inflammatory drug, GPM – gastro-protective medication.

Table 4– Median Duration in Months of Transitions between Key Events for Indicator I2. IQR in [square] brackets. Number of transitions in (round) brackets. NSAID – non-steroidal anti-inflammatory drug, GPM – gastro-protective medication.

Next event / Event	Bleed	NSAID (no GPM)	NSAID (GPM)	NSAID Stopped	GPM Started	GPM Stopped
Initial Bleed	3 [0,60] (603)	154 [46,283] (564)			2 [1,71] (2102)	
Bleed	2 [1,7] (752)	57 [17,107] (125)	18 [4,52] (88)	1 [0,8] (21)	1 [0,24] (729)	5 [1,24] (299)
NSAID (no GPM)	1 [0,6] (14)			1 [1,2] (2388)	0 [0,1] (861)	
NSAID (GPM)	5 [1,27] (11)			1 [1,3] (1416)		2 [1,5] (112)
NSAID Stopped	15 [4,38] (157)	14 [7,30] (1437)	11 [6,25] (695)		13 [4,36] (910)	1 [0,11] (912)
GPM Started	4 [1,19] (501)		13 [4,33] (819)	1 [1,3] (630)		2 [1,7] (6884)
GPM Stopped	9 [3,27] (226)	13 [6,33] (1147)		1 [0,3] (305)	6 [3,14] (5551)	

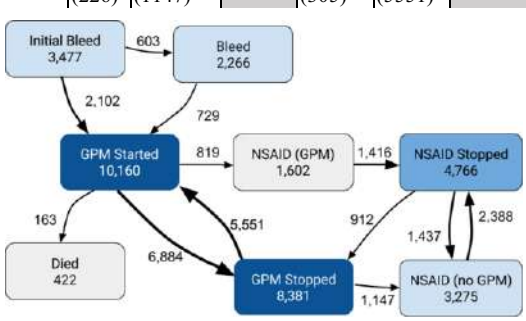


Figure 2– Process Diagram for Indicator I2. NSAID – non-steroidal anti-inflammatory drug, GPM – gastroprotective medication.

Similar results were found in indicator I2 when out of 6,368 NSAIDs, 25% (1,602) were for patients with a pre-existing GPM, 14% (861) were followed almost instantly by a GPM, and 38% (2,388) were for a short-term prescription without a GPM. The results for indicator I3 suggest that GPMs are more frequently co-prescribed with APs with only 22% of AP start events (513 out of 2309), neither having a pre-existing GPM or immediately followed by a GPM.

The event most likely to precede a GiB or peptic ulceration is a previous GiB or peptic ulceration. This is to be expected as it is known that a strong predictor of gastro-intestinal adverse events is a previous bleed or ulceration.

Table 5– Median Duration in Months of Transitions between Events for Indicator I3. IQR in [square] brackets. Number of transitions in (round) brackets. AP – antiplatelet, GPM – gastro-protective medication.

Next event / Event	Bleed	AP (no GPM)	AP (GPM)	AP Stopped	GPM Started	GPM Stopped
Initial Bleed	3 [0,77] (621)	78 [1,287] (390)			4 [1,120] (2197)	
Bleed	2 [1,7] (758)	39 [2,122] (81)	12 [4,54] (52)	11 [2,40] (59)	1 [0,35] (761)	5 [1,24] (286)
AP (no GPM)	7 [1,36] (33)			2 [1,7] (451)	0 [0,6] (851)	
AP (GPM)	7 [2,24] (51)			5 [1,25] (482)		8 [2,31] (160)
AP Stopped	7 [1,23] (30)	6 [2,13] (326)	6 [2,14] (353)		8 [3,29] (181)	0 [0,6] (480)
GPM Started	5 [1,25] (525)		14 [2,47] (540)	7 [2,26] (445)		2 [1,7] (7337)
GPM Stopped	10 [3,33] (246)	10 [2,31] (566)		4 [1,21] (144)	7 [4,18] (6165)	

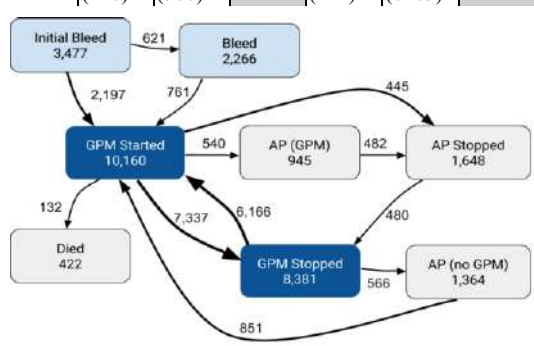


Figure 3– Process Diagram for Indicator I3. AP – antiplatelet. GPM – gastro-protective medication.

The median duration of NSAID and GPM prescriptions is 1 month, suggesting that these medications are typically short-term. Antiplatelets are prescribed at a median length of 5 months and 2 months for patients with and without a pre-existing GPM respectively, suggesting longer term prescriptions.

Discussion

Summary of findings and comparison to existing literature

Little published work on process mining in primary care exists [2]: Dagliati et al. [17] used primary care data to investigate care pathways related to cardiovascular risk of Type II diabetes patients. However, the majority of their data was obtained from secondary care. Another paper used primary care data, but didn't report any results [18]. A further 4 papers used insurance data [19–22] which probably included primary care data, but also secondary care and tertiary care data. Also, the level of data included in insurance datasets is different to that which is routinely collected in primary care for the provision of care. To the best of our knowledge, the process mining performed for this paper is the first performed exclusively using primary care electronic health data.

A GPM was more likely to be co-prescribed to patients receiving a course of antiplatelets than it was to those receiving NSAIDs. The difference in prescription lengths is one possible explanation. When a clinician prescribes a short-term NSAID course, perhaps in response to an acute injury or minor illness, he/she may decide the risk is small enough that co-prescription of a GPM is unnecessary. However, when prescribing a longer-term course, the risk is increased. This might also be true for those on longer courses of NSAIDs for chronic pain conditions. Stratifying medications depending on whether they are short or long term might give further insight into clinicians' behaviour.

GPMs such as proton pump inhibitors can be prescribed for a variety of reasons. For treating an active bleed, to manage the symptoms of gastrointestinal irritation of reflux, or prophylactically for patients at high risk of a bleed – especially when increased because of other medications. Attempting to stratify the GPM events accordingly could again lead to more understanding of prescribing behaviour.

Implications for practice and research

In order to achieve our results, there were several challenges that needed to be overcome which could explain why process mining in secondary care is more prevalent.

Data quality

The quality of healthcare data is limited for many reasons. Events can be incorrectly recorded, unrecorded, or uncoded. All of which limit the confidence and utility of any results generated. Researchers must try and understand the limitations in their data to make best use of it. Primary care data can be thought of as snapshots of coded information that are generated on every contact with the health system. This is different for inpatient secondary care where the entire duration of treatment can be observed and recorded.

While many events recorded in a primary care system may have uncertain veracity, the generation of a prescription is an event we can mine with confidence because, in the UK, virtually all prescriptions are electronically generated in primary care. This is not true for the adverse event of bleeding which may occur in secondary care and may not be coded in the primary care record. To mitigate against this, linked primary and secondary data would be required, and is another reason why process mining exclusively in primary care is not done. Future work should focus on pathways that occur almost exclusively in primary care such as the diagnosis, monitoring and treatment of certain chronic conditions such as hypertension.

Start and end points

Within secondary care, the start of a process can clearly be defined as the admission to hospital, while the end of the process is discharge or death. A patient visiting hospital more than once can be treated as two separate pathways. Within primary care, processes are often cyclical and entangled with other processes. Taking indicator I2 as an example, should the start event be the first instance of peptic ulceration, or should it be the first prescription of an NSAID in a patient with previous peptic ulceration? The former means that each patient only has one pathway, but with potentially multiple cycles, while the latter separates each NSAID prescription into a separate process but then doesn't take the patient's history into account. Detailed consideration must be given to determine whether the primary care processes under analysis have well defined end points, and queries structured to separate data into these individual processes.

Event granularity

The coded events within primary care are not necessarily the events that should make up the event log on which process mining depends. An example is medications where the patient record contains the prescription events, whereas the events of interest on a care pathway would be when the clinician has started, stopped, or altered the dosage of a medication. This is also true for measurements, where a series of blood pressure (BP) values do not necessarily constitute events, but the occurrence of two systolic BPs >140 mmHg within two weeks might be a trigger to investigate a diagnosis of hypertension and could therefore be considered an event to process mine. However, this introduces a bias as the researcher must decide *a priori* what constitutes an event. Is it that a BP was taken, that the BP was high or that some combination of values was measured over a certain period of time? Careful consideration must be made to convert the raw data into an event log, but this is not straightforward and is largely subjective.

Medications

The lack of a stop event for medications requires an extra processing step to determine when a patient's medication has expired. There is also no way of knowing whether a medication once collected is in fact consumed by the patient, or if the patient is using over the counter medications. This is less of an issue in inpatient secondary care when the both the prescription and administration of medication can be monitored and recorded.

Memory

The process mining diagrams that we have produced are heuristic nets which are memory-less: each transition in the process map is taken in isolation without consideration of prior events. By redefining the start events of NSAID and antiplatelets to take into account whether a GPM was already prescribed allowed us to introduce an element of memory into the system. This is useful to better understand the various pathways, however future work using other process mining modelling techniques, such as causal nets [23], might produce better results.

Conclusions

Primary care data in the UK has reliably coded prescribing information and process mining can be successfully applied leading to results that may be useful for clinical decision support systems and improving patient safety.

However primary care data presents several unique challenges. Careful pre-processing must first be undertaken, but this is subjective and must therefore be sensibly performed and meticulously recorded in order to facilitate scrutiny and reproducibility. The use of a clinical reference group to review and confirm data quality and provide insight into the direction of research would be beneficial. Other more powerful process mining and machine learning techniques could be applied now that the initial problems with primary care data have been considered and to some extent addressed.

Acknowledgements

This article was funded by the National Institute for Health Research Greater Manchester Patient Safety Translational Research Centre (NIHR Greater Manchester PSTRC). The views expressed are those of the author and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

References

[1] W.M.P. van der Aalst, Process Mining: Discovery, Con–formance and Enhancement of Business Processes, Springer, 2011. doi:10.1007/978-3-642-19345-3.

[2] R. Williams, E. Rojas, N. Peek, and O.A. Johnson, Process mining in primary care: A literature review, 2018. doi:10.3233/978-1-61499-852-5-376.

[3] A.J. Avery, M. Ghaleb, N. Barber, B.D. Franklin, S.J. Armstrong, B. Serumaga, S. Dhillon, A. Freyer, R. Howard, O. Talabi, and R.L. Mehta, The prevalence and nature of prescribing and monitoring errors in English general practice: A retrospective case note review, *Br. J. Gen. Pract.* **63** (2013) e543-53. doi:10.3399/bjgp13X670679.

[4] R.A. Elliott, E. Camacho, F. Campbell, D. Jankovic, M. Martyn St James, E. Kaltenthaler, R. Wong, M.J. Sculpher, R. Faria, and M. Rita Faria, Prevalence and Economic Burden of Medication Errors in the NHS in England: Rapid evidence synthesis and economic analysis of the prevalence and burden of medication error in the UK, *Policy Res. Unit Econ. Eval. Heal. Care Interv.* (2018).

[5] Ekaterini Blaveri, Mapping the Landscape of UK Health Data Research & Innovation A snapshot of activity in 2017, 2017. https://mrc.ukri.org/documents/pdf/mapping-the-landscape-of-uk-health-data-research-and-innovation-report/ (accessed November 13, 2018).

[6] A.R. Tate, S. Dungey, S. Glew, N. Beloff, R. Williams, and T. Williams, Quality of recording of diabetes in the UK: How does the GP's method of coding clinical data affect incidence estimates? Cross-sectional study using the CPRD database, *BMJ Open*. **7** (2017) e012905. doi:10.1136/bmjopen-2016-012905.

[7] C.W. Günther, and W.M.P. van der Aalst, Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics, in: 2007: pp. 328–343. doi:10.1007/978-3-540-75183-0_24.

[8] R. Spencer, B. Bell, A.J. Avery, G. Gookey, and S.M. Campbell, Identification of an updated set of prescribing-safety indicators for GPs, *Br. J. Gen. Pract.* **64** (2014) e181-90. doi:10.3399/bjgp14X677806.

[9] A.J. Avery, S. Rodgers, J.A. Cantrill, S. Armstrong, K. Cresswell, M. Eden, R.A. Elliott, R. Howard, D. Kendrick, C.J. Morris, R.J. Prescott, G. Swanwick, M. Franklin, K. Putman, M. Boyd, and A. Sheikh, A pharmacist-led information technology intervention for medication errors (PINCER): A multicentre, cluster randomised, controlled trial and cost-eff ectiveness analysis, *Lancet*. **379** (2012) 1310–1319. doi:10.1016/S0140-6736(11)61817-5.

[10] R. Williams, R. Keers, W.T. Gude, M. Jeffries, C. Davies, B. Brown, E. Kontopantelis, A.J. Avery, D.M. Ashcroft, and N. Peek, SMASH! The Salford medication safety dashboard, *J. Innov. Heal. Informatics*. **25** (2018) 183. doi:10.14236/jhi.v25i3.1015.

[11] O. Corporation, OptimizeRx, (n.d.). https://www.optimizerx.com/ (accessed November 13, 2018).

[12] R. Williams, B. Brown, N. Peek, and I. Buchan, Making Medication Data Meaningful: Illustrated with Hypertension, *Stud. Health Technol. Inform.* **228** (2016) 247–51. doi:10.3233/978-1-61499-678-1-247.

[13] R. Williams, E. Kontopantelis, I. Buchan, and N. Peek, Clinical code set engineering for reusing EHR data for research: A review, *J. Biomed. Inform.* **70** (2017) 1–13. doi:10.1016/j.jbi.2017.04.010.

[14] Node.js, (n.d.). https://nodejs.org/en/ (accessed December 19, 2017).

[15] Fluxicon, Process Mining and Automated Process Discovery Software for Professionals - Fluxicon Disco., (n.d.). https://fluxicon.com/disco/ (accessed October 31, 2018).

[16] Ministry of Housing Communities & Local Government, English indices of deprivation 2015, 2015. doi:10.1016/j.diamond.2017.05.012.

[17] A. Dagliati, L. Sacchi, C. Cerra, P. Leporati, P. De Cata, L. Chiovato, J.H. Holmes, and R. Bellazzi, Temporal data mining and process mining techniques to identify cardiovascular risk-associated clinical pathways in Type 2 diabetes patients, in: IEEE-EMBS Int. Conf. Biomed. Heal. Informatics, IEEE, 2014: pp. 240–243. doi:10.1109/BHI.2014.6864348.

[18] T. Meneu, V. Traver, S. Guillen, B. Valdivieso, J. Benedi, and C. Fernandez-Llatas, Heart Cycle: Facilitating the deployment of advanced care processes, *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*. (2013) 6996–6999. doi:10.1109/EMBC.2013.6611168.

[19] R.P. Hilton, N. Serban, and R.Y. Zheng, Uncovering Longitudinal Healthcare Utilization from Patient-Level Medical Claims Data, (2016).

[20] K. Tóth, I. Kósa, A. Vathy-Fogarassy, and D. Hayn, Frequent Treatment Sequence Mining from Medical Databases, *Stud. Health Technol. Inform.* **0** (2017) 211–218. doi:10.3233/978-1-61499-759-7-211.

[21] J. Lismont, A.S. Janssens, I. Odnoletkova, S. vanden Broucke, F. Caron, and J. Vanthienen, A guide for the application of analytics on healthcare processes: A dynamic view on patient pathways, *Comput. Biol. Med.* **77** (2016) 125–134. doi:10.1016/j.compbiomed.2016.08.007.

[22] G.G.T. Lakshmanan, S. Rozsnyai, and F. Wang, Investigating clinical care pathways correlated with outcomes, *Bus. Process Manag.* (2013) 323–338. doi:10.1007/978-3-642-40176-3_27.

[23] W. Van Der Aalst, A. Adriansyah, and B. Van Dongen, Causal nets: A modeling language tailored towards process discovery, in: Lect. Notes Comput. Sci., Springer, Berlin, Heidelberg, 2011: pp. 28–42. doi:10.1007/978-3-642-23217-6_3.

Address for correspondence

Richard Williams - richard.williams@manchester.ac.uk

G303a, JMF Building, University of Manchester, M13 9PL, UK

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190262

A Proficient Spelling Analysis Method Applied to a Pharmacovigilance Task

T. Elizabeth Workman [a], Guy Divita [b], Yijun Shao [a], Qing Zeng-Treitler [a]

[a]*Biomedical Informatics Center, George Washington University, and Washington DC VA Medical Center, Washington, D.C.*
[b]*University of Utah School of Medicine, Division of Epidemiology, and VA Medical Center, Salt Lake City, UT*

Abstract

Misspellings in clinical free text present potential challenges to pharmacovigilance tasks, such as monitoring for potential ineffective treatment of drug-resistant infections. We developed a novel method using Word2Vec, Levenshtein edit distance constraints, and a customized lexicon to identify correct and misspelled pharmaceutical word forms. We processed a large corpus of clinical notes in a real-world pharmacovigilance task, achieving positive predictive values of 0.929 and 0.909 in identifying valid misspellings and correct spellings, respectively, and negative predictive values of 0.994 and 0.333 as assessments where the program did not produce output. In a specific Methicillin-Resistant Staphylococcus Aureus use case, the method identified 9,815 additional instances in the corpus for potential inaffective drug administration inspection. The findings suggest that this method could potentially achieve satisfactory results for other pharmacovigilance tasks.

Keywords:
Natural Language Processing; Machine Learning

Introduction

Pharmacovigilance is "The process and science of monitoring the safety of medicines and taking action to reduce the risks and increase the benefits of medicines" [1]. Pharmacovigilance is "relevant for everyone whose life is touched in any way by medical interventions" [2] and was initiated in a systemic method on a global level by the World Health Organization in 1961 in response to manifested congenital deformities in infants caused by thalidomide. In the United States, pharmacovigilance efforts are conducted by many agencies, including the Food and Drug Administration (FDA) (e.g., [3]), the Centers for Disease Control (CDC) (e.g., [4; 5]), and many state and regional organizations. An example pharmacovigilance task is monitering for the use of ineffective treatments for pathogens like Methicillin-Resistant Staphylococcus Aureus (MRSA). Treating MRSA with antibiotics to which there is an established resistance can exacerbate infections, thus compounding an already serious condition [6].

EMR data is a common source of information for pharmacovigilance efforts. Text mining this data typically consists of rule-based systems that utilize vocabularies from sources such as the Unified Medical Language System (UMLS) to locate named entities, comparing terms from the clinical narrative to vocabulary entries [7]. One shortcoming of this approach is that UMLS vocabularies generally consist of drug terms in a standard form, yet clinical text is often rife with misspellings, in some instances constituting 5% of all content [8], and over 17% of content addressing the pharmaceutical domain [9]. Misspellings of drug terms have a proportional relationship to the character length of their respective standard forms [10]. Differences between the correct and misspelled forms can be measured by the Levenshtein edit distance [11], which is the quantitative difference between two strings or words in terms of their characters. For example, the Levenshtein distance between "amoxicillin" and "amoxycilline" is 2, because there is a "y" in the second form where there is an "i" in the first, plus the second form has an added "e" on the end.

Related Work

Several groups have addressed misspellings in electronic medical record (EMR) data. Danielsson-Ojala analyzed ICU Finnish clinical text addressing wound care [12], finding many instances of misspellings, and abbreviations. Ruch, Baud, and Geissbühler built a spellchecker consisting of three modules [13]. These modules sequentially applied edit distance analysis, part of speech tagging, contextual morpho-syntactic analysis, and word sense filtering (utilizing UMLS) in assessing misspelled terms and candidate corrections. As part of an ontology mapping task, Dziadek et al. used journal articles and EMR text in Swedish as input for a system that employed a commercial spell checker, Levenshtein distance, and trigrams to identify and correct misspellings [14]. Meystre et al. used fuzzy string matching and a constrained Levenshtein edit distance to identify misspelled medications in 3000 randomly selected clinical notes from a larger annotated corpus [15]. Levin et al. annotated clinical text to train and test an algorithm for normalizing misspelled drug terms in clinical text [16] utilizing RxNorm and LVG Metaphone, NLP products from the National Library of Medicine.

Word Embedding (a technique mapping words to real number vectors) facilitated by Word2Vec models [17], holds promise in identifying words with both correct and incorrect spellings. Word2Vec models implement simple neural networks to create word vectors, using either a skip-gram or continuous bag of words (CBOW) approach. The skip-gram approach identifies multiple words in a contextual window, given a single word. The CBOW approach applies the opposite logic, identifying a single word given the other words in the window. The end-products of either method are word embedding vectors that can be used to identify words that have similar word embeddings, or that are found in similar contexts. Word embeddings and similar methodologies have been used as one of several components in identifying spelling errors in consumer-generated text [18]. They could potentially contribute to spelling analysis research involving EMR data.

The Veterans Health Administration (VHA) is one of the largest integrated health care systems in the world [19], providing care to over 9 million patients each year, at 1,243 facilities [20]. Efforts to computerize VHA data began in the 1970's, resulting in the creation of VistA, one of the first EMR systems [21]. This has resulted in the creation of a vast electronic clinical data resource, which VHA maintains in their Corporate Data Warehouse (CDW). These data are made available for research activities through the Veterans Affairs Informatics and Computing Infrastructure (VINCI), a secure platform enabling data research.

Goals of this Work

The primary goal of this study was to identify misspellings of drug names for a specific pharmacovigilance task, namely to identify the potential use of antibiotics for which certain infections have developed a resistance. As an additional exploratory task, we sought to identify the corrected forms of misspellings also used in such a task. We created an application (see https://qtzeng.smhs.gwu.edu/redmine/projects/drug-term-word-analysis for documentation) that identifies misspellings and correct spellings of words in clinical text. This approach leverages the hypothesis that comparing words that share the same contexts in text, utilizing Levenshtein edit distance [11] constraints and a lexicon containing the correctly spelled terms could enable identification of misspellings or correct spellings, depending on the need. By operationalizing this hypothesis, our system identified both the correct and incorrect spellings of terms used in the pharmacovigilance task, given the original input forms, according to predictive value, the evaluative focus of this work

As a test case, we used a list of both correct spellings and misspellings of antibiotic terms. This list had previously been used in a CDC drug-resistance pharmacovigilance task. Two annotators evaluated the results, measuring positive predictive value (PPV) for application output, and negative predictive value (NPV) for when the application produced no output.

Methods

Data Procurement

We used a corpus of 400,000 clinical notes randomly pulled from the VA's TIU tables (i.e. text note tables) through VINCI, of document types addressing infectious diseases. These records were created from October 1999 to May 2018, representing 138190 patients, 131681 males, and 6509 females. The mean age was 65.4 years old.

We obtained a list of 112 antibiotic terms and phrases from the CDC that are regularly monitored in an antibiotic resistance surveillance task to use as input. This list contained both numerous misspellings and a select number of correct spellings of antibiotic terms. Because the focus of this study was individual words, we split phrases such as sulfamethoxazole/trimethoprim into two separate words, and then removed duplicates, producing a list of 108 individual terms

Method Process, including Corpus Treatment

We trained a Word2Vec model, implementing the Genism Word2Vec library in Python [22], for the task of locating words found in the same contexts within the corpus as the 108 pharmaceutical terms. Preliminary work indicated that a model with a feature vector dimensionality of 600, using the Continuous Bag of Words algorithm, with a context window of 8 words, implementing the softmax function with no negative sampling, and 10 iterations produced the most effective model for this given task. In producing the final model, we built word embeddings for all words in the corpus, regardless of frequency. This last measure ensured that we could observe the final efficiency of this method even for terms that occurred just once in the corpus. The Word2Vec model contained a vocabulary of 354,559 terms.

To determine frequency of useful terms in the corpus, we subjected it to a separate, independent process, tokenizing the content, removing many common words (e.g., articles, prepositions, ordinal numbers), specific punctuation patterns that were typical of noise in VA notes, and transformed the remaining tokens to lower case. Superfluous punctuation was also removed. This resulted in a set of 285,909 tokens that were particularly information-bearing, for which we calculated the frequency of each and stored the results.

We built a customized lexicon of standard spellings using the RXNCONSO file from RxNorm [23] for use in the method. The goal of this process was to harvest standardized RxNorm terms for both trade and generic drug names. This automated process placed all such RxNorm terms in a separate resource that contained 19,732 unique drug terms.

The application processed each term from the prepared CDC list of antibiotics using the following methodology. First, the program determined if the input term was correctly spelled, using the customized lexicon, and if the word was present in the Word2Vec model vocabulary. Properties such as dosage were not relevant to the analysis process. If the input term was not in the vocabulary, this fact was recorded for output. If the input term was in the vocabulary, rules facilitated its comparison to the word embeddings of words in the corpus or words in the customized lexicon of standard drug terms, specific to one of three possible scenarios:

Scenario 1: If the input term was correctly spelled, the program retrieved other corpus words that occurred in similar contexts, by ranking the similarity scores between the target input word and other words in the Word2Vec model vocabulary, using the default cosine measurement, retrieving a maximum of 1000 similar terms. For each of these "context" terms, the program calculated the Levenshtein edit distance between it and the input term. If the edit distance was equal or less than the integer floor value of the character count of the input term divided by three (minimum threshold of three after division), and the context word did not end in "s" (i.e., a probable plural form), and did not occur in the customized lexicon (i.e., a probable standardized spelling of a different drug), the program identified it as a misspelling of the input term.

Scenario 2: If the input term was a misspelling, the program first identified candidates from the retrieved context terms also appearing in the customized lexicon that met one of these conditions:

1. The edit distance between the input term and the context term was equal or less than the absolute value of the character count of the context term minus that of the input term (or alternatively four, if this length was less than four). This cast a sufficient yet not overly large net, based on Kilicoglu et al. observations [18].
2. The input term and the context term began with the same three letters.

All these candidate terms were weighted with a simple metric that computed the given candidate's word embedding similarity score, divided by the inverse of the frequency's $\log_2$ value, with

the highest scoring candidate term selected as the logical correct spelling.

Scenario 3: If the input term was a misspelling and the previous step did not identify a logical correct spelling, the input term was compared to terms in the lexicon where the edit distance was three or less. The term(s) with the smallest edit distance that also began with the same first three letters was chosen as the correct form(s).

The edit distances used in Scenarios 2 and 3 were determined through common heuristics (i.e., an optimal maximum edit distance will be a proportion of the input term's length) and testing. The weight metric used in Scenario 2 was guided by two additional heuristics: (a) the correct form will have a high similarity score, and (b) will occur frequently in the corpus. Taking the log value of the frequency scales its influence so that it proportionately affects the value. The metric used in Scenario 3 was guided by the heuristic that the alternative correct form would otherwise be the closest in character variation to the misspelling and begin with the same characters. Edit distances were also determined through preliminary testing on another dataset.

Figure 1 provides a graphical representation of this process.

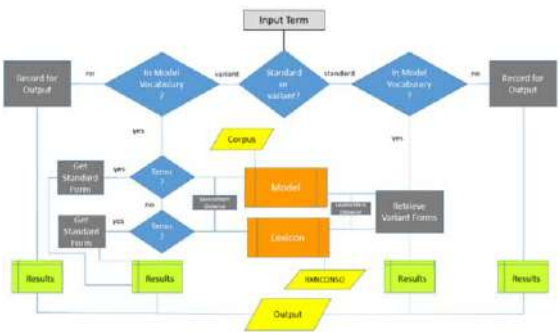


Figure 1. Application process

Evaluation of Output

Two authors (GD, TEW) reviewed the method's output. For each input term, they considered the following questions to identify valid output values:

Where the program produced output:

- For correctly spelled input words, how many candidate terms identified by the program are logical misspellings of the associated term (true positives)? How many are not (false positives)? (*evaluation Oa*)
- For misspelled input words, did the program identify a logical correct spelling (true positive), or not (false positive)? (*evaluation Ob*)

Where the program did not produced output:

- For correctly spelled input words, for each context word, could it logically be a misspelling of the word (false negative), or not (true negatives)? (*evaluation Na*)
- For misspelled input words, did a logically correct spelling of the term appear in the corpus or lexicon (false negative) or not (true negative)? (*evaluation Nb*)

To answer these questions for each input term, the annotators could use a reference standard, such as a dictionary or lexicon. Additionally, occurrences of output terms were analyzed in the corpus to assure they were likely used as drug terms.

We calculated Cohen's Kappa to assess inter-annotator agreement. Disagreements were settled by consensus. We calculated the positive predictive value (*True Positives/(True Positives + False Positives*)) and negative predictive value (*True Negatives/(True Negatives + False Negatives)*) to evaluate performance. Positive predictive value is used to evaluate the application's performance in terms of its output (*evaluation Oa* and *evaluation Ob*). Negative predictive value is used to determine how the program performed when it produced no output (*evaluation Na* and *evaluation Nb*), in terms of not identifying valid output (i.e., false negatives).

Results

Raw Application Output

Input terms were organized by six classifications, in accordance with the application. Input words that were correctly spelled drug terms (a) were in the corpus, and the application produced misspellings (33 input terms), (b) were in the corpus, and the application did not produce misspellings (20 input terms), or (c) were not in the corpus (37 input terms). Similarly, input words that were misspelled drug terms (d) were in the corpus, and the application identified a potentially correct spelling (10 input terms), (e) were in the corpus, and the application did not identify a potentially correct spelling (3 input terms), or (f) were not in the corpus (5 input terms). Terms were considered correct if they appeared in the customized lexicon.

For the 33 correctly spelled input terms in the corpus where the program did produce output, there was a total of 184 misspellings, prior to annotator review. For the 10 misspellings where the program produced output, there was a total of 11 potential correct spellings identified (there were 2 identified for "septran").

Inter-Rater Agreement

In terms of Cohen's Kappa (Table 1), agreement was fair for *evaluation Na*, substantial for *evaluation Oa* and *evaluation Ob*, and essentially perfect for *evaluation Nb* [24], with percentages ranging from 90% to 100% for all. For *evaluation Na*, the Cohen's Kappa agreement was borderline fair / moderate, yet the percentage of overall agreement was 99.68%. This discrepancy is likely due to Cohen's Kappa's vulnerability to skewed data [25].

Method Performance

For *evaluation Oa*, of the total 184 misspellings output by the application, 171 were determined to be logical misspellings of their given correctly spelled input forms, thus producing a PPV of 0.929 for this task. For *evaluation Ob*, of the 11 correct forms output by the application, 9 were determined to be logical representations of their given input misspellings. This produced a PPV of 0.909 for this task (Table 1).

For *evaluation Na*, there were 619 true negatives, and four false negatives, producing an NPV of 0.994. For *evaluation Nb*, there were 1 true negative and 2 false negatives, producing an NPV of 0.333. For these input misspellings we checked if a logical correct spelling occurred in the corpus or lexicon. Of the three terms, logical corrections appeared in the corpus for two, therefore there were two false negatives, and one true negative (Table 1).

Table 1. Inter-Rater agreement and application performance

	Raw(Valid) Terms	Disagreements (Agreement %)	Cohen's Kappa	Performance Metric(Value)
Oa	184 (171)	6 (96.74%)	0.733	PPV (0.929)
Ob	11 (9)	1 (90.90%)	0.621	PPV (0.909)
Na	623 (4)	3 (99.68%)	0.398	NPV (0.994)
Nb	3(1)	0 (100%)	1	NPV (0.333)

Examples of how some terms were used in the corpus follow, with indications of the correct and incorrect spellings:

- **Bactrim/bactim:** "the pt will be best tx with bactim i will order this drug as well"
- **Clindamycin/clindamcyin:** "was initally started on vancomycin and switched to clindamcyin po 600mg tid"
- **Zosyn/zozyn:** "agree with outpatient treatment with zozyn and clindamycin for 7 days"

Misspelled Antibiotic Terms Identified

For the task of monitoring clinical text for mentions of the antibiotics on the list, the program identified 171 logical misspellings after annotator review and consensus. This equals an average of 5 additional representations per correct spelling, of the 33 correctly spelled words from the CDC antibiotics list.

For example, the correctly spelled input term Cefazolin occurred in the corpus. It was matched to the contextual misspellings "cafazolin", "cefazoin", "cefazoline", "cefezolin", "ceftazolin", "cefzolin", and "cephazolin" (all true positives for *evaluation Oa)* using the edit distance constraint and lack of presence of these contextual words in the customized lexicon, as described in Scenario 1.

Correct Spellings Identified

There were 18 misspellings in the original CDC list. All but 5 of the 18 misspellings were also found in the corpus. Of those 13, the application identified logical correct spellings for 10, including two for "septran". After annotator review and consensus, 9 terms were identified as valid corrections for 9 misspellings.

To illustrate a Scenario 2 output for the input misspelling "zithromycin", Azithromycin had a similarity score of 0.3403 and occurred 22556 times in the corpus, producing a score of 4.9212:

$$(0.3403) / (1/ \log_2(22556)) = 4.9212$$

The highest score of all candidate corrections, and is a true positive for *evaluation Ob*.

To illustrate Scenario 3, the application identified "Tetracycline" as the logical correct spelling of the misspelled input term "tetraycycline", because among terms in the customized lexicon, it had the smallest edit distance from "tetraycycline" and began with the first three letters. This is another true positive for *evaluation Ob*.

Discussion

The application identified 171 valid misspellings of the correctly spelled antibiotic terms, thus providing additional material for review in a pharmacovigilance task. As a specific use case, consider Penicillin G, Tetracycline, and Amoxicillin, to which MRSA demonstrates resistance [26]. The correct spellings of these exact terms occur 15,317 times in the corpus. The application found a total of 19 misspellings of penicillin, tetracycline, and amoxicillin. In the corpus, these misspellings (in the case of "penicillin", appending the misspelled variations with "g") occurred 9,815 times, representing 9,815 additional occurrences where these drugs may have been used to treat MRSA, thus meriting scrutiny.

Method Output Performance

The application achieved PPV of 0.929 and 0.909 in identifying misspellings of correctly spelled input terms, and logical corrections of misspelled input terms, respectively. By combining similarity scores provided by the word embeddings with the edit distance constraints, the application was able to identify misspellings from the many words that occur in the same context as a given correctly spelled word. For the correctly spelled input word "tetracycline", the application identified "tetracyclin", but not "tegecycline" as a logical misspelling. The latter term, which occurs in the corpus, is within the edit distance constraint, but did not occur in the same context as tetracycline and is more likely a misspelling of Tigecycline. For the misspelling "amoxicilline" the program identified "amoxicillin" as the correct spelling instead of the other candidate "amoxi" because it produced a score of 4.8849, whereas the score for "amoxi" was 2.2499. A visual comparison of the three terms suggests that "amoxicillin" is the valid correct spelling. For the two misspelled input terms "septran" and "tetraycycline" for which no context word met the Scenario 2 criteria, the method of finding the customized lexicon term with the shortest edit distance that began with the same three letters identified logical corrections for each.

The application performed well in terms of NPV where correct spellings were used as input, but not as well for input misspellings, where no output was produced. *Evaluation Nb* determined that in most cases, among all context words, there were no valid logical misspellings for the twenty correctly spelled input terms for which the method did not produce output. In this evaluation, there were only four false negatives and 619 true negatives. For the three input misspellings with no output, we identified two false negatives and one true negative. This has been targeted as an area of improvement in future work.

Overall, the application demonstrated good performance for identifying both misspelled and correctly spelled antibiotic terms for the pharmacovigilance task, using the randomly selected 400,000 VA notes as a corpus. Although this is a small study addressing antibiotic use, these results suggest that this method could achieve satisfactory results for other pharmacovigilance tasks.

Limitations

For testing purposes, we designed the application to produce word embeddings for terms that occurred at any frequency in the corpus, but generally Word2Vec is not as efficient in modeling highly infrequent words. The task of identifying logically correct forms of input misspellings should be interpreted as such, and not as an attempt to identify the exact intended term. For example, it would be impossible to truly ascertain from the text whether or not a clinician meant "Amoxicillin" and not "Amoxil" when using a similar misspelling. This is a small study addressing 108 antibiotics but demonstrates the program's potential value in a pharmacovigilance task. The VHA serves a predominantly older adult male population. However, because the VHA is so large, with so many patients as well as healthcare practice specialties, and significant populations of both adult male and

female patients of various ages, these results can be generalized to a certain extent.

Conclusions

We developed a novel method to identify misspellings and correct spellings of pharmaceutical terms in clinical text using Word2Vec, Levenshtein edit distance constraints, and a customized lexicon, evaluating it for a real-world pharmacovigilance task using a corpus of VHA clinical notes and a CDC surveillance list of pathogen-resistant antibiotics, achieving PPV of 0.929 for identification of misspellings and 0.909 for identification of correct spellings in output, and NPV of 0.994 and 0.333 for these tasks where there was no output. As a specific use case, the method provided 9815 additional occurrences in the corpus of three MRSA-resistant drugs.

Acknowledgements

This work was supported by the VA IDEAS 2.0 HSRD Research Center and CREATE: A VHA NLP Software Ecosystem for Collaborative Development and Integration project, grant CRE 12–315, and Award Numbers UL1TR001876 and KL2TR001877 from the NIH National Center for Advancing Translational Sciences. The views expressed are those of the authors and do not necessarily reflect those of the Department of Veterans Affairs, the United States Government, or the academic affiliate organizations. We also thank Dr. Halil Kilicoglu for his wisdom and advice in preparing the manuscript.

References

[1] E.U. European Commission. Pharmacovigilance. https://ec.europa.eu/health/human-use/pharmacovigilance_en. [accessed 31 March 2019]

[2] World Health Organization, *The Importance of Pharmacovigilance: Safety Monitoring of Medicinal Products*, World Health Organization, 2002.

[3] D.C. Throckmorton, S. Gottlieb, and J. Woodcock, The FDA and the Next Wave of Drug Abuse - Proactive Pharmacovigilance, *N Engl J Med* **379** (2018), 205-207.

[4] J. Huang, J. Du, R. Duan, X. Zhang, C. Tao, and Y. Chen, Characterization of the Differential Adverse Event Rates by Race/Ethnicity Groups for HPV Vaccine by Integrating Data From Different Sources, *Front Pharmacol* **9** (2018), 539.

[5] S. Reagan-Steiner, D. Yankey, J. Jeyarajah, L.D. Elam-Evans, J.A. Singleton, C.R. Curtis, J. MacNeil, L.E. Markowitz, and S. Stokley, National, Regional, State, and Selected Local Area Vaccination Coverage Among Adolescents Aged 13-17 Years--United States, 2014, *MMWR. Morbidity and mortality weekly report* **64** (2015), 784-792.

[6] S.J. Dancer, The Effect of Antibiotics on Methicillin-Resistant Staphylococcus Aureus, *J Antimicrob Chemother* **61** (2008), 246-253.

[7] R. Harpaz, A. Callahan, S. Tamang, Y. Low, D. Odgers, S. Finlayson, K. Jung, P. LePendu, and N.H. Shah, Text Mining for Adverse Drug Events: The Promise, Challenges, and State of the Art, *Drug Saf* **37** (2014), 777-790.

[8] W.R. Hersh, E.M. Campbell, and S.E. Malveau, Assessing the Feasibility of Large-Scale Natural Language Processing in a Corpus of Ordinary Medical Records: a Lexical Analysis, *Proc AMIA Annu Fall Symp* (1997), 580-584.

[9] L. Zhou, L.M. Mahoney, A. Shakurova, F. Goss, F.Y. Chang, D.W. Bates, and R.A. Rocha, How Many Medication Orders are Entered through Free-Text in EHRs?--a Study on Hypoglycemic Agents, *AMIA Annu Symp Proc* **2012** (2012), 1079-1088.

[10] C. Senger, J. Kaltschmidt, S.P. Schmitt, M.G. Pruszydlo, and W.E. Haefeli, Misspellings in Drug Information System Queries: Characteristics of Drug Name Spelling Errors and Strategies for their Prevention, *Int J Med Inform* **79** (2010), 832-839.

[11] V.I. Levenshtein, Binary Codes Capable of Correcting Deletions, Insertions, and Reversals, in: *Soviet physics doklady*, 1966, pp. 707-710.

[12] R. Danielsson-Ojala, H. Lundgren-Laine, and S. Salanterä, Describing the Sublanguage of Wound Care in an Adult ICU, *Stud Health Technol Inform* **180** (2012), 1093-1095.

[13] P. Ruch, R. Baud, and A. Geissbühler, Using Lexical Disambiguation and Named-Entity Recognition to Improve Spelling Correction in the Electronic Patient Record, *Artificial intelligence in medicine* **29** (2003), 169-184.

[14] J. Dziadek, A. Henriksson, and M. Duneld, Improving Terminology Mapping in Clinical Text with Context-Sensitive Spelling Correction, *Informatics for Health: Connected Citizen-Led Wellness and Population Health* **235** (2017), 241.

[15] S.M. Meystre, Y. Kim, J. Heavirland, J. Williams, B.E. Bray, and J. Garvin, Heart Failure Medications Detection and Prescription Status Classification in Clinical Narrative Documents, *Studies in health technology and informatics* **216** (2015), 609.

[16] M.A. Levin, M. Krol, A.M. Doshi, and D.L. Reich, Extraction and Mapping of Drug Names from Free Text to a Standardized Nomenclature, in: *AMIA Annual Symposium Proceedings*, American Medical Informatics Association, 2007, p. 438.

[17] T. Mikolov, K. Chen, G. Corrado, and J. Dean, Efficient Estimation of Word Representations in Vector Space, *arXiv preprint arXiv:1301.3781* (2013).

[18] H. Kilicoglu, M. Fiszman, K. Roberts, and D. Demner-Fushman, An Ensemble Method for Spelling Correction in Consumer Health Questions, in: *AMIA Annual Symposium Proceedings*, American Medical Informatics Association, 2015, p. 727.

[19] J.B. Perlin, R.M. Kolodner, and R.H. Roswell, The Veterans Health Administration: Quality, Value, Accountability, and Information as Transforming Strategies for Patient-Centered Care, *Am J Manag Care* **10** (2004), 828-836.

[20] Veterans Heatlh Administration: About VHA. https://www.va.gov/health/aboutvha.asp. [accessed 31 March 2019]

[21] S.H. Brown, M.J. Lincoln, P.J. Groen, and R.M. Kolodner, VistA—US Department of Veterans Affairs National-Scale HIS, *International journal of medical informatics* **69** (2003), 135-156.

[22] R. Rehurek and P. Sojka, Software Framework for Topic Modelling with Large Corpora, in: *In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, Citeseer, 2010.

[23] S. Liu, W. Ma, R. Moore, V. Ganesan, and S. Nelson, RxNorm: Prescription for Electronic Drug Information Exchange, *IT professional* **7** (2005), 17-23.

[24] M.L. McHugh, Interrater Reliability: The Kappa Statistic, *Biochem Med (Zagreb)* **22** (2012), 276-282.

[25] S. Xu and M.F. Lorber, Interrater Agreement Statistics with Skewed Data: Evaluation of Alternatives to Cohen's Kappa, *J Consult Clin Psychol* **82** (2014), 1219-1227.

[26] M.A. Al-Ashmawy, K.I. Sallam, S.M. Abd-Elghany, M. Elhadidy, and T. Tamura, Prevalence, Molecular Characterization, and Antimicrobial Susceptibility of Methicillin-Resistant Staphylococcus aureus Isolated from Milk and Dairy Products, *Foodborne Pathog Dis* **13** (2016), 156-162.

Address for correspondence

lizworkman@gwu.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190263

Outcome-Driven Clustering of Acute Coronary Syndrome Patients Using Multi-Task Neural Network with Attention

Eryu Xia[a], Xin Du[b], Jing Mei[a], Wen Sun[a], Suijun Tong[a], Zhiqing Kang[c], Jian Sheng[c], Jian Li[c], Changsheng Ma[b], Jianzeng Dong[b,d], Shaochun Li[a]

[a] *IBM Research – China, IBM, Beijing, China*
[b] *Department of Cardiology, Beijing Anzhen Hospital, Capital Medical University, National Clinical Research Center for Cardiovascular Diseases, Beijing, China*
[c] *AstraZeneca China*
[d] *Department of Cardiology, The First Affiliated Hospital, Zhengzhou University, Zhengzhou, Henan, China*

Abstract

Cluster analysis aims at separating patients into phenotypically heterogenous groups and defining therapeutically homogeneous patient subclasses. It is an important approach in data-driven disease classification and subtyping. Acute coronary syndrome (ACS) is a syndrome due to sudden decrease of coronary artery blood flow, where disease classification would help to inform therapeutic strategies and provide prognostic insights. Here we conducted an outcome-driven cluster analysis of ACS patients, which jointly considers treatment and patient outcome as indicators for patient state. Multi-task neural network with attention was used as a modeling framework, including learning of the patient state, cluster analysis, and feature importance profiling. Seven patient clusters were discovered. The clusters have different characteristics, as well as different risk profiles to the outcome of in-hospital major adverse cardiac events. The results demonstrate cluster analysis using outcome-driven multi-task neural network as promising for patient classification and subtyping.

Keywords:

Acute Coronary Syndrome, Cluster Analysis, Computer

Introduction

Precision medicine is a healthcare approach which aims at developing more effective ways to improve health and treat disease by taking individual traits into account [1]. One attempt toward precision medicine is to provide the best available care for patients based on their disease subtypes within a disease of common biological basis. Patient cluster analysis comprises a solid step towards precision medicine, which fulfils the task of disease classification and subtyping [2]. Cluster analysis has been used for subgroup analysis of type 2 diabetes [3], accurate phenotyping of heart failure and related syndromes [4,5], as well as identifying meaningful patient clusters for developing specific treatment programs in geriatric stroke patients [6]. The results support cluster analysis as a useful tool to discover disease classes and subtypes, which can inform therapeutic strategies like individualizing treatment regimens and providing prognosis insights.

Cluster analysis is performed based on a similarity or distance measure. Commonly used similarity measures include Euclidean distance, cosine similarity, Jaccard similarity, and so on. Traditionally, as no associating outcome measure is available for cluster analysis, the methods are unsupervised, and thus the similarity measure takes all patient characteristics as equally important, the results of which are less desired when we target the clustering results at reflecting specific patient traits. It has been recognized that patient similarities for cluster analysis are commonly context-based and are sometimes associated with clinical outcomes of interest. Outcome-driven clustering (sometimes referred to as 'semi-supervised clustering' or 'supervised clustering') is applied when outcome measures are available and can serve as a noisy surrogate for the (unobserved) target cluster [7], which has been proven useful in patient cluster analysis for precision cohort finding [8] and clinical decision support [9].

Neural network has been increasingly used as a successful data modeling paradigm, which solves tasks such as pattern recognition and classification through a learning process and has been recently used in medical informatics research for representation learning [10]. Multi-task learning is a strategy where multiple learning tasks are solved at the same time to benefit from their commonalities and contrasts. Neural networks adapt to multi-task learning intuitively by designing specific network structure and cost function [11]. Though useful, neural network is often criticized for lack of interpretability. Attention mechanism is thus introduced to neural network to increase model interpretability as well as performance [12] and has been applied to healthcare research [13]. Thus, a joint use of the three techniques would facilitate representation learning, leveraging information from different tasks in an interpretable manner.

Acute coronary syndrome (ACS) is a syndrome due to sudden decreased coronary artery blood flow. A treatment objective of ACS is to prevent major adverse cardiac events (MACE) during hospitalization. ACS can be classified into ST elevation myocardial infarction (STEMI), non-ST elevation myocardial infarction (NSTEMI), and unstable angina (UA) by cardiac marker and manifestation of ST-elevation in electrocardiogram. However, exploration of biomarkers for disease classification and subtyping has never stopped [14,15]. In this study, we present outcome-driven clustering of ACS patients based on biomarkers as well as clinical indicators. We desired using patient state (which is an abstract characterization of patient traits regarding the disease) for clustering and decided on four outcome measures as surrogates to indicate the patient state: antiplatelet treatment, beta-blockers treatment, statins treatment, and in-hospital MACE. The four measures are supposed to reflect different facets of the patient state. Therefore, a joint consideration would enable a more comprehensive and targeted depiction of the patient state. Cluster analysis has been conducted on ACS patients to discover symptom clusters [16], assess the differences in mortality between symptom clusters [17], discover clusters of different lifestyle risk factors [18], and to detect critical patients using medical parameter time series [19]. However, all the above studies are unsupervised and none of them use neural network as the modeling framework.

In this study, we conducted outcome-driven patient clustering on hospitalized ACS patients, identified underlying patient clusters, and profiled the cluster characteristics, especially risk factors to in-hospital MACE. Novelty of our study includes:

(1) using outcome-driven cluster analysis to guide cluster analysis; (2) using multi-task neural network to learn a multi-faceted representation of patient characteristics; and (3) attention mechanism was introduced to the neural network model to increase model interpretability and facilitate feature importance profiling.

Methods

Cohort Construction

The multi-center, retrospective cohort study was conducted at 38 urban and rural hospitals in China. Adult hospitalized patients (aged $\geqslant$18 years) with a final diagnosis of ACS identified at the time of death or discharge were included. Each hospital enrolled the first five consecutive patients on a monthly basis from January 1, 2008 through December 31, 2015. We excluded patients who: (1) had potentially lethal diseases (e.g., incurable cancer, decompensated cirrhosis, multisystem organ failure); (2) had an expected life span below 12 months; or (3) died within 10 minutes of arrival at the hospital. A patient needs to have age, gender, ACS type recorded to be included in the analysis. A total of 26,986 patients were included in this study.

Feature Construction

Patient data was identified and reviewed by trained investigators to record clinical information. We included 41 patient characteristics as features, including: disease types (ACS type and Killip class), demographics, personal disease history, comorbidities, habits, laboratory test results, and procedures. Data outliers determined based on clinical knowledge were removed and represented as missing data. Missing values were imputed by multiple imputation utilizing the 'mice' package in R [20]: continuous variables by predictive mean matching, binary variables by logistic regression, and a proportional odds model for ordinal variables. We conducted one-hot encoding on categorical variables with more than two categories, and standardized continuous values by removing the mean and scaling to unit variance.

Classification with Multi-task Neural Network Model

Multi-task neural network with attention was used as the framework. A schematic representation of the neural network design is shown in Figure 1. The attention layer was implemented as a hidden layer with softmax activation, with the same number of nodes as the shared input layer. Attention was added to the attention layer by multiplying element-wise with the shared input layer.

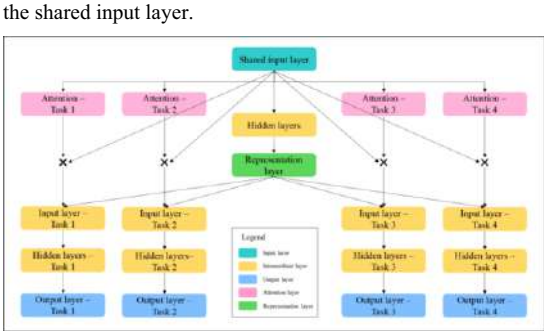


Figure 1. *Illustration of the Neural Network Design.*

Following the shared input layer are hidden layers, then a representation layer which learns a joint patient representation. The four classification tasks are optimized simultaneously. For each task, the input for the task was formed by concatenating the representation layer with the features in the shared input layer after adding attention to each feature. It then reaches the final output layer after adding hidden layers in between.

During training, features mentioned in the 'Feature construction' section were used as input for in the shared input layers, and the ground truth of the four patient traits (MACE, antiplatelet treatment, beta-blockers treatment, and statins treatment) were used as output in the four output layers for each task.

Binary cross entropy loss was used for each classification task. For task j (j in 1, 2, 3, and 4), the task weight is denoted as w_j, the ground truth and predicted probability for an instance i are denoted as y_{ij} and $\widehat{y_{ij}}$ respectively, and $H\left(y_{ij}, \widehat{y_{ij}}\right)$ is the binary cross entropy loss. The cost function for neural network we used is defined as:

$$C \triangleq \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{4} w_j H\left(y_{ij}, \widehat{y_{ij}}\right)$$

where n is the number of training samples. In our experiments, we assigned equal weight to all classification tasks.

Parameters were optimized using 'Adam'. Training was conducted with a batch size of 512 and 50 epochs. Class weights were added to balance the biased proportion of positive and negative cases respectively for all four tasks.

To validate the performance of the neural network for the classification tasks, cross validation was conducted 10 times by each randomly splitting data into training set and validation set at a ratio of 4:1. For patient clustering, all samples were used for neural network training.

Post-Classification Analysis Workflow

Analysis after classification with multi-task neural network model included three steps: (1) evaluating neural network classification performance; (2) clustering patients using values from the representation layer; and (3) profiling risk factors for in-hospital MACE in each patient cluster using the attention values.

Patient Clustering

For each patient, values of the representation layer after training were used as the vector for clustering, which is a 32-dimension vector. K-means was used for patient clustering. Model selection was conducted using Bayesian Information Criteria to choose the model from a range of different K (number of clusters) settings (2 to 15). We selected K = 7 for K-means clustering.

Implementation

Cohort construction, feature construction and post-classification analysis were conducted using R 3.4.1. Neural network training and analysis were conducted using Python 2.7.14, Keras 2.2.4, and Theano 1.0.3.

Results

Neural Network Classification Performance

We used the multi-task neural network for the four selected classification tasks. Proportions of positive cases in four classification tasks are shown in Table 1, which are largely imbalanced. Classification performances are evaluated by AUROC (area under the receiver operating characteristics) and AUPRC (area under the precision recall curve) on the validation set in a cross validation setting (Table 1). From the results, MACE and antiplatelet treatment were best classified while beta-blockers treatment had the lowest classification performance. The results suggest that the learned neural network model is a greater reflection of the patient states corresponding to MACE and antiplatelet treatment.

Table 1. Neural network classification performance

Task	Positive case	Performance		
		AUROC	AUPRC (class 0)	AUPRC (class 1)
MACE	3.54%	0.8602 (0.0141)	0.9926 (0.0007)	0.2924 (0.0551)
Antiplatelet treatment	80.50%	0.8634 (0.0078)	0.5799 (0.0197)	0.9640 (0.0038)
Beta-blockers treatment	68.87%	0.6881 (0.0131)	0.5035 (0.0199)	0.8184 (0.0097)
Statins treatment	89.24%	0.7725 (0.0167)	0.2842 (0.0275)	0.9635 (0.0045)

Note: In performance cells, the numbers denote mean value (standard deviation).

K-means Clustering

K-means was conducted on the representation layer to cluster patients into seven groups. Clustering results are visualized in Figure 2, where t-SNE was conducted to reduce the 32-dimension data in the representation layer to 2 dimensions and used color to denote the assigned cluster membership. We see clear separation of the clusters in the low dimension representation.

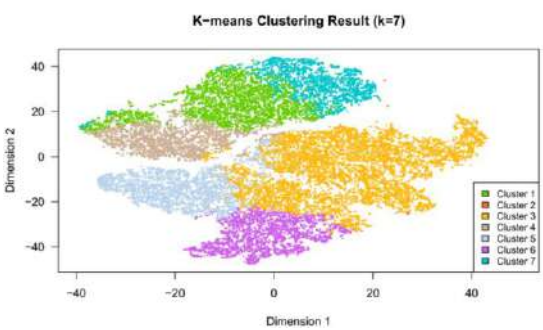


Figure 2. *Visualization of Patient Clustering using t-SNE plot.*

Distribution of cluster size, and the properties relating to classification tasks are shown in Table 2. Clusters were organized based on descending MACE onset rate. The seven clusters have different cluster size, MACE rate and treatment rate. Cluster 2 has only 5 samples and is thus not included in later comparison. The largest cluster (9,889/26,986; 37%) had the highest MACE rate, and lower treatment rate compared to the overall cohort.

Table 2. Distribution of Cluster Size and Classification Labels

	Size	MACE	Anti-platelet treatment	Beta-blockers treatment	Statins treatment
Overall	26,986 (100%)	955 (3.5%)	21,708 (80.5%)	18,572 (68.9%)	24,066 (89.2%)
Cluster **3**	9,889 (37%)	760 (7.7%)	7,768 (78.6%)	6,210 (62.8%)	8,496 (85.9%)
Cluster **4**	3,163 (12%)	74 (2.3%)	3,139 (99.2%)	2,361 (74.6%)	3,064 (96.9%)
Cluster **1**	4,353 (16%)	82 (1.9%)	4,346 (99.8%)	3,225 (74.1%)	4,284 (98.4%)
Cluster **7**	2,709 (10%)	23 (0.8%)	2,708 (100.0%)	2,292 (84.6%)	2,690 (99.3%)
Cluster **6**	3,212 (12%)	14 (0.4%)	1,119 (34.8%)	1,788 (55.7%)	2,214 (68.9%)
Cluster **5**	3,637 (13%)	2 (0.1%)	2,623 (72.1%)	2,691 (74.0%)	3,313 (91.1%)
Cluster **2**	5 (0%)	0 (0.0%)	5 (100.0%)	5 (100.0%)	5 (100.0%)

Note: In each cell, the numbers denote count (proportion). In column 'Size', the proportion is the proportion in the overall patient cohort. In other columns, the proportion is the proportion in the specified cluster.

Profiles of patient clusters were analyzed. Notable features of each patient cluster are presented in Table 3. Specifically, Cluster 3 has more severe conditions as shown by the highest average age, proportion of patients with elevated cardiac enzyme levels, and the lowest average left ventricular ejection fraction (LVEF [%]). Cluster 4 also has comparatively severe condition as is shown by the Killip class. Cluster 1 does not show severe disease, but the MACE rate is still high, which is potentially associated with bad living habits (highest proportion of current smoker and current alcohol drinker) of this cluster. Cluster 7 is featured by the highest proportion of STEMI patients and lowest proportion of UA patients. Though STEMI patients are far more prone to in-hospital MACE compared to UA and NSTEMI, this cluster is not associated with a high MACE rate, potentially as a combined effect of the less complicated disease manifestation and the high level of treatment. Cluster 6 has the highest proportion of UA patients, and are less prone to MACE even though they have the lowest treatment rates. Cluster 5 is featured by the low disease severity, and correspondingly, the lowest MACE rate.

Risk Factors for MACE in Each Patient Cluster

For each patient, a feature's attention value for the MACE classification task from the neural network is used as its importance in predicting MACE. For each patient cluster, a feature's importance is calculated as the average attention value of all patients in the cluster. Feature importance in each patient cluster is shown in Table 4. Features with different importance or high clinical relevance are selectively listed. The largest value in each row is shown in bold. Different clusters have different feature importance, indicating different risk profiles. As an example, smoking is a more important risk factor for MACE in Cluster 1 and Cluster 7, current comorbidity of hypertension more important for Cluster 4, 5 and 6, age and systolic blood pressure are more important in Cluster 3 than in other clusters.

Table 3. Profile of Patient Clusters

Cluster	Has the highest	Has the lowest
3	MACE rate Age Proportion of NSTEMI patients Cormobidity of atrial fibrillation Elevated cardiac enzyme levels	LVEF (%)
4	Killip class History of myocardial infarction	
1	Current smoker and current alcohol drinker	
7	Antiplatelet treatment Beta-blockers treatment Statins treatment Proportion of STEMI patients	Proportion of UA patients Killip class Proportion of patients with cormorbidity Proportion of patients with disease history
6	Proportion of female patients Proportion of UA patients	Antiplatelet treatment Beta-blockers treatment Statins treatment
5	LVEF (%) Cormobidity of hypertension History of vascular disease History of established coronary artery disease History of percutaneous coronary intervention History of coronary artery bypass grafting History of other conditions confirmed by computed tomography angiography	MACE rate Elevated enzyme levels

Discussion

Attention Mechanism and Feature Importance

Attention model in neural network is inspired by brain's neural mechanism of attention and is simplified here as: in each sample, including a numerical weight ('attention value') for each predictor associated with each outcome. When we normalize the weights for each sample to be all larger than 0 and have a sum of 1, the attention values look similar to a probability distribution to show the feature importance. In our study, we considered the outcome of in-hospital MACE. For each feature, we calculated the average attention value of patients in a cluster and used it as its importance in this cluster. When a feature has an importance larger than 0, we regard it as a risk factor to the outcome, where the feature importance is taken as the importance of the risk factor. The attention mechanism makes the neural network models, otherwise 'black boxes', interpretable to some degree, but is still less clear compared to logistic regression models, where both the feature importance and the action directionality (positive or negative impact) are shown with odds ratio, confidence intervals, and coefficient p-values.

Choice of Outcomes

Multi-task learning is an approach to transfer domain knowledge contained in related outcomes and learn in parallel using a shared representation [21]. The incentive for this approach is that the outcomes are reflections of different facets of a common latent representation. To better know the latent representation, we learn from different outcomes. When thinking of the patient state as the latent representation, we need to choose outcomes that are reflections of the patient state. In our study, antiplatelet, beta-blockers, and statins treatment are prescribed based on doctors' perception of the patient based on domain knowledge. MACE outcome is a direct result of the patient disease state. We thus include all four outcomes to better represent the patient. After assessing the classification performance, the patient representation better characterizes the patient state regarding MACE and antiplatelet treatment than the patient state regarding beta-blocker treatment.

Table 4. Feature Importance in Each Patient Cluster

Cluster	**3**	**4**	**1**	**7**	**6**	**5**
Ethnic group (Han)	0.046	0.070	0.048	0.050	**0.127**	0.088
ACS Type (NSTEMI)	0.029	0.015	0.030	**0.039**	0.016	0.008
ACS Type (STEMI)	0.002	0.002	0.003	**0.003**	0.002	0.002
ACS Type (UA)	0.085	0.134	0.059	0.045	0.138	**0.208**
Current smoking	0.046	0.052	0.091	**0.095**	0.032	0.030
Current Hypertension	0.032	0.103	0.036	0.033	0.127	**0.146**
Current Diabetes	0.032	0.033	**0.053**	0.034	0.034	0.028
Current atrial fibrillation	0.082	0.055	0.078	**0.085**	0.056	0.025
Current Percutaneous coronary intervention	0.064	0.071	**0.101**	0.086	0.018	0.032
Elevated enzyme levels	0.047	0.043	0.050	0.060	**0.080**	0.061
Killip Class	**0.048**	0.029	0.035	0.036	0.005	0.011
Age (years)	**0.045**	0.029	0.033	0.027	0.008	0.013
Systolic blood pressure (mmHg)	**0.033**	0.017	0.018	0.019	0.010	0.012
White blood cell count (×10^9/L)	0.029	**0.031**	0.020	0.020	0.019	0.029

Pitfalls in Interpretation of the Results

Two points need to be addressed regarding interpretation of the results. First, algorithmically meaningful clusters are not necessarily clinically meaningful clusters. Though the clustering result has implication for disease prognosis, whether it can inform clinical practice needs further clinical research. Second, risk factors cannot be directly translated to clinical intervention. As an example, though comorbidity of hypertension and high systolic blood pressure are risk factors for in-hospital MACE, the results are not sufficient to claim the intensity of hypertension treatment required for different clusters.

Suggestions for Future Study

Our suggestions for future study using similar approach include: (1) carefully select clinically meaningful outcomes to be used; (2) use fewer features and easily acquired ones would make the results more applicable; and (3) identify cluster-specific interventions considering treatment effectiveness would add extra clinical value to similar studies.

Conclusions

In this study, we used multi-task neural network with attention as a modeling framework to support learning of patient state representations, cluster analysis of patients, and profiling of feature importance. Seven patient clusters were discovered, which have different characteristics and risk profiles to in-hospital MACE. The results demonstrate cluster analysis using outcome-driven multi-task neural network as a promising approach for ACS patient classification and subtyping.

Acknowledgements

This work was supported by the grants from National Science Foundation of China: 81530016 to Changsheng Ma, and from Beijing Municipal Commission of Science and Technology: D151100002215003 and D151100002215004 to Jianzeng Dong.

References

[1] F.S. Collins, and H. Varmus, A new initiative on precision medicine., *N. Engl. J. Med.* **372** (2015), 793–5. doi:10.1056/NEJMp1500523.

[2] M.X. Eryu Xia, Haifeng Liu, Jing Li, Jing Mei, Xuejun Li, Enliang Xu, Xiang Li, Gang Hu, Guotong Xie, Gathering Real World Evidence with Cluster Analysis for Clinical Decision Support, *Stud. Health Technol. Inform.* (2017), 1185–1189. doi:10.3233/978-1-61499-830-3-1185.

[3] E. Ahlqvist, P. Storm, A. Käräjämäki, M. Martinell, M. Dorkhan, A. Carlsson, P. Vikman, R.B. Prasad, D.M. Aly, P. Almgren, Y. Wessman, N. Shaat, P. Spégel, H. Mulder, E. Lindholm, O. Melander, O. Hansson, U. Malmqvist, Å. Lernmark, K. Lahti, T. Forsén, T. Tuomi, A.H. Rosengren, and L. Groop, Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables, *Lancet Diabetes Endocrinol.* (2018). doi:10.1016/S2213-8587(18)30051-2.

[4] S.J. Shah, D.H. Katz, S. Selvaraj, M.A. Burke, C.W. Yancy, M. Gheorghiade, R.O. Bonow, C.C. Huang, and R.C. Deo, Phenomapping for novel classification of heart failure with preserved ejection fraction, *Circulation* (2015). doi:10.1161/CIRCULATIONAHA.114.010637.

[5] T. Ahmad, M.J. Pencina, P.J. Schulte, E. O'Brien, D.J. Whellan, I.L. Piña, D.W. Kitzman, K.L. Lee, C.M. O'Connor, and G.M. Felker, Clinical implications of chronic heart failure phenotypes defined by cluster analysis, *J. Am. Coll. Cardiol.* (2014). doi:10.1016/j.jacc.2014.07.979.

[6] B.I. Buijck, S.U. Zuidema, M. Spruit-van Eijk, H. Bor, D.L. Gerritsen, and R.T.C.M. Koopmans, Is patient-grouping on basis of condition on admission indicative for discharge destination in geriatric stroke patients after rehabilitation in skilled nursing facilities? The results of a cluster analysis, *BMC Health Serv. Res.* **12** (2012) 443. doi:10.1186/1472-6963-12-443.

[7] E. Bair, Semi-supervised clustering methods, *Wiley Interdiscip. Rev. Comput. Stat.* (2013). doi:10.1002/wics.1270.

[8] H. Liu, X. Li, G. Xie, X. Du, P. Zhang, C. Gu, and J. Hu, Precision cohort finding with outcome-driven similarity analytics: A case study of patients with atrial fibrillation, *Stud. Health Technol. Inform.*, (2017). doi:10.3233/978-1-61499-830-3-491.

[9] J. Sun, F. Wang, J. Hu, and S. Edabollahi, Supervised patient similarity measure of heterogeneous patient records, *ACM SIGKDD Explor. Newsl.* **14** (2012) 16. doi:10.1145/2408736.2408740.

[10] R. Miotto, F. Wang, S. Wang, X. Jiang, and J.T. Dudley, Deep learning for healthcare: review, opportunities and challenges, *Brief. Bioinform.* (2018). doi:10.1093/bib/bbx044.

[11] M.X. Eryu Xia, Jing Mei, Guotong Xie, Xuejun Li, Zhibin Li, Learning Doctors' Medicine Prescription Pattern for Chronic Disease Treatment by Mining Electronic Health Records: A Multi-Task Learning Approach, *AMIA*, 2017.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, and I. Polosukhin, Attention Is All You Need [Transformer], in: 31st Conf. Neural Inf. Process. Syst. (NIPS 2017), 2017. doi:10.1017/S0952523813000308.

[13] E. Choi, M.T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. Stewart, Retain: An interpretable predictive model for healthcare using reverse time attention mechanism, *Adv. Neural Inf. Process. Syst.*, 2016.

[14] R.H. Christenson, and H.M. Azzazy, Biochemical markers of the acute coronary syndromes., *Clin. Chem.* (1998).

[15] Y. Zhu, Y. Lin, W. Yan, Z. Sun, Z. Jiang, B. Shen, X. Jiang, and J. Shi, Novel biomarker MicroRNAs for subtyping of acute coronary syndrome: A bioinformatics approach, *Biomed Res. Int.* (2016). doi:10.1155/2016/4618323.

[16] H.A. DeVon, C.J. Ryan, S.H. Rankin, and B.A. Cooper, Classifying subgroups of patients with symptoms of acute coronary syndromes: A cluster analysis, *Res. Nurs. Health*. (2010). doi:10.1002/nur.20395.

[17] B. Riegel, A.L. Hanlon, S. McKinley, D.K. Moser, H. Meischke, L. V. Doering, P. Davidson, M.M. Pelter, and K. Dracup, Differences in mortality in acute coronary syndrome symptom clusters, *Am. Heart J.* (2010). doi:10.1016/j.ahj.2010.01.003.

[18] D. Monzani, M. D'Addario, F. Fattirolli, C. Giannattasio, A. Greco, F. Quarenghi, and P. Steca, Clustering of Lifestyle Risk Factors in Acute Coronary Syndrome: Prevalence and Change after the First Event, *Appl. Psychol. Heal. Well-Being*. (2018). doi:10.1111/aphw.12141.

[19] K. Magoev, V. V. Krzhizhanovskaya, and S. V. Kovalchuk, Application of clustering methods for detecting critical acute coronary syndrome patients, *Procedia Comput. Sci.* (2018). doi:10.1016/j.procs.2018.08.277.

[20] S. van Buuren, and K. Groothuis-Oudshoorn, Buuren, S. V., & Groothuis-Oudshoorn, K. (2010). mice: Multivariate imputation by chained equations in R, *J. Stat. Softw.* **1** (2010).

[21] R. Caruana, Multitask Learning, *Mach. Learn.* (1997). doi:10.1023/A:1007379606734.

Address for correspondence

Eryu Xia, PhD, IBM Research – China, 28, Zhongguancun Software Park, Haidian District, Beijing, China. Email: xerbj@cn.ibm.com.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190264

Stratified Mortality Prediction of Patients with Acute Kidney Injury in Critical Care

Zhenxing Xu[a], Yuan Luo[b], Prakash Adekkanattu[a], Jessica S. Ancker[a], Guoqian Jiang[c], Richard C. Kiefer[c], Jennifer A. Pacheco[b], Luke V. Rasmussen[b], Jyotishman Pathak[a], Fei Wang[a]

[a] *Weill Cornell Medicine, Cornell University, New York, New York, USA*
[b] *Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA*
[c] *Mayo Clinic, Rochester, Minnesota, USA*

Abstract

Acute Kidney Injury (AKI) is the most common cause of organ dysfunction in critically ill adults and prior studies have shown AKI is associated with a significant increase of the mortality risk. Early prediction of the mortality risk for AKI patients can help clinical decision makers better understand the patient condition in time and take appropriate actions. However, AKI is a heterogeneous disease and its cause is complex, which makes such predictions a challenging task. In this paper, we investigate machine learning models for predicting the mortality risk of AKI patients who are stratified according to their AKI stages. With this setup we demonstrate the stratified mortality prediction performance of patients with AKI is better than the results obtained on the mixed population.

Keywords:

Acute Kidney Injury, Critical Care, Forecasting

Introduction

Acute Kidney Injury (AKI) is defined as a sudden (within hours) decrease in nephritic function, including both renal structural damage and functional disorder [1]. It has been shown AKI is the most common cause of organ dysfunction in the critical care setting, and AKI is associated with a significant increase on the morbidity and mortality risk [2]. Early prediction of the mortality risk for AKI patients can help clinical decision makers better understand patient conditions in time and take appropriate actions.

There has been prior research on building mortality risk prediction models for AKI patients. For example, Luo et al. [3] created a scoring model based on multivariate logistic regression to predict the 90-day mortality risk of AKI patients. Skarupskienė et al. [4] adopted logistic regression to predict the mortality risk of AKI patients requiring renal replacement therapy after cardiac surgery. Demirjian et al. [5] used logistic regression to predict the 60-day mortality risk of the AKI patients who are enrolled in the Veterans Affairs/National Institutes of Health Acute Renal Failure Trial Network study. Ohnuma et al. [6] compared the ability of various mortality prediction models in a retrospective, multi-centric cohort of 343 Japanese patients with AKI requiring continuous renal replacement therapy in 14 ICUs captured between January and December 2010.

All of these existing research tried to build a unique predictor for the entire AKI patient cohort. However, it has been demonstrated AKI is a heterogeneous disease with various kinds of causes and clinical manifestations [7]. In this case, it would be challenging to build a single model that is capable of predicting the mortality risk over the entire AKI patient population.

With the above considerations and previous studies [8,9], we propose to perform a stratified prediction in this paper. Specifically, we will first stratify the AKI patients according to their disease stage (I, II, and III). Then, a mortality risk prediction model will be built for each patient strata. In addition to logistic regression, we will also test the performance of other machine learning models including random forest and gradient boosting tree. The Medical Information Mart for Intensive Care III (MIMIC-III) database [10] is used for empirical evaluations. Our results demonstrate the performance of stratified mortality prediction considering different AKI stages is higher than the mortality prediction on the entire mixed population. Additionally, the experiment results show different features play different roles for mortality prediction in patients with different AKI stages.

Methods

Data source

The Medical Information Mart for Intensive Care III (MIMIC-III) database is employed to extract patient data [10], which includes 58,976 admissions of patients. This database has comprehensive information (e.g., patient demographics and vital signs) regarding ICU admissions, which is open-source and freely accessible. The MIMIC-III dataset is passive and de-identified [11], which is in compliance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule and does not make a significant impact on patient privacy.

AKI stages

The criteria of AKI stages has experienced some development progress based on the studies of researchers and clinicians. In particular, there are four AKI criteria: Risk-Injury-Failure-Loss-End (RIFLE) criteria [12], pediatric RIFLE (pRIFLE) criteria [13], Acute Kidney Injury Network (AKIN) criteria [14], and Kidney Disease: Improving Global Outcomes (KDIGO) criteria [15]. KDIGO criteria unified previous criteria in 2012, which improves the sensitivity of AKI diagnostic criteria and has been widely used by researchers and physicians [1]. In this study, we employed the KDIGO criteria to stratify patients into different AKI stages:

Stage 1: 1.5-1.9 times baseline, which is known or presumed to have occurred within the prior 7 days; or not less than 0.3 mg/dL (not less than 26.5 mol/L) absolute increase in serum creatinine (SCr); or urine volume less than 0.5 mL/kg/h for 6-12 hours.

Stage 2: SCr not less than 2.0-2.9 times baseline; or urine volume less than 0.5 mL/kg/h for not less than 12 hours.

Stage 3: SCr not less than 3.0 times from baseline; or increase in SCr not less than 4.0 mg/dL (not less than 353.6 mol/L); or

initiation of renal replacement therapy; or urine volume less than 0.3 mL/kg/h for not less than 24 hours; or anuria for not less than 12 hours; or renal replacement therapy required.

Patient features

A large number of features from MIMIC-III are extracted, which is shown as follows. Additional details about features can be found at: https://github.com/xuzhenxing2018/amia/blob/master/features_name.xlsx.

1. Demographics: Gender, age, and ethnicity.
2. Medications: Medications the patient took from the patients' ICU admission until prediction time. We focused on the following categories: diuretics, non-steroidal anti-inflammatory drugs (NSAID), radiocontrast agents, and angiotensin.
3. Comorbidities: We considered all comorbidities (e.g., congestive heart failure, peripheral vascular, hypertension, diabetes, liver disease, myocardial infarction (MI), coronary artery disease (CAD), cirrhosis, and jaundice) that patients already had. There were a total of 31 comorbidities in our dataset.
4. Chart-events: Vital signs measured at the bedside. We mainly focused on diastolic blood pressure (DiasBP), glucose, heart rate, mean arterial blood pressure (MeanBP), respiration rate, blood oxygen saturation level (SpO2), systolic blood pressure (SysBP), and temperature.
5. Lab-events: Laboratory test results. We considered bicarbonate, blood urea nitrogen (BUN), calcium, chloride, creatinine, hemoglobin, international normalized ratio (INR), platelet, potassium, prothrombin time (PT), partial thromboplastin time (PTT), and white blood count (WBC).

Data pre-processing

We mainly pre-processed two types of features: time-dependent continuous features and discrete features. For continuous features (e.g., lab and chart-events), we computed statistics including the first, last, average, minimum, maximum, slope, and the count based on observations during the observation window. The final encoding of the continuous features was represented by a vector representation with real numbers. The discrete features (e.g., medication and comorbidities) were encoded as zero-one multi-hot vectors. If there were missing demographics and discrete features for an ICU stay record, we deleted that ICU stay. Finally, each ICU stay was represented by a feature vector with a 224 dimension, which was indexed by icustay_id.

Experimental setting

In this study, we predicted whether a patient with AKI in different stages would die within the next seven days. Specifically, a predictive modeling setting with a rolling observation window design was adopted. We supposed t was the elapsed time (in hours) after a patient was admitted to the ICU, and the patient records in t were used to predict the mortality risk for the next seven days, where t took the value of 24, 48, 72, 96, 120, and 144 hours. An illustration of this rolling window design is demonstrated in Figure 1.

For each ICU stay, the AKI severity stage was identified based on whether the criteria in KDIGO stages were satisfied in an observation window, and then we put this ICU stay into corresponding cohorts (AKI Stage-I cohort, AKI Stage-II cohort, and AKI Stage-III cohort). If an ICU stay experienced multiple AKI stages during the observation window (from ICU admission to prediction point), we only chose the last AKI stage. Based on the built three cohorts, we predicted whether a patient with AKI in different stages would die within the next seven days. If an ICU stay did not meet one of the three AKI stage criteria in the observation window, we excluded the ICU stay from the prediction.

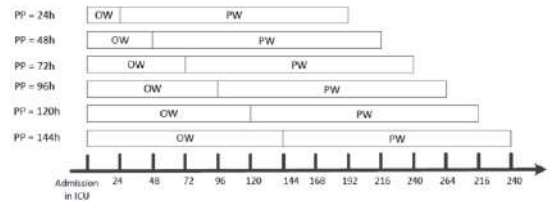


Figure 1 - An illustration of mortality prediction for patients with AKI in different stages. PP: Prediction Point; OW: Observation Window; PW: Prediction Window

Predictive models

Several popular machine learning methods including Logistic Regression (LR) [16], L2 norm regularized Logistic Regression (Ridge) [17], Random Forest (RF) [18], and Gradient Boosting Decision Tree (GBDT) [19] were utilized to build predictive models using free and publicly available software. Specifically, for the implementations of LR, Ridge, and RF, we employed the Scikit-learn software library [20]. For the GBDT, we used the XGBoost software library [21]. For each predictive model, we used the 5-fold cross validation to assess the performance of these algorithms, and employed several popular and important metrics (AUC, recall, and precision) to evaluate the performance of these models.

Note that, the number of cases (mortality patient with AKI) and controls (alive patient with AKI) was imbalanced (e.g., an approximate 1:10 case to control ratio in 24 hours for AKI Stage-I) in our dataset. More statistics for all AKI stages are shown in Table 1. With such an imbalanced dataset, most classifiers had a potential to support the majority class (alive patient with AKI) because they were considered to maximize the overall number of correct predictions, thus resulting in poor performance in the minority class (mortality patient with AKI) prediction [22].

To address this imbalanced issue, we used case-control matching techniques [23] in this study, which matched each case with a control by considering the APACHE II score [24], Charlson comorbidity index [25], and demographic information. A matched control met two conditions: (1) has the same gender as the case and the age difference is within 5 years; (2) has the highest similarity score with the case measured by Manhattan distance on the basis of APACHE II and Charlson comorbidity index. By this technique, a resampled balanced training set was constructed with matched cases and controls.

Results

Comparison of different methods with all patient features

We tested the performance of different predictive models with different AKI stages using all patient features with varying data observation windows. Figure 2 shows the performance of these methods in terms of AUC. In order to compare the performance of mortality prediction on subpopulations completely, we constructed a mixed cohort, which combined all the patients with different AKI stages together. This mixed cohort is represented as Stage-I-II-III in Figure 2. For the performance in terms of precision and recall, please refer to https://github.com/xuzhenxing2019/MedInfo.

Table 1 - The number of mortality and alive patients with AKI in different stages

		24h	48h	72h	96h	120h	144h
Stage-I	Total	7506	8181	5801	4252	3222	2416
	Mortality	756	816	715	647	567	493
	Alive	6750	7365	5086	3605	2655	1923
	Mortality Rate	10.07%	9.97%	12.33%	15.22%	17.6%	20.41%
Stage-II	Total	3216	2895	2025	1576	1267	1007
	Mortality	514	514	421	386	347	310
	Alive	2702	2381	1604	1190	920	697
	Mortality Rate	15.98%	17.75%	20.79%	24.49%	27.39%	30.78%
Stage-III	Total	408	414	404	398	380	367
	Mortality	188	167	176	159	158	161
	Alive	220	247	228	239	222	206
	Mortality Rate	46.08%	40.34%	43.56%	39.95%	41.58%	43.87%

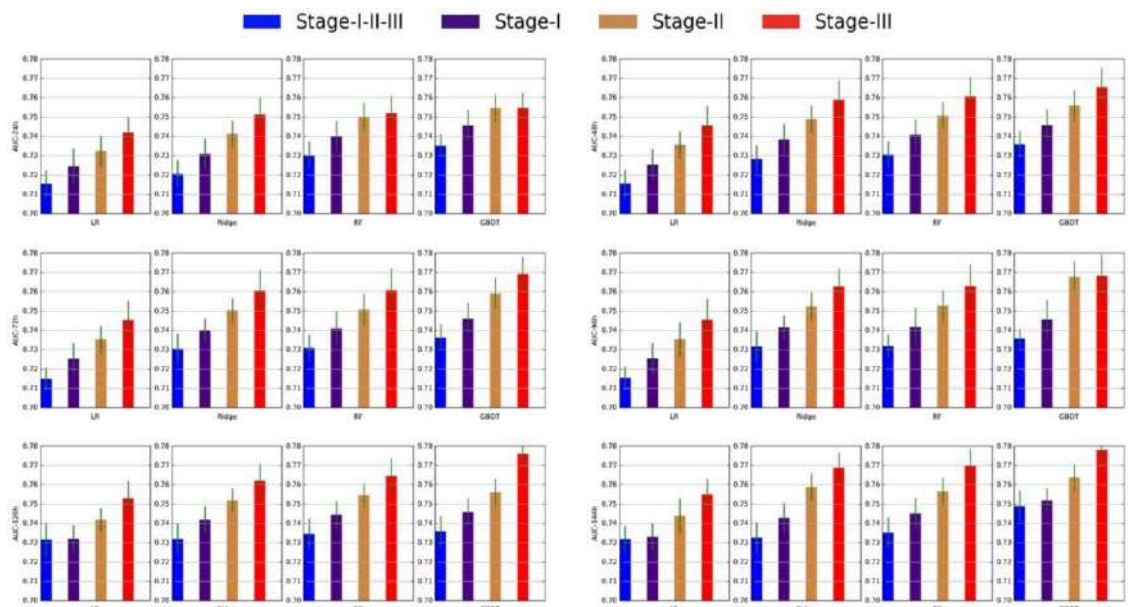


Figure 2 - The AUC of different methods with 24-144 hours of data observation window in terms of different AKI stages

From Figure 2, we observed that:

1. The technique of stratifying patients with AKI into different stages obtained a better performance compared to the mixed cohort. This provided some evidence for our assumption that stratifying patients with AKI into different stages has the potential to improve the performance of mortality prediction because of AKI's complex etiology and pathophysiology.
2. The technique of constructing a mortality predictor based on different cohorts has the potential to consider different AKI causes. This predictive modelling strategy can be viewed as a case for local learning [26], which first splits the data space into multiple local areas and constructs different predictors for the areas.
3. Comparing the Stages, Stage-III shows a higher mortality predictive performance. One implication is the AKI patients in Stage-III who are going to die had more distinctive clinical manifestations compared to the control patients.
4. Comparing the performance of four predictors, GBDT acquires better results.

Comparison of feature group predictability with GBDT

Five different groups of features were employed to build mortality predictors for patients with different AKI stages. In this section we explored the prediction performance of GDBT with different feature groups within a 24 hour data observation window. The mortality prediction results are demonstrated in Table 2.

We observed feature groups played different roles when patients were in different AKI stages. For example, the laboratory features acquired better performance than other feature groups when patients were diagnosed with AKI Stage-

I. Chart features played a more important role in predicting mortality when patients were in Stage-II and III. One potential reason for this is the chart events had a finer time resolution, and for patients with a later stage of AKI, these events provided more in-time information of the patients' condition which progressed rapidly. In addition, there were no obvious differences between comorbidities and medications feature groups. The demographic feature group contributed the least to mortality prediction in patients with different stages.

Table 2 - The performance of different group features in all AKI stages on the basis of 24 hours data

	Demographics	**Medications**	**Comorbidities**	**Chart-events**	**Lab-events**
Stage-I	0.6823±0.0219	0.6971±0.0218	0.7027±0.0233	0.7031±0.0241	**0.7237±0.0223**
Stage-II	0.7029±0.0222	0.7128±0.0221	0.7139±0.0237	**0.7339±0.0246**	0.7331±0.0243
Stage-III	0.7134±0.0238	0.7236±0.0247	0.7239±0.0251	**0.7424±0.0255**	0.7334±0.0249

The important features chosen from all feature groups

There were 224 features extracted in total for mortality prediction in different AKI stages. The importance of each factor in mortality prediction was further explored in our experiments. The GBDT model was employed to acquire the importance score of each feature based on 24 hour data. The range of the feature importance score for mortality prediction in AKI Stage-I, Stage-II, and Stage-III were 0.0015-0.0457, 0.0015-0.0556, and 0.0019-0.0316, respectively. The top ten features are shown in Table 3.

From the table, we can observe HeartRate, CREATININE, CHLORIDE, WBC, HEMOGLOBIN, and RespRate were more important because there were strong associations between these features and mortality in patients with AKI. These results aligned well with some previous studies [27-32]. For example, the magnitude of increase in serum creatinine levels can determine the severity of AKI [27], which was associated with worse survival rates [28]. Chloride levels are also associated with the severity of AKI [29] and Shaw et al. demonstrated an association between higher intravenous chloride loads and hospital mortality [30].

Table 3 - The top 10 features selected from all feature groups on the basis of importance score

	Stage-I		**Stage-II**		**Stage-III**	
	Features	**Importance**	**Features**	**Importance**	**Features**	**Importance**
1	POTASSIUM_count	0.0457	HEMOGLOBIN_slope	0.0556	HeartRate_last	0.0316
2	metastatic_cancer	0.0365	HeartRate_last	0.0278	MeanBP_last	0.0247
3	CREATININE_avg	0.0350	CHLORIDE_slope	0.0200	HEMOGLOBIN_slope	0.0237
4	BUN_last	0.0289	WBC_min	0.0185	RespRate_last	0.0237
5	PTT_avg	0.0274	CREATININE_slope	0.0173	BICARBONATE_slope	0.0221
6	WBC_min	0.0243	CHLORIDE_max	0.0170	CREATININE_slope	0.0217
7	PLATELET_last	0.0223	SysBP_last	0.0168	BUN_slope	0.0198
8	HeartRate_max	0.0213	HeartRate_avg	0.0164	Glucose_min	0.0198
9	Temp_avg	0.0213	Temp_last	0.0156	DiasBP_last	0.0198
10	RespRate_avg	0.0182	HeartRate_max	0.0154	age	0.0178

Discussion and Conclusions

Accurate prediction of the mortality risk for AKI patients is helpful for clinicians to understand the condition of the patients and take appropriate actions. AKI has a complex etiology and pathophysiology, so it is challenging to construct a single model that can accurately predict the mortality risk over the entire AKI patient cohort. This study investigated the impact of different AKI subpopulations on the performance of mortality prediction. We stratified the AKI patients on the basis of their disease stage (I, II, and III), and then a mortality risk predictor was constructed for each patient cohort. Several popular machine learning models (e.g., logistic regression, RF, and GBDT) were built based on these subpopulations for the mortality prediction in patients with AKI. GBDT showed a better performance than other methods for the mortality prediction in this study.

Prior models of risk prediction of mortality in critically ill patients with AKI typically employed mixed cohorts [5,6]. The utilization of stratification mechanisms performed in this study provides a chance to predict mortality in patients with AKI specific to each stage. This technique of constructing a mortality predictor based on different cohorts has the potential to consider different AKI causes. In addition, dynamic real-time data may allow for clinical monitoring of mortality risk beyond a solely static prediction upon ICU arrival. The utilization of dynamic clinical monitoring algorithms for mortality prediction may allow for its incorporation into the electronic medical record [33].

With the ability to identify patients at high risk of mortality by employing real-time data may allow for initiation of earlier interventions to prevent or reduce mortality [34]. It might be useful to incorporate more information (e.g., imaging and genomic biomarkers) to further improve the performance of dynamic clinical mortality risk prediction models. In the future, we will consider some advanced models (e.g., recurrent network) with stronger learning capabilities to improve the performance of prediction.

Acknowledgements

This work was supported in part by NIH Grants 2R01GM105688-06 and 1R21LM012618-01.

References

[1] K. Makris and L. Spanou, Acute kidney injury: Definition, pathophysiology and clinical phenotypes, *The Clinical Biochemist Reviews* **37** (2016) 85-98.

[2] J.F. Doyle and L.G. Forni, Acute kidney injury: Short-term and long-term effects, *Critical Care* **20** (2016) 188.

[3] M. Luo, Y. Yang, J. Xu, et al., A new scoring model for the prediction of mortality in patients with acute kidney injury, *Scientific Reports* **7** (2017) 7862.

[4] I. Skarupskiené, D. Adukauskiené, J. Kuzminskiené, et al., Mortality prediction in patients with acute kidney injury requiring renal repalcement therapy after cardiac surgery, *Medicina (Kaunas, Lithuania)* **53** (2017) 217-223.

[5] S. Demirjian, G.M. Chertow, J.H. Zhang, et al., Model to predict mortality in critically ill adults with acute kidney injury, *Clin J Am Soc Nephrol* **6** (2011) 2114-2120.

[6] T. Ohnuma, S. Uchino, N. Toki, et al., External validation for acute kidney injury severity scores: A multicenter retrospective study in 14 Japense ICUs, *American Journal of Nephrology* **42** (2015) 57-64.

[7] R. Bellomo, J.A. Kellum, and C. Ronco, Acute kidney injury, *Lancet (London, England)* **380** (2012) 756-766.

[8] C.C. Miller, M.J. Reardon, and H.J. Safi, Risk Stratification: A Practical Guide for Clinicians, 1st ed., Cambridge University Press, Cambridge, 2001.

[9] V.B. Kraus, F.J. Blanco, M. Englund, et al., Call for standardized definitions of osteoarthritis and risk stratification for clinical trials and clinical use, *Osteoarthritis and Cartilage* **23** (2015) 1233-1241.

[10] A.E. Johnson, T.J. Pollard, L. Shen, et al., MIMIC-III, a freely accessible critical care database, *Scientific Data* **3** (2016) 160035.

[11] H. Mohamadlou, A. Lynn-Palevsky, C. Barton, et al., Prediction of acute kidney injury with a machine learning algorithm using electronic health record data, *Can J Kidney Health Dis* **5** (2018) 2054358118776326.

[12] R. Bellomo, C. Ronco, J.A. Kellum, et al., Acute renal failure - Definition, outcome measures, animal models, fluid therapy and information technology needs: The second international consesnuss conference of the acute dialysis quality initiative (ADQI) group, *Critical Care (London, England)* **8** (2004) R204-R212.

[13] A. Akcan-Arikan, M. Zappitelli, L.L. Loftis, et al., Modified RIFLE criteria in critically ill children with acute kidney injury, *Kidney Int* **71** (2007) 1028-1035.

[14] J.W. Pickering and Z.H. Endre, GFR shot by RIFLE: Errors in staging acute kidney injury, *Lancet (London, England)* **373** (2009) 1318-1319.

[15] J.A. Kellum, N. Lameire, P. Aspelin, et al., Kidney disease: Improving global outcomes (KDIGO) acute kidney injury work group. KDIGO clinical practice guideline for acute kidney injury, *Kidney International Supplements* **2** (2012) 1-138.

[16] D.W. Hosmer, S. Lemeshow, and R.X. Sturdivant, Applied Logistic Regression, 3rd ed., John Wiley & Sons, Inc., Hoboken, 2013.

[17] S. Le Cessie and J.C. Van Houwelingen, Ridge estimators in logistic regression, *Journal of the Royal Statistical Society, Series C (Applied Statistics)* **41** (1992) 191-201.

[18] L. Breiman, Random forests, *Machine Learning* **45** (2001) 5-32.

[19] J.H. Friedman, Stochastic gradient boosting, *Computational Statistics & Data Analysis* **38** (2002) 367-378.

[20] F. Pedregosa, G. Varaoquaux, A. Gramfort, et al., Scikit-learn: Machine learning in python, *Journal of Machine Learning Research* **12** (2011) 2825-2830.

[21] T. Chen and C. Guestrin, XGBoost: A scalable tree boosting system, In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, San Francisco, CA, USA, 2016, 785-794.

[22] P. Cheng, L.R. Waitman, Y. Hu, et al., Predicting inpatient acute kidney injury over different time horizons: How early and accurate?, In *AMIA 2017 Annual Symposium Proceedings*, Washington, D.C., USA, 2017, 565-574.

[23] J.B. Cologne and Y. Shibata, Optimal case-control matching in practice, *Epidemiology (Cambridge, Massachussetts)* **6** (1995) 271-275.

[24] W.A. Knaus, E.A. Draper, D.P. Wagner, et al., APACHE II: A severity of disease classification system, *Critical Care Medicine* **13** (1985) 818-829.

[25] M.E. Charlson, P. Pompei, K.L. Ales, et al., A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation, *Journal of Chronic Diseases* **40** (1987) 373-383.

[26] L. Bottou and V. Vapnik, Local learning algorithms, *Neural Computation* **4** (1992) 888-900.

[27] J.A. Kellum, F.E. Sileanu, R. Murugan, et al., Classifying AKI by urine output versus serum creatinine level, *J Am Soc Nephrol* **26** (2015) 2231-2238.

[28] A. Linder, C. Fjell, A. Levin, et al., Small acute increases in serum creatinine are associated with decreased long-term survival in the crticially ill, *Am J Respir Crit Care Med* **189** (2014) 1075-1081.

[29] H.J. Oh, S. Kim, J.T. Park, et al., Baseline chloride levels are associated with the incidence of contrast-associated acute kidney injury, *Scientific Reports* **7** (2017) 17431.

[30] A.D. Shaw, K. Raghunathan, F.W. Peyerl, et al., Association between intravenous chloride load during resuscitation and in-hospital mortality among patients with SIRS, *Intensive Care Medicine* **40** (2014) 1897-1905.

[31] J. Bouchard, S.B. Soroko, G.M. Chertow, et al., Fluid accumulation, survival and recovery of kidney function in crticially ill patients with acute kidney injury, *Kidney International* **76** (2009) 422-427.

[32] Y.C. Chen, F.C. Tsai, C.H. Chang, et al., Prognosis of patients on extracorporeal membrane oxygenation: The impact of acute kidney injury on mortality, *The Annals of Thoracic Surgery* **91** (2011) 137-142.

[33] J.L. Koyner, R. Adhikari, D.P. Edelson, et al., Development of a multicenter ward-based AKI prediction model, *Clin J Am Soc Nephrol* **11** (2016) 1935-1943.

[34] J. Keim-Malpass, K.B. Enfield, J.F. Calland, et al., Dynamic data monitoring improves predictive analytics for failed extubation in the ICU, *Physicological Measurement* **39** (2018) 075005.

Address for Correspondence

Fei Wang: Division of Health Informatics, Department of Healthcare Policy and Research, Weill Cornell Medical College, Cornell University. 425 E 61 St. New York City. NY 06032.

Email: few2001@med.cornell.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190265

Mapping the Hyperlink Structure of Diabetes Online Communities

Hongyi Shi[a], Fabien Pfaender[b], Marie-Christine Jaulent[a]

[a] INSERM, UMR_S 1142, LIMICS, F-75006, Paris, France
Sorbonne Université, Paris 06, UMR_S 1142, LIMICS, F-75006, Paris, France
Université Paris 13, Sorbonne Paris Cité, LIMICS, (UMR_S 1142), F-93430, Villetaneuse, France
[b] Fabien Pfaender, UTSEUS, Shanghai University, Shanghai, China
Costech EA2223, Université de Technologie de Compiègne, Compiègne, France

Abstract

Diabetes is one of the largest global health emergencies of the 21st century. As a chronic disease, diabetes requires continuous medical care and constant patient self-management. Such care involves several stakeholders to improve health outcome and patient quality of life. This paper makes use of World Wide Web network analysis to highlight how stakeholders, providing information about online diabetes communities, link to each other. To achieve this, we capture the network of diabetes related websites as a digital trace of a non-digital phenomenon. Furthermore, this helps us to understand the current situation of diabetes organizations from a digital perspective. The methodology involves state-of-the-art tools to crawl (Hyphe) and visualize (Gephi) topic-sensitive networks. While neither of these tools is new in itself, their combination provides a promising way to analyze chronic disease stakeholders, organizations and communities, representing a large proportion of the knowledge and support diabetes patients have access to nowadays.

Keywords:

Diabetes Mellitus, Data Analysis, Data Visualization

Introduction

Diabetes is a chronic disease that occurs either when the pancreas does not produce enough insulin or when the body cannot effectively use the insulin it produces. According to a WHO report, an estimated 1.6 million deaths were directly caused by diabetes in 2015. Diabetes is also a major cause of blindness, renal failure, heart attacks, stroke, and lower limb amputations. The global economic cost of diabetes in 2014 was estimated to be US$612 billion. In the past two years, there were 415 million people with diabetes and this figure could rise to 642 million by 2040 [1].

Therefore, the recent dramatic rise of diabetes occurrences raises several questions that need to be addressed: how do people react after they are diagnosed with diabetes? What will they do for long term self-management? Where can they find relevant information and how easy is it to access it? How can they get psychological support and community help?

Diabetes can be treated and its consequences could be avoided or delayed with a specific diet, physical activity, medication and regular screening and treatment for complications. This helped us establish a preliminary list of possible key stakeholders: hospitals, physicians, research laboratories, pharmaceutical companies, medical and social relay structures, associations, health-related NGOs (non-governmental organizations), patients and their families, publications around diabetes care. All of these key diabetes actors play their own roles in the health care system but the nature of their relationship with each other, if any, remains unexplored outside of a patient-centric paradigm. Our hypothesis is that all stakeholders involved in diabetes have connections of some sort that can be revealed on the World Wide Web (abbreviated WWW or the Web) as an organized world of communities rather than a randomly organized network. This study aims to visualize the network of diabetes in the digital world represented by the Web to find key health web players connections with each other and to figure out the nature of their relationships.

Since its invention at the beginning of the 1990's, the Web has gradually become a major media platform itself hosting billions of resources accessible from a URL (Uniform Resource Locator) and dozens of sub-media in the form of specific websites hosting social media (like Facebook or Twitter), scientific knowledge or any other niche [2]. The Web is also the source of scientific investigation in itself and the network structure of a hyperlinked environment can be a rich source of information for whom looking for an effective means to collect, analyze and understand it [3; 4].

Exploring web structures is a proxy to human organizational structures [5]. Thus, we propose to map the hyperlink structure of diabetes stakeholders online. As a matter of fact, the web contains crucial aspects of the embodiment of social factors: personal blogs, institutional websites, health-oriented media, etc. To address the relationship with stakeholders in the diabetes area, we have examined how classical studies of network explorations can be brought to use in the context of diabetes to answer the following questions: Who is connected to whom by which means? Which organizations receive support from which ones? What resources or information are published through which platforms? What is the relationship between charities and governmental agencies? How individuals, companies, and organizations interact together? What is the ecosphere of the diabetes world? Therefore, our objective is to provide a replicable practical methodology to visualize the diabetes network of websites to make sure all key health web players have connections with one and another. And then, the previous questions can be addressed by studying the resulting map of diabetes online communities. A corollary expected impact is that this map can help the diabetics to access useful information regarding the neighboring context (such as information on diet or physical activity) that search engines do not display at first when looking for information about diabetes. At the same time,

this can also help pharmaceutical companies and NGOs to find influencers or key opinion leaders and help governmental agencies to take effective actions that reflect more efficiently the needs of diabetes communities. The visual representation will be helpful to quickly and effectively identify areas on the graph where improvements can be made.

To achieve our objectives, we take advantage of two state-of-art tools to assist the creation of website networks maps on diabetes. These two tools are combined in a robust methodology to collect the websites links and visualize it in detail. They are presented in the next section.

Background

Web networks belong to a complex network family that display properties such a power law distribution of the node degrees, high clustering coefficient or small world phenomenon [6]. While these properties often contribute to networks analysis, they are less effective to explore local structures inside the network. In order to do so, one can use mature large graphs visualization software. Graph visualization is an effective approach to spatialize the complex networks [7]. Using a visual display helps identify features in a network structure and data while relying on human very efficient perceptual abilities.

Creating and analyzing topic specific websites network is a methodology that has been applied to various contexts, although not yet in the healthcare field. In order to produce a visualization map of diabetes online communities which does not exist, we propose to combine two separate existing tools that could support the creation of such map in the context of diabetes. The map will provide proxies to generate key insights about the community.

Hyphe[1] is a web crawler[2], developed by the Médialab department of Sciences Po[3] in Paris, France, which is a tool to build web corpus by crawling web pages and generating networks between "web entities"[4] [8], connected with each other using hyperlinks. Hyphe has been successfully used in many published papers in social science. Hyphe allows to methodically gather web entities and visualize the network aiming at data journalists [9]. A team of librarians and sociologists delineated the debate on climate change on the web, notably measuring the unexpected large presence of climate skeptics [8]. Even more, people tried to study through the "Alternative for Germany" (AFD) Facebook wall which has already become one of the largest right-wing forums on the German-speaking internet to see how social media plays a crucial role for the party's mobilization strategy [10]. Yet, Hyphe has never been used in public health care to assess chronical disease online communities' structure. Moreover, even if Hyphe does include basic network visualization capabilities, it quickly reaches its limits as the networks grows beyond one hundred nodes.

Gephi[5] is an open source software for network visualization and analysis. It uses a 3D render engine to display large networks in real-time allowing researchers to explore their network. Gephi embeds a large variety of algorithms, filters and layouts to provide fine tuning over the network analysis and display [11]. It can deal with large network (i.e. over 20,000 nodes) and, because it is built on a multi-tasks model, it takes advantage of multi-core processors. In addition, nodes and edges designs can be personalized to create map-like visualization for communication purposes, ideal for a diabetes community map.

In this context, we have developed a methodology to uncover the online organization of diabetes websites. We leveraged and made best out of these tools to collect information using Hyphe and then visualize their properties and links using Gephi to produce the network on diabetes websites and finally provide key insights about this community.

Methods

The proposed methodology consists in a workflow embedding the following milestones:

- collecting diabetes community websites
- analyzing the resulting websites' structure
- visualizing the diabetes community network as a navigational map

These milestones can be decomposed into a more detailed process potentially applicable to any online community analysis (see figure 1). The methodology thus follows the following steps:

1. gathering a list of starting websites
2. creating a pre-crawl corpus
3. consolidating the pre-crawl corpus
4. launching and performing a full-scale crawling
5. extracting the websites network
6. visualizing the network structure

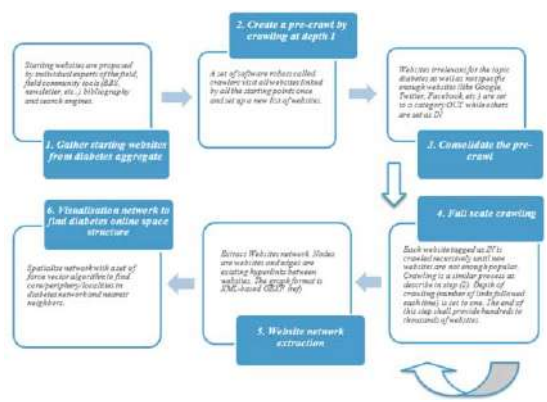


Figure 1– Illustration of the Workflow

These six steps are based on the following decision criteria.

Collection of starting websites: In order to explore an online community consisting of topic related websites, we need to define entry points into the community. These entry points will serve as a gate into the community by following the points links with a robot crawler carefully controlled. As the objective is to

[1] http://hyphe.medialab.sciences-po.fr (accessed on March 28th, 2019)

[2] A web crawler, sometimes called a spider, is an Internet bot that systematically browses the World Wide Web, typically for the purpose of Web indexing.

[3] Sciences-Po is an international research university located in Paris, ranking among the finest institutions in the fields of humanities and social sciences. Médialab is the tenth research center of Sciences-Po to help social sciences and humanities take full advantage of the huge amount of data made available by digitization.

[4] An entity of homogenous content often assimilated to a website hence the use of websites (itself a collection of smaller entities) but not always.

[5] https://gephi.org (accessed on March 28th, 2019)

explore all aspects of the community, the entry points shall reflect different points of view, opinions or topics the community might address so as to make sure to capture enough gates leading to the potential different parts of the final network. To ensure the diversity of entry points we selected a variety of sources starting with (a) two or more experts of the topic suggesting websites (for the diabetes topic, these experts were found in people who are living with diabetes and also working in the diabetes area for several years); (b) using various queries about diabetes on Google to ensure different results and (c) using diabetes community social network pages suggestions.

Choice of depth and the criteria to stop crawling: Once the starting web entities[6] are set, Hyphe offers the possibility to follow their hyperlinks (html markup <a> for anchor) automatically by detecting them in web pages. Hyphe have three depths for crawling, from 1 hyperlink away from the original web entity to 3 hyperlinks away to (3 clicks) depending on how deep the web entities shall be explored. As the web network follows a power law degree distribution, a crawl at depth 3 can potentially gather thousands of websites making it overwhelming for a human to properly assess each and every one of their individual membership to the targeted diabetes community. In this study, to avoid too much noise, we decided to consider only a slower depth level of 1, repeated several times if needed but only for targeted websites whose contribution to the topic is substantial. Moreover, the criteria to stop crawling new web entities depends on how much websites related to diabetes showing up in the final results. If no new diabetes-related entities show up after the crawling, then we considered having reached a local frontier in the diabetes community and proceed with others web entities until no more entities are left to be un-crawled. Then all topic frontiers are discovered. At last, considering the huge numbers of web entities, we did not consider in this first study the entities mentioned only once. It is a web network property that noticeable entities should be linked to by several neighbors. A web entity mentioned only once have little chance to be of importance [12]. We considered a minimum of 3 hyperlinks to an entity which we call it "popular" website for it to be eligible for an additional depth 1 crawling.

Cleaning process to feed each iteration of step 4 with an IN database: To clean the database after each one depth level iteration, we used two ways to make a decision and classify the websites as: IN, OUT or UNDECIDED. In this study, we classified the contents of the web entities strictly related to diabetes as IN; when nothing indicates a potential diabetes-related content, we classified it as OUT[7]; finally, ambiguous websites that mention diabetes among other topics were categorized as UNDECIDED waiting for further in-depth analysis to lift the veil on their classification. The first way to classify the entities is to automatically select in the database the URL string containing the word diabetes and default them as IN. The second way to classify the entities is to select manually the class after a review of the content of each website. This is done by opening its URL in a browser. After the classification, we can filter the database to only retain the IN websites.

Choice of the final pattern in step 6 (Gephi): Once the "pre-crawl" was created by crawling the starting websites at depth 1 and cleaning them, we applied the crawl & clean process again several times to get the full-scale crawling (step 4). Each web entity regarded as IN is crawled recursively until no new websites are popular enough which are cited above 3 times to be included or are all off-topic (OUT or UNDECIDED). The end of this step provides with hundreds of IN websites and their hyperlinks. Then we exported the websites network as a graph in the XML-based GEXF format. In the final network, nodes are websites and edges are existing hyperlinks found between websites (step 5) weighted by count of occurrences. Lastly, we imported the graph to Gephi to visualize the hyperlink structure of diabetes online communities (step 6). In order to do so, after importing the network, we used a force directed "Force Atlas 2" [13] layout well adapted to scale-free networks to spatialize the nodes and their edges. The edges represent hyperlinks only. Therefore, the proximity of the nodes is the result of the force vector-based layout algorithm only and do not demonstrate any physical proximity whatsoever.

Results

As the starting websites were proposed by individual experts of the field, by field community tools (BBS, newsletters, etc.) bibliography and search engines like Google, we combined all the resources and picked up most well-known 15 websites which represent different aspects in the diabetes world as the starting points to feed Hyphe. The following table describes these 15 websites.

Table 1– Starting Websites for Crawling

Different Aspects	Websites
1 portal website	http://www.childrenwithdiabetes.com
1 well-known patient's blogger	https://yogafordiabetesblog.com
1 local association focus on diabetes educator	https://www.diabeteseducator.org
1 study focus on diabetes globally beyond Novo nordisk	https://www.dawnstudy.com
2 diabetes publications	https://asweetlife.org https://diatribe.org
2 leading pharmaceutical companies	https://www.myomnipod.com https://www.dexcom.com
2 international associations	https://www.jdrf.org https://www.idf.org
5 charities	https://www.t1international.com https://beyondtype1.org https://www.worlddiabetesfoundation.org https://www.diabetessisters.org https://diabetesdestiny.com

After the web crawler Hyphe visited all websites linked by all the starting points once, we set up a new list of websites and

[6] A web entity is a web page (a unique URL leading to a media displayed as a unique page) or a collection of pages relevant for a topic. It can range from a single page to a sub domain of a website to the whole website itself. Website is the web entity by default on Hyphe but this can be tuned to adapt to the web heterogeneity of content structure.

[7] Websites very often provide many links to tools or software or popular social network that are not related to the topic at hand but rather offer another medium opportunity (ex linkedin.com, Instagram.com, google.com, etc.).

use the cleaning process to make sure we only keep the IN websites which are strictly talking about diabetes. Once Hyphe discovers as many web entities as possible and their hyperlinks, we get the final number of the DISCOVERED web pages in Hyphe overview part. Each of the discovered websites then is classified as IN/OUT/UNDECIDED. In the end, we successfully captured 430 IN websites with 6587 hyperlinks, each website contains information about diabetes or some aspects of it together with some basic tags on the nature and content of these websites. Figure 2 is the diabetes hyperlink structure as visualized by Hyphe directly. This graph is messy and while it shows a strong core component, there is no visible underlying structure.

Figure 2– Diabetes Hyperlink Structure with Basic Force Directed Layout

In order to visualize the diabetes online communities in more detail, we created the visualization in Gephi. Then to provide a first view at the sub-communities inside the diabetes community we applied a community detection algorithm aimed at detecting clusters of similarly connected nodes [14]. It does so by maximizing the modularity quality metric in all possible partitions. Modularity measures the difference between the edge density in the partition and a randomized graph with the same number of nodes and the same degree distribution [15]. Then a color map is associated with the 5 discovered communities class resulting in the figure 3.

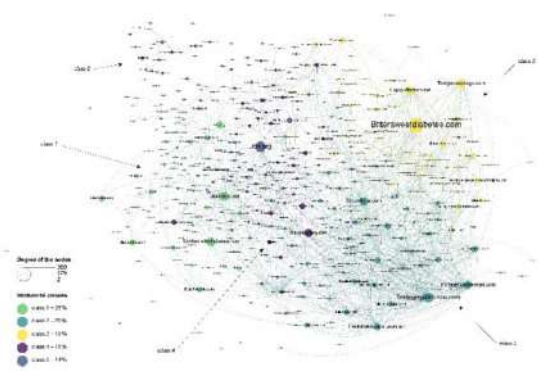


Figure 3– Diabetes Hyperlink Structure with Force Atlas 2 Layout and Communities

The result shows that diabetes online community has its own space and is organized with clusters and distribution of websites of relative importance. More analysis is necessary to show exactly what is the topic but we get a general idea that they don't belong together by chance. It is remarkable that the algorithm working only on node probability highlight these communities that in real life share a common interest when we read their content. If we want to have some general ideas about which community mainly talking about which topic, we need to manually or semantically annotate of websites using terminology of the diabetes fields would be a good way towards explanation the communities' contents. Moreover, we used a degree centrality algorithm [15] which assigns an importance score based on the number of links held by each node to get several central websites in our 430 websites diabetes digital. Node Degree reveals the number of IN and OUT hyperlinks which are from 1 to 303 related to one web entity. Among them:

- bittersweetdiabetes.com is the most popular blog.
- medtronicdiabetes.com, dexcom.com, novonordisk.com, lilly.com, myomnipod.com, tandemdiabetes.com are the main pharmaceutical companies in diabetes industry.
- jdrf.org is the most authoritative association.
- diabetesdaily.com is the mainstream media website talking about diabetes.
- childrenwithdiabetes.com is the largest diabetes online community.
- diabeteseducator.org is the biggest diabetes society.
- integrateddiabetes.com is one consulting institution which offers integrated diabetes consulting services.

Discussion

This approach demonstrates the potential in analyzing and visualizing the diabetes network online. Using Hyphe to extract 430 specific and relative websites, people with diabetes can easily navigate most relevant websites to obtain information and knowledge about their conditions or different aspects of it. Furthermore, such a strong community can offer support from a knowledge and a psychology point of view. Search engine like Google can find global websites by considering their authority in the network such as general organizations like International Diabetes Federation (IDF), American Diabetes Association (ADA), Diabetes Australia, etc. But they cannot offer people with diabetes detailed information to help people manage their diabetes and improve the quality of their lives efficiently. 80% of the iceberg is still under the sea and is difficult to reveal. Hyphe and our methodology is proven reliable to explore the localities, and we were able to find more interesting websites such as personal blogs sharing the stories of the real diabetes daily lives including their struggles to combat diabetes, we can also find different diabetes online communities focusing on the different aspects, nutrition, sports, camping, etc. In addition, some websites are more difficult to detect when they talk about some rare issues, such as diabulimia which is an eating disorder related to type 1 diabetes. However, this study is still limited due to the initial websites where English is solo language to feed hyphe. As a result, it leads to a majority of English-domain websites in the final map despite having some multilingual websites coming from US (using English and Spanish) or Canada (using English and French). Due to the crawler depth limitation and the relatively small number of websites chosen to be presented in the final map, it cannot show the whole world with diabetes but just part of it. A deeper crawl and a more web page lists by reviewing would undoubtedly have more information covered. Future work will focus on the semantic analysis on URLs to indicate to which extend the community is well focused on diabetes topics in general with some underlying trends and identify areas where improvements can be made in the diabetes real world.

Conclusion

We have successfully shown that all stakeholders involved in diabetes have connections on the World Wide Web as an organized world of communities rather than a randomly organized network. In this study, we also provide a replicable practical methodology which combines crawl and visualize tools to analyze a web-based network of diabetes. It is a novel promising way to analyze further chronic disease stakeholders' organizations and communities networking relationships.

Acknowledgements

This paper was made possible by the cooperation program by Sorbonne Université and Chinese Scholarship Council. I would like to thank LIMICS, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé, UMRS_1142 for the support and this work is also supported partly by Complex City Laboratory, Shanghai.

References

[1] International Diabetes Federation*, IDF Diabetes Atlas, 8th Edition,* accessed at http://www.diabetesatlas.org on March 28th , 2019.

[2] D.H. Shutz, What is the difference between the Web and the Internet, *W3C Help and FAQ. W3C*. **200** (2015).

[3] J.M. Kleinberg, Authoritative sources in a hyperlinked environment, *Journal of the ACM (JACM)* **46** (1999), 604-632.

[4] F. Pfaender, M. Jacomy, and G. Fouetillou, Two Visions of the Web: from globality to localities, in: *2006 2nd International Conference on Information & Communication Technologies*, IEEE, 2006, pp. 566-571.

[5] P. Pirolli, J.E. Pitkow, and R. Rao, Silk from a sow's ear: Extracting usable structures from the web, in, Georgia Institute of Technology, 1996..

[6] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, *Reviews of Modern Physics* **74** (2002), 47.

[7] O. Sporns, D.R. Chialvo, M. Kaiser, and C.C. Hilgetag, Organization, development and function of complex brain networks, *Trends in Cognitive Sciences* **8** (2004), 418-425.

[8] M. Jacomy, P. Girard, B. Ooghe-Tabanou, and T. Venturini, Hyphe, a curation-oriented approach to web crawling for the social sciences, in: *Tenth International AAAI Conference on Web and Social Media*, 2016.

[9] T. Venturini, M. Jacomy, L. Bounegru, J. Gray, Visual network exploration for data journalists, *S. E. I. and B. Franklin (Ed.). Abingdon: Routledge,* (2018).

[10] K. Arzheimer, The AfD's Facebook wall: A new hub for far-right mobilisation in Germany, in: *American Political Science Association Conference Annual Meeting*, 2015.

[11] M. Bastian, S. Heymann, and M. Jacomy, Gephi: an open source software for exploring and manipulating networks, in: *Third International AAAI Conference on Weblogs and Social Media*, 2009.

[12] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, *Computer Networks and ISDN Systems* **30**, (1998), 107–117.

[13] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian, ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software, *PloS one* **9** (2014), e98679.

[14] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, Fast unfolding of communities in large networks, *Journal of statistical mechanics: theory and experiment* **2008** (2008), P10008.

[15] M. Chen, K. Kuzmin, and B.K. Szymanski, Community detection via maximization of modularity and its variants, *IEEE Transactions on Computational Social Systems* **1** (2014), 46-65

[16] M. G. Everett, S.P. Borgatti, P.J. Carrington, J. Scott, and S. Wasserman, *Models and methods in social network analysis*, Cambridge University Press, 2005.

Address for correspondence

Hongyi SHI

LIMICS, Escalier D, 2eme étage

15 rue de l'école de médecine, 75006, Paris

maryshi@live.cn

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190266

Rich Text Formatted EHR Narratives: A Hidden and Ignored Trove

Zexian Zeng[a*], Yuan Zhao[a*], Mengxin Sun[b], Andy H Vo[c], Justin Starren[a], Yuan Luo[a]

[a]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
[b]Hospital Medicine, Northwestern Memorial Hospital, Chicago, IL, USA
[c]Committee on Developmental Biology and Regenerative Medicine, The University of Chicago, Chicago, IL, USA
** Authors contributed equally*

Abstract

This study presents an approach for mining structured information from clinical narratives in Electronic Health Records (EHRs) by using Rich Text Formatted (RTF) records. RTF is adopted by many medical information management systems. There is rich structural information in these files which can be extracted and interpreted, yet such information is largely ignored. We investigate multiple types of EHR narratives in the Enterprise Data Warehouse from a multisite large healthcare chain consisting of both, an academic medical center and community hospitals. We focus on the RTF constructs related to tables and sections that are not available in plain text EHR narratives. We show how to parse these RTF constructs, analyze their prevalence and characteristics in the context of multiple types of EHR narratives. Our case study demonstrates the additional utility of the features derived from RTF constructs over plain text oriented NLP.

Keywords:

Electronic Health Records, Natural Language Processing, Information Management

Introduction

Nationwide adoption of Electronic Health Records (EHRs) has given rise to a large amount of digital health data, which can be used for secondary analysis [1]. The data within these records are expected to improve efficiency and overall healthcare. In addition to structured data, a large amount of clinically relevant data remains present in an unstructured free-text format. These narratives are largely variable between institutions and even healthcare professionals within them making it difficult for data extraction. In order to optimize research to improve quality or clinical needs, a systematic representation of EHRs for both structured and unstructured data is necessary. For instance, computational phenotyping has been used to mine or predict both clinically and scientifically significant phenotypes from structured EHR data, unstructured clinical narratives, and their combinations [2]. Computational phenotyping calls for efficient approaches to mine a large volume of clinical narratives for structured information on patient pathophysiology. To this end, natural language processing (NLP) is widely utilized as an effective tool for turning unstructured narratives into structured information [3].

Applications of language processing methods to clinical data in many projects have achieved reasonable success in analyzing various types of biomedical data, including clinical narratives [4, 5]. However, when it comes to EHRs, there is a gap between what data can support and what the method asks for. NLP generally assumes plain text as inputs, but clinical narratives in EHR often possess rich-text format (RTF) such as tables, different fonts, and font sizes. The majority of studies ignore the informative RTF information in EHR narratives. To our knowledge, there are very few, if any, studies that process RTF EHR narratives and utilize the formatting for additional information. When evaluating numerous clinical note repositories for shared task challenges (e.g., challenges from BioNLP [6-8], i2b2 [9, 10], SemEval [11, 12], and BioCreative [13]), there are few RTF EHR narratives.

The task of processing RTF documents is relatively straightforward though non-trivial. Despite the existence of standard packages to parse RTF or convert it to other formats such as HTML, research in the field of biomedical NLP largely ignores RTF and still uses plain-text as their raw input. Typical NLP processes include sentence breaking, word tokenization, part-of-speech tagging, constituency or dependency parsing, named entity detection, semantic role labeling, event frame extraction, etc. These processes largely ignore the formatting information. As a result, EHR databases (e.g., Enterprise Data Warehouse) for medical institutions often additionally store the plain text format of EHR narratives in addition to their RTF counterparts, or simply store in plain text format instead of RTF. Furthermore, information loss can sometimes occur during the conversion of RTF to plain-text formats.

Interestingly, researchers who focus on studying structured data do not treat EHR narratives as structured and often rely on NLP to extract structured pieces out of EHR narratives from plain text. Many studies utilize generic medical NLP systems such as MedLEE [14], MetaMap [15], cTakes [16], or GATE [17]. Other studies take advantage of special purpose NLP systems such as MedEx [18] for medication detection or TEES [19] for medication event frame extraction. Finally, many authors developed their NLP components or even systems to serve specific data mining tasks such as computational phenotyping [20-27].

Our study addresses the gap between structured EHR and clinical NLP research regarding the usage of rich-text formatted EHR narratives. In the current study we demonstrate how to extract the structured pieces (e.g., tables and section headings) from RTF EHR narratives and investigate the prevalence of such structures in EHR narratives. Our objective is to expose the need to start from rich-text formatted EHR narratives, making it suitable for NLP or structured EHR research and demonstrate that this can be done without much hassle. We also perform a case study to show the efficacy of mined structured information in improving computational phenotyping tasks.

Methods

A. Dataset

We constructed multiple rich-text formatted corpora consisting of different types of EHR narratives by querying Northwestern Medicine Enterprise Data Warehouse (NMEDW). The NMEDW is a joint initiative between the Northwestern University Feinberg School of Medicine and Northwestern Memorial HealthCare [28]. The NMEDW is a multi-center data repository with heterogeneous types of data. We constructed multiple patient corpora over a specific period from breast cancer patients. In this study the progress notes were queried, which are records of events during a patient's office visit or hospitalization and can be used to communicate opinions, findings, and plans between healthcare professionals. A well-documented progress note is complete, accurate, and concise for the care delivered, including diagnosis and treatments [29]. We extracted progress notes that were dated between 1 July 2015 and 1 October 2015. In order to test performance evaluation we extracted progress notes for breast cancer patients from a previous study [30], matching their plain text counterparts as a use case.

B. RTF processing

Many word processors as well as many of the medical information management systems support the reading and writing of rich-text format (RTF) files. In this study we found and identified the important RTF structural information from medical narratives, namely tables and section headings. A medical history table may contain several columns with headings stressed in bold, underlined, or inserted as the first line of the table. These tables often summarize important clinical information such as prescriptions, allergies, or medical histories. The format of RTF text is defined by groups, control-words, and delimiters. Groups are enclosed in braces, and control-words start with backslash (\) characters. For example, a 2-by-2 table and the corresponding RTF code is shown in Figure 1.

cell 11	cell 12
cell 21	cell 22

(a)

```
1  code{\rtf1\ansi\deff0
2  \trowd
3  \cellx1000
4  \cellx2000
5  cell 11\intbl\cell
6  cell 12\intbl\cell
7  \row
8  \trowd
9  \cellx1000
10 \cellx2000
11 cell 21\intbl\cell
12 cell 22\intbl\cell
13 \row}
```

(b)

Figure 1: A 2-by-2 table and the corresponding RTF code. (a) A 2-by-2 table defined by the RTF code (b) RTF code snippet that defines a 2-by-2 table.

As illustrated in Figure 1, the codes are within closed braces with control-words in the first line indicating the format and encoding used in this group. Each row in the table begins with a "\trowd" tag and ends with a "\row" tag. Thus, lines 2-7 define the first row and lines 8-13 defines the second row. The "\cellx###" tag declares the position of the right side of a cell and the "\cell" tag denotes the end of a cell. The above code generates a 2-by-2 table with cell widths equal to 1000. Searching for the "\trowd" tag in RTF files allows for table discovery. In addition to "\trowd" tags, many tables in medical notes are also defined by tab-delimited-tables, i.e., cells in a table row are separated by tab characters. To identify these tab characters, RTF uses the "\tab" tag.

Section headings in medical notes often describe the topic of that section. The information within a section is usually cohesive and can easily be interpreted by both NLP programs and manual curation. As a result, the Common Data Model of Observational Medical Outcomes Partnership specifies that the sections be stored as annotations to facilitate portable computational phenotyping (https://github.com/OHDSI/CommonDataModel) [accessed on March 20, 2019]. Section headings tend to share some characteristics such as bold or underlined formatting font. The headings may or may not occupy a whole line and may or may not be capitalized or be followed by a colon. Heading levels are often indicated by indentation. In summary, section headings provide important information and the RTF format allows us to apply specific searching criteria to accurately and efficiently identify section headings and their levels.

However, the syntax of RTF is relatively more obscure and lacks powerful software packages for data parsing compared to more popular formats such as HTML, XML, and LaTex. On the other hand, HTML is a much more commonly used format with many existing tools for parsing and extraction of structured information, such as Python's BeautifulSoup package (https://www.crummy.com/software/BeautifulSoup/bs4/doc/) [accessed on March 20, 2019]. To address this, we convert RTF to HTML using LibreOffice prior to parsing any data and then analyze the structures using Python's BeautifulSoup web scraping package. For regular tables we searched for the HTML <table> tag, and summarized the rows by counting the row tag <tr>. For tab-delimited-tables, we matched them using the regular expression "\S+ *\t *\S+", i.e., two non-space characters separated by a tab and zero or more space characters. The section headings in the documents are usually stressed in bold (HTML tag of <b>) or with an underline (HTML tag of <u>). The stressed phrase occupy a whole line.

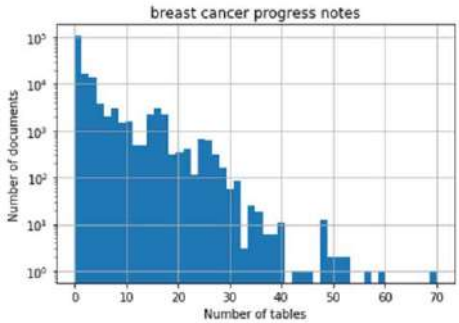
Results

A. Parsing the tables

In our study, we extracted and used 158,948 breast cancer progress notes. Given that column names are often listed in the first row of a table, we also extracted the cells in the first rows to summarize the topic of the tables. The distribution of the number of tables and the most common first rows of these tables are summarized in Figure 2. A Histogram indicating table frequency is shown on the upper panel of the figures. Note that the y-axis is in log-scale. The number of tables in a medical note approximately follows an exponential distribution – the probability decreases by a factor of 10 for roughly every 10 additional tables. The most common first rows are summarized on the lower panel of the figures. Adjacent cells are separated by a comma.

In breast cancer progress notes, the most common first row contains a single cell with the most common terms being "Allergies" and "Past Medical History", making up 16.3% and

16.2% of the notes respectively. The 16.3% of the medical notes have an Allergies table. In addition to these terms, we also identified terms involving lab results, vital information, complaints, current prescriptions, illness history, medication list, and physical exam information (Figure 2). It is evident that the related tables belong to the pre-formatted template.



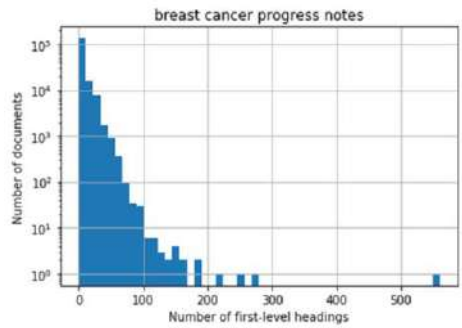
Most common First Rows	Counts	Ratio
Allergies	25855	16.3%
Past Medical History	25750	16.2%
Past Surgical History	20419	12.8%
Lab Results	11444	7.2%
Filed Vitals	11076	7.0%
Chief Complaint	10773	6.8%
Current Outpatient Prescrip-tions	10506	6.6%
History of Present Illness	9671	6.1%
Medication List	9650	6.1%
Physical Exam	9642	6.1%

Figure 2: Histogram of the number of tables and the most common first rows in breast cancer progress notes. The "Ratio" column is the count of the first rows over the total number of notes of this type.

B. Parsing the section headings

Section headings often contain critical information that can be used to identify key features. We searched the section headings in progress notes of patients with breast cancer from the NMEDW. Among the 158,948 breast cancer progress notes we found 46.6% files have first level headings and 2.4% files have higher level headings. Histograms of the number of first-level section headings (section headings with the smallest indentation) in the notes are summarized in Figure 3. The number of headings in a note generally follows an exponential distribution. The lower panel of the figures shows the most common first-level section headings.

In the breast cancer progress notes, we found the most common heading to be "Allergies" similar to the most common first row found in these notes. Interestingly, we also identified headings involving Physical Exams, Lab Results, and Medication Lists, which are also the most common first rows of tables in these notes (Figure 3). These results suggest that section headings are important too. In addition, the structure of a progress note can be better understood by combining data from tables and section headings. Note that the vital signs (which is the first row of a table) are also deemed as the section headings because they are usually emphasized in bold with occupying a whole line.



Most common First Level Headings	Counts	Ratio
Allergies	11599	7.3%
Ref Range	11395	7.2%
Physical Exam	10942	6.9%
Impression and Plan	10829	6.8%
Diagnosis	10725	6.7%
Physical Exam	10712	6.7%
Lab Results	10383	6.5%
History of Present Illness	10018	6.3%
Medication List	9972	6.3%
Chief Complaint	9964	6.3%

Figure 3: Histogram of the number of first-level section headings, and the most common first-level headings in the breast cancer progress notes. Note that the y-axis is in log-scale. The "Ratio" column is the count of headings over the total number of notes of this type.

C. Case study on breast cancer phenotyping

The structured data extracted from the medical notes can be used as features in machine learning models to improve their performances. As a case study, we added the structure information extracted to identify breast cancer patients with contralateral events. The problem was investigated in our previous study [30], with a cohort of 1063 patients who were diagnosed with breast cancer. In the cohort, 33 patients were identified with contralateral events (the case group). We previously identified the breast cancer contralateral events from the cohort using progress notes and counts of pathology reports. In our previous study, the progress notes were treated as plain text; structured information was ignored. This previous study was used as a baseline because it utilized an 8-step, relatively sophisticated NLP system and applied extensive feature engineering (e.g., filtered powerset method) to build the classification model. Thus, our previous system provides competitive plain-text oriented NLP baselines. In this case study, we pulled out the corresponding RTF versions of the same progress notes for the 1063 patients. We extracted the past medical history, past surgical history, and lab test results from the tables in the progress notes and included them as features in the classification model. For lab test results, we used the reference ranges presented in the test result tables to determine whether the measured results are higher or lower than the ranges. We chose common features that occured in at least 10 patients' progress notes, resulting in 37 medical history entries, 33 surgical history entries, and 10 abnormal lab test features. The most common medical/surgical history entries and abnormal lab tests are summarized in Table 1.

We encoded the structured features as whether a patient had a condition in the past medical/surgical history and whether a lab test is below or above the reference range. We adopted the same train-test split (7:3) and the same evaluation metric (AUC) as

in our previous study [30]. We used logistic regression (LR) model and 5-fold cross-validation to tune the regularization factor of LR, and obtained an AUC of 0.89 on the held-out test set. By building an ensemble model that averages the predicted probabilities from the structured-feature model and the NLP-feature model, an AUC of 0.92 was obtained. The model performance was compared with previous results in Table 2. Using much fewer structured features we can achieve a comparable classification performance using unstructured MetaMap CUI features (83 vs. 1285). Combining the two models further improves the classification performance. Note that the pathology report count is essentially a type of structured feature, though not derived from RTF tables [30]. For a fair comparison, the top 3 models in Table 2 (including the 3rd model that was reported as the best model in our previous study [30]) all use the pathology report count. In this case study, including the structured features improved the performance from 0.89 to 0.92. The overall results suggest that structured features extracted from RTF tables can be effective features in computational phenotyping tasks and may work even better when combined with unstructured features extracted by NLP.

Table 1: Most common medical/surgical history entries and abnormal lab tests.

Common Medical History entries (# of occurrences)	Common Surgical History entries (# of occurrences)	Common Abnormal Lab test (# of occurrences)
Breast cancer (313)	Lumpectomy (213)	MCHC low (112)
Hypertension (213)	Right breast (81)	Blood urea nitrogen high (32)
Hyperlipidemia (134)	Colonoscopy (78)	Hemoglobin low (23)
Disorder of refraction and accommodation (77)	Removal of tonsils and adenoids (78)	Albumin low (21)
Hypothyroidism (72)	Left breast (73)	MCH low (20)

Table 2: Comparison of classification performance on the held-out test set using different features. Shaded parts indicate the results from our previous study[30] that treated clinical narratives as plain text.

Model	Number of Features	Test AUC
Combined MetaMap + Structured Features + Pathology Report Count	1365	0.92
Structured Features + Pathology Report Count	83	0.89
Combined MetaMap + Pathology Report Count	1285	0.89
Combined MetaMap	1282	0.68
Pathology Report Count	3	0.67
Full MetaMap without Combination	42	0.30
Bag of Words	55192	0.70

Discussion

This study is a preliminary work that tries to introduce the informatics community to the rich-text formatted EHR narratives but it comes with limitations. First and foremost, we note that different institutions may have different clinical documentation systems, templates, and styles, especially across different healthcare chains on a national scale. Although our study uses multi-site data (a mixture of academic medical center and community hospitals) from regional healthcare chains, we expect challenges may arise when generalizing our study on a national scale. In addition, different healthcare chains may use different EHR systems with possible variations in the prevalence of the rich-text structures. Thus, we plan to expand our study in the future to include more types of EHR narratives, from different EHR systems, and systematically study their variations to better inform the data mining practice for EHR. We will also study the potential impact of rich-text format on improving the interoperability in health information exchange and portable computational phenotyping, as it brings more regularity to the otherwise "free" text.

Conclusion

Electronic Health Records (EHR) narratives are not equivalent to plain text but contain important structural information. One common medical note file format for archived EHR narratives is rich text format (RTF). We have utilized the low-level syntax tags in RTF files to mine the structured information of multiple types of EHR narratives. We ran our parsing systems across RTF clinical narratives from multiple community hospitals and an academic medical center, and successfully extracted tables and section headings from them. The tables contain patient allergies, medical histories, lab test results, prescriptions, vital signs, and even follow-up appointment information. A long EHR narrative may be divided into several sections, with each section containing cohesive information of a perspective and context of a patient's pathophysiologic profile. Properly recognizing the tables and sections is useful to extract structured information from EHR and prevents unnecessary input chunks of text to NLP systems. In addition, targeted NLP algorithms can be further applied to individual sections to extract information more precisely. Our case study on breast cancer phenotyping shows that tasks such as computational phenotyping can benefit from properly extracted structural information (e.g., tables) and more accurate NLP, thanks to the usage of RTF structural information – a hidden and often ignored trove in EHR.

Acknowledgements

This project is supported in part by NIH grant R21 LM012618.

References

[1] D. Blumenthal and M. Tavenner, "The “meaningful use” regulation for electronic health records," *N Engl J Med,* vol. 2010, no. 363, pp. 501-504, 2010.

[2] C. Shivade *et al.*, "A review of approaches to identifying patient phenotype cohorts using electronic health records," *Journal of the American Medical Informatics Association,* vol. 21, no. 2, pp. 221-230, 2013.

[3] Y. Luo *et al.*, "Natural Language Processing for EHR-Based Pharmacovigilance: A Structured Review," (in eng), *Drug Saf,* vol. 40, no. 11, pp. 1075-1089, Nov 2017.

[4] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: an introduction," *Journal of the American Medical Informatics Association,* vol. 18, no. 5, pp. 544-551, 2011.

[5] Z. Zeng, Y. Deng, X. Li, T. Naumann, and Y. Luo, "Natural Language Processing for EHR-based computational phenotyping," *IEEE/ACM Transactions on Computational Biology and Bioinformatics,* pp. 1-1, 2018.

[6] C. Nédellec *et al.*, "Overview of BioNLP shared task 2013," in *Proceedings of the BioNLP Shared Task 2013 Workshop*, 2013, pp. 1-7: Association for Computational Linguistics.

[7] J.-D. Kim, Y. Wang, T. Takagi, and A. Yonezawa, "Overview of genia event task in bionlp shared task 2011," in *Proceedings of the BioNLP Shared Task 2011 Workshop*, 2011, pp. 7-15: Association for Computational Linguistics.

[8] J.-D. Kim, T. Ohta, S. Pyysalo, Y. Kano, and J. i. Tsujii, "Overview of BioNLP'09 shared task on event extraction," in *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task*, 2009, pp. 1-9: Association for Computational Linguistics.

[9] Ö. Uzuner, Y. Luo, and P. Szolovits, "Evaluating the state-of-the-art in automatic de-identification," *Journal of the American Medical Informatics Association,* vol. 14, no. 5, pp. 550-563, 2007.

[10] Ö. Uzuner, B. R. South, S. Shen, and S. L. DuVall, "2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text," *Journal of the American Medical Informatics Association,* vol. 18, no. 5, pp. 552-556, 2011.

[11] N. Elhadad, S. Pradhan, W. Chapman, S. Manandhar, and G. Savova, "SemEval-2015 task 14: Analysis of clinical text," in *Proc of Workshop on Semantic Evaluation. Association for Computational Linguistics*, 2015, pp. 303-10.

[12] I. Segura Bedmar, P. Martínez, and M. Herrero Zazo, "Semeval-2013 task 9: Extraction of drug-drug interactions from biomedical texts (ddiextraction 2013)," 2013: Association for Computational Linguistics.

[13] M. Krallinger *et al.*, "Evaluation of text-mining systems for biology: overview of the Second BioCreative community challenge," *Genome biology,* vol. 9, no. 2, p. S1, 2008.

[14] C. Friedman, L. Shagina, Y. Lussier, and G. Hripcsak, "Automated encoding of clinical documents based on natural language processing," *Journal of the American Medical Informatics Association,* vol. 11, no. 5, pp. 392-402, 2004.

[15] A. R. Aronson, "Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program," in *Proceedings of the AMIA Symposium*, 2001, p. 17: American Medical Informatics Association.

[16] G. K. Savova *et al.*, "Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications," *Journal of the American Medical Informatics Association,* vol. 17, no. 5, pp. 507-513, 2010.

[17] H. Cunningham, "GATE, a general architecture for text engineering," *Computers and the Humanities,* vol. 36, no. 2, pp. 223-254, 2002.

[18] H. Xu, S. P. Stenner, S. Doan, K. B. Johnson, L. R. Waitman, and J. C. Denny, "MedEx: a medication information extraction system for clinical narratives," *Journal of the American Medical Informatics Association,* vol. 17, no. 1, pp. 19-24, 2010.

[19] J. Björne and T. Salakoski, "TEES 2.1: Automated annotation scheme learning in the BioNLP 2013 Shared Task," in *Proceedings of the BioNLP Shared Task 2013 Workshop*, 2013, pp. 16-25: Association for Computational Linguistics.

[20] R. L. Richesson *et al.*, "Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory," *Journal of the American Medical Informatics Association,* vol. 20, no. e2, pp. e226-e231, 2013.

[21] G. Hripcsak and D. J. Albers, "Next-generation phenotyping of electronic health records," *Journal of the American Medical Informatics Association,* vol. 20, no. 1, pp. 117-121, 2013.

[22] J. Pathak, A. N. Kho, and J. C. Denny, "Electronic health records-driven phenotyping: challenges, recent advances, and perspectives," ed: The Oxford University Press, 2013.

[23] Z. Zexian *et al.*, "Using Clinical Narratives and Structured Data to Identify Distant Recurrences in Breast Cancer," in *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, 2018, pp. 44-52: IEEE.

[24] Z. Zeng *et al.*, "Using natural language processing and machine learning to identify breast cancer local recurrence," (in eng), *BMC Bioinformatics,* vol. 19, no. Suppl 17, p. 498, Dec 28 2018.

[25] Y. Luo, A. R. Sohani, E. P. Hochberg, and P. Szolovits, "Automatic lymphoma classification with sentence subgraph mining from pathology reports," *Journal of the American Medical Informatics Association,* vol. 21, no. 5, pp. 824-832, 2014.

[26] Y. Luo, Y. Xin, E. Hochberg, R. Joshi, O. Uzuner, and P. Szolovits, "Subgraph augmented non-negative tensor factorization (SANTF) for modeling clinical narrative text," *Journal of the American Medical Informatics Association,* vol. 22, no. 5, pp. 1009-1019, 2015.

[27] Y. Luo, Y. Xin, E. Hochberg, R. Joshi, O. Uzuner, and P. Szolovits, "Subgraph augmented non-negative tensor factorization (SANTF) for modeling clinical narrative text," *Journal of the American Medical Informatics Association,* p. ocv016, 2015.

[28] J. B. Starren, A. Q. Winter, and D. M. Lloyd-Jones, "Enabling a learning health system through a unified enterprise data warehouse: the experience of the Northwestern University Clinical and Translational Sciences (NUCATS) Institute," *Clinical and translational science,* vol. 8, no. 4, pp. 269-271, 2015.

[29] R. Dick and E. Steen, "The computer-based patient record. Institute of Medicine," ed: Washington, DC: National Academy Press, 1991.

[30] Z. Zeng *et al.*, "Contralateral breast cancer event detection using Nature Language Processing," in *AMIA Annual Symposium Proceedings*, 2017, vol. 2017, pp. 1885-1892: American Medical Informatics Association.

Address for correspondence

Yuan Luo

Department of Preventive Medicine

Northwestern University, Chicago, IL 60611, USA.

E-mail: yuan.luo@northwestern.edu.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190267

A Deep Learning-Based Approach for Gait Analysis in Huntington Disease

Shisheng Zhang[a], Simon K. Poon[c], Kenny Vuong[b], Alexandra Sneddon[a], Clement T Loy[b]

[a] School of AMME, Faculty of Engineering & IT, The University of Sydney, Sydney, Australia,
[b] Sydney School of Public Health, The University of Sydney, Sydney, Australia,
[c] School of Computer Science, The University of Sydney, Sydney, Australia

Abstract

Huntington Disease (HD) is a genetic neurodegenerative disease which leads to involuntary movements and impaired balance. These changes have been quantified using footstep pressure sensor mats such as Protokinetics' Zeno Walkway. Drawing from distances between recorded footsteps, patients' disease severity have been measured in terms of high level gait characteristics such as gait width and stride length. However, little attention has been paid to the pressure data collected during formation of individual footsteps. This work investigates the potential of classifying patient disease severity based on individual footstep pressure data using deep learning techniques. Using the Motor Subscale of the Unified HD Rating Scale (UHDRS) as the gold standard, our experiments showed that using VGG16 and similar modules can achieve classification accuracy of 89%. Image pre-processing are key steps for better model performance. This classification accuracy is compared to results based on 3D CNN (82%) and SVM (86.9%).

Keywords:

Huntington Disease, Deep Learning, Diskynesias, Gait Analysis

Introduction

Huntington disease (HD) is a rare genetic neurodegenerative disease in which patients can have different levels of motor impairment and abnormal gait patterns [3]. In this project, deep learning is used to develop a novel evaluation method to classify patients' severity of disease, based on their footstep patterns.

Degeneration of motor function experienced by Huntington disease patients results in impaired balance and involuntary movements, which can be recorded using electronic pressure sensitive channels. All previous studies in this field analyzed high level gait characteristics, such as stride and tempo [9], calculated from the sensor mat based on distances between recorded footsteps. However, no one has studied whether the information contained in the pressure data of individual footsteps can provide a deeper understanding of the nature of the disease. This paper presents a novel approach to classify HD severity based on foot pressure data obtained from Protokinetics' ZenoTM Walkway.

Patients diagnosed with HD were grouped according to their UHDRS motor score, which is currently the gold standard in measuring motor disease severity. A classifier was then developed, and its parameters were optimized using cross-validation. The classifier was trained on the formation of the footprints.

Huntington Disease

Huntington disease is a rare autosomal dominant disease, also known as Huntington's chorea. Patients generally become unwell from middle age and have motor, cognitive, and mental health symptoms. The clinical symptoms of Huntington's disease are complex and variable, but patients invariably worsen over time [4]. The main features are dance-like involuntary movements with a progressive dementia. Huntington Disease is caused by expanded polynucleotide (CAG) repeats on the *Huntingtin* gene, which affect different molecular pathways and ultimately lead to neurological dysfunction and degradation [8].

Clinical Assessment of HD

The Unified Huntington's Disease Rating Scale (UHDRS) was developed as a clinical rating scale to assess four domains of clinical performance and capacity in HD: motor function, cognitive function, behavioral abnormalities, and functional capacity. The motor subscale is currently regarded as the gold standard in motor disease severity assessment, and is used as the endpoint in a number of clinical trials. Other clinical tests commonly used to determine the degree of motor function are the Timed Up and Go (TUG) test, the functional coverage test (FRT) and the Tinetti Performance Oriented Mobility Assessment (POMA) [1].

HD's Impact on Gait

The gait abnormalities in Huntington disease can be summarized as: (1) involuntary movements typically seen as chorea (sudden movements of the hands, feet, and facial muscles), and (2) worsening of balance and voluntary movement control [6]. The latter includes general slowing movements as well as impaired balance and gait.

High level gait charactertistics, such as gait width and stride length, have been extensively studied in HD. As can be seen from the Figure 1, three consecutive footprints form a gait cycle and include information such as stride width, stride length, and step size. These features can be measured and used for classification.

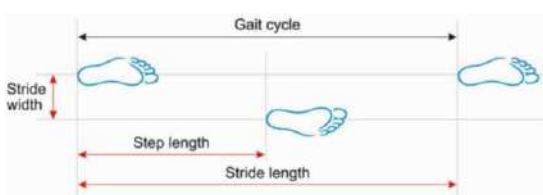


Figure 1 – Gait Cycle[5]

This study proposes to examine the pressure data of individual footsteps for the first time, rather than the high level gait pattern data.

Methods

Hardware for Pressure Data Collection (ProtoKinetics ZenoTM Walkway System)

The ProtoKinetics' ZenoTM walkway system consists of three layers - the pressure sensor and the carpet of the circuit are embedded on top of the protective substrate, hidden beneath the top layer of the linoleum. The pressure sensor grid has a high spatial resolution and is spaced 1.27 cm along the two axes. Due to the protective layered structure, sidewalks are compatible with auxiliary equipment, such as walking sticks and walking racks, without distorting the gait analysis results. The ProtoKinetics Motion Analysis Software (PKMAS) was used to collect and collate footstep pressure data.

Data Acquisition

Subjects and conditions

Participants included individuals who underwent rehabilitation at the Huntington Disease rehabilitation service at St. Joseph's Hospital, Auburn. Each subject has been clinically diagnosed with HD prior to starting the study. All participants gave their informed consent for the HD-FALLS study, which has been approved by the St Vincent's Hospital (Sydney) Human Research Ethics Committee (Reference HREC/13/SVH/378).

Clinical Assessment

Clinical and Zeno Walkway data were collected from participants of HD-FALLS. Although the study collected a broader data, the most relevant to this study is the UHDRS motor score.

Data format and visualization

Patients' pressure data were measured by Zeno Walkway and collected as text files, as indicated below:

Time (sec.);X (sensor number);Y (sensor number);Level
3.3417;5;6;2
3.3417;4;7;3
3.3417;4;8;2
3.3417;5;8;2
3.3417;5;9;2
3.3417;6;9;2
3.3500;5;6;3
3.3500;6;6;2
3.3500;3;7;2
3.3500;4;7;3
3.3500;5;7;2
3.3500;6;7;2
3.3500;7;7;2
3.3500;4;8;2
3.3500;5;8;2
3.3500;6;8;3
3.3500;7;8;3
3.3500;5;9;2

Figure 2 – Pressure Data in Text File

Text files were converted to video using python code. The basic idea of this code was to save the active sensor with the same time information as one frame. All saved frames were then animated as footprint videos. The video speed was 30 fps.

Figure 3 and 4 show the visualization of healthy subject's walking and HD patient's walking.

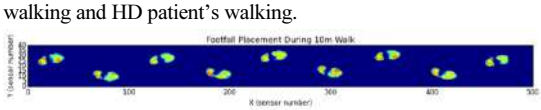


Figure 3 – Visualization of a Healthy Subject's Walk Along the ZenoTM Walkway

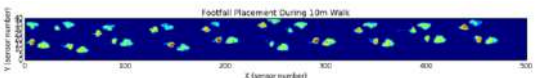


Figure 4 – Visualization of an HD Subject's Walk Along the ZenoTM Walkway

Data Preprocessing

The data preprocessing for novel method is discussed here which included denoise and normalization, and data augmentation. The preprocessing for the other two comparison methods is discussed in corresponding result parts.

Denoise and normalization

Single footprint pictures help better study the formation of the footprint. Four angular coordinates of the footprint were collected, and corresponding images were cut as shown in Figure 5.

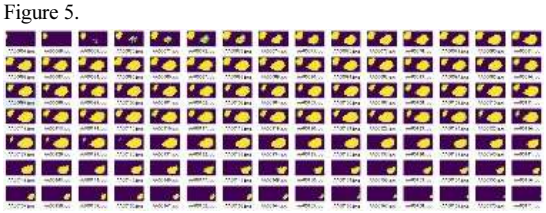


Figure 5 –Collected Footprint

After collecting the footprint, the color image was turned into a grayscale image with a range of pixel values from 0 to 255. To eliminate background noise, the pixel value below 50 was set to 255 to remove the background, but the footprint pixel was saved as original value. The resulting images of individual footprints were concatenated into a large mosaic image as shown in Figure 6.

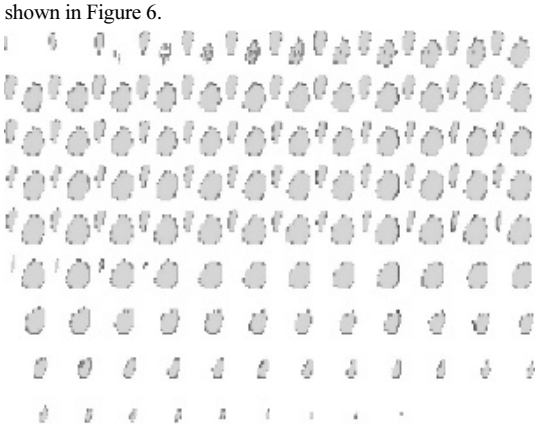


Figure 6 – Footprint Formation Process in Greyscale (Denoised)

The picture shows that the initial touch position (upper left corner) and the last touch (lower right corner). Most grey areas have the same level of pressure, but some dark points on the edges indicate lower pressure levels.

Data augmentation

A cohort of 12 patients included six patients each with low and high severity disease for our best approach. A 6-fold validation was used for model training. To clarify, five patients were randomly picked from one group as training data and 1 patient was left for model validation. 18 footprints were collected from each patient. In total, 180 mosaic images were used for training and 36 images were used for validation.

180 images were still not enough for deep learning training, especially for model with many parameters such as VGG16. However, as a rare disease, it is often hard to obtain many HD patients for research. To enlarge the dataset without collecting new footprint, the built-in function ImageDataGenerator was used. The ImageDataGenerator is a matured function that specializes in data augmentation. There are some parameters to set before it can create new images. In our project, shear range and zoom range were both set at 0.2. Horizontal flip was set as True. A schematic image in Figure 7 shows the basic operation in ImageDataGenerator.

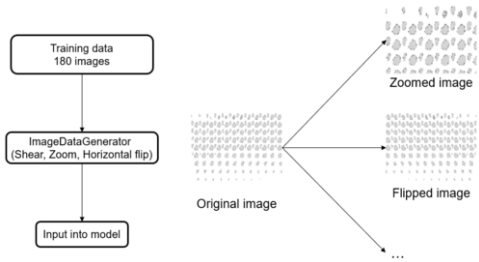


Figure 7 – Workflow for ImageDataGenerator

Modeling

The model for the pressure data of individual footsteps was easy and simple but models for control methods were advanced. It showed that pressure data of individual footsteps have rich features and can achieve good performance even without combining high level features.

For our novel method, a pretrained VGG16 was used for the purpose of feature extraction. VGG16 is a convolutional neural network architecture named after the Visual Geometry Group developed by Oxford University. It was used to win the 2014 ILSVR (ImageNet) competition [7]. So far, it is still considered as an excellent visual model and has outperformed more advancements in some areas, such as Inception and ResNet [2]. Since we have a small training data set, too many parameters can lead to overfitting. Compared to Inception and ResNet, VGG16 has fewer parameters and shorter training time.

A 3D convolution neural network was trained on the same patients data for comparing performance. The 3D convolution calculates features from spatial and temporal dimensions. 3D convolution is achieved by convolving a 3D kernel to a cube formed by stacking multiple consecutive frames together. With this configuration, the feature map in the convolutional layer is connected to consecutive frames in the previous layer, thereby capturing gait information.

3D Convolution neural network is designed to analyse videos or consecutive frames. In this project, normal pressure images were captured at 10Hz from patients' gait pressure video.

The inputs consisted of three types of images: Footprint pressure images, x gradient and y gradient images (Figure 9).

The x and y gradient images were obtained from normal images using the Prewitt operator. 30 contiguous frames were used as an input unit for each type of image group.

$$G_x = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} \quad G_y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

Figure 8 – Prewitt Operator for (x,y) Gradient

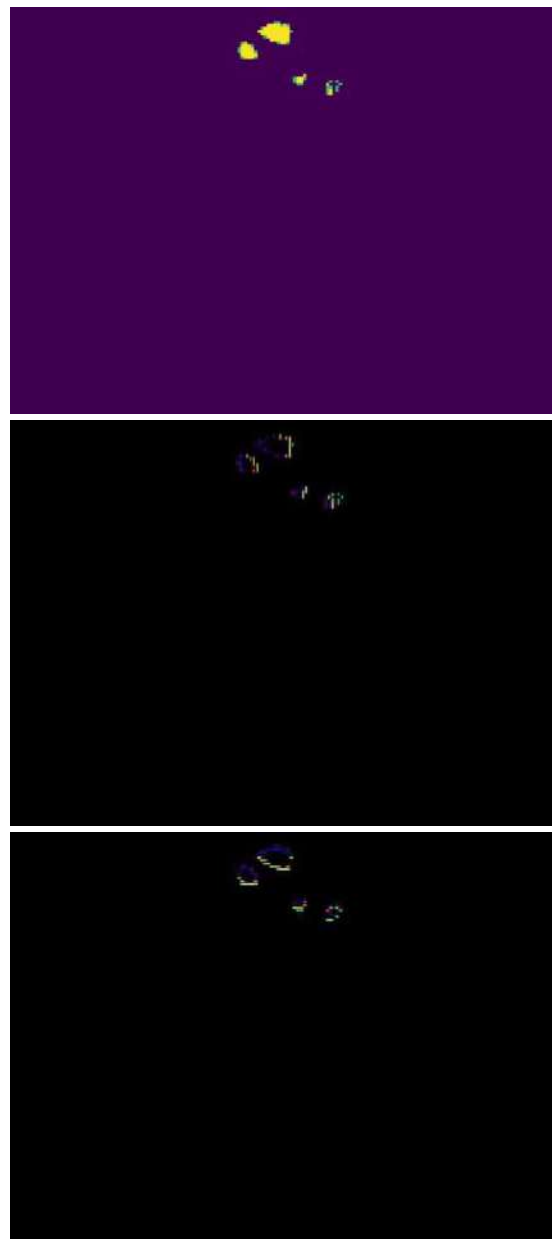


Figure 9 – Normal Pressure Image, (x,y) Gradients

SVM was also used for comparison. The SVM learning algorithm has several parameters that include the kernel function (used to transform the dataset into a higher dimensional feature space), the kernel parameters (which depend on the choice of kernel), and the margin parameter. The values of these parameters can significantly impact classification accuracy and thus they need to be tuned to develop the optimal model for a given problem. A grid search was therefore undertaken to optimize the combination of values for these three parameters.

The datasets used for SVM were pressure pattern feature set and gait pattern feature set (Figure 10). The pressure pattern feature set represented the footfall over a 30 cm x 20 cm area within a 3 second duration. These features, which represented the temporal variation in footfall pressure distribution for each step,

comprised the feature set that shall hereby be referred to as the pressure pattern feature set.

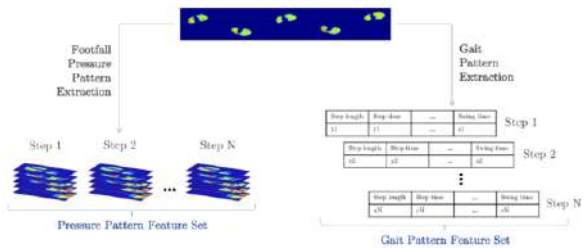


Figure 10 – Overview of SVM Data Preprocessing Approach

The gait pattern feature set that included only the features affected by the symptoms of HD were extracted, as partly listed in Table 1.

Table 1 – Spatiotemporal Features Extracted by PKMAS That Comprise the Gait Pattern Feature Set

Feature	Definition
Step length	Distance between heel contacts of consecutive footfalls.
Absolute step length	Diagonal distance between heel contacts of consecutive footfalls.
Stride length	Distance between heel contacts of the same foot (equivalent to the summation of consecutive step lengths)
Stride width	Distance between consecutive heel contacts, taken perpendicularly to the direction of gait.
Step time	The period of time taken between the first contacts of consecutive footfalls.
Stride velocity	Equal to stride length divided by stride time.
Stance time	The period of time during which the foot is in contact with the ground.

Results

Result of Novel Method

The formation of the footprint was used as an input to the experiment. The model used had two parts. The first part was VGG16, which was used to extract features such as edges and corners from the image. The second part worked as a weighting matrix, and different weights can be arranged for different types of features. The final output was the severity level (high or low).

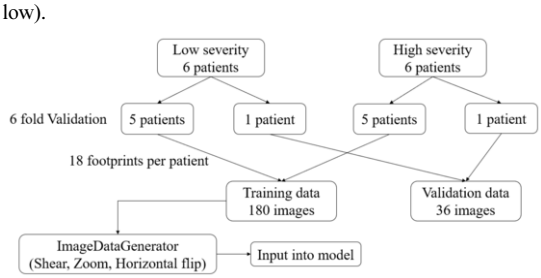


Figure 11 –Work flow for Footprints Formation Model

The grouping method and data structure are shown in Figure 11. With 6-fold validation, the classification accuracy was 89%. To avoid the problem of small sample size, we used ImageDataGenerator to create larger data sets based on the formation of footprints, data augmentation, and cross-validation to achieve accuracy.

Result of 3D CNN Method

18 patients' walking pressure videos were used for model training including 9 low severity patients and 9 high severity petients. 9-fold validation was used to evaluate our model on a limited data sample. The size of captured frames was 2700x200 and rearranged as 800x600. 3D convolution parts shown in Figure 12 were of the same architecture. The features were extracted and concatenated followed by classification layer. Our 3D CNN model achieved a classification accuracy of 82%.

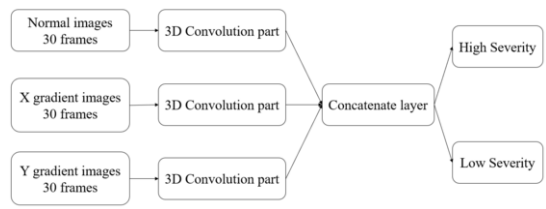


Figure 12 –Work flow for 3D CNN Model

Result of SVM Method

For comparision, two SVM classifiers were developed; one trained on raw footfall pressure data and the other on high-level gait features (computed by the Protokinetics' Movement Analysis Software package) for each footstep recorded for 65 subjects diagnosed with HD. The objective was to classify HD severity as either `high' or `low', which was compared against the Unified Huntington's Disease Rating Scale score of the HD patient to whom the step belonged.

The SVM trained on only the footfall pressure data achieved a classification accuracy of 76.9% (while that trained only on high-level gait features classified HD severity with an accuracy of 82.4%). When the predictions of both the classifiers were combined using an optimized weighting system, classification accuracy increased to 86.9%.

Discussion

Although deep learning method seems to be a blackbox to researchers, we can find clues by comparing different data input and model configurations.

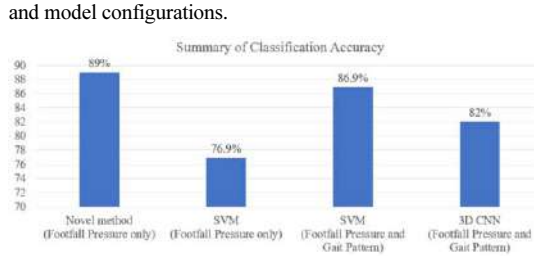


Figure 13 – Classification Accuracy of Four Methods

Our novel method used the pressure data of individual footstep formation process as temporal feature. Compared to 3D CNN which has a strong spatial information input and a designed structure for temporal information input, the novel method has a better performance. It shows that focusing on formation of individual footprint gives more representative features for model training.

During our work with 3D CNN, we found that including x,y gradient of images helps to boost the performance. Convolution layers were capable of extracting features such as edges or corners, but when edges became important factors in our project, it was better to extract them before providing input into the model.

Footprint pressure frames were not working properly with SVM as its accuracy ranked the last but these frames work well with 3D CNN model. When the results of the 3D CNN and the SVM trained on high level features were compared, we found that 3D CNN seemed to rely on footprint's local area but failed to get high level features. High level features (stride length, stride width, step length) were more recognizable and useful for the SVM model based on the performance results.

A combined SVM model ranked second in our performance summary showing that an optimized weighting system of deep learning model for footprint images and the SVM for high level features might be able to achieve better performance.

We managed to achieve a classsification accuracy at 89% in this research by using only objective pressure data without considering the widely used gait pattern data.

Limitations

Being the first study to examine individual footprint pressure data, our study has a number of limitations. First, the preprocessing module can be further optimized. Footprints were processed individually. However, this approach did not consider whether the steps involved were from left or right foot. Analyzing the left and right foot separately may improve our results since HD may affect gait asymmetrically. Additional measures should be taken in future processing to ensure that the left and right foot are equally represented. Second, in terms of category labels, separating the UHDRS motor score into "low" and "high" severity could be a simplification. Third, the number of patients in this study was limited and did not include any healthy subjects as controls. Models derived from small samples may have biases that can be overcome with larger data sets.

Conclusions

This study uses an innovative perspective to explore the nature of Huntington Disease and its impact on motor function. By analyzing individual foot pressure data obtained from a pressure sensor mat, we found that the formation of footprint and temporal change in foot pressure is potentially a feasible classifier for HD motor disease severity.

Future work

Gait features

For the first time, we have demonstrated the feasibility of classifying disease severity using individual footprint pressure data. After further verification, we will explore how this approach can be integrated with exisiting gait measures (such as gait width and stride length) to improve classification accuracy.

Deep learning method

The novel model only includes convolutional layers that cannot extract features between two consecutive footprints. Although 3D CNN model can work with consecutive frames, the feature extracted seem to be less representative which leads to worse performance compared to novel model result. An idea is to combine advantages of both models. First, we will design a new model structure with input portals for x, y gradient images. Second, the inputs will not only include the current existing footprint formation, but also the relationship between the high-level features and the low-level features of the footprint.

Acknowledgements

HD-FALLS was supported by a St Vincent's Clinic Multidisciplinary Patient Focused Grant. CTL is supported by an Australian National Health and Medical Research Council-Australian Research Council Dementia Research Development Fellowship (APP1107657)

References

[1] J.H. Ansai, L.P. de Andrade, T.H. Nakagawa, F.A.C. Vale, M.J.D. Caetano, S.R. Lord, and J.R. Rebelatto, Cognitive Correlates of Timed Up and Go Subtasks in Older People With Preserved Cognition, Mild Cognitive Impairment, and Alzheimer's Disease, *American journal of physical medicine & rehabilitation* **96** (2017), 700-705.

[2] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[3] K. Kieburtz, J.B. Penney, P. Corno, N. Ranen, I. Shoulson, A. Feigin, D. Abwender, J.T. Greenarnyre, D. Higgins, and F.J. Marshall, Unified Huntington's disease rating scale: reliability and consistency, *Neurology* **11** (2001), 136-142.

[4] J.-M. Lee, V.C. Wheeler, M.J. Chao, J.P.G. Vonsattel, R.M. Pinto, D. Lucente, K. Abu-Elneel, E.M. Ramos, J.S. Mysore, and T. Gillis, Identification of genetic factors that modify clinical onset of Huntington's disease, *Cell* **162** (2015), 516-526.

[5] W. Pirker and R. Katzenschlager, Gait disorders in adults and the elderly : A clinical guide, *Wien Klin Wochenschr* **129** (2017), 81-95.

[6] R. Reilmann, B.R. Leavitt, and C.A. Ross, Diagnostic criteria for Huntington's disease based on natural history, *Movement Disorders* **29** (2014), 1335-1341.

[7] M. Rezaee, Y. Zhang, R. Mishra, F. Tong, and H. Tong, Using a VGG-16 Network for Individual Tree Species Detection with an Object-Based Approach, in: *2018 10th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS)*, IEEE, 2018, pp. 1-7.

[8] A.H. Schapira, C.W. Olanow, J.T. Greenamyre, and E. Bezard, Slowing of neurodegeneration in Parkinson's disease and Huntington's disease: future therapeutic perspectives, *The Lancet* **384** (2014), 545-555.

[9] D. Trojaniello, A. Ravaschio, J.M. Hausdorff, and A. Cereatti, Comparative assessment of different methods for the estimation of gait temporal parameters using a single inertial sensor: application to elderly, post-stroke, Parkinson's disease and Huntington's disease subjects, *Gait & posture* **42** (2015), 310-316.

Address for Correspondence

A/Professor Simon Poon, Faculty of Engineering and Information Technologies, the University of Sydney.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190268

Identifying Cardiomegaly in ChestX-ray8 Using Transfer Learning

Sicheng Zhou[a], Xinyuan Zhang[b], Rui Zhang[a,c]

[a] *Institute for Health Informatics, University of Minnesota Twin Cities, Minneapolis, Minnesota, USA*
[b] *School of Biomedical Informatics, University of Texas Health Science Center, Houston, Texas, USA*
[c] *Department of Pharmaceutical Care & Health Systems, University of Minnesota Twin Cities, Minneapolis, Minnesota, USA*

Abstract

Recently, the National Institutes of Health (NIH) published a chest X-ray image database named "ChestX-ray8", which contains 108,948 X-ray images that are labeled with eight types of diseases. Identifying the pathologies from the clinical images is a challenging task even for human experts, and to develop computer-aided diagnosis systems to help humans identify the pathologies from images is an urgent need. In this study, we applied the deep learning methods to identify the cardiomegaly from the X-ray images. We tested our algorithms on a dataset containing 600 images, and obtained the best performance with an area under the curve (AUC) of 0.87 using the transfer learning method. This result indicates the feasibility of developing computer-aided diagnosis systems for different pathologies from X-rays using deep learning techniques.

Keywords:

X-rays, Cardiomegaly, Machine Learning

Introduction

Cardiomegaly is a medical condition that indicates the enlarged heart. It can be an important sign of potentially severe heart diseases, such as high blood pressure, heart valve disease, heart muscle disease, pericardial effusion, and heart attack. Also, the cardiomegaly can increase the risks of several heart diseases, including the heart failure, cardiac arrest, blood clots in heart, and heart murmur [1]. It is estimated that there are more than 200,000 new cases of cardiomegaly every year in the US, which is an important health issue for the middle-aged and senior population. The cardiomegaly is easier to treat if detected early, thus its early diagnosis is important. The chest X-ray, as one of the most frequently used examination in radiology, has been applied for several decades to detect and visualize the abnormalities of body organs [2-3]. It has been proven to be an effective diagnostic tool for pathological alterations, with additional advantages due to its noninvasive and low-cost characteristics [4]. Also, the X-ray is an important clinical diagnosis method for cardiomegaly [5].

Distinguishing the different types of pathologies from the X-ray is a challenging and tedious task even for domain experts [6]. Figure 1 shows the examples of X-rays in healthy and cardiomegaly condition. Only trivial differences exist between the two images. Thus, it is meaningful to develop a computer-aided detection method to facilitate the clinicians to accurately identify the X-ray images with cardiomegaly. Many related works have been done in developing such techniques in pathology detection field. Traditional image classification models have been applied in medical image classification and achieved satisfying performances. For instance, the Local Binary Patterns (LBP) [7-8] and the Bag-of-Visual-Words (BoVW) [9; 10] are two models widely applied to retrieve or classify the radiology images. In the LBP model, the images are usually divided into small cells of pixels, then the values of pixels in each cell are binarized. Next, the LBP histograms over each cell can be calculated as features passed to the classifier [10]. The BoVW is also called dictionary learning, it was proposed to mimic the visual processing by the human's brain [10]. The BoVW is similar to the bag-of-words representation of text, it regards an image as a distribution of local descriptors. Those local patch descriptors can be color, texture, shape, and the most popular descriptor is the scale-invariant feature transform (SIFT) [9-10].

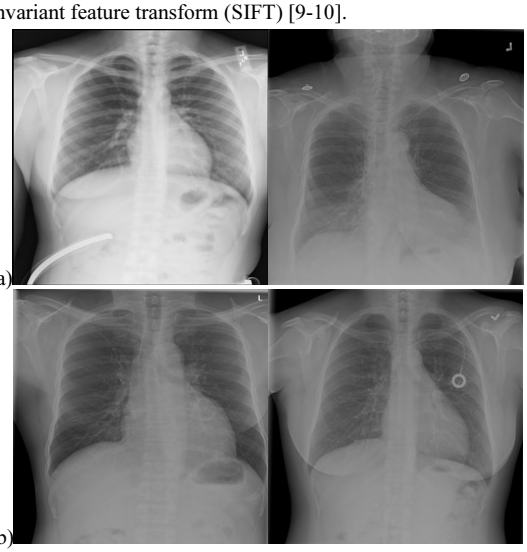


Figure 1–(a)X-ray Images for Patients without Cardiomegaly. (b) X-ray Images for Patients with Cardiomegaly.

The deep neural networks have been widely applied in image processing area due to the development of new variant convolutional neural networks (CNNs) models [11-12] and modern hardware [13]. These models have reached promising performances for image classification tasks for the ImageNet Large Scale Visual Recognition Competition (ILSVRC) [14]. Also, these methods have been successfully applied in the processing of the medical images, including the detection of pleural effusion assessment on chest radiography, the lymph node detection on CT, brain segmentation, and assessing diabetic retinopathy [15]. Thus, our study explores the feasibility of using deep learning methods to identify the cardiomegaly from the X-ray images.

The deep learning methods are most effective when applied on large data sets, however, such large datasets are often not available in the medical field due to privacy issues [13]. Recently, the National Institutes of Health (NIH) published a new chest X-ray database called "ChestX-ray8", which contains 108,948 frontal view X-ray images taken from 32,717 unique patients. Natural language processing (NLP) techniques were used to annotate each image with eight disease labels from the associated radiological reports. The disease labels include atelectasis, cardiomegaly, effusion, infiltration, mass, nodule, pneumonia, and pneumothorax [16]. This annotated large dataset provided opportunities to developing computer-aided diagnosis (CAD) systems based on the deep learning methods. Initial experiments of image classification were conducted using this dataset [16]. They built a Deep Convolutional Neural Network (DCNN) architectures that used different ImageNet pre-trained models, i.e., AlexNet [11], GoogLeNet [12], VGGNet-16 [17] and ResNet-50 [18]. It is found that using the ResNet-50 pre-trained model achieved the best results in identifying the cardiomegaly from the X-rays (AUC = 0.814).

There are three pre-trained models we will use in this study, including the ResNet-50, InceptionV3, and Xception [19]. The ResNet-50 is a type of CNN model, and it was proposed to solve the problem that when the traditional CNN model has deeper layers, the performance would decrease in training and test set but not due to overfitting. Basically, ResNet-50 is a residual net with 50 layers. It can allow the model to go deeper, and enable smooth propagation and optimization [18]. This model won the first place on the ImageNet detection challenge. Compared to the traditional CNN models, the InceptionV3 has advantages that it can improve the overfitting problems of complex neural networks and reduce the parameters of the neural networks to reduce the computing cost. The core idea of InceptionV3 is to use convolution kernels with different sizes to extract features with different granularities from the images [12]. Xception was developed based on the InceptionV3 model, the depthwise separable convolution was introduced into the Xception model, which can process the spatial dimension and channel dimension of the image separately. Xception has roughly the same number of parameters as InceptionV3 and obtained better results on classification dataset consisting of 350 million images and over 17,000 classes [19]. In this work, we tried to identify the cardiomegaly from X-rays by combing three above-mentioned state-of-art pre-trained models based on ImageNet. We implemented the transfer learning method to integrate the three pre-trained models for identifying the cardiomegaly from X-rays.

Methods

Data Sources

We used the "ChestX-ray8" database described previously. It comprises 108,948 frontal-view X-ray images from the year 1992 to 2015. In this project, we selected 21,966 images in total. In 767 of the images they were labeled as "cardiomegaly" and the others were labeled as "healthy". The whole dataset was divided into a training and a testing set. The training set contains 467 images labeled with "cardiomegaly" and 20,899 images labeled with "healthy". The testing set contains 300 images for both "cardiomegaly" and "healthy" categories. We split the training and testing set in this way due to the limited number of images labeled with "cardiomegaly".

Image Pre-processing

Three groups of images were used in our experiments. The first group contains the original image set. For the second group, two steps were implemented to preprocess the original image set. The first step is to crop the image based on the characteristics of cardiomegaly. Since cardiomegaly is closely related to the shapes and sizes of heart and breastbone. We tried to crop the unnecessary background to highlight the heart area. The dimension of the original image is 1024 × 1024. We crop all the images into the dimension of 800 × 800. The second step is to do the histogram equalization for all images [20]. Histogram equalization is a frequently used method for image contrast enhancement. It uses the cumulative distribution function to map the original image histogram to a new image histogram. It stretches the grayscale of the image by normalization. For the third group of images, we transformed the second group of images that contain three color channels into gray-scale images with one color channel.

Transfer Learning

We implemented transfer learning for the image classification task. It is usually unpractical to train the CNN from the beginning because of the insufficient data size and computing power. Thus, the transfer learning becomes an effective method in the image processing domain that could take the advantages of the well-developed models to solve new tasks [21]. Transfer learning mainly has two types when dealing with imaging processing problems. First, a pre-trained model can be used as the feature extractor for the new dataset. Once the features are extracted, a linear classifier can be trained for the new task. The second type is to retrain the fully-connected layer on top of the CNN on the new dataset while to fine-tune the weights of the pre-trained network through backpropagation. Considering the limited computing power and the size of our dataset, we used the first method in our study. Three models with weights pre-trained on ImageNet, including the InceptionV3 [10], ResNet50 [12] and Xception [19], were applied to extract the features from the training set of X-ray images and trained a classifier based on the different combinations of the features. The pipeline of our experiment was shown in Figure 2. Basically, for each experiment group,

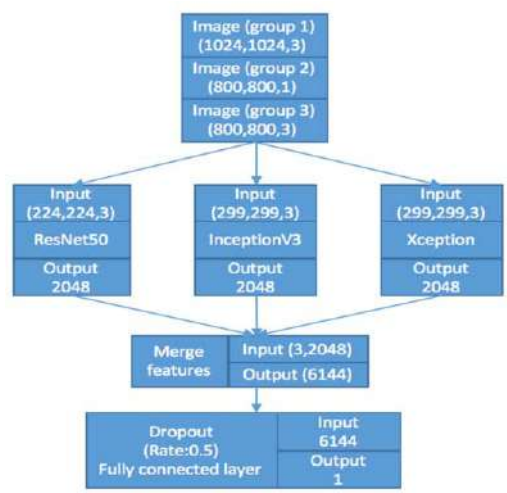


Figure 2–The Experiment Pipeline for Transfer Learning.

we used the three pre-trained models to extract the features, and trained the classifiers based on features extracted by one

model, different combinations of two models and the combination of three models.

Evaluation

We evaluated the performances of the trained classifiers using different combinations of features extracted by the pre-trained models as illustrated in Figure 2. We also tested the scaled and non-scaled features as the comparison. We used the accuracy, F-1 score and the area under the curve (AUC) to evaluate the classification results using the test set mentioned previously.

Results

Images after Pre-processing

Examples of images labeled healthy and cardiomegaly from three experiment groups are shown in Figure 3. The original images, cropped images that focus on the heart area, images after histogram equalization based on cropped images, and one-channel gray-scale image based on all the previous operations are shown in order. After cropping, many unrelated areas could be removed. The histogram equalization highlights the heart area, which could help the following image classification task.

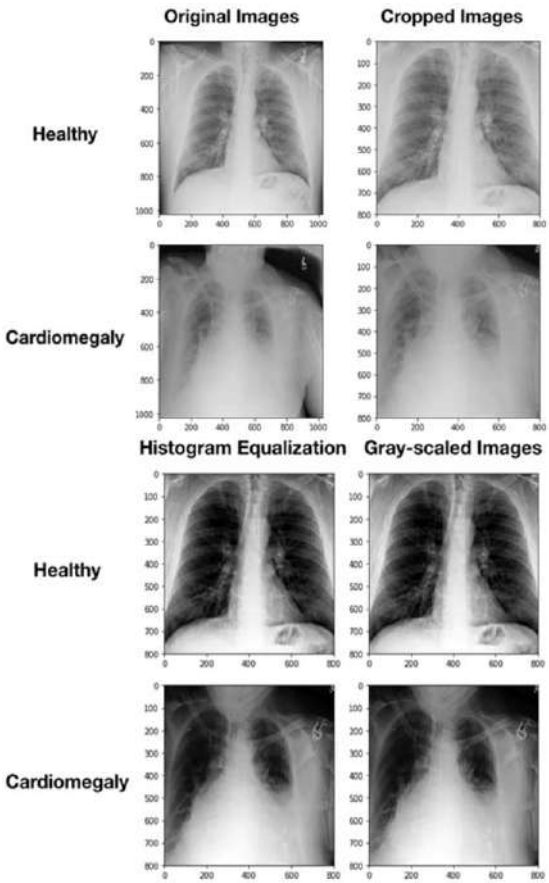


Figure 3– X-ray Images after Cropping and Histogram Equalization.

Classification Results Using Different Pre-trained Models

The results for identifying cardiomegaly from X-ray images are shown in table 1-3. For each experiment group, the results using separate and combined pre-trained models are listed. The scaled features were compared with non-scaled features.

Table 1– Image Classification Results for Group 1

Model	Accuracy	F-1 Score	Scaled-feature
InceptionV3	0.725	0.720	Yes
InceptionV3	0.792	**0.790**	No
ResNet-50	0.695	0.690	Yes
ResNet-50	0.747	0.740	No
Xception	0.673	0.670	Yes
Xception	0.712	0.700	No
InceptionV3 &ResNet-50	0.750	0.750	Yes
InceptionV3 &ResNet-50	0.778	0.780	No
InceptionV3 &ResNet-50 &Xception	0.735	0.730	Yes
InceptionV3 &ResNet-50 &Xception	0.782	0.780	No

Table 2– Image Classification Results for Group 2

Model	Accuracy	F-1 Score	Scaled-feature
InceptionV3	0.723	0.720	Yes
InceptionV3	0.795	0.790	No
ResNet-50	0.693	0.690	Yes
ResNet-50	0.765	0.760	No
Xception	0.655	0.650	Yes
Xception	0.708	0.700	No
InceptionV3 &ResNet-50	0.737	0.740	Yes
InceptionV3 &ResNet-50	0.797	**0.800**	No
InceptionV3 &ResNet-50 &Xception	0.726	0.720	Yes
InceptionV3 &ResNet-50 &Xception	0.792	0.790	No

After initial experiments, we found that the Xception model always had the worst performances and did not improve the performances in combined features. Thus, we stopped using Xception model separately in group 3.

Table 3– Image Classification Results for Group 3

Model	Accuracy	F-1 Score	Scaled-feature
InceptionV3	0.715	0.710	Yes
InceptionV3	0.782	0.780	No
InceptionV3 &ResNet-50	0.748	0.750	Yes
InceptionV3 &ResNet-50	0.793	**0.790**	No
InceptionV3 &ResNet-50 &Xception	0.715	0.710	Yes
InceptionV3 &ResNet-50 &Xception	0.790	**0.790**	No

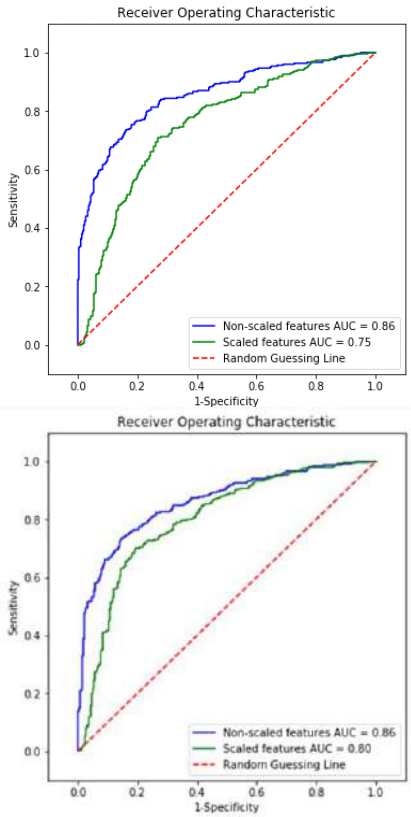


Figure 4–The Selected Best ROC Curves in Experiments.

The group 1 used original images, the features extracted by the combination of InceptionV3obtained the best results (F-1 = 0.79). The images in group 2 were cropped and stretched through histogram equalization. The combination of InceptionV3 and ResNet-50 models obtained the best results (F-1 = 0.800). The group 3 used gray-scale image with only one color channel. The combination of InceptionV3 and ResNet-50 models and the combination of InceptionV3, ResNet-50, and X-ception both reached the best results (F-1 = 0.790).

The selected best ROC curves are shown in Figure 4. The upper figure shows the ROC curve for using the only InceptionV3 model in group 3, and the lower figure shows the ROC curve for using the combination of InceptionV3 and ResNet-50 models in group 2. Our best performing experiment obtains the AUC = 0.860. Both the individual InceptionV3 model and the combination of InceptionV3 and ResNet-50 models reach the best results. The performances of experiments using the Xception model are not ideal, usually with F-1 around 0.700.

Discussion

This study explores the feasibility of applying transfer learning methods to identify the cardiomegaly from X-ray images. We used the pre-trained CNN models based on the ImageNet to extract the features from the X-ray images and obtained promising results (AUC = 0.860). In Wang's et al. work [16], they used the ResNet-50 model and tried to finish the similar task. They obtained the best result with AUC = 0.814 for identifying the cardiomegaly disease from X-ray images. Our method out-performed their results. Although the experimental scenarios are not identical, our task is binary classification while their task is multi-classification.

The three pre-trained CNN models in our study are InceptionV3, ResNet-50, and Xception. When using the model separately, the features extracted by the InceptionV3 obtained the best performances in all experiment groups, while the Xception is not proper for the X-ray images compared to others. The combination of InceptionV3 and ResNet-50 models slightly improve the performances. The best F-1 score is 0.800 and AUC is 0.860 using the combination of the two models. However, using the combined features does not mean better results. In group 1, the InceptionV3 itself has the highest accuracy than other combined models.

We used three experiment groups in the study, each group has images with different pre-processing methods. All the pre-processing steps used classic methods and obtained better images visually. However, the results show that compared with the average accuracy and F-1 score, there's no much difference among three groups, which indicates the pre-trained CNN models extract the features in a way that not influence by the image pre-processing process used in the study. Considering the property of X-ray images and the cardiomegaly disease, we also tried other pre-processing methods. For instance, to highlight the edges of the images, however, due to the unstable qualities of the images, the method cannot apply to all the images. Figure 5 shows the effect of the edge detection, the upper images show successful examples of pre-processing to highlight the breastbone and heart area, while the lower images show the bad examples. Since the quality of the X-ray images varies a lot, more powerful image pre-processing methods would be necessary to improve the results in future studies. Besides, the normalization of extracted features was conducted in all experiment groups. Usually, the feature normalization would improve the performances of neural networks in image classification, since it could reduce the noise of the images by bringing the intensity values to the normal distribution. We used the 0-1 normalization in the study, however, it has side-effects and harms the performances of the model.

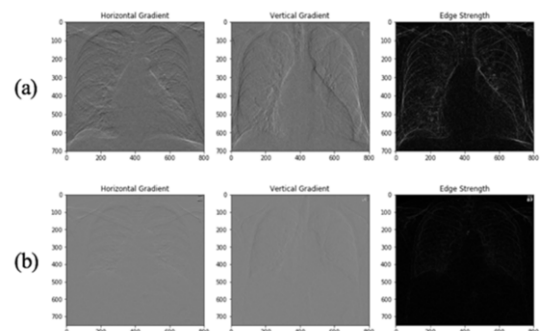


Figure 5–Successful and Failed Edge Detection Examples.

Limitations & Future Work

There are mainly two limitations for the study. First, is the unbalanced dataset. We have 767 X-ray images labeled with "cardiomegaly" and 20,899 images labeled with "healthy". Considering the difficulty of obtaining the annotated X-ray images, it is already the most comprehensive dataset we could obtain. The second limitation is related to the transfer learning method. We used the pre-trained CNN models based on the

ImageNet. Theoretically, these models only work best when our dataset is similar to ImageNet data. The ImageNet data contains over 14 million images in our daily life, but the X-ray image is a special type of image. Currently, there's no pre-trained model based on X-ray images. In the future, we will try to improve the performance by retraining the Inception V3 model. Also, we plan to work with the clinical department to evaluate if our method could improve the efficiency of manual review of X-Rays in clinical settings.

Conclusions

In this study, we applied the transfer learning method to identify the cardiomegaly from X-ray images. We obtained the best performance using the combination of InceptionV3 and ResNet-50 pre-trained models, with AUC of 0.860 that surpass the state-of-the-art results in related work. This result shows the feasibility of using transfer learning methods to develop the CAD system for X-rays.

Acknowledgments

The authors declare that they have no competing interest. Thanks Dr. Luca Giancardo for preparing the image dataset.

References

[1] Mayo Clinic, Enlarged Heart, [Internet]. 2017 [cited 15 November 2018]. Available from : https://www.mayoclinic.org/diseases-conditions/enlarged-heart/symptoms-causes/syc-20355436

[2] O. Er, N. Yumusak, and F. Temurtas, Chest diseases diagnosis using artificial neural networks, *Expert Systems with Applications* **37** (2010), 7648-7655.

[3] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, Chest pathology detection using deep learning with non-medical training, in: *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 2015, 294-297.

[4] A.A. El-Solh, C.-B. Hsiao, S. Goodnough, J. Serghani, and B.J. Grant, Predicting active pulmonary tuberculosis using an artificial neural network, *Chest* **116** (1999), 968-973.

[5] W.H. Frishman, J. Nadelmann, W.L. Ooi, S. Greenberg, M. Heiman, S. Kahn, H. Guzik, E.J. Lazar, and M. Aronson, Cardiomegaly on chest x-ray: Prognostic implications from a ten-year cohort study of elderly subjects: A report from the Bronx Longitudinal Aging Study, *American Heart Journal* **124** (1992), 1026-1030.

[6] R.H. Abiyev and M.K.S. Ma'aitah, Deep Convolutional Neural Networks for Chest Diseases Detection, *J Healthc Eng*, (2018).

[7] L. Sorensen, S.B. Shaker, and M.d. Bruijne, Quantitative Analysis of Pulmonary Emphysema Using Local Binary Patterns, *IEEE Trans Med Imaging* **29** (2010), 559-569.

[8] L. Sorensen, S.B. Shaker, and M. de Bruijne, Texture classification in lung CT using local binary patterns, *Med Image Comput Comput Assist Interv* **11** (2008), 934-941.

[9] W. Yang, Z. Lu, M. Yu, M. Huang, Q. Feng, and W. Chen, Content-based retrieval of focal liver lesions using bag-of-visual-words representations of single- and multiphase contrast-enhanced CT images, *J Digit Imaging* **25** (2012), 708-719.

[10] M.D. Kumar, M. Babaie, S. Zhu, S. Kalra, and H. Tizhoosh, A comparative study of CNN, BoVW and LBP for classification of histopathological images, *arXiv preprint arXiv:1710.01249* (2017).

[11] A. Krizhevsky, I. Sutskever, and G.E. Hinton, ImageNet classification with deep convolutional neural networks, *Commun. ACM* **60** (2012), 84-90.

[12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, Going deeper with convolutions, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, 1-9.

[13] Y. Bar, I. Diamant, L. Wolf, and H. Greenspan, *Deep learning with non-medical training used for chest pathology identification*, 2015.

[14] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, and M. Bernstein, Imagenet large scale visual recognition challenge, *International Journal of Computer Vision* **115** (2015), 211-252.

[15] P. Lakhani, Deep Convolutional Neural Networks for Endotracheal Tube Position and X-ray Image Classification: Challenges and Opportunities, *J Digit Imaging* **30** (2017), 460-468.

[16] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R.M. Summers, Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, IEEE, 2017, 3462-3471.

[17] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556* (2014).

[18] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, 770-778.

[19] F. Chollet, Xception: Deep learning with depthwise separable convolutions, *arXiv preprint* (2017), 1610.02357.

[20] Y.C. Hum, K.W. Lai, and M.I. Mohamad Salim, Multiobjectives bihistogram equalization for image contrast enhancement, *Complexity* **20** (2014), 22-36.

[21] S.J. Pan and Q. Yang, A survey on transfer learning, *IEEE Transactions on knowledge and data engineering* **22** (2010), 1345-1359.

Address for correspondence

Rui Zhang, PhD
MMC 912, 420 Delaware Street SE
Minneapolis, MN 55455, USA
zhan1386@umn.edu;
612-626-4209

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190269

Analysis of the Health Information Needs of Diabetics in China

Xiaofeng Zhou[a], Yuan Ni[a], Guotong Xie[a], Wei Zhu[a], Cai Chen[a], Tianhao Wang[b], Zhigang Pan[b]

[a] Ping An Medical Technology, Shanghai, China
[b] Department of General Practice, Zhongshan Hospital affiliated to Shanghai Fudan University, Shanghai, China

Abstract

According to the latest statistics of the China National Health Protection Commission, the prevalence of adult diabetics in China has reached 11.6%, and the number of patients has exceeded 114 million. Understanding the needs of diabetics and what kind of problems they are anxious about are crucial for doctors, hospitals and other health care providers, which can be used to ameliorate patient education services and help patients to improve their disease management skills. Hence we have conducted a study to analyze the questions about diabetes collected from a Chinese health website; the number of which is 151,589. We have divided these questions into 9 categories using a convolutional neural network. The shocking results showed that the questions about prevention only account for 1.23%. And a Chinese patented drug Xiaoke pill, the main component of which is glibenclamide, ranks fourth among the drugs the user cares most about due to the cheap price. However, patients know very little about its side effects.

Keywords:

Diabetes Mellitus, Machine Learning, Natural Language Processing

Introduction

According to the China Internet Network Information Center, the number of Internet users in China reached 802 million in 2018, and the Internet penetration rate was 57.7% [1]. The Internet has become an important carrier for consumers to express their health information needs and search for health information. Furthermore, patients with chronic diseases are more willing to share experiences and seek help via the Internet [2]. In order to meet the needs of the market, there are dozens of companies in China that started offering online disease counseling services for patients, by constructing a connection between patients and doctors through the internet. As a result, these sites have accumulated a large number of disease-related questions, from which we can mine valuable information about what the patients are concerned. Health providers like hospitals can make use of these questions raised by patients to improve their services such as patient health education, and patient follow-up to consumers.

In 2000, the International Diabetes Federation (IDF) predicted that the number of diabetics worldwide would reach 366 million by 2030 [3]. But just 10 years later, in 2011, diabetes data released by IDF showed that the number of people with diabetes worldwide had increased to 366 million [4]. The spread of diabetes is faster than expected. Moreover the number of patients with diabetes in China has ranked first in the world [5].

Nevertheless, people with diabetes usually have no medical background. It's natural for them to feel overwhelmed and scared when faced with diabetes, which is incurable. Hence studying their needs and analyzing their anxieties are fundamental to health care providers to comprehend their needs, which can break the information gap between the patients and health care providers. As a result, health care providers can provide better patient education services and improve self-management ability of diabetics. In order to achieve this goal, we have chosen to study diabetes-related questions on the internet to analyze the needs of diabetics.

In previous studies, Haihong Guo et al. analyzed hypertension-related questions, which studied a very limited amount of data of 2000 questions [6]; Zongcheng Jia et el. studied cancer-related questions with a data of 1000 questions [7]. As far as we are concerned, there is occasionality in the distribution of such a small amount of data, which in general cannot represent the reality. And with the development of artificial intelligence, some traditional machine learning algorithms like support vector machines (SVMs) and Naive Bayesian have been used to classify questions by feeding some features as Bag of Words (BOW) or TF-IDF extracted from the dataset into the algorithm [8], which is not using state of the art technology, meaning that better models can be used to improve the performance.

To analyze the diabetes-related questions, we used a Convolutional Neural Network to classify these questions, which is capable of capturing semantic level information. Based on the results of the classifier, we have further explored the hidden information in the data. This method can be generalized to analyze other chronic diseases by means of feeding labeled data to our model.

Methods

Data Collection

In order to reflect the real distribution of the diabetes-related questions, we designed a spider specific for a Chinese health website (http://www.39.net/) to crawl all the diabetes-related questions without any selection or filter, of which the total number is 151,589. With such a large amount of data, we believe that the results obtained are highly convincing.

Since these questions are all raised by patients, most of which don't have a medical background, these questions about diabetes are very colloquial and there are a lot of misspellings. For instance, Many patients may type "二甲双胍"(metformin) as "二甲双瓜"(metformin). And some consumers may merge several questions into one single question like "Do I have diabetes? Is it necessary to take medicine? Which treatment is better?". The diversity and complexity of the dataset is also a huge challenge for us, which put forward higher requirements for the robustness of our model.

Classification Schema

According to the classification schema of the references [6,9], combined with the specific characteristics of diabetes itself, all diabetes-related questions are classified into 9 categories: diagnosis, treatment, lifestyle, complication, maternity related, prognosis, health provider choosing, prevention, and others. The "others" category represents some description irrelevant to disease. Using such a schema, each question can be classified into one of the classifiers.

Word2vec

Word2vec is an efficient model for learning high-quality, distributed vector representations that capture a large number of precise syntactic and semantic word relationships [10]. Unlike traditional Bag of Words representations, words are projected from a sparse, 1-of-V encoding using one-hot encoding (here V is the vocabulary size) onto a lower and dense dimensional vector space via a hidden layer. In general, the dimensions of vectors are set between 100 and 500. In such dense representations, semantically close words are likewise close in Euclidean or Cosine Distance in the lower dimensional vector space. Since Word2vec uses the unsupervised learning method, it's very efficient and easy to use.

It has been shown that performance can be improved by training the Word2vec model using domain-specific data before training a downstream neural network [11]. To train our own word2vec model, we fed a large corpus of text into the model, including all the diabetes-related questions and a lot of other medical-related documents. In our model, the dimension of the vector is set at 300.

As a result, each word can be represented in a distributed vector, containing syntactic and semantic information. Therefore, we can calculated the similarity that can be represented by the cosine similarity between the 2 vectors correspondent. Here are some interesting results:

Table 1. Word Similarity

Word1	Word2	Similarity
糖尿病 (diabetes)	高血压 (hypertension)	0.62
糖尿病 (diabetes)	苹果 (apple)	0.18
糖尿病 (diabetes)	水果 (fruit)	0.17
糖尿病 (diabetes)	蔬菜 (vegetable)	0.12
水果 (fruit)	高血压 (hypertension)	0.14
水果 (fruit)	苹果 (apple)	0.63
水果 (fruit)	蔬菜 (vegetable)	0.87

As is shown in the table above, the semantic information is well captured by the model trained. Such a well-trained model can improve the performance of our neural network model.

Convolutional Neural Network

There are two, main deep, neural network architectures in NLP (natural language processing): convolutional neural network (CNN) and recurrent neural network (RNN). Generally speaking, CNN is mainly used for classification tasks like sentiment classification since sentiment is usually determined by some key phrases; RNNs is mainly used for a sequence modeling task like language modeling as it requires flexible modeling of context dependencies [12]. Since our task is to classify questions, CNN is chosen as our basic network architecture.

Convolutional neural network, originally invented for computer vision tasks, which uses convolving filters to extract local features from images, have also been shown to perform well in natural language processing. With the help of the Word2vec model, sentences can be embedded in a matrix. After that, sentences can be treated just as an image which is originally represented by a matrix.

In this study, we used a multi-channel CNN which is largely used in our other classification tasks, and have also been proven to achieve excellent results. The main model architecture is shown in Figure 1.

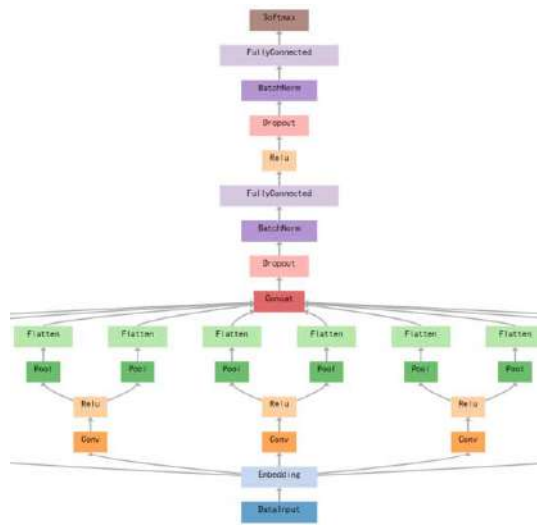


Figure 1. Main Model Architecture

Note: Each "Conv"(convolutional) module in this graph represents a convolving filter with different kernel size. Due to limitations of image size, not all the "Conv" module are presented in the figure. Max Pooling layer and Average Pool layer are used at the same time, represented by "Pool" module in the graph. To avoid overfitting problem, 0.5 is set as the dropout rate in the "Dropout" module.

Annotation and Training Procedure

To train a deep neural network, it's essential to feed labeled data into the deep model. We manually annotated 1000 questions, randomly selected from the dataset in the first place. Using such a small amount of data, we have been able to train a model that performs not very well but can already identify certain patterns according to the limited training data.

What we did next was to use the convolutional neural network to predict these unlabeled data and use the results as references to help us speed up the annotation process. The output of the model represents the probabilities that one sentence is classified into each category. Those sentences with low probabilities mean that the model fails to identify the patterns in the questions, which in most cases, represents that there is no similar pattern in the labeled data. So we need to annotate more data with new patterns for the model to learn. In this study, we defined the uncaptured pattern threshold at 0.7, and the captured pattern threshold at 0.95, meaning that if prediction probability of a question is below 0.7, we judge that the pattern of this sentence cannot be recognized by the model. In contrast, the prediction probability of a question above 0.95 means the sentence is well recognized.

Next, we manually annotated another 500 questions, randomly selected from the unrecognized questions. By repeating the training and annotation process above 4 times, we were able to train a well performing convolutional neural network.

Results

Neural Network Performance

We have annotated 3000 questions in total by means of the annotation and training strategy mentioned in the method section. All labeled data is divided into 2 sets, the training set occupying 80% for training our model, and the test set occupying 20% for evaluating our model, which is invisible to the model. The accuracy of our neural network has reached 96.7% in the test set, representing the performance of our convolutional model.

Using such an annotation and training strategy, we avoided annotating the same patterns all the time and were able to feed more new patterns to our neural network, which can largely improve the performance and robustness. Here is an example of the classification result to prove the robustness of our deep neural network:

- Question without misspelling:
  - 长期吃二甲双胍对健康都有些什么副作用？(What are the side effects of long-term consumption of metformin on health?)
- Question with misspelling:
  - 长期服用二甲双瓜有什么副作用(What are the side effects of taking metformin for a long time?)

Both questions are properly classified into the same class: treatment. The classifier works fine even if there is a spelling mistake in the sentence.

Classification Result

The classification result of the number of 151,189 questions is presented in Figure 2. We can conclude that the most concerned aspects of diabetics are diagnosis, lifestyle, and treatment. However, little attention has been paid to how to prevent diabetes, occupying just 1.23 percent, suggesting that patients rarely care about diabetes before actually getting diabetes.

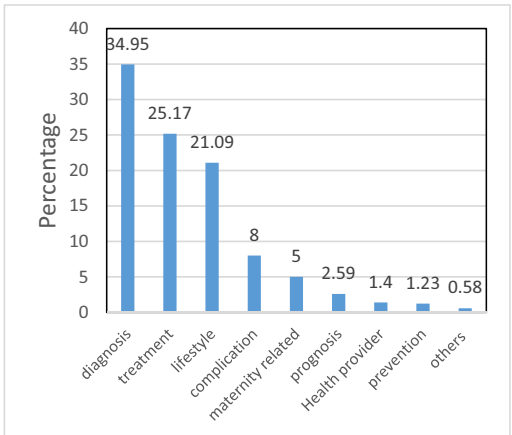


Figure 2. Classification Results

Study of Prevention Class

We carefully studied the questions of diabetes prevention. What we have found is that under most circumstances, people will start asking questions about how to prevent diabetes only when a relative or a friend has diabetes, here are some typical examples:

- 我的几个好朋友都得了糖尿病。怎么样才能预防糖尿病呢？(Several friends of mine have got diabetes. How can I prevent diabetes?)
- 身边有不少的亲戚朋友得了糖尿病，他们都很痛苦。所以我心里很担心，也很害怕。请问我该怎样预防糖尿病？(There are quite a few friends and relatives around me who suffer a lot from diabetes. So I am distressed and scared. How can I prevent diabetes?)

On the basis of the result, education about diabetes prevention needs to be spread to more people to help everyone recognize the importance of diabetes prevention.

Study of Lifestyle Class

Comparing questions before 2010 and after 2010 in Figure 3, the proportion of patients' questions about lifestyle increased from 13.71% to 22.59%, which shows a significant increase. A reasonable explication is that consumers are now more aware that lifestyle plays a very import role in diabetes.

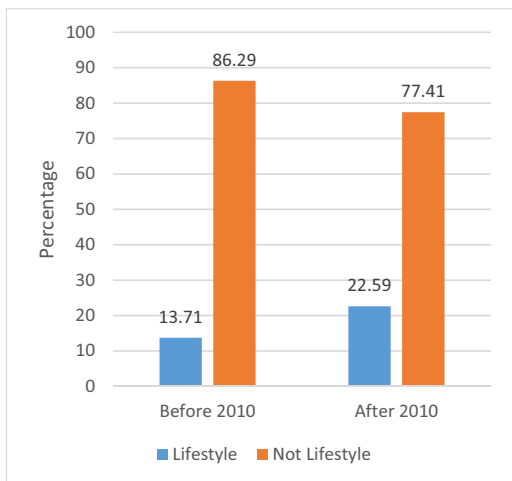


Figure 3. Comparison of Lifestyle Related Questions Before and After 2010

Knowing that diet plays a crucial role in diabetes, we conducted a further study about the most cited food by diabetics in the lifestyle class, shown in Figure 4. As for Chinese, eating has always been one of the most important things in daily life. Medical professionals can provide a better dietary recommendation for patients as a reference by means of the results obtained.

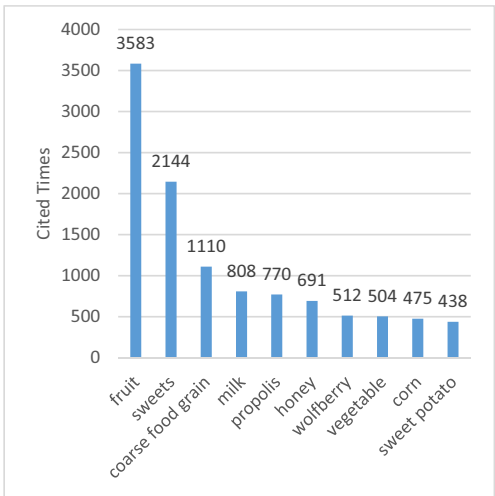


Figure 4. Top 10 Cited Foods

Study of Different Age Groups

It is also fundamental to analyze the different needs of patients of different ages, the result of which is presented in Figure 5. We can see significant differences in the aspects that people of different ages are concerned about.

Young people are more concerned about maternity-related issues, who are at an age suitable for childbearing. They may worry about problems like whether diabetes will be passed on to children or whether people with diabetes have fertility.

While older people are more concerned about complications and lifestyle to alleviate the impact of diabetes on life. Many people at this age are using oral medication or insulin therapy. They are more eager to know the drug usage, efficacy and adverse reactions.

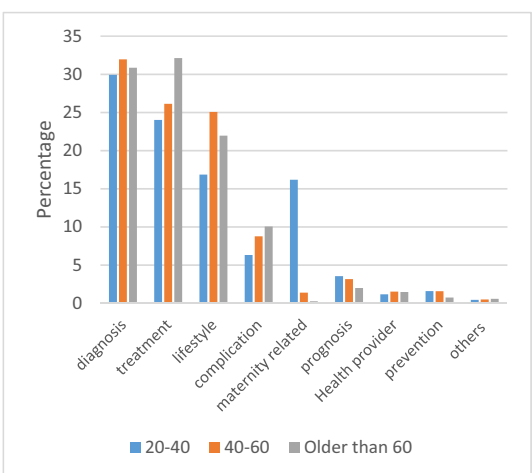


Figure 5. Classifcation Results of Different Age Groups

Study of Drugs

About the medication for diabetes, consumers raised more questions about insulin, metformin, gliclazide, and the Xiaoke Pill. Despite the import role of insulin in blood sugar control, the application of insulin is far from sufficient. Many diabetics, even medical staff, especially non-endocrinologists, have a variety of concerns and hesitation about the use of insulin, which suggests that diabetes professionals should teach patients and relevant medical staff more knowledge of insulin, improve their attitude towards diabetes.

The questioning rate of metformin is high, and that may be related to metformin being recommended as the first choice for diabetes treatment.

Xiaoke Pill is a kind of Chinese patent medicine, which is favored by many consumers because of the cheap price. Also, in China, many patients believe that Chinese traditional medicine has fewer side effects than western medicine, which may also be a reason why Xiaoke Pills are very popular in China. However, Xiaoke Pill contains glibenclamide, which can cause severe hypoglycemia and may lead to liver and kidney damage with a long-term application. Therefore, medical professionals should fully explain to consumers the side effects of Xiaoke Pills, and guide consumers to abandon the use of this drug.

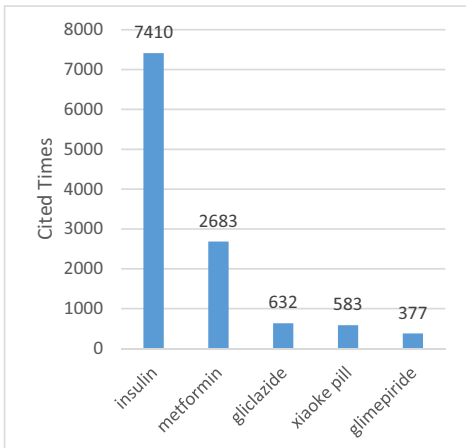


Figure 6. Top 5 Cited Drugs

Discussion

In this study, we have applied Word2vec and a state of art convolutional neural network to classify diabetes questions. In addition, we have used a more advanced annotating strategy, which can feed more patterns to the deep neural network to improve the performance. To achieve the same performance using the normal annotating strategy without pattern selection, we need a larger amount of annotated data.

Compared to the previous works presented in the introduction section, which tried to classify and analyze disease-related questions, the results of our work is more convincing. At the same time, we further explored more hidden information based on the results of the classifier.

Doctors, hospitals, and government medical institution can use the results of this analysis to better understand the needs of diabetics and discover the aspects in need of improvement.

For example, there are only a few problems related to the prevention of diabetes, meaning that people are not paying enough attention to the prevention of diabetes, which may be a reasonable cause of the growing number of diabetics in China.

There are also many shortcomings in our research. First, we have annotated 3,000 diabetes-related questions, which requires extensive work and takes a long time. Our model cannot be directly generalized to other diseases. To do a similar analysis of another disease, we have to redesign the

classification schema and annotate the new data. The classification schema we used is not perfect. For example, questions about sport, diet, and living habit are classified into the same class: lifestyle.

In the future, we should aim at improving the limitations of our work. Some more advanced deep neural networks like ELMO [13] and BERT [14] have achieved state of the art performance in many NLP tasks. Maybe these networks can be used to improve the annotating and training process. And, we can redesign a better classification schema that can be easily generalized to another chronic disease. We hope that more advanced natural language processing technologies can be applied to the medical domain.

Conclusions

In this study, we identified the diabetes information needs for consumers by analyzing a large amount of data. We have developed a classification schema for diabetes-related questions and used a convolutional neural network to classify the questions according to the schema designed. And, we found that the prevention of diabetes has not received enough attention in China and the popularity of Xiaoke Pill in China also needs to be reviewed.

Acknowledgements

This study is supported Department of General Practice and Department of Endocrinology, Zhongshan Hospital affiliated to Shanghai Fudan University.

References

[1] China Internet Network Information Center. China Internet Network Development State Statistic Report. 2018.8.
[2] Susannah F, Kristen P. Chronic Disease and the Internet. [http://www.pewinternet.org/2010/03/24/chronic-disease-and-the-internet/, Accessed November 25, 2018]
[3] International Diabetes Federation: Diabetes Atlas 2000. Brussels, International Diabetes Federation, 2000
[4] International Diabetes Federation: Diabetes Atlas 2011. Brussels, International Diabetes Federation, 2011
[5] International Diabetes Federation. IDF Diabetes Atlas, 8th edn. Brussels, Belgium: International Diabetes Federation, 2017.
[6] Haihong G, Xu N, Jiao L., Consumer health information needs and question classification: analysis of hypertension related questions asked by consumers on a Chinese health website, *Studies in Health Technology and Informatics* **216** (2015), 810-814.
[7] Zongcheng J, Yaoyun Z, Jun X, Xiaoling C, Yonghui W, and Hua X., Comparing Cancer Information Needs for Consumers in the US and China, *Studies in Health Technology and Informatics* **245** (2017), 126-130
[8] Haihong G, Xu N, Jiao L., Automatically Identifying Topics of Consumer Health Questions in Chinese, *Studies in Health Technology and Informatics* **245** (2017), 388-392.
[9] Zhang J, Zhao Y., A user term visualization analysis based on a social question and answer log, *Information Processing & Management* (2013) 1019–1048.
[10] Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In Proceedings of NIPS 2013.
[11] Yoon Kim. Convolutional Neural Networks for Sentence Classification. In Proceedings of EMNLP 2014
[12] Yin, Wenpeng; Kann, Katharina; Yu, Mo; Schütze, Hinrich. (2017). Comparative Study of CNN and RNN for Natural Language Processing. eprint arXiv:1702.01923.
[13] Peters, Matthew E.; Neumann, Mark; Iyyer, Mohit; Gardner, Matt; Clark, Christopher; Lee, Kenton; Zettlemoyer, Luke. (2018). Deep contextualized word representations. eprint arXiv:1802.05365.
[14] Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. eprint arXiv:1810.04805.

Address for correspondence

Corresponding author: Xiaofeng ZHOU, Ping An Medical Technology, 316rd Laoshan Road, Pudong District,Shanghai, China
Email: 9437852@qq.com.
Phone: +86-18672842160.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190270

Challenges and Opportunities in Changing Data Structures of Clinical Document Archives from HL7-V2 to FHIR-Based Archive Solutions

Jochen Zohner [a], Kurt Marquardt[a], Henning Schneider[a], Achim Michel Backofen[a]

[a] *Department of Medical Informatics, Justus Liebig University, Giessen, Hesse, Germany*

Abstract

In 2018, a major replacement of clinical applications took place at the University Hospital of Giessen. One key part was the clinical document archive containing a vast collection of clinical data from the last 30 years. The aim of this sub-project was to move all data to a new system without any loss, while maintaining all functionality and all communication interfaces. This project successively resulted in a complete paradigm change in document storage. While the legacy clinical data repository (LCDR) was designed according to HL7-V2 principles, the replacement resulted in an HL7-FHIR implementation. The aim of this work is to discuss the differences between both approaches, the obstacles that appeared during migration, but also the opportunities resulting from the new philosophy, especially as far as the impact on the use of scientific data is concerned.

Keywords:

Information Storage and Retrieval**;** Database Management Systems**;** Electronic Health Records

Introduction

In current clinical network infrastructures, a central document archive has become more and more essential to prevent medical staff from spending too much time seeking relevant patient information. The resulting data aggregation not only leads to more efficient clinical workflows but produces new opportunities for data analysis, ranging from user optimized cumulative data views, to support for accounting information generation, to diagnosis prediction and assistance on scientific studies. The University Hospital of Giessen has a long tradition in the use of clinical document archives, resulting in a vast collection of data, starting with lab values in the late 1980s today, offering a spectrum of more than 50 data sources and a volume of about 300 million datasets. To cope with these large amounts of data, the HL7-V2 standard was consequently used not only for data transmission, but also for storing the data in databases using the HL7-segment structures as table definitions. The outcome was a standardized Entity-Attribute-Value (EAV) model [1,2], which has served clinical demands well during the last decades.

The emerging requirements for mass data storage, data security and protection, as well as legal requirements, forced the University Hospital of Giessen to replace the proprietary LCDR with a modern commercial clinical document archive solution by Synedra, an information technologies company. Synedra's approach hereby entirely differs from the LCDRs EAV-model, since all data is stored as monolithic data in database BLOBs, compensating for the disadvantages of this approach by using the HL7-FHIR [3] standard in JSON-format.

Fundamental Concepts and Data Types

The principal purpose of a clinical document archive (CDA) comprises data collection, storage, and display. Additionally, the CDA has to provide stored information for various medical subsystems, especially for the clinical information system (CIS) itself. The export of anonymized data for the support of clinical studies is another purpose that a CDA has to satisfy. In all of these cases, the CDA acts as a uniform abstraction layer for data originating from a broad range of different medical subsystems, like laboratory information systems (LIS), radiology information systems (RIS), or department specific information systems like for cardiology, pneumonology, etc. For different sources of information, the structure of the corresponding data is also different. Basically, we can distinguish the following classes of data, each requiring entirely different treatment:

1. **Monolithic documents**, which are most often stored as PDF or image formats like JPEG;
2. **Plain text** or formatted text (like HTML), which are transmitted inside an HL7-ORU message;
3. **Structured data** with multiple observation values;
4. **Reference pointers** to documents or web pages stored in medical subsystems;
5. **DICOM** references from the PACS; and
6. Additional data from **clinical studies** most often stored in Excel spreadsheets.

The LCDR was principally designed for presentation and as a broker server without the demand of storing all data inside its own database, but leaving monolithic data, webpages, and DICOM data on external databases of the originating subsystems and referring to them via web linkage. According to this, the preferred data type of the LCDR was structured data, having the advantage of flexible display and the best usability for statistical evaluation. In contradiction, the new Synedra archive is intended to store all information in its own database, for the sake of data persistence, and treats all information like monolithic data.

Therefore, Synedra creates a HL7-FHIR resource containing all the structured information and metadata concerning the clinical finding. These data are stored in JSON format into a Postgres database, making it reasonably searchable by creating appropriate indices, as well as in XML format, to the file system for display purposes. This closes a major gap between the functionality of the LCDR and the Synedra archive and enables Synedra to provide customized data presentation

layers, like cumulative lab display and selective access to structured data. Also, Synedra uses the XML resources to deliver entire documents via RESTful web services, allowing the CIS and other subsystems to participate in the data stored in the Synedra archive on demand.

Another improvement provided by the Synedra archive is the complete integration of DICOM data. Since 1996, Giessen was lucky to successfully manage the linkage between DICOM images and corresponding radiological findings by transmission of study instance UIDs within HL7-ORU messages. However, the information from DICOM image tags was not searchable in the LCDR because it only possessed a display link to the DICOM series but not the images themselves. This problem is fixed in the Synedra archive by receiving all DICOM series from the PACS. Similarly, the monolithic PDF documents and links to external web interfaces, which could not be searched by the LCDR, became searchable by Synedra by importing its contents into its database. Finally, data spreadsheets provided by clinical studies, including questionnaire responses or observations, were able to be imported and stored in the Synedra archive as appropriate FHIR resources [3].

Methods and Results

Challenges during Data Migration

It turned out that the migration of data from the LCDR to the Synedra archive was even more demanding than expected. The very first approach, to generate download lists with all the metadata of the documents and have Synedra download all the documents using the usual web interface of the LCDR, led to an undefinable set of crippled data due to the complexity of the data landscape and the many subsystems involved, making appropriate exception handling impossible. Therefore the transmission via HL7-MDM using the Orchestra communication server as middleware was considered. This resulted in a vast range of different errors, depending on the data's document classes (as mentioned in the section above), requiring individual adjustment.

Structured data

The Synedra archive was not intended for the use of structured data, initially. When the question arose, how to handle these data, maintaining both the original presentation of the data and the structured information itself, the idea of using the HL7-FHIR standard was born. Since a FHIR resource is suggested to not only contain patient information and observation result sets but an HTML representation of the data for display purposes [3], this format turned out to exactly fit the existing requirements.

Monolithic documents

The architecture of the Synedra archive was primarily designed for storing documents without caring about their inner structure. These documents are most often stored with a fixed representation like PDF. The transport of such documents can be performed either by storing the data to a shared folder and transmitting a reference to the document's storage location or transmitting the document within the corresponding HL7-message. The latter approach allows the distribution of documents to several recipients without interference and so was preferred by Synedra. This approach impacted on the Orchestra communication server like inflating its log files to exceed the file system's limits and overburden the debugging tools whenever errors occurred, and so was abandoned.

The most common errors with these data were due to subsystems deleting or changing the documents without notifying the LCDR of the changes. These problems also were solved by Synedra's storage philosophy for the future.

Plain text

The LCDR behaved rather fault tolerant. When transmitting plain text data in HL7-OBX segments, special characters, wrong coding information and misused HL7-escape-characters spoiled quite a few data transmissions either when passing through the Orchestra communication server or when received by Synedra. These errors had to be fixed in the LCDR data stock one by one.

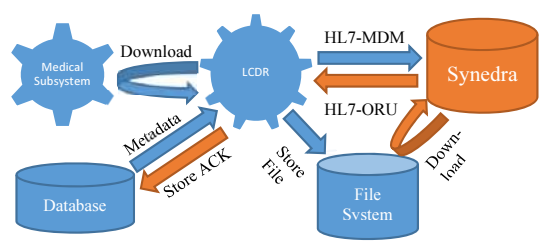
Reference pointers

Synedra stores all data persistently in their database and does not hold references to external resources. The LCDR held several references to different subsystems, some of them providing interactive web pages that display data only after a user action was performed. These resources could not be downloaded like usual documents and needed to be rebuilt or exported in different ways, directly from the source of information. Another obstacle arose by access control using tokens with time limitations.

Communication Approaches

The remedy for the latter problem led us to the general approach of making the LCDR download the documents, store them to a shared folder and send an HL7-MDM notification for each one to the Synedra archive. Since the Synedra archive always sends HL7-ORU messages with reference pointers to access recently stored documents along with the original document identifiers to the communication server after successful storage, we used this mechanism like an "acknowledge with regard to contents" notification and stored these response messages along with the original data into the LCDRs database. This procedure made the export status completely transparent (*Figure 1*) (*Table 1*).

Figure 1 – Document Export



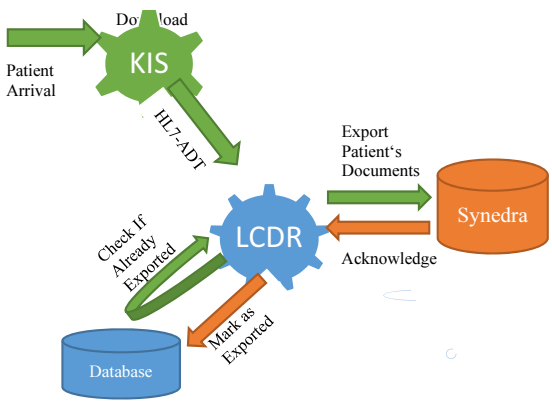
Another problem proved to be the timely manner. Data export would take longer than acceptable due to performance limitations of the LCDR. On the other hand, one could not expect the clinical staff to search two archives for the documents. To solve this, an additional process, which we called "export on demand," was established. Whenever a patient visits the hospital, an HL7-ADT message is released by the CIS. This message, received by the LCDR, triggers an export mechanism for all documents concerning this patient (*Figure 2*).

Table 1 – Sample Data from the LCDR including Links to Notification for and Acknowledgement Messages from Synedra

Document ID	Description of Finding	Date of Finding	Notification Msg. ID	Notification Msg. Date	Stored File Name	Acknowledge Msg. ID	Acknowledge Msg. Date
339583885	Herzkatheteruntersuchung	02.05.2013 09:40	519210180	23.10.2018 19:13		519210202	23.10.2018 19:14
327421949	OPG-P1-Standard	02.02.2012 09:58	519210180	23.10.2018 19:13		519210198	23.10.2018 19:14
353465629	Belastungs-EKG	19.08.2014 09:23	519157188	22.10.2018 08:21	0237270.353465629.pdf	519157213	22.10.2018 08:21
353463516	Ruhe-EKG	19.08.2014 08:47	519157173	22.10.2018 08:20	0237270.353463516.pdf	519157207	22.10.2018 08:21
230778364	Belastungs-EKG	18.06.2009 10:28	519135847	20.10.2018 06:45	3132846.230778364.pdf	519135904	20.10.2018 06:45
309785671	IMMUNPATHOLOGIE: Befundbericht	08.03.2011 13:00	519135847	20.10.2018 06:45	3132846.309785671.pdf	519135892	20.10.2018 06:45
245594337	Urin	20.10.2010 14:04	519135847	20.10.2018 06:45	3132846.245594337.html	519135951	20.10.2018 06:45
321944955	Pathologisch-anatomische Begutachtung	07.07.2011 12:20	519135847	20.10.2018 06:45	3132846.321944955.html	519135931	20.10.2018 06:45
237326059	Anztrum, Corpus, Z-Linie	13.01.2010 14:04	519135847	20.10.2018 06:45	3132846.237326059.html	519135953	20.10.2018 06:45

Thus, instantly all of a patient's documents seemed to be present in the Synedra archive to the clinical users, although the very last transmission took place almost one year after the production startup of Synedra.

Figure 2 – Export on Demand



Used HL7-FHIR Resources and Versions

HL7-FHIR is a standard currently evolving at a rapid pace. FHIR consists of several resources, each one designed for appropriate clinical demands. Hereby, both the set of resources and the structure of the individual resource itself depend on the version of FHIR used. At the project's start, FHIR version "STU3", the current version at that time, was used. In December 2018, "R4" was officially released [3]. The Synedra CDA today supports both versions, STU3 and R4, as far as the interfaces are concerned, but stores new data according to version R4 and successively migrates legacy data from STU3 to R4.

Concerning the FHIR resources, "DiagnosticReport" and "Observation," supported by "Patient", "Organization" and "Practitioner," were the first to be implemented. Further resources like "MedicationAdministration", "Condition", "Procedure", "Claim", "AllegyIntolerance", "FamilyMemberHistory", "Immunization" and "QuestionaireResponse" [3] were drafted and are currently being implemented. The latter resources especially play an important role for the use of scientific data.

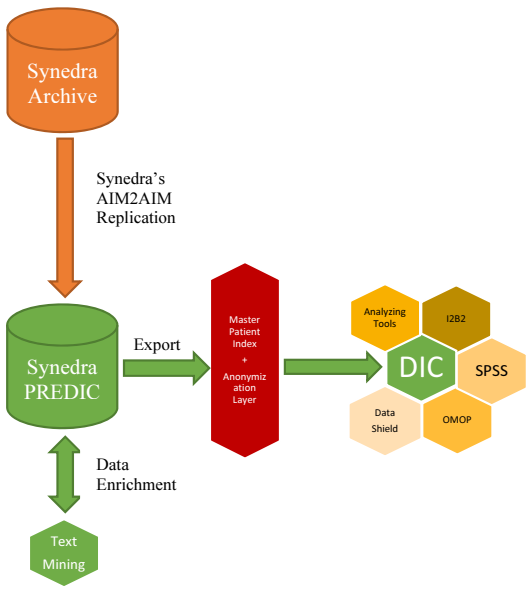
The FHIR standard allows the implementation of extensions, offering the possibility of storing additional information not covered by the standard, if necessary. For the Synedra CDA, the standard proved to be sufficient except for one single extension of the "DiagnosticReport," which holds the original HL7-ORU message for the sake of reproducibility of the "HL7-V2 to FHIR" conversion step. While no major problems occurred using FHIR, it may be noteworthy to mention that the flexibility of FHIR, like the substantial number of optional elements and the extensibility, makes the beginning steps with FHIR easy but leads to a lot of work later on.

Impact on Scientific Data Use

A significant benefit from centralized CDAs is the possibility to use a broad range of clinical data for the purpose of clinical research. The LCDR supported such attempts by providing a number of searching facilities based on the selection of patients having some observations exceeding user-specified values. Of course this only worked with structured data and so was mostly limited to lab values and some similar data sources. Along with the Synedra archive's architecture and its capabilities, the idea was raised to also use the archive in a more extensive way to support clinical research. As we have already seen, Synedra made not only structured data but all data searchable, including even monolithic and DICOM data. This allows us to specifically export selected anonymized data to be stored into a so-called data integration center (DIC) [6], which can in turn be used afterwards with analyzing tools like SPSS, OMOP [7] or I2B2 [4,5].

For non-structured data, the use of text-mining tools becomes sensible and allows us to include a much broader range of data than ever before, accompanied by the possibility to add the so-gained structured information to the original document. In order to avoid confusing the clinical data with scientific analysis data, a second instance of the Synedra archive was implemented, acting like a preprocessing facility for the DIC and therefore being called PREDIC. By Synedra's event-based replication capabilities, the Synedra PREDIC becomes a complete mirror of the data collected in the Synedra archive for the sake of scientific evaluation (*Figure 3*), expanded by study data, additional subsystems database contents, and document-related analysis results. Therefore, again Synedra's HL7-FHIR interface is used for injecting the additional information. These features open the door to standardized data export of patient-related medical observation data, surpassing the rudimentary possibilities of the LCDR by far.

Figure 3 – PREDIC Architecture



Discussion

The replacement of an old system by a new one can either be regarded from the perspective of the old system or from the new one. After two decades of the LCDR running in this manner, its replacement at first was planned like a one-to-one exchange of one archive for the other. Thus, the first replacement approaches were based very much on the LCDR structures, like using the abstraction layer of the LCDR for downloading all data sets in a uniform way. The power of this abstraction layer in disguising the documents' origin and the delivery method turned out to be more appropriate for human usage than for archiving purposes. Even taking random samples is not an adequate method for discovering problems that appear in some thousands of 300 million cases. One has to face all possible exceptions and errors that the legacy system tolerates. The generic approach with trial and error results not only in delays but in vague states, forcing data exports to be repeated several times. On the other hand, the method of storing acknowledgement messages directly related to the original document (*Table 1*) slows down the export process for every single document, but proves to accelerate the entire operation, since it makes selective resending of failed data possible without exploiting log files or performing manual corrections. This procedure also made it possible to establish the export on-demand function, instantly eliminating the user's need to fall back to the use of the LCDR for legacy documents.

When considering the functionality of the new CDA, the opposite mistake was made in the beginning. The view of the present Synedra archive dominated, reducing the expectations to simply store and display monolithic documents. The emerging need for dealing with structured data augmented the view and merged the pure archive with the repository approach by involving the HL7-FHIR standard.

Conclusions

The original attempt of renewing a clinical document archive primarily designed to maintain the status quo became a mutual process of enhancement. While the data transfer became a rather complex task, the Synedra archive evolved into a versatile platform including solutions for all the repository capabilities of the LCDR by the use of HL7-FHIR, accompanied by a remarkable expansion of Synedra's archive functionalities. We not only saved 30 years of precious clinical data, we even gained an overall solution for both clinical document archiving and a universal support of clinical studies.

Acknowledgements

The authors thank Michael Maier (University Hospital of Giessen), the MIRACUM Team (Justus Liebig University) and the Synedra Information Technologies Co. for supporting this work and the permission to publish specific details.

References

[1] S.B. Johnson. Generic data modeling for clinical repositories, *J Am Med Inform Assoc.* **3** (1996), *328-339.*

[2] J. Anhøj. Generic design of Web-based clinical databases, *J Med Internet Res.* **4** (2003), e27.

[3] HL7 FHIR Foundation, (Accessed 03/26/2019) http://www.hl7.org/fhir/directory.cfml

[4] T. Ganslandt, S. Mate, K. Helbing, U. Sax, H.U. Prokosch. Unlocking Data for Clinical Research - The German i2b2 Experience. *Appl Clin Inform.* **2** (2011), 116-27.

[5] R.W. Majeed, R. Röhrig. Automated realtime data import for the i2b2 clinical data warehouse: introducing the HL7 ETL cell. *Stud Health Technol Inform*. **180** (2012), 270-4.

[6] H.U. Prokosch, T. Ganslandt. Perspectives for medical informatics. Reusing the ectronic medical record for clinical research. *Methods Inf Med.* **48** (2009), 38-44.

[7] Y. Si, C. Weng. An OMOP CDM-Based Relational Database of Clinical Research Eligibility Criteria. *Studies in health technology and informatics.* **245** (2017), 950-954.

Address for Correspondence

Jochen Zohner, GB IT, Universitätsklinikum Gießen, Klinikstr. 23, D-35392 Gießen, Germany.

jochen.zohner@akad.med.uni-giessen.de

This page intentionally left blank

II. Supporting Care Delivery

This page intentionally left blank

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190272

The Role of Electronic Health Records in Improving Communication Between Health Professionals in Primary Healthcare Centres in Riyadh: Perception of Health Professionals

Bander Alanazi[a], Kerryn Butler-Henderson[b], Mohammed R. Alanazi [c]

[a] *Tasmanian School of Business & Economics, Launceston, Tasmania, Australia*
[b] *Tasmanian School of Business & Economics, Launceston, Tasmania, Australia,*
[c] *King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia,*

Abstract

Improving communication among healthcare providers is one of the critical components of safe and quality patient care. The study objective is to examine how occupation and training of health professionals in Riyadh, Saudi Arabia influence professionals' perception of the role of electronic health records (EHRs) in improving communication between healthcare providers in primary healthcare centres. A survey-based study method employing a descriptive, cross-sectional design was used. Health professionals' occupation and training were found to influence their perception of the role of the EHR in improving interprofessional communication. Physicians and professionals with training on EHRs had the highest rating for the system's role in improving communication between healthcare professionals. All healthcare providers should embrace EHR systems in their practice to promote interprofessional communication and collaboration in the patient care process.

Keywords:
Electronic health records, Primary health care, Health personnel

Introduction

Effective communication among health professionals is crucial for the improvement of patient safety, and outcomes with ineffective communication noted to result in several adverse events such as medical errors that put patient's life at risk [1]. Electronic health records (EHRs) have been demonstrated to be useful systems for improving communication not only among the healthcare providers themselves but also between patients and their caregivers [2-5]. In promoting the exchange of patient's health information among providers, the EHR creates an opportunity for care coordination between health professionals and medical departments that enhances the provision of safe and quality patient care marked with reduced medical errors and other adverse events [6].

Despite the implementation of EHR systems resulting in these benefits due to improved communication, it has been established that sufficient communication among healthcare providers is still a challenge in the healthcare system with much focus having been placed on provider-patient interaction more than in communication between physician and professional colleagues [7]. The challenge is even bigger in Saudi Arabia where English is a second language but considered the official language used in communication in Saudi healthcare organisations. In addition, the Saudi healthcare organisations have health professionals from all around the world with different backgrounds and of course different mother tongue languages. Most of the research on the EHR role in improving healthcare outcomes has also only been reported in western countries with little evidence in the Saudi context. This could be due to the low adoption of EHR systems in Saudi Arabia due to various factors such as lack of perceived usefulness by the health professionals and hospital size [8]. To the best of the authors' knowledge, there are no previous studies that have discussed the role of the EHR in improving communication in Saudi Arabia especially from the perceptions of health professionals in primary health care (PHC). Factors influencing the use of EHRs by healthcare providers in these settings have also not been adequately investigated. Therefore, this study aimed to evaluate the perceptions of health professionals in relation to their occupation and training on the role of EHRs in improving communication among healthcare providers for better patient care in primary health care centers (PHCCs) in Riyadh, Saudi Arabia. In so doing, the study had three specific objectives: 1) to investigate the healthcare professionals' perception towards the EHR role in improving communication between healthcare providers, 2) to investigate the correlation between the occupation of health professionals and perception of use of EHRs in improving communication between health professionals, and 3) to investigate the relationship between training on EHRs and perception of use to improve communication between health professionals.

It was hypothesized that several factors could influence the users' perception and use of the EHR to facilitate communication with colleagues. In this study, the occupation of the healthcare professionals in Saudi primary care and training on EHR systems were expected to influence the professionals' perception of the role of the EHR in improving communication among health practitioners.

Methods

This cross-sectional study was designed to investigate the perception of healthcare professionals working in PHCCs in Riyadh city on the role of EHRs in improving communication among health professionals leading to high-quality and safe healthcare in relation to the professionals' occupation and training on EHRs.

Study setting and sample

The study was conducted in Riyadh City which has the highest number of PHCCs since it is the capital of Saudi Arabia. The region has 1710 healthcare professionals comprising physicians,

nurses, pharmacists, laboratory technicians and those from allied health professions such as physiotherapists working in PHCCs [9]. Therefore, a survey involving all the healthcare professionals was undertaken to evaluate their perceptions on the role of EHRs in regards to the quality of service provision. This population was chosen because they interact with EHR on a daily basis and are considered to be in the best position to gauge the application of these tools in the provision of care. Other employees in the primary health care sector in Riyadh falling in various categories such as clerks, drivers, office assistants, administrative personnel, and any other group of supporting staff were not included in the study.

Instruments

The data collection instrument was an online-based questionnaire using Research Electronic Data Capture (REDCap) to ensure the participants' convenience. The questionnaire included both closed and open-ended questions. It was adapted from the survey tool developed by Secginli and peers to evaluate the perception of Turkish healthcare professionals towards the EHR in Family Health Centres (FHCs) which are primary health care settings [10]. The tool had 14 statements related to EHR benefits of which "improves communication between health professionals" was one of them. In addition, it had 9 items for perceived obstacles to implementing the EHR in PHCCs and 10 related to the quality of healthcare services due to the use of EHR in PHCCs. The measurements were based on a Likert scale of 1 to 5 representing strongly disagree to strongly agree. The survey tool was tested for reliability and validity using a pilot study done online in the same way it was to be used in the actual study. This pilot test involved 23 participants drawn from the same population with the sample illustrating a normal distribution with Cronbach's alpha of 0.828 and 0.969 for the combined 23 items for perceived benefits and obstacles, and 10 items for quality of health services respectively compared to the total population as shown in Table 1. The validity of the results was enhanced by including the entire population in the study which improves the response rate. For clarity, the questions in the questionnaire were carefully worded and properly ordered to eliminate any form of ambiguity or bias.

Table 1 – Reliability Tests

Dimension	Cronbach's alpha	No of items
EHR benefits and obstacles to implementation	0.828	23
Quality impact on healthcare services	0.969	10

Data collection procedures and ethical issues

The data was collected for a period of two months, that is between 30th November 2017 and 30th January 2018. All the healthcare professionals working in PHCCs in Riyadh City were invited, through the General Directorate of Health Affairs in Riyadh Region and Human Resource (HR) departments in each of the PHCCs, to complete the survey that was made available in an online format using REDCap for the participants' convenience. The participation was entirely voluntary and by completing the survey, one was considered to have given informed consent. After completing the survey, the data was retrieved from the database for analysis. The study was approved by the University of Tasmania Social Science Human Research Ethics Committee and the Ministry of Health of Saudi Arabia.

Data analysis procedure

The Statistical Packages for Social Sciences (SPSS) (version 20) was used for data analysis. Descriptive statistics were mainly calculated for demographics and overall responses, and they included frequencies, percentages and means. A chi-square test of independence was used to determine the relationship between the healthcare professionals' perceptions of EHR benefit in improving inter-professional communication with respect to their occupation or previous training on EHR. The p-value was set at 0.05.

Results

Participants' characteristics

Out of the 1710 questionnaires distributed, 1127 were returned and found to be eligible for analysis. This represents a response rate of 66%. Most of the respondents were nurses representing 32.6% of the sample. Majority of the respondents (72%) were Saudi nationals. Female health professionals also comprised more than half of the health professionals representing 55.4%. The majority were 20 – 34 years old and had worked for less than 10 years or less in PHCCs in Riyadh. Moreover, almost 60% had no previous experience outside Saudi Arabia, training on EHR and experience on EHR in primary healthcare. This data is presented in Table 2.

Table 2 – Demographic Data of the Whole Population (n = 1710) and the Respondents (n = 1127)

Demographic characteristics		All (N = 1710) n (%)	Respondents (N = 1127) n (%)	Response rate per demographic (%)
Occupation	Physician	369 (21.5)	209 (18.5)	56.6
	Nurse	543 (31.7)	367 (32.6)	67.6
	Pharmacist	256 (14.9)	208 (18.5)	81.3
	Technician	368 (21.5)	228 (20.2)	62.0
	Other	174 (10.1)	115 (10.2)	66.0
Gender	Male	795 (46.5)	503 (44.6)	63.3
	Female	915 (53.5)	624 (55.4)	68.2
Nationality	Saudi	1103 (64.5)	811 (72.0)	73.5
	Non-Saudi	607 (35.5)	316 (28.0)	52.0
Age	20 - 34	803 (47.0)	608 (53.9)	75.7
	35 – 49	665 (38.9)	471 (41.8)	70.8
	50+	242 (14.2)	48 (4.3)	19.8
Length of working years in PHCCs in Riyadh	0 – 10	1065 (62.2)	870 (77.2)	81.7
	11-20	420 (24.6)	226 (20.1)	53.8
	21+	225 (13.2)	31 (2.8)	13.8
Previous health experience outside Saudi Arabia	No	978 (57.2)	686 (60.9)	70.1
	Yes	732 (42.8)	441 (39.1)	60.4
Previous training on EHR in primary healthcare	No	966 (56.5)	674 (59.8)	69.8
	Yes	744 (43.5)	453 (40.2)	60.9
Previous EHR experience in primary healthcare	No	978 (57.2)	686 (60.9)	70.1
	Yes	732 (42.8)	441 (39.1)	60.2

Perceptions of healthcare providers about the benefits of EHR systems

The majority of healthcare providers perceived EHR systems to be useful in improving communication between health professionals with the statement having a high agreement level of 73.6% as shown in both Tables 3 and 4. On an average scale, the statement had a mean agreement level of 2.65 (Standard deviation = 0.63, range: 2.61 - 2.69) at 95% confidence level.

Correlation between perceptions of EHR role in improving communication and health professional's occupation

There was a significant difference between health professional's perception of EHR role in improving communication between health professionals and other benefits in general ($p<0.0001$). Physicians had the highest agreement level of 83.7% with EHR benefit in improving communication between healthcare providers compared to other professionals. They were followed by pharmacists, nurses, other professions and lastly technicians with an agreement level of 63.2%. This data is presented in Table 3.

Table 3 – Correlation between Perceptions of EHR Benefits and Occupation Categories

Benefit variable	**Occupation**						
	Total	**Physicians**	**Nurses**	**Pharmacists**	**Technicians**	**Others**	**p-value**
Agreement level	**%**	**%**	**%**	**%**	**%**	**%**	
Provides quick and reliable access to scientific research	76.0	89.5	79.0	79.3	63.6	60.0	<0.05
Enables easy access to information from past medical records	73.9	86.6	76.8	75.5	61.4	63.5	<0.05
Provides access to patient data and analysis	76.8	91.4	77.4	80.8	65.4	64.3	<0.05
Provides better data	74.5	89.0	74.7	76.9	62.7	67.0	<0.05
Makes it easy to transfer data	74.4	85.6	73.0	78.8	66.2	66.1	<0.05
Provides access to practice standards	73.6	86.6	74.4	75.0	63.2	65.2	<0.05
Enables following test results	75.3	86.1	77.7	78.8	62.7	67.0	<0.05
Saves time in documenting health data	75.2	83.7	76.8	78.4	66.7	66.1	<0.05
Decreases paper-based documentation	77.1	89.5	78.5	79.3	67.5	65.2	<0.05
Improves the feeling of professionalism	76.2	85.6	77.7	79.3	66.2	68.7	<0.05
Improves communication between health professionals and patients	73.7	83.7	74.4	78.4	64.0	64.3	<0.05
Contributes to health professionals 'ability to make patient care decisions	72.3	81.8	73.6	73.6	64.0	65.2	<0.05
Improves communication between health professionals	**73.6**	**83.7**	**74.9**	**76.9**	**63.2**	**65.2**	**<0.05**
Reduces medical errors	63.5	73.7	65.5	63.0	54.4	58.3	<0.05

Correlation between perceptions of EHR role in improving communication between health professionals and training

A significant difference was found in the perception of the EHR role in improving communication between healthcare professionals with training on the EHR and those without. Approximately 81% of the health professionals with previous training on the EHR and its benefits agreed that the EHR improves communication among healthcare providers compared to about 69% who did not have any previous training as shown in Table 4.

Table 4 – Correlation between Perceptions of EHR Role in Improving Communication and Health Professionals' Training in EHRs

Benefit variable	**Previous training on EHR in primary healthcare**			
	Total	**No**	**Yes**	**p-value**
Agreement level	**%**	**%**	**%**	
Provides quick and reliable access to scientific research	76.0	70.8	83.7	<0.001
Enables easy access to information from past medical records	73.9	68.5	81.9	<0.001
Provides access to patient data and analysis	76.8	71.7	84.5	<0.001
Provides better data	74.5	71.8	78.6	<0.001
Makes it easy to transfer data	74.4	69.9	81.0	<0.001
Provides access to practice standards	73.6	68.7	80.8	<0.001
Enables following test results	75.3	70.6	82.3	<0.001
Saves time in documenting health data	75.2	70.3	82.6	<0.001
Decreases paper-based documentation	77.1	72.6	83.9	<0.001
Improves the feeling of professionalism	76.2	69.9	85.7	<0.001
Improves communication between health professionals and patients	73.7	69.3	80.4	<0.001
Contributes to health professionals 'ability to make patient care decisions	72.3	67.7	79.2	<0.001
Improves communication between health professionals	**73.6**	**68.7**	**80.8**	**<0.001**
Reduces medical errors	63.5	59.6	69.3	0.002

Discussion

The aim of this study was to investigate the influence of health professionals' occupation and training in EHRs on the perceived role of EHR systems in improving communication between health professionals in Saudi PHCCs.

The findings suggest that healthcare providers consider EHRs to play a key role in the provision of healthcare in primary healthcare. Moreover, the study shows that the occupation of healthcare providers as well as their training in the EHR determine how they perceive EHR benefits including improving communication among healthcare professionals.

In general, the majority of all healthcare providers agreed that EHRs have benefits in improving communication among healthcare teams. Previous studies have also shown that the majority of the health care providers perceive the EHR to be useful in improving provider-provider interactions [3; 6]. This might be due to their experience in interacting with EHR systems at different levels of care including diagnosis, posting lab results, ordering and issuing of prescriptions. However, it could also be a common belief that such systems are effective in improving communication due to their interlinked nature of different departments in a hospital or organisational setting that allows sharing of patient's medical information. EHRs could be critical in these scenarios by providing the patient's medical and treatment history where a particular intervention is required [4]. It has been observed that EHRs serve as an important platform for sharing of patient information to guide the provision of better care that could significantly improve the efficiency and safety of health care and result in cost savings [11]. Gordon and colleagues also asserted that there is a need for effective communication among members of a healthcare team to not only improve the quality of patient care but also to reduce costs of treatment as well as improve patient experience [2; 12].

The benefits of EHRs in improving communication among health professionals were perceived most by physicians, pharmacists, and nurses. This could be explained by the close working relationships of these three categories of professions. They usually interact in many areas of patient care including diagnosis of diseases and prescription by the physician, drug selection and monitoring by the pharmacist, and administration and monitoring by the nurse. Hence, the need for constant communication. Vitari and Ologeanu-Taddei similarly showed in their study that the intent of use of EHRs varied among three categories of clinical staff including physicians, paraprofessionals, and administrative staff with physicians having a higher intent mainly due to professional autonomy and medical responsibility [13]. Another study involving health workers in Central Malawi similarly found that job title influenced EHR usage; however, its finding showing that about two-thirds of the clinicians preferred paper-based records to electronic ones was surprising [14].

Those who have training in EHRs in Saudi PHCCs agreed that the EHR plays a major role in improving communication among healthcare providers. A possible explanation for this result is that training improves understanding of the benefits of EHRs including improved communication among health professionals. Another study also observed that training of staff about EHRs significantly improved their use of the systems in the provision of healthcare services [15]. It had also been established that EHR training resulted in more usage of the system possibly due to increased user confidence [14]. For instance, a literature review by Wang and colleagues established that team training is one of the effective strategies that could, to some extent, improve communication between intensive care unit physicians and nurses [16]. It could, therefore, be argued that effective use of the system should go beyond the training to involve implementation where communication among healthcare providers remain key to achieve the desired patient outcomes.

Strengths and limitations of the study

A major strength of this study is that it is the first available published study in Saudi Arabia and covers a large geographical area making it generalisable to a large population in similar settings. However, the study also had some limitations. First, the findings could be biased as the study was based on participants' reports about their feeling of the impact of EHR systems in improving communication between health professionals. The target population was also not representative of health professionals working in Saudi public hospitals because the survey was done in only PHCCs. The population was also limited to healthcare professionals, but it could be extended to include other healthcare personnel such as administration staff. Furthermore, it did not include other facilities that are privately owned, managed by non-governmental organisations (NGOs) or other health service providers. It was not established whether other factors such as level of education of the participants might have an influence on their perception of the role of EHRs on improving provider-provider communication as there is evidence suggesting that these factors could influence electronic systems usage [14]. Lastly, the research mainly reports the findings from primary healthcare centres in an urban setting with well-equipped facilities that may not be generalisable to other geographical areas.

Implications for practice

The study impacts the current practice by highlighting the need for increased implementation and awareness of EHRs and their roles in healthcare among Saudi healthcare providers in PHCCs. In particular, it emphasizes the need to embrace the use of EHRs to facilitate communication among health professionals in the process of providing patient care. Considering this is the first point of care for the majority of patients, it would reduce the potential medical errors and harm to the patient thus ensuring the provision of cost-effective treatments to achieve the best outcomes. The staff should be regularly trained in the use of EHRs to improve their knowledge, skills, and confidence in using the systems so as to facilitate communication. However, the implementation of EHRs in the Saudi context should take into consideration the potential hindrance factors such as the language of communication. The working of the systems is mainly based on English, which is a second language to a large Saudi national population. This highlights the need for further training not only in using the EHR systems but also the language of communication.

Implications for research

The influence of occupation and training of health professionals on their perception of the EHRs' role on improving communication among healthcare providers should be investigated further in a larger population from urban, peri-urban, and rural areas as the geographical location may affect the results. Moreover, future studies should target all Saudi hospitals as this study targeted healthcare professionals in only PHCCs. The study also implies that other factors other than occupation and training that could influence EHR implementation in primary healthcare settings should be explored.

Implications for education

Training on EHRs should be included in teaching curriculum for all students undertaking health courses to improve their

knowledge about these systems and their role in facilitating communication for better patient care. Saudi hospital organisations should also conduct regular training of their health workers on EHRs and their benefits.

Conclusions

The study demonstrates that the occupation of Saudi healthcare providers plays a significant role in the providers' perception of the EHR role in improving communication between health professionals covering a wide range of issues related to patient care such as diagnosis, medications, and follow-up among others. However, the finding that the role of EHRs in improving communication among health professionals among other benefits varies across professions is a call for effective strategies that ensure that all healthcare providers embrace EHR systems for improved healthcare delivery. Training on EHR systems similarly influenced this providers' perception implying the tremendous benefit of training.

In this view, the study highlights the need for all PHCCs in Saudi Arabia and beyond to have EHR systems to facilitate effective communication. This is necessary to prevent errors, harms, and adverse events that may arise from ineffective communication. Moreover, all healthcare providers irrespective of occupation should be encouraged to use EHRs. This could be achieved through training that ensures that they are equipped with the necessary skills and knowledge to use them effectively.

Acknowledgements

The authors highly appreciate the healthcare professionals working in PHCCs in Riyadh city for their precious responses. We are also grateful to the University of Tasmania and the Ministry of Health of Saudi Arabia for its support.

References

[1] Gluyas, H., 2015. Effective communication and teamwork promotes patient safety. Nursing Standard (2014+), 29(49), p.50.

[3] Gordon, J., Deland, E. and Kelly, R., 2015. Let's talk about improving communication in healthcare. Col Med Rev. 1(1), 2015, 23-27. doi: 10.7916/D8RF5T5D.

[4] Chase, D.A., Ash, J.S., Cohen, D.J., Hall, J., Olson, G.M. and Dorr, D.A., 2014. The EHR's roles in collaboration between providers: a qualitative study. In AMIA Annual Symposium Proceedings (Vol. 2014, p. 1718). American Medical Informatics Association.

[5] Wang, Y.Y., Wan, Q.Q., Lin, F., Zhou, W.J. and Shang, S.M., 2017. Interventions to improve communication between nurses and physicians in the intensive care unit: An integrative literature review. International journal of nursing sciences.

[6] Vermeir, P., Vandijck, D., Degroote, S., Peleman, R., Verhaeghe, R., Mortier, E., Hallaert, G., Van Daele, S., Buylaert, W. and Vogelaers, D., 2015. Communication in healthcare: a narrative review of the literature and practical recommendations. International journal of clinical practice, 69(11), pp.1257-1267.

[7] Robinson, G., Electronic Health Record Use: Health Care Providers' Perception at a Community Health Center.

[8] Rose, D., Richter, L.T. and Kapustin, J., 2014. Patient experiences with electronic medical records: Lessons learned. Journal of the American Association of Nurse Practitioners, 26(12), pp.674-680.

[9] Secginli, S., Erdogan, S. and Monsen, K.A., 2014. Attitudes of health professionals towards electronic health records in primary health care settings: a questionnaire survey. Informatics for Health and Social Care, 39(1), pp.15-32.

[10] Vitari, C. and Ologeanu-Taddei, R., 2018. The intention to use an electronic health record and its antecedents among three different categories of clinical staff. BMC health services research, 18(1), p.194.

[11] Hoover, R., 2016. Benefits of using an electronic health record. Nursing2018, 46(7), pp.21-22.

[12] Msiska, K.E.M., Kumitawa, A. and Kumwenda, B., 2017. Factors affecting the utilisation of electronic medical records system in Malawian central hospitals. Malawi Medical Journal, 29(3), pp.247-253.

[13] Walsh, C., Siegler, E.L., Cheston, E., O'donnell, H., Collins, S., Stein, D., Vawdrey, D.K., Stetson, P.D. and Informatics Intervention Research Collaboration (I2RC), 2013. Provider to provider electronic communication in the era of meaningful use: A review of the evidence. Journal of hospital medicine, 8(10), pp.589-597.

[14] Bredfeldt, C.E., Awad, E.B., Joseph, K. and Snyder, M.H., 2013. Training providers: beyond the basics of electronic health records. BMC health services research, 13(1), p.503.

[15] Pérez, A., Connelly, C. D., Baum, T., Bush, R. A., and Etland, C., 2018. A systematic review of the use of the electronic health record for patient identification, communication, and clinical support in palliative care. JAMIA Open, 1(2), 294-303.

[16] Ministry of Health, 2017. Annual Statistical Yearbook 2017. Available at: https://www.moh.gov.sa/en/Ministry/Statistics/book/Documents/ANNUAL-STATISTICAL-BOOK-1438H.pdf (accessed 17 January 2019.).

Address for correspondence

Bander Dpwahi Alanazi,Tasmanian School of Business & Economics, Launceston, TASMANIA, Australia.

Email: bander12@live.com.

Dr. Kerryn Butler-Henderson, Tasmanian School of Business & Economics, Launceston, TASMANIA, Australia.

Email: Kerryn.Butlerhenderson@utas.edu.au

Dr. Mohammed R. Alanazi, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia

Email: mratoubi@yahoo.com.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190273

Health Professionals' Experience with Patients Accessing Their Electronic Health Records: Results from an Online Survey

Monika A. Johansen[a,b], Per Egil Kummervold[c], Tove Sørensen[d], Paolo Zanaboni[a]

[a] *Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway*
[b] *Telemedicine and E-health Research Group, Arctic University of Norway*
[c] *Norut Northern Research Institute, Tromsø, Norway*
[d] *Helse Nord IKT HF, Tromsø, Norway*

Abstract

The aim of this study was to investigate hospital professionals' experience and attitude with patients accessing their own electronic health records. The study was conducted one year after service establishment. Data was collected through an online survey. In total, 457 replies were received. The results revealed a quarter of the administrative staff received feedback from patients or relatives regarding mistakes or missing information in their EHR. In addition, 67.5% of health professionals expected more patients to have basic knowledge of their health status in the future, and 21.4% found patients already gained better knowledge about diagnosis, treatment, or follow-up. The results also revealed some challenges with the service, especially for health professionals working in psychiatry, with some scepticism on whether the service is suitable for the sickest and most vulnerable patients.

Keywords:

Electronic Health Records, Patient Access to Records, Telemedicine

Introduction

Patient-accessible electronic health records (PAEHR) are developing in many countries, including Norway [1]. In Norway, the Electronic Health Record (EHR) has been fully established for many years now, and the patient is both the object and the owner of the health record. Patients have, since 2001, had the right by law to access their health record [2] and, in 2013, a White Paper stated patients should have digital access [3]. In accordance with the Norwegian Patient Right Act § 5.1 [2], health professionals can deny patients' access to either the full record or to specific documents in the journal if it is "necessary to avoid endangering the patient's life or serious damage to the patient's health", or if access to the information is inadvisable for persons close to the patients.

Currently, two of four health regions in Norway offer patients ages 16 or older, and parents of children under the age of 12, digital access to their hospital's EHR via the national health portal Helsenorge.no. Northern Norway was the first health region that offered all patients in the region digital access to their own EHR. In general, all documents available in digital format, including psychiatry reports, are made available for the patient as soon as they are approved/signed by health professionals, unless health professionals decide to deny access. In addition, a log list, showing those who accessed the record was made available to the patient at the time of this survey.

Before PAEHR was established, the University Hospital North Norway (UNN) sent their patients epicrisis by post after each consultation or hospitalisation. The other hospitals only sent it to the patient by request. In general, access to full or specific parts of the health record was given upon request where print or CD was sent via ordinary mail.

Patients reported to be generally satisfied with the PAEHR, recommended it to others, and generally understood the content [4]. At the same time, as the patients received this new service, a variety of tasks previously performed by secretaries or other support staff were transferred to the doctors [5]. Media claimed the time Norwegian doctors could use to direct patient care declined from about 60% in 2004 to less than 43% in 2012 [6]. It is not known whether PAEHR will increase the health professionals' workload and impact their work practice [7], create a different impact between health professionals (e.g. doctors and nurses) [8], and differ between somatic care and psychiatry, as others have been reluctant to provide open access to psychiatric records, considering it too sensitive [8].

The primary aim of this study was to investigate hospital health professionals' experiences and attitudes with patients accessing their own EHRs. The secondary aim was to explore whether there were differences in experiences and attitudes based on the implemented practices between hospitals, between doctors and nurses, and between psychiatry and somatic care.

The study was carried out in collaboration with the Northern Norway Regional Health Authority and the Norwegian Directorate of eHealth.

Methods

Data collection was performed through an online survey by sending a link via e-mail. The survey was distributed to all employees through a common e-mail list for the four hospitals in Northern Norway. The study was conducted in December 2016, after one year of experience with the service.

The respondents received different questions based on their professional background. The questionnaire to the health professionals and the administrative staff consisted of 25 and 14 items, respectively, including background variables such as job position, employment fraction, hospital, main working field (somatic/psychiatry/both), duration of practice, gender, and age. The questionnaire comprised of a combination of multiple choice questions, follow-up questions depending on the choices, and free text fields. Questions and quotes have been translated from Norwegian to English.

The questionnaire was anonymous. A questionnaire used in a pilot study was used as a template for the survey development. The questionnaire was pilot-tested by four researchers several times until no suggestion for modification came up.

No questionnaires were excluded from the analysis due to incomplete answers. The research objectives were investigated by descriptive statistical analyses using the statistical program R, version 3.4.2. For calculating the mean, the questions "Not applicable" and "I do not know" were omitted, and the response options were assumed to be at the interval level. A two-sample Student's t-test was used to test whether the differences in attitudes between health professionals from somatic care and psychiatry were statistically significant.

When presenting frequencies in the Results section, the "Totally agree" and "Quite agree" categories were merged into "Agree", and the "Yes, sometimes" and "Yes, quite often" categories were merged to "Yes". When comparing experiences and attitudes between doctors and nurses, both psychiatrists, psychologists, and physicians were included under the term doctor.

The free text responses were subjected to a content analysis aimed at identifying dominant themes related to the specific questions. Responses from health professionals within psychiatry and somatic fields were analyzed separately.

Some overall results from this study have been previously reported in Norwegian via oral presentations or the web.

Results

Demographics

In total, 457 replies were received. The responses were divided among the four hospitals as follows: 212 from UNN, 194 from Nordland Hospital (NH), 39 from Finnmark Hospital, and 12 responses from Helgeland Hospital.

As many as 77.7% of the respondents were female, and 80.7% of the respondents had a full-time position. The age distribution was quite uniform: 29.3% between 50 and 59 years old, 24.3% between 40 and 49, 24.1% between 30 and 39, 10.7% younger than 30, and 11.6% older than 59.

Most of the respondents worked in the somatic field (65.2%), while 27.4 worked in the psychiatry field, 3.5% in both fields, and 3.9% in other fields. Among respondents there were nurses (29.5%), doctors or psychiatrists (17.9%), other clinical positions (13.6%), psychologists (5.5%), social workers (2.2%), physiotherapists (1.8%), ergotherapists (1.3%), and radiographers (0.9%) in addition to administrative positions (27.4 %).

General Experiences Among Health Professionals and Administrative Staff

There were 332 responses from health professionals, and 125 responses from administrative staff members.

The main finding was that more than a quarter of the administrative staff received feedback from patients and/or their relatives regarding mistakes or missing information in their EHR (25.6%).

More than one third of both clinical (36.4%) and administrative (36.8%) staff received questions from the patients and/or their relatives related to use of the PAEHR. The same number of administrative staff forwarded requests from patients and/or their relatives to responsible health personnel (36.8%). Among health professionals, 15.4% received feedback from patients and/or their relatives regarding mistakes or missing information in their EHR. 72.8% of clinicians and 54.3% of administrative staff knew where to find information about the service.

Health Professionals' Experiences

The main finding was that 67.5% of the health professionals expected more patients to have a basic knowledge of their health status in the future, and 21.4% found patients were better informed about diagnosis, treatment, or follow-up than before (Table 1).

In addition, 28.3% experienced that the patients or their relatives referred to information from their EHR, and 19.6% of the health professionals planned to use the PAEHR in future follow-ups, for example, by adding more information in the EHR. 26.5% of the health professionals reported they had changed their way of writing in the EHR, while 71.4% meant that they should, in principle, complete the EHR documentation earlier, regardless of the patient's access. More results are presented in Table 1 below.

Table 1 - Health Professionals' Experience with PAEHR N=332

Do you agree with the following statements?	Yes[a]	No[a]
I expect that more patients will have basic knowledge of their health status in the future.	67.5	17.2
I find that patients are better informed about diagnosis, treatment, or follow-up than before.	21.4	36.8
I experience that patients/relatives refer to information they have found in their EHR.	28.3	69.6
I receive requests where I think: "Patients could easily find this information in their EHR, if they access it through the service".	26.8	69.9
I would use the patient's access in the follow-up, for example by adding more information to the patient in the EHR.	19.6	38.3
I changed my way of writing in the EHR.	26.5	61.8
In principle, I mean that we should write the EHR documentation earlier, regardless of the patient's access.	71.4	17.7
I inform patients that they can read their own EHR and check their referrals through the service.	50.8	38.3
I am worried that I need to spend a lot of time explaining journal content.	17.5	56.9
I am worried that I need to spend a lot of time reassuring patients because they have read their EHR.	26.5	46.4
During the past year, I discussed with colleagues whether there is a basis for denying a patient access to their EHR.	26.2	64.5

[a] "Yes" and "No" is presented in %. The "I do not know" category is not included in the table.

Differences in Practices Among Hospitals

The quantitative results showed no significant difference in attitude and experience between UNN (the only hospital that for several years sent epicrisis to the patients by mail after a consultation) and the other hospitals. However, by limiting the results to the psychiatry field, only, for the two hospitals with the most responses, 47.4% of the respondents from NH claimed they changed their way of writing in the EHR, while the corresponding figure for UNN was 30.8% (less than 40 clinicians responded at both institutions, p=0.139).

Differences in Experiences and Attitude Between Health Professionals in Psychiatry and Somatic Care

The results showed statistically significant differences in experiences and attitude between health professionals in psychiatry and somatic care regarding the use of PAEHR (Table 2).

The main finding was that as many as 43.9% of the health professionals in psychiatry reported they changed the way they wrote in the EHR after the service was established, compared to 23.6% from the somatic field. On the other hand, 77.4% of health professionals from psychiatry and 83.0% from somatic care expressed they should, in principle, write the EHR documentation earlier, regardless of whether the patient has online access or not.

In general, 27.2% of health professionals had discussed with a colleague whether to deny a patient access to information in their EHR or not. The problem was much more relevant in psychiatry, where as many as 60%, compared to 15.2% in somatic care, discussed this issue.

There was also a small difference between psychiatry and somatic care with regard to how often patients refer to information they find in their journal (18.4% vs. 25.4%).

Table 2 - Differences in Experiences Between Psychiatry and Somatic Care, 95% KI, p<0.05

Related to Use of the PAEHR-Service:	Somatic Care	Psychiatry	p
I experience that patients/relatives refer to information they have found in their EHR[a]	1.25 (1.20 - 1.31)	1.43 (1.28 - 1.48)	0.013
I changed my way of writing in the EHR[b]	3.24 (3.11 - 3.36)	2.70 (2.49 - 2.90)	<0.000
In principle, I mean that we should write the EHR documentation earlier, regardless of patient's access[b]	1.82 (1.70 - 1.94)	2.06 (1.88 - 2.24)	0.028
I discussed with colleagues whether there is a basis for denying a patient access to their EHR[a]	1.17 (1.12-1.23)	1.72 (1.58 - 1.86)	<0.000

[a] 1-no, 2-yes, sometimes, 3-yes-often
[b] 1-totally agree, 2-quite agree, 3-quite disagree, 4-totally disagree

Differences in Experiences and Attitude Between Doctors and Nurses

While 37.8% of the doctors were worried they would have to spend a lot of time reassuring patients or their relatives after reading their EHR, only 15.1% of the nurses expressed the same concerns (p<0.000).

While 37.8% of the doctors claimed they changed the way they wrote in the EHR, only 24.8% of the nurses changed their practice (p=0.045).

Qualitative Feedback on the Service

A total of 99 respondents provided additional comments in the open text field. There were 58 comments from health professionals from somatic care and 38 comments from health professionals in psychiatry, while three were categorised as both/other. From the somatic field, 39 comments came from UNN, 14 from NH, and five from the other hospitals. Of the 38 comments from psychiatry, 17 came from UNN, 15 from NH, and six from the other hospitals. Some comments included both support and criticism of the service, while others raised several concerns. Comments that did not contribute any specific experience with the service were categorized as neutral (Table 3).

Table 3 - Comments from the Respondent

Comments	Psychiatry			Somatic		
	UNN	NH	F	UNN	NH	F
Number of comments	17	15	6	39	14	5
Neutral	3			17	8	1
Positive	5[a]		2	6[a]	1	4
Critical	11	15	4	20	5	
Content of critical[b] comments regarding the PAEHR service						
Not suitable for any mentally ill patients	2	1	1			
Not suitable for all patient groups	8	5	1	1		
Patients should only be able to access parts of the EHR		2			2	
Patients might misunderstand		1		7	2	
Need to deny access	3	2	1			
Omit information	1	4		1		
Write a hidden journal	1	2				
Suggest delaying the information		2	1	1		
Complicates their work	2	4	3	8	2	
Worry for their own security		1		1		
Skeptical of the new logging functionality	1	2		1	1	
Other comments	1			7		

[a] Two of these also include critical comments, [b] Some critical comments raised several concerns

The positive comments from both fields mainly support the establishment of the PAEHR as a service.

The frequency of comments from psychiatry (38/125=0.304) was higher than that from the somatic field (58/298=0.195). In addition, 79% (30/38) of the comments from psychiatry were critical, compared to only 43% (25/58) from somatic care. Looking at the content of the critical comments, we also found more concerns were raised from psychiatry, compared to the somatic field (52 vs. 34).

The respondents' main concern was that PAEHR was not considered suitable for mentally ill patients (4), while 14 respondents considered it unsuitable for all patient groups in psychiatry. Examples of unsuitable patient groups were the sickest patients with psychosis, delusions, unrest, and utterance, where reading the EHR could worsen the situation. In addition, concern was raised towards vulnerable children with parents accessing the EHR on their behalf. Parents in a conflict situation, or if the child is a victim of abuse, could cause further

problems to the child. Vulnerable adult patients could also be threatened to show their EPR to others.

Some respondents commented patients should not be able to read the whole EHR, but maybe only the epicrisis. They worried patients with severe illness might misunderstand information, especially in the middle of a therapy period. Respondents referred to experiences where patients refused to speak with health professionals based on what they had read in the journal.

Some health professionals denied access to information they worried might harm the patient or their relationship with the patient. However, some pointed out that the functionality for denying access to information was complicated to use and little known. Other respondents reported they omitted some information from the EHR, and a few others reported they wrote a "hidden" journal containing the information they did not want the patient to read. Other suggested to delay displaying the journal documents, for example, until after the therapy period in psychiatry was completed. A number of respondents, from both fields, commented that PAEHR complicated their work, caused more work, and worsened the treatment. They felt they had to spend more time to evaluate what they should write or not. They had to write in a manner that the patient would not find offensive, and had to consider who they wrote to, a young person, old person, or a very sick person. If they decided to deny access to information, this could harm the patient-therapist relationship, as the patient might become suspicious and mistrustful. The service could also make it complicated to reflect on patients' symptoms through documenting "suspected" illness. Other comments focused on the difficulty to use the service for elderly patients, who might not receive any information from the hospital unless the epicrisis is sent by ordinary post as before.

Some of the respondents worried about the new functionality which provides patients with a log list of those who access their EHR. The respondents worried the service could complicate their work and patients could question their motives since they often accessed other journals to look for similar symptoms, used them as a template, or used them for teaching purposes.

A doctor working in psychiatry stated the PAEHR *"<…>complicates my work and worsens the treatment and alliance I will build with the patients*". Another said: "*I think it is not right that patients in psychiatry should have access to their journal. In fear of writing something "offensive" I think many therapists unfortunately have to do double-journal entries. Which again is vulnerable to getting lost. I deny the access when I know it's information that can be a trigger for the patient, but frankly, within psychiatry, there's a lot to be offended by, especially if you're mentally ill, and it could be impossible to predict what someone sometimes can take offence at".*

Discussion

The results of this online survey demonstrates several positive effects of the PAEHR. The fact that administrative staff received feedback from patients or relatives regarding mistakes or missing information in their EHR might improve the quality and correctness of the journal content and hence ensure patient safety and quality of the health service. This is very important for the patient in order to receive a correct diagnosis and a correct treatment. Two of three health professionals expected patients to gain more knowledge of their health status in the future, and more than one fifth found patients were already better informed about diagnosis, treatment, or follow-up than before. Despite some questions may be interpreted to cover more than the PAEHR, we believe the results of this survey demonstrate the potential clinical relevance of this service.

More than one fourth of the health professionals reported they changed their way of documenting in the EHR, as reported in other studies [8], while more than two thirds reported that, in principle, they should write the EHR documentation earlier. Future studies might explore potential changes over time, for instance, if more health professionals will alter their way of documenting due to the patients' access.

The results showed no significant difference in attitude and experience between UNN, the only hospital that for several years had sent the patients epicrisis by post after the consultation, and the other hospitals. However, looking at the psychiatry field, health professionals from NH claimed to have changed their way of writing to a greater extent than the health professionals from UNN. The results were not statistically significant, but supported by the comments from the two hospitals. There were more positive and neutral comments, and less critical comments from UNN compared to NH, where all the comments were critical (Table 3). These comments might support that UNN's earlier practice of sending out the epicrisis making the transition to the PAEHR service easier for their employees working in the psychiatry field.

The results revealed some challenges with using or adapting to the service, especially for health professionals working in psychiatry. Almost twice as many respondents from psychiatry than from somatic care reported they have changed the way they write in the journal, and nearly four times as many health professionals from psychiatry compared to somatic care have discussed with their colleagues whether to deny patients access or not. When comparing the frequency and content of the text comments, the PAEHR might have put an additional burden on some health professionals, especially those dealing with psychiatric patients. Many respondents questioned if the service was actually suitable for the sickest and most vulnerable patients. The health professionals could deny patients' access, but some respondents commented they did not have a strong enough reason to hide information, while others found the functionality to hide the information not user-friendly. Health professionals could benefit from information on, and continuous training on, how and when to deny patients' access to journal records.

Health professionals in the somatic filed received more questions from the patients regarding information in the EHR than health professionals in the psychiatry field (25.4% vs. 18.4%). There is therefore little reason to believe that higher levels of patient demand was the reason why health professionals within psychiatry seemed to raise more concerns towards the PAEHR.

More doctors than nurses claimed they changed their way of reporting, and twice as many doctors than nurses worried that they will have to spend more time reassuring patients, or their relatives, after they read their journal. This finding has probably more to do with the fact that the doctors who diagnose the patient, and have extensive knowledge, also have the overall responsibility for the patient, thus implying a stronger relationship. This might lead to a more negative attitude to the PAEHR as they feel a greater threat to their autonomy [9,10].

Norway and Sweden have comparable healthcare systems, and comparable PAEHR solutions in the way that patients have access to mainly all the information in the EHR system. Patients reported they are satisfied with the service [4,11], and the service fills important needs for them [11]. In Sweden, criticism has been raised from the clinical professions, and mainly from

physicians [8,9,12,13]. Physicians are mainly negative toward patients reading their health record online, while nurses are mostly positive, and nurses with some experience from patients using this service are more positive than nurses with no experience [9]. Physicians who used the PAEHR for themselves, as a patient, had a more positive attitude compared to physicians without such an experience [9]. Physicians were concerned patients could misunderstand the information in the EHR, something that would affect their work process and workload in a negative way [9,10]. A systematic review of studies from primary care indicated patients increased convenience and satisfaction, while health professionals were concerned about impact on workload and that the information would cause worry for the patient [7]. No changes in health outcomes were reported, but medical errors were detected, which may improve patient safety, and uptake of prevented care improved [7]. Several studies reported concern over increased workload, while only some demonstrated an actual increase in workload, and then in e-mail or online messaging, with face-to-face contact staying the same or falling [7]. In another study including 105 General Practitioners (GPs) and 13,564 patients, the GPs expected to increase their workload, while no significant increase in workload was showed after one year of experience. At the same time, patients reported an increased sense of control, better understanding of their medical issues, and they felt better prepared for future visits, in addition to improved adherence to medications [14].

Today, the PAEHR is well integrated in the two health care regions, and we believe that the positive benefits will compensate for the possible additional workload for health professionals. However, the issues raised by the health professionals from psychiatry should be followed up.

There are some limitations with this study which should be acknowledged. The online survey was sent to all employees through a third party (the IT-support organisation), using a common e-mail list, regardless if they had journal access or not. As a consequence, it was impossible to calculate an accurate response rate. However, the response rate was relatively low. We acknowledge that this type of recruitment for research easily leads to responses from people with strong opinions, very positive or very negative, more often than those who have not made an opinion. It is therefore important to be aware that the data material may be subject to this bias in the interpretation of the results.

More knowledge on how the service will influence both patients and health professionals in the future is necessary through further studies.

Conclusions

Health professionals' experiences and attitude with patients accessing their own EHRs was investigated through an online survey. The results revealed several positive findings, including patients identifying mistakes in the EHR and being better informed about diagnosis, treatment, or follow-up than before. The results indicate minor differences in experiences and attitudes based on the different practices existing at the different hospitals, and between doctors and nurses. On the other hand, major differences in experience and attitude were found between psychiatric and somatic care.

Health professionals working in psychiatry questioned if the service was suitable for the sickest and most vulnerable patients. Some adaptions, instructions, or training might be necessary to make the service more suitable for the psychiatry field.

Acknowledgements

This study has been partly funded by the Northern Norway Regional Health Authority and the Norwegian Centre for E-health Research. We would like to thank all the respondents for their contribution, and Helse Nord IKT and the hospitals for distributing the online survey by e-mail.

References

[1] A. Essén, I. Scandurra, R. Gerrits, G. Humphrey, M.A. Johnasen, P. Kiergegaard, et al., Patient access to eletronic health records: Differenes across ten countries, *Health Policy and Technology* **7** (2018) 44-56.
[2] HOD, The Norwegian Patient Right Act (LOV-1999-07-02-63), (2001). https://lovdata.no/dokument/NL/lov/1990-07-02-63#KAPITTEL_6.
[3] Stoltenberg-government-II, Meld. St. 9 (2012-2013) Èn innbygger – én journal: Digitale tjenester I helseog omsorgssektoren, (2012). Only in Norwegian. [One Inhabitant- One Journal]. www.regjeringen.no/no/dokumenter/meld-st-9-2012203/id708609/.
[4] T. Sørensen and M.A. Johansen, Developing and implementing patients' full-scale electronic access to their health record, *Stud Health Technol Inform* **228** (2016) 85-89.
[5] I.H. Amundsen, Vi har I det stille overtatt sekretærjobben, in, VG, (2013). http://www.vg.no/nyheter/innenriks/artikkel.php?artid=10119286 (accessed November 23, 2018).
[6] I.H. Amundsen, Sykehuslegene fanget I papirmølla, in, VG, (2013). http://www.vg.no/nyheter/innenriks/artikkel.php?artid=10119284 (accessed November 23, 2013).
[7] S. de Lusignan, F. Mold, A. Sheikh, A. Majeed, J.C. Wyatt, T. Quinn, et al., Patients' online access to their electronic health records nad linked online services: A systematic interpretative review, *BMJ Open* **4** (2014), e006021.
[8] L. Petersson and G. Erlingsdottir, Open notes in swedish psychiatric care (part 2): Survey among psychiatric care professionals, *JMIR Ment Health* **5** (2018) e10521.
[9] T. Alander and I. Scandurra, Experiences of healthcare professionals to the introdccution in Sweden of a public eHealth service: Patients' online access to their eletornic health records, *Stud Health Technol Inform* **216** (2015) 153-157.
[10] C. Grunloh, A. Cajander, and G. Myreteg, "The record is our work tool!" – Physicians' framing of a patient portal in Sweden, *J Med Internet Res* **18** (2016) e167.
[11] J. Moll, H. Rexhepi, A. Cajander, C. Grunloh, I. Huvila, M. Hagglund, et al., Patients'experiences of accessing their electronic health records: National patient survey in Sweden, *J Med Internet Res* **20** (2018) e278.
[12] M. Hagglund and I. Scandurra, Patients' online access to electronic health records: Current status and experiences from the implementation in Sweden, *Stud Health Technol Inform* **245** (2017) 723-727.
[13] L. Petersson and G. Erlingsdottir, Open notes in Swedish psyhiatric care (part 1): Survey among psychiatric care professionals, *JMIR Ment Health* **5** (2018) e11.
[14] T. Delbanco, J. Walker, S.K. Bell, et al., Inviting patients to read their doctors' notes: A quasi-experiemntal study and a look ahead, *Ann Intern Med* **157** (2012) 461-470.

Address for Correspondence

Monika.johansen@ehealthresearch.no, +47 90684407

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190274

Service-Oriented Device Connectivity: Device Specialisations for Interoperability

Björn Andersen, Simon Baumhof, Josef Ingenerf

Institute of Medical Informatics, University of Lübeck, Lübeck, Schleswig-Holstein, Germany

Abstract

There are IEEE 11073 standards for foundational, structural, and semantic point-of-care medical device interoperability, but the first devices with this interface have yet to enter the market. One of the missing pieces for implementation and approval are Device Specialisations that specify how to use information and service models to represent a specific type of device on the network. Required and optional metrics need to be standardised as well as nomenclature terms, units of measure, and extension points. Finally, device-to-device interaction at runtime has to be defined for automatic verification during testing and approval. Applications include C-arm fluoroscopes used in different clinical settings.

Keywords:

Health Information Interoperability; Reference Standards; Computer Communication Networks

Introduction

Modern, sophisticated medical devices still lack the ability to exchange data with devices from different manufacturers despite the increasing demand from clinicians and operators [1]. Until recently, the options for integration were indeed limited: Next to proprietary solutions, the only open alternative was the *ISO/IEEE 11073 Point-of-Care Medical Device (PoCD) Communication Standard* that enabled simple point-to-point interaction between an agent and a manager. The requirement for a loosely-coupled system of networked medical devices, however, was only met by the introduction of the ISO/*IEEE 11073 Service-Oriented Device Connectivity (SDC)* sub-series of standards. Driven by the research project OR.NET [2], this series now provides point-of-care medical and surgical devices with a contemporary communication protocol based on web service technology.

In SDC, the information exchange is based on the *IEEE 11073-20702-2016 Standard for Medical Devices Communication Profile for Web Services (MDPWS)* that defines safety, streaming, and compression features on top of previously existing web service standards. In addition, the *IEEE 11073-10207-2017 Standard for Domain Information & Service Model for Service-Oriented Point-of-Care Medical Device Communication* specifies a participant model, which allows for the structured network representation of a medical device's capabilities, and a message model, which defines communication endpoints for these capabilities. Both models are represented in XML Schema (XSD), whereas a device capability description must be expressed in eXtensible Markup Language (XML). This allows for the validation of a device's description against the standardised model.

The endpoints defined therein are bound to MDPWS through the *IEEE 11073-20701-2018 Standard for Service-oriented Medical Device Exchange Architecture & Protocol Binding* [3]. For semantic interoperability, every item of a device description is annotated with a code from the IEEE 11073-1010X series of nomenclature standards. Furthermore, the SDC standards provide mechanisms for authentication, authorisation, and encryption in order to maintain the confidentiality of personal and associated medical data [3].

Whereas the components of a device description are well-defined, different manufacturers may yet model their functionally equivalent devices differently. For personal health devices (PHD), the introduction of *Device Specialisations* facilitated interchangeability of devices from different manufacturers in the patient's home environment. These formal definitions of device types specify the structure of the hierarchical *containment tree* that constitutes the capability description.

Device Specialisations are also necessary for SDC devices, but the specifications must allow for a wider range of applications than their PHD counterparts. In addition to the mostly static containment tree, it is necessary to describe the dynamic interaction and the requirements towards communication partners to allow for automated testing and validation procedures, which are currently under development within the scope of the research project *Modular Validation Environment for Medical Device Networks (MoVE)* [4]. These Device Specialisations will not only be required for regulatory issues and type approval, but also for actual plug-and-play of medical devices.

Methods

Requirements Analysis

The most important difference between PHD and SDC is the device-to-device interaction and remote control that is one of the key benefits of device interoperability at the point-of-care. For patient safety, it is paramount to describe the underlying interaction patterns in a way that allows for automatic verification. This includes the definition of non-trivial safe states and fall-back mechanisms in case of a communication error or breakdown.

Furthermore, medical devices may need other network participants in order to provide their own functionality. A universal foot switch, for example, requires some device that can be controlled. Deploying a component within a medical device network can be assisted by automatic assessment of

compatibility by validating the functionality that is offered by a system of medical devices against the required capabilities.

Device Specialisations for Service-Oriented Device Connectivity

An SDC Device Specialisation combines knowledge of the composition and usage of a medical/surgical device with expertise in device modelling. Manufacturers of a certain type of device agree on a containment tree structure that is then brought into a machine-readable representation using XSD. This enables the validation of a device's description not only against the standardised data model but also against this schema in order to determine the correctness of the device model and the compliance with the standard.

However, the underlying semantics of the device description are of equal if not greater importance. Therefore, besides the semantic information that is conveyed implicitly through the structure of the containment tree, every description and state element is annotated with a *type* using standardised nomenclature codes, the same applies for units of measure. These codes are either taken from the preferred IEEE 11073-1010X series of nomenclature standards or from other controlled vocabularies that can be referenced by the device description. Which term(s) to use is either strictly defined by the Device Specialisation or can be chosen from a limited set for a specific metric.

Whereas items that are defined in a Device Specialisation cannot be omitted in a device that is to fulfil the standard, adding specific functionality is allowed. It is thus possible to integrate manufacturer-exclusive innovations into the network representation or to fulfil more than one Device Specialisation at a time. For example, a complex patient monitor may serve as an electrocardiograph *and* a pulse oximeter *and* a blood pressure monitor.

Oftentimes, elements of the containment tree depend on one another, e.g. the dose area product (DAP) calculated by a C-arm fluoroscope depends on settings such as tube voltage, current, and exposure time. These dependent metrics are expressed in the specification as well as (safety) requirements towards devices that exercise remote control. In this example, a controlling device that modifies critical parameters of the fluoroscope may have to visualise the estimated DAP to the human operator for confirmation. In the same way, quality-of-service (QoS) parameter boundaries or technical infrastructure requirements can be expressed.

Finally, the dynamic interaction of a medical device with other devices and components needs to be defined and verified. Therefore, the runtime behaviour of the device is specified in a machine-readable way. The MoVE project currently explores the usage of *Testing and Test Control Notation version 3 (TTCN-3)* [5] for this purpose. It allows for the definition of simple and complex test cases involving an arbitrary number of participants including the device-under-test (DUT).

Regulatory Issues and Type Approval

SDC Device Specialisations simplify the development and testing process for medical device manufacturers, but they also play an important role in type approval. In addition to conformance testing against the IEEE 11073 SDC communication protocol, Notified Bodies are also expected to validate the functionality of a device against the Device Specialisation for its type. This kind of integration testing involves interoperability with other (simulated) devices as well as *intraoperability* – the correct representation of the actual (physical) device state on the network [6].

Ultimately, medical devices that perform a given task as part of an ensemble of components will need to obtain certification for the precise role they play in the ensemble. This role and the interaction capabilities thus need to be explicitly stated in the intended use description of the device and may modify the classification of the device if it is, for example, intended to control a device of a higher class [6]. Referring to roles and capabilities that have been standardised in the form of a Device Specialisation significantly simplifies this procedure for both the manufacturer and the Notified Body.

These benefits have also been identified by the U.S. Food and Drug Administration (FDA) in a guidance document on interoperable medical devices [7]. Whereas the European Union's 2017 Medical Device Regulation (MDR) [8] also offers a definition of interoperability, it remains vague with regard to its regulatory impact and the benefits of communication standards.

Results

In this section, we present how specifying the characteristics of a C-arm fluoroscope facilitates device interchangeability in two example use cases. Figure 1 shows a simplified containment tree with channels for the operational parameters of the fluoroscope, dosage information, and the motion of the C-arm.

Surgical Navigation: Collision Avoidance

For this use case, consider the task of collision avoidance, for example in the operating room: For intraoperative radiography or fluoroscopy, a C-arm is often used as it can move around the operating table to acquire image data. Obviously, this motion should not cause a collision with the table. Through the provision of positional data from the C-arm, a surgical navigation system could warn the user before a collision occurs or could even remotely stop the repositioning completely. In addition to the containment tree, the interaction pattern for this control operation is also part of a Device Specialisation.

Synchronisation of Fluoroscopy and Ventilation

Another use case that greatly benefits from standardised devices is the synchronised fluoroscopy. It would allow this procedure to be carried out with any two devices from different manufacturers. The fluoroscopy can be synchronised with the breathing cycle of a patient who is connected to a ventilator if both devices provide the respective data. For projectional radiography, it would also be possible to stop the ventilator for the duration of the X-ray image acquisition in order to minimise motion artefacts and continue the breathing cycle afterwards.

Note that ventilators at the point-of-care (e.g. in the ICU or for anaesthesia) require a different Device Specialisation than a Home Healthcare Environment Ventilator from the PHD domain. The latter, which is under development as IEEE P11073-10426, targets a different device category and is not expected to include device-to-device interaction or remote control capabilities.

Discussion

The benefit of a standardised medical device communication protocol in general and an SDC Device Specialisation in particular lies in facilitating real-world applications. The manufacturer-independent interchangeability reflects the clinical reality of heterogeneous devices that need to interoperate.

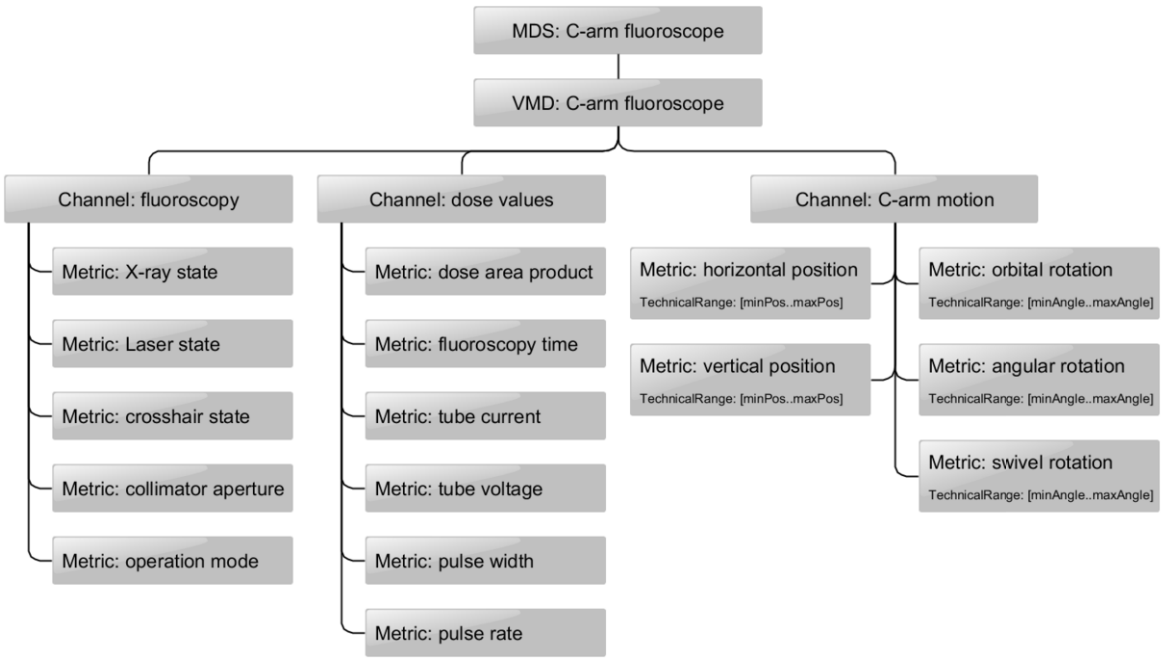


Figure 1 - Simplified model of a C-arm fluoroscope as a hierarchical containment tree; operations, contextual information, and alerts are omitted for brevity.

Whereas the PoC Device Specialisations currently under development refer to the SDC series and are going to be standardised within the new IEEE 11073-107XX sub-series, the payload of all IEEE 11073 communications is semantically described using the same vocabulary. Therefore, physiological measurements and other data can be combined from multiple sources, e.g. the home environment and the clinical workplace.

Conclusions

The standardisation of SDC Device Specialisations will further the implementation of open communication interfaces into actual medical devices, support the approval process, and facilitate plug-and-play. Devices that are interconnected using SDC will provide their operators with better assistance, perform more complex tasks in an ensemble, and ultimately increase patient safety.

References

[1] K. Lesh, S. Weininger, J. M. Goldman, B. Wilson and G. Himes, Medical Device Interoperability-Assessing the Environment, 2007 Joint Workshop on High Confidence Medical Devices, Software, and Systems and Medical Device Plug-and-Play Interoperability (HCMDSS-MDPnP 2007), Boston, MA, 2007, pp. 3-12. doi: 10.1109/HCMDSS-MDPnP.2007.22

[2] OR.NET – Sichere dynamische Vernetzung in Operationssaal und Klinik. http://ornet.org/wp-content/uploads/2017/10/Broschure2016_ENG_final.pdf (accessed on March 29, 2019).

[3] M. Kasparick, M. Schmitz, B. Andersen, M. Rockstroh, S. Franke, S. Schlichting, F. Golatowski, and D. Timmermann. OR.NET: A Service-Oriented Architecture for Safe and Dynamic Medical Device Interoperability. *Biomedical Engineering/Biomedizinische Technik* **63, no. 1** (2018): 11-30.

[4] Modular Validation Environment for Medical Device Networks (MoVE). http://www.ornet.org/move/ (accessed on March 29, 2019).

[5] ETSI ES 201 873 Methods for Testing and Specification (MTS); The Testing and Test Control Notation version 3, 2016-2017.

[6] A. Janß, J. Thorn, M. Schmitz, A. Mildner, J. Dell'Anna-Pudlik, M. Kasparick, M. Leucker, and K. Radermacher. Extended device profiles and testing procedures for the approval process of integrated medical devices using the IEEE 11073 communication standard. *Biomedical Engineering/Biomedizinische Technik* **63, no. 1** (2018): 95-103.

[7] U.S. Department of Health and Human Services, Food and Drug Administration. Design Considerations and Premarket Submission Recommendations for Interoperable Medical Devices. Guidance for industry and food and drug administration staff, FDA, Sep 2017.

[8] European Parliament, Council of the European Union. Regulation 2017/745 of the European Parliament and of the Council on medical devices. Official Journal L117, Apr 2017.

Address for Correspondence

Björn Andersen, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany; E-mail: andersen@imi.uni-luebeck.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190275

Face to Face Appointment vs. Telemedicine in First Time Appointment Orthopedic Oncology Patients: A Cost Analysis

Luis Alberto Aponte-Tinao[a], Germán Luis Farfalli[a], José Ignacio Albergo[a], Fernando Plazzotta[b], Janine Sommer[b], Daniel Luna[b], Fernán González Bernaldo de Quirós[b]

[a] Institute of Orthopedics, "Carlos E. Ottolenghi" Hospital Italiano de Buenos Aires, Ciudad de Buenos Aires, Argentina,
[b] Department of Health Informatics, Hospital Italiano de Buenos Aires, Ciudad de Buenos Aires, Argentina

Abstract

Medicine has evolved considerably in recent decades in part thanks to information and communication technologies in health (ICTs). However, face-to-face consultations continue to be the predominant model, since alternatives such as telemedicine are still the subject of debate. On the other hand, in some very specific specialties, centralization is relevant, mainly due to the low frequency and prevalence of diseases, as well as the need to have highly specialized professionals, causing problems in terms of accessibility and costs for the health system. In this study we have analyzed the first consultations to an orthopedics oncology service at a tertiary institution and performed an analysis of economic costs was carried out between 2 possible scenarios: face-to-face consultations versus telemedicine. Analyzing the 2 scenarios, there would be a cost-benefit in the use of telemedicine leading to a decrease in healthcare cost between 12.2% and 72%.

Keywords:

Telemedicine, Cost Analysis, Personal Health Records

Introduction

Medicine has evolved considerably in the last decades, mainly based on advances in technology [1; 2]. Some examples include imaging studies that have increases significantly in terms of definition and accuracy [3; 4]. Regarding the field of Surgery, advances have also been remarkable, including navigation-assisted surgery for oncologic musculoskeletal tumors [5-8]. However, the way to perform medical appointments, has not changed in the last decades. Personal interviews continue to be the prevalent model and although new alternatives have been proposed, such as telemedicine, its benefits and acceptance by the medical community continue to be debated.

For very specific medical specialties, centralization is relevant, mainly due to the low frequency and prevalence of the diseases as well as the need to have highly specialized professionals [9; 10]. However, in large countries, centralization can be a problem in terms of distance, transfer and healthcare costs.

Tele-orthopedics involves the provision of specialized services in orthopedics remotely. This, according to the taxonomy of telemedicine regarding the participants of the consultation, is usually carried out between an orthopedic surgeon and a patient [11]. Benefits have been reported regarding the use of tele-orthopedics services such as: significant savings in travel time; reduction of work times and increase in patient satisfaction regarding postoperative care compared to traditional methods [12].

The equipment used for telemedicine improves quickly as technology evolves in terms of quality, costs and ease of use. However, it is important to evaluate the quality and safety of the care provided, including patient satisfaction and its economic impact in detecting and reducing events before using this type of technology in medical care. Some studies suggest that telemedicine can be used as an alternative to face to face consultations for orthopedics patients in an outpatient setting. It was also proven that the use of videoconferencing systems was adequate to provide orthopedic care in rural areas, but most but most studies lack adequate assessment methods and cost-benefit analysis [13].

We analyzed first-time appointments for appointments of orthopedic oncology patients at a tertiary complexity centre and carried out an economic cost analysis between two possible scenarios: face-to-face appointments vs. telemedicine consultations. We attempted to determine if the use of telemedicine for first-time patients could save costs to the healthcare system.

Methods

Setting

Hospital Italiano de Buenos Aires (HIBA) is a non-profit organization with 165 years of history in Argentina. Its healthcare network includes a Universitary hospital of high complexity that covers health care for outpatients, inpatients, emergencies, critical care, home care, chronic care and medical and surgical specialties. It has its own medical insurance service (health maintenance organization), with more than 160,000 affiliates, and provides health services to 1,500,000 people with other health insurances. Annually, more than 45,000 patients are admitted to their hospitals, and 45,000 surgical procedures and 3,000,000 outpatient visits take place.

Since 1998, the HIBA has its own health information system (in-house) that includes the management of clinical and administrative information. Its Electronic Health Record (EHR) is an integrated, modular, problem-oriented and patient-centered system, used in the different clinical scenarios (ambulatory, hospitalization, emergency center and home care) [14; 15].

As part of the information system, an integrated Personal Health Records (PHR) called POPES is available to all outpatients since 2007. PHR allows patients receiving medical care in the hospital network to access and verify clinical and administrative information, and to interact with the health system. Among its main functionalities, the PHR allows users to update their personal information, share information, manage scheduled appointments, view test results, check, order and buy prescribed medication, to consult with the healthcare team through Telemedicine tools, and it has a messaging service for

communication with the general practitioner". At present, POPES has approximately 400,000 registered users [16].

Design

We carried out a prospective study. All the patients attended at an outpatient clinic by the orthopedics oncology team during the period of time between January and June of 2018 were analyzed. Relevant information of each patient was collected in a structured form in the EHR, which was completed by each professional. A total of 4 orthopedic oncology surgeons participated in the research. Only first-time appointments were analyzed.

An estimated economic cost analysis was carried out, considering: a) transfer cost b) face to face appoint cost c) telemedicine cost. For transfer costs, we took as reference a round trip airplane ticket for patients who lived at least 400 kilometers away from our hospital. The national airline (Aerolineas Argentinas) was used as reference for price and always the cheapest option was chosen. The cost of the face-to-face appointment according to our Hospital data was 49 U.S. dollars. The cost of the telemedicine appointment was set by our institution in US $80. Accommodation and travel expenses were not considered in the analysis. For patients under 18 years old, an extra flight ticket was added for cost analysis because they must be accompanied by an adult. The objective of the study was to determine whether the use of telemedicine for first-time consultations in orthopedics oncology would reduce the costs to the health system.

Results

A total of 1,224 consecutive appointments for orthopedic oncology were analyzed, of which 207 (17%) were first-time patients (see figure 1).

108 out of the 207 first-time patients (52%) were male and the mean age was 48 years (range 2-90).

The mean time between referral and face-to-face appointment in our institution with an orthopedic oncology specialist was 1.6 weeks (range 0-11). In 107 patients (52%), the diagnosis was made at the first-time appointment without the need of medical treatment, thus solving the problem with a single appointment.

36 patients out of 207 patients of 207 (18%) had to travel at least 400 km to access the hospital (table 1). These patients came from 14 different provinces (figure 2).

Estimated cost analysis

The equivalent cost for telemedicine for the 207 first-time patients would have been US $16,560. The estimated cost of transportation for this group of patients was $8,572 and the cost of the regular face-to-face consultation US $10,143 (total value: 18,715). This represents a mean value of US $90 per first time patient. Analyzing both scenarios, there would be a cost-benefit in the use of telemedicine for first-time patients at our orthopedic oncology unit, which would lead to a decrease in healthcare costs of 12.2%.

For the select group of patients (n = 36) who had to travel at least 400 km for the appointment, the economic benefit increases significantly, (US $2,880 vs US $10,336), representing a 72% reduction in costs.

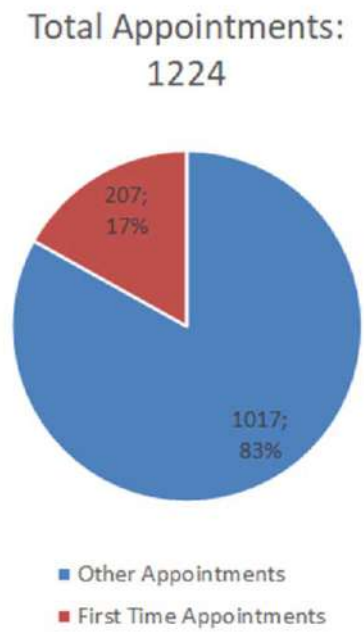


Figure 1: Total Appointments and First Time Appointments

Discussion

The present study, through a cost analysis, seeks to demonstrate potential savings with the use of telemedicine for first-time appointments to orthopedic oncology specialists for patients with suspected bone tumors.

Bone tumors are of low prevalence, with a wide variety of lesions that have a high risk of local recurrence and metastases, which is why timely diagnosis by a team of specialists has a direct impact on patient outcomes. However, the diagnosis is not a simple task and is often delayed due to the insidious presentation and high diagnostic error rate, since there are similar lesions that simulate these tumors [17]. In all cases, the formulation of differential diagnoses must be made with the integration of clinical signs and symptoms and diagnostic images. Based on differential diagnoses and staging studies, it is determined how and where to perform a biopsy.

Table 1: Characteristics of Patients in First Time Appointments

Variable	Value
Patients	107
Age average (yo)	48 (2-90)
Female	48% (N=99)
Male	52% (N=108)
Distance > 400 km.	18% (N=36)
Diagnosis in First Time Appointment	52% (N=107)
Mean waiting time from referral	1.6 Weeks

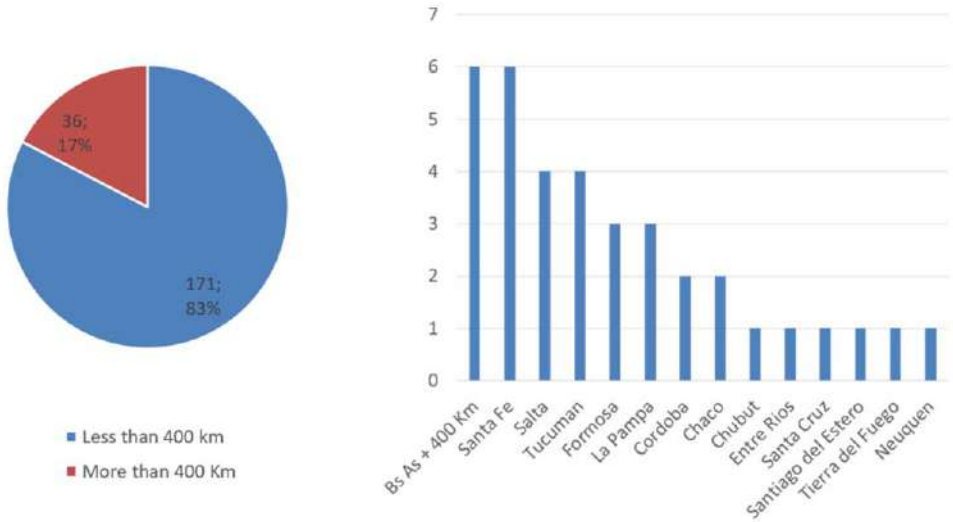


Figure 2: Patients' precedence

The average delay for an appointment with a specialist can be up to 10 months, and sometimes patients are attended by up to 5 medical teams with 30% of inadequate procedures carried out [17]. In the United Kingdom delayed diagnosis is the main cause of sarcoma litigation [18].

The wide geographical distribution of the patients that are treated the Hospital Italiano de Buenos Aires means that the use of Telemedicine has an enormous potential when it comes to improving accessibility and healthcare costs, since 50% of consultations can be solved in the first appointment. This first appointment is usually centered in the review of patients' medical records and all their tests and studies. This information is sent asynchronously, reviewed and analyzed by the professional team and then a synchronous meeting with the patient is scheduled for notification of the resolution of the team in each case, and to schedule the steps to follow (either the need of referral to perform more studies or schedule long-term controls).

The present study was limited to comparing estimated costs for the healthcare system, establishing a standardized cost of travel ticket and cost of care, without taking into account the real costs according to the distance for each case, nor the cost derived from the need to repeat studies, as well as the loss of profit of patients and their families. Additionally, we did not consider differential costs according to patient age, despite pediatric and elderly patients could have higher costs given the need of a companion.

At the end of this study, we did not find evidence of a cost analysis for these patients. However, there is plenty of evidence for other groups. For example, a study that evaluated the impact of electronic consultation on a Canadian tertiary care pediatric specialty referral system concluded that similar to what happens with adults, electronic consultations improves primary care physician access and timeliness to elective pediatric specialist advice and influences their care decisions, while reporting high end-user satisfaction [18].

A study that evaluated the cost of a speech pathology synchronous telepractice service for patients with head and neck cancer concluded that such telepractice service provides cost-efficiency over standard care through the provision of a remote specialist [19]. Another group that studied the use of telemedicine for the same pathologies reported that this service allows timely access to surgical care along with considerable savings for patients with head and neck cancer [20].

The main limitation of our study is the estimated costs methodology: it is not an analysis of real costs, since we used the same value for transfer costs for all patients. Another limitation was that other variables that can influence costs were not taken into account.

In future lines we intend to make qualitative evaluations to patients whose diagnosis were negative in the first consultation to make a long-term follow-up of the final diagnosis and an adequate calculation of costs taking into account all the variables, so we can account for the real cost-effectiveness of the system.

Conclusions

In our experience with the implementation of a telemedicine tool for first-time appointments with orthopedic oncology specialists, the estimation of costs in a limited period of time suggests that this intervention may reduce costs to the healthcare system.

Acknowledgements

This work was carried out with the collaboration of the Institute of Orthopedics and the Department of Health Informatics of Hospital Italiano de Buenos Aires. We thank the entire interdisciplinary team (doctors, nurses, designers and developers) and the patients who trusted us and made their consultations at a distance.

References

[1] L. Jeys, G. Morris, S. Evans, J. Stevenson, M. Parry, and J. Gregory, Surgical innovation in sarcoma surgery, *Clinical Oncology* **29** (2017), 489-499.

[2] L.E. Ritacco, F.E. Milano, G.L. Farfalli, M.A. Ayerza, D.L. Muscolo, and L.A. Aponte-Tinao, Accuracy of 3-D planning and navigation in bone tumor resection, *Orthopedics* **36** (2013), e942-e950.
[3] K.C. Wong, S.M. Kumta, G.E. Antonio, and L.F. Tse, Image fusion for computer-assisted bone tumor surgery, *Clinical orthopaedics and related research* **466** (2008), 2533-2541.
[4] P.G. Thacker and E.Y. Lee, Advances in multidetector CT diagnosis of pediatric pulmonary thromboembolism, *Korean journal of radiology* **17** (2016), 198-208.
[5] R.L. Satcher, How intraoperative navigation is changing musculoskeletal tumor surgery, *Orthopedic Clinics* **44** (2013), 645-656.
[6] D. Cheong and G.D. Letson, Computer-assisted navigation and musculoskeletal sarcoma surgery, *Cancer Control* **18** (2011), 171-176.
[7] T.Y. So, Y.-L. Lam, and K.-L. Mak, Computer-assisted navigation in bone tumor surgery: seamless workflow model and evolution of technique, *Clinical Orthopaedics and Related Research®* **468** (2010), 2985-2991.
[8] L. Aponte-Tinao, L.E. Ritacco, M.A. Ayerza, D.L. Muscolo, J.I. Albergo, and G.L. Farfall, Does intraoperative navigation assistance improve bone tumor resection and allograft reconstruction results?, *Clinical Orthopaedics and Related Research®* **473** (2015), 796-804.
[9] O. Derbel, P.E. Heudel, C. Cropet, P. Meeus, G. Vaz, P. Biron, P. Cassier, A.-V. Decouvelaere, D. Ranchere-Vince, and O. Collard, Survival impact of centralization and clinical guidelines for soft tissue sarcoma (A prospective and exhaustive population-based cohort), *PloS one* **12** (2017), e0158406.
[10] A. Bhangu, J. Beard, and R. Grimer, Should soft tissue sarcomas be treated at a specialist centre?, *Sarcoma* **8** (2004), 1-6.
[11] R. Bashshur, G. Shannon, E. Krupinski, and J. Grigsby, The taxonomy of telemedicine, *Telemedicine and e-Health* **17** (2011), 484-494.
[12] L.J. Caffery, M. Taylor, J.B. North, and A.C. Smith, Tele-orthopaedics: a snapshot of services in Australia, *Journal of telemedicine and telecare* **23** (2017), 835-841.
[13] A. Buvik, E. Bugge, G. Knutsen, A. Småbrekke, and T. Wilsgaard, Patient satisfaction with remote orthopaedic consultation by using telemedicine: A randomised controlled trial, *Journal of telemedicine and telecare* (2018), 1357633X18783921.
[14] D. Luna, F. Plazzotta, C. Otero, F. González Bernaldo de Quirós, A. Baum, and S. Benítez, Incorporación de tecnologías de la información y de las comunicaciones en el Hospital Italiano de Buenos Aires, (2012).
[15] F. Plazzotta, D. Luna, and F.B. de Quirós González, Health information systems: integrating clinical data in different scenarios and users, *Revista peruana de medicina experimental y salud publica* **32** (2015), 343-351.
[16] M.I. Smith, G. Garcia, M. Simon, L. Bruchanski, J. Frangella, J.A. Sommer, M.V. Giussi, and D.R. Luna, Lessons Learned After Redesigning a Personal Health Record, in: *MEDINFO 2017: Precision Healthcare Through Informatics: Proceedings of the 16th World Congress on Medical and Health Informatics*, IOS Press, 2018, p. 216.
[17] G. Johnson, G. Smith, A. Dramis, and R. Grimer, Delays in referral of soft tissue sarcomas, *Sarcoma* **2008** (2008).
[18] L. Lai, C. Liddy, E. Keely, A. Afkham, J. Kurzawa, N. Abdeen, T. Audcent, M. Bromwich, J. Brophy, and S. Carsen, The impact of electronic consultation on a Canadian tertiary care pediatric specialty referral system: A prospective single-center observational study, *PloS one* **13** (2018), e0190247.
[19] C.L. Burns, S. Kularatna, E.C. Ward, A.J. Hill, J. Byrnes, and L.M. Kenny, Cost analysis of a speech pathology synchronous telepractice service for patients with head and neck cancer, *Head & neck* **39** (2017), 2470-2480.
[20] D.M. Beswick, A. Vashi, Y. Song, R. Pham, F.C. Holsinger, J.D. Rayl, B. Walker, J. Chardos, A. Yuan, and E. Benadam–Lenrow, Consultation via telemedicine and access to operative care for patients with head and neck cancer in a Veterans Health Administration population, *Head & neck* **38** (2016), 925-929.

Address for correspondence

Luis Alberto Aponte-Tinao MD, Institute of Orthopedics, Hospital Italiano de Buenos Aires, Argentina. email: luis.aponte@hospitalitaliano.org.ar

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190276

Developing an Electronic Record Tool Representative of Primary Health Care in the Public Health Care System of Buenos Aires City

F. Faretta[a], D. Levi[a], L. Marques[a], D. Ferrante[b], M.V. Giussi Bordoni[a], A. Baum[a], F. Gonzalez Bernaldo de Quirós[c]

[a] Health Information Systems Office, Ministry of Health, Buenos Aires City, Argentina
[b] Strategic Planning Office, Ministry of Health, Buenos Aires City, Argentina
[c] Government Health Advisor, Buenos Aires City, Argentina

Abstract

The Primary Health Care strategy is based on organization of interdisciplinary teams and comprehensive approach to health, disease and health care processes. To strengthen information systems so that they represent primary health care complexities, participatory meetings were held with primary care practitioners from the public health care system of Buenos Aires City. Terms for the record tool and its components were chosen using consensus methodologies. This process involved 300 practitioners from 49 centers, and submission of 21 proposals. It was decided to change the term "Electronic Medical Record" with "Comprehensive Health Record." It was also agreed that, apart from "Reason for Consultation," the field "Problem Situation" would be added, that "Care Service" would be replaced with "Care Act," and that a new module "Health Team Management and Education Activities" would be included to document practitioners team activities.

Keywords:

Public health informatics, primary health care, electronic health record

Introduction to this Article

Use of information technologies (IT) in complex adaptative health care systems remain a challenge for design, development, implementation, and evaluation of interventions. The comprehensive socio-technical model proposed by Dean F. Sittig [1] provides useful analysis framework for considering complex interaction between social and technical factors coming into play in any implementation.

An electronic record system was implemented following 2016-2019 Health Plan for the public health care system of Buenos Aires City, which included a digital health strategy. As the system failed to reflect specific characteristics of primary care work processes and primary health care (PHC) strategy [2,3], it encountered practitioners' resistance, dissatisfaction, and dissent.

It is well-known that users' involvement and engagement in design of health information systems is key to software adjustment to work processes, lower resistance to use, tool adoption, and successful implementation [4-6]. This approach proves to be even more significant with in-house development capabilities. However, achieving meaningful participatory processes in health care organizations remains a challenge. The case under study is molded by characteristics of work processes in the primary health care strategy, whose foundational principles include interdisciplinary teamwork and comprehensive community-based approach to health, disease, and health care processes [7-10].

In this context, the question is how to introduce adjustments into the record system in a participatory way, given characteristics of primary health care practice and PHC strategy in public health care system of Buenos Aires City. The objective of this paper is to present consensus-based experience and participatory decision-making employed to develop record tools representative of characteristics of primary health care.

Background

Primary health care professionals are responsible for all actions and services delivered in basic specialties and outpatient modalities. Primary health care centers are people's main entry point to and follow-up area in health care networks. Primary health care centers are organized around the PHC strategy. According to WHO, this strategy is based on certain essential elements, such as universal health care access and coverage, comprehensive and integrated care, emphasis on health promotion and disease prevention, family, and community guidance, active participation mechanisms, organization of multidisciplinary teamwork, territory-based approaches, among others [11].

Implementation of electronic information systems in Buenos Aires City started June 2016 in primary health care setting. Its first phase was completed in 2017. This implementation phase included system based on three modules: patient identification, appointments, and simple-format electronic medical record (EMR) for documenting health consultations between patients and practitioners. The EMR had two mandatory fields: free text progress report and reason for consultation using terminology service whereby list of patient problems is generated [2,12]. Figure 1 shows EMR components used to document a medical act.

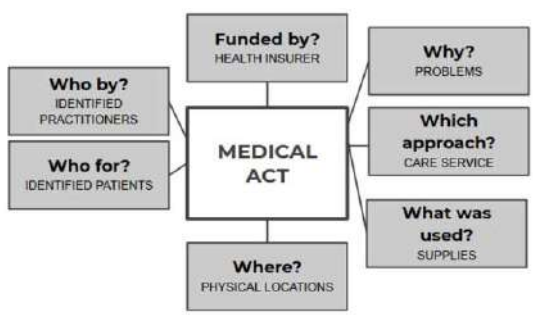


Figure 1-Medical Act Components in the EMR

Given that it is crucial to monitor how the system is used at early implementation stages in order to ensure immediate

response to users' comments and requests [9], in November and December 2017, a Health Information Systems Office team conducted a survey to make qualitative assessment of users´ perceptions. Group interviews were carried out with 130 system users in 11 primary health care centers with at least a 6-month system implementation. Results concerning EMR design and format revealed difficulty in documenting some primary care practices, such as interdisciplinary health care and community activities. Some limitations were also observed in trying to represent complexity of health, disease, and health care processes with available terminology services, particularly in connection with social determinants of health as described in literature [13].

In early 2018, based on results of this assessment, the entire terminology used in the EMR was examined. It was observed that over 50% of total documented terms (n=39,492) were not health-related problems, but referred to other components and processes that the system was not adequately capturing. Furthermore, medical act structure was duly adjusted to primary care documentation needs. Figure 2 shows results of this analysis. The care act, not only medical act is central, and components that need to be represented in the system have been added. These components include different care modalities, other physical locations, other types of care provision, and terminology for comprehensive health care.

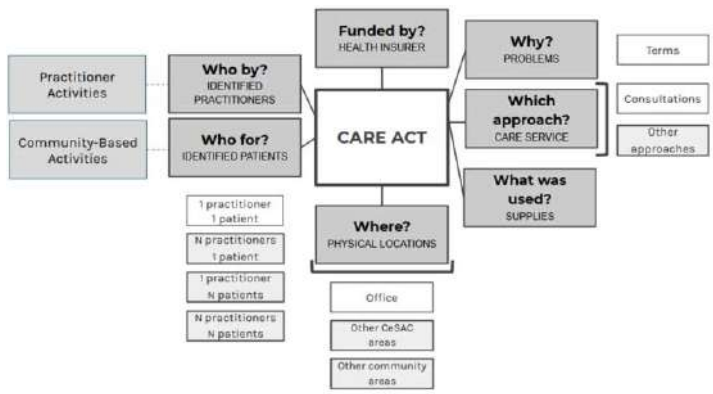


Figure 2-Care Act Components in Primary Health Care

Methods

Setting

Buenos Aires City healthcare network is composed of Ministry of Health, 35 Hospitals, 74 Primary Care Centers (CESAC, CMB), 1 Ambulatory Reference Medical Facility (CEMAR), 2 children´s dental centers, and 2 mental health centers. It is structured into 12 geographical areas to organize health care delivery. The health system employs a total of 41,000 people. Since June 2016, Electronic Medical Record (EMR) has been gradually implemented in outpatient settings [2]. By October 2018, 60 healthcare facilities were using EMRs, over 1 million EMRs had been opened, and 2.5 million clinical notes registered.

Experience Systematization

An interdisciplinary team of medical practitioners, psychologists, educators, anthropologists, and communicators was created in order to plan the strategy. Starting point was a number of meetings convened by the organizing team. It was decided that the objective of systematization was to document the process of reaching participatory consensus and its ensuing agreements. All records and documents produced during planning and development were collected and analyzed, their history being traced (see *Meeting Methodology*). *Results* were systematized, and subsequently all information was examined and organized according to predetermined structuring component, to eventually arrive at critical interpretation of the whole process (see *Discussion*). Finally, *conclusions* and lessons were drawn and documented in writing.

Meeting Methodology

Design of system evincing complexity of PHC strategy, as in care act structure, required thorough reflection by its users. First, organizing team reached their own conceptual agreements, consulted literature on consensus methodologies [10], and decided to hold participatory meetings with end users for tool design as first step in a series of iterative cycles of interaction and agreements. Agenda of these first sessions was to find: a) most appropriate denomination for patients' individual health record tool, commonly referred to as Electronic Medical Record, b) appropriate denomination for each of the care services provided, c) how to organize and denominate problems and reasons for consultation, and d) how to organize and denominate professionals´ activities which do not entail direct contact with patients. Four activities were planned, and work proposals were pilot tested before implementation with a team consisting of 6 primary care practitioners. Meeting development was supported by active strategy of communication with end-users.

First On-Site Meeting

Primary health care centers´ heads and practitioners comprising maximum 4 people per institution were invited. It was suggested attendees should represent different disciplines. Invitation was made via official communication, e-mail, and telephone calls. Previous registration was also requested using structured form with mandatory fields including institution, full name, discipline, contact information, and available time for attendance. The first meeting aimed at introducing some concepts about information systems based on care act structure and work on components that needed to be represented. Work was organized in small groups of about 6 people from different

disciplines and institutions. These groups were previously formed by the organizing team. Each group was given a list of 100 terms frequently used in the EMR. These terms had been previously selected out of the total terminology (n=39,492) to collect a purposive sample representing term heterogeneity. Sample was made of 300 terms divided in 3 groups of 100 terms. Each team was asked to analyze and divide terms into categories, choosing a name for each of them and determining their scope. These ideas were visually represented in a map and shared among all attendees. Findings were pooled, similarities and differences identified. At end of the meeting, working materials were handed over to each institution, so that activities be replicated locally with rest of the team and new proposals generated. A virtual semi-structured template was developed for motions to be submitted. Structured fields had to contain suggested terms for each component, and free text fields for rationale, comments, and so on.

Local Work

Each institution was allowed to work on its own or with the organizing team. New proposals were presented within 3 weeks, as established. Organizing team systematized information in a comparative chart. Structured data were collected, eliminating duplicates. Data in free text format were examined based upon conceptual recurrence.

Second On-Site Meeting – Panel of Representatives

Two representatives of each institution were invited. Proposals resulting from local work instances and their analysis were shared. By means of scoring system, all attendees on one hand agreed on alternatives to be excluded and, on the other, accepted those to be submitted to open voting. Conceptual definitions were discussed, agreed on, and put in writing. In addition, lists of practitioners willing to participate in later work cycles were drafted.

Open Virtual Voting

The consensus-based options were submitted to voting using web platform known as *Portal APS en Red*, used by primary health care institutions for news, instructions, and system access. Voting was open for 3 weeks.

Results

First On-Site Meeting

Meeting was attended by 120 users from different disciplines (see Table 1), representing 49 primary health care institutions. At debriefing, following initial consensus was made:

1. Having common information system helps institutions perceive themselves as part of an integrated network.
2. The way of representing population´s health problems is heterogeneous across institutions.
3. It is necessary to work together in order to enrich terminology services, particularly in connection with social determinants of health.
4. It is vital for the system to represent interdisciplinary activities.
5. Care acts may be health care-related, disease preventive, health promotional, or educational. They may also be individual or group/community-based.
6. Care acts may take place either in health care institutions or outside them, that is, in community institutions.
7. Reasons for consultations are not always health-related problems.
8. Information documented in the system may be epidemiological or administrative, for planning and management purposes.
9. The EMR is a tool for documenting people's health. It is necessary to count on other tools to document practitioner education and management activities.

Local Work

Twenty-one proposals were received. The organizing team was called in by 3 institutions. Structured data systematization resulted in following alternative names for each component: a) EMR: Single Comprehensive Health Record, Comprehensive Health Record, Health Care Electronic Record Instrument (IRES, by its Spanish acronym), Health Record, Electronic Health Record, Diagnostic Summary, Medical Practice Record, Health Practice Record, Personal Health Record, and Comprehensive Health Record. b) Care service: Provision, Intervention, Approach, Care Act, Team Practice and Response. c) Problems: Problems, Problem Situation, and Reason for Consultation. d) Practitioner activities: Special Practitioner Activities, Professional Interventions, Health Team Management and Education Activities, Practitioner Activities, PHC Team Professional Activity.

Analysis of free text showed no new findings and reinforced some of the initial agreements reached during first on-site meeting, particularly need to differentiate reasons for consultations from health problems, and existence of different types of care acts.

Second On-Site Meeting – Panel of Representatives

Meeting was attended by 32 practitioners from 21 institutions and different disciplines (see Table 1). The consensus-based voting options were: a) EMR: Comprehensive Health Record and Health Record; b) Care Service: Care Act and Intervention; c) Problems: it was unanimously decided not to submit it to voting and resort to a different record for Reasons for Consultation and Problem Situations; and d) Practitioner Activities: Health Team Management and Education Activities and Practitioner Activities. The definitions for each component were agreed in writing (see Table 2).

Table 1-Disciplines of On-Site Meeting Attendees

Discipline	Meeting 1	Meeting 2
Medicine/Dentistry	55%	43.74%
Social work	10%	21.87%
Psychology/Educational Psychology	6.66%	15.72%
Speech therapy	2.5%	-
Administration	4.17%	-
Nutrition	4.17%	9.37%
Nursing	1.67%	-
Obstetrics	1.67%	-
Anthropology	0.83%	-
Unspecified	11.67%	9.37%

Open Virtual Voting

Total of 153 voters participated. Results were as follows: a) EMR: Comprehensive Health Record (81.05%), b) Care Service: Care Act (52.24%), c) Reason for Consultation and Problem Situation was unanimously agreed on during on-site meetings and, d) Practitioner Activities: Health Team Management and Education Activities (79.58%).

Discussion

The complexity of primary health care in public health care system of Buenos Aires City was evidenced in the number of health care centers participating in the meetings, multiplicity of disciplines involved, and heterogeneity of terms used to refer to the population's health problems. In this context, reflecting

upon design of information systems which capture this complexity turned out to be quite challenging [8].

Table 2-Consensus-based Definitions for Voting

Component	Consensus-based Definitions
Electronic Medical Record	Mandatory, unified, personal, electronic record, documenting all actions performed on individuals by health care professionals and teams (Law 26.529/09 and Law 5.669/16). It was agreed that the denomination of Health Record to be used across whole network should be representative of complexity of the health, disease and health care processes.
Care Service	Every act performed by health care professionals or teams (Law 5.669/16). This refers to every contact between the health care team and individuals in the community, be it individually or in groups, and for care, prevention, or promotion purposes. It may adopt different modalities, such as consultation, advice, recreation, workshops, talks, and so on.
Problems	Reason for consultation: Everything patients express as reasons for presenting to, visiting or contacting a health care professional, this being their subjective perception and including the way they express themselves. Problem situation: Everything that triggers action by the health care professional, both in terms of education, promotion, and prevention, as well as for diagnosis, treatment, and rehabilitation purposes. These are assessments and findings based on professional expertise and knowledge of patients. It may be individual or family-related, and it may be a social or environmental determinant.
Practitioner Activities	Individual or group activities carried out by health care professionals without direct contact with community or patients including education, planning, evaluation, team meetings, case reviews, supervisions, and so on.

Design and implementation of information systems are impacted by technical aspects. However, social and cultural dimensions within institutions play fundamental role [1]. In primary health care, it was necessary to engage in in-depth discussion as to how health, disease, and health care processes are understood and how work processes are represented within the framework of PHC strategy. The consensus methodology used during meetings paved way to conceptual agreements, which are necessary starting points. These overall agreements were attained owing to the fact that participants got involved in the process and effort was made to set transparent rules [6,10].

At onset of implementation in primary health care, the information system was resisted by users. With time, different opportunities for interaction and joint work facilitated progressive appropriation of the tools by users, which was apparent in increased meeting participation. Once certain degree of system maturity was gained, it was a paramount goal to convene meetings to enable users make thorough assessment which might lead to ideas for real improvement.

Considering characteristics of health care organizations, two fundamental challenges arise when trying to manage change: striking adequate balance between practitioners' autonomy and clear definition of responsibilities; and at same time, implement change smoothly and strategically whilst coping with resistance [14,15]. Thus, it was key to promote opportunities for practitioners to work autonomously, with probable support of organizing team, open communication and interaction channels, clear goals and scope when managing meetings, and to underscore shared responsibility for introducing tool improvement motions as well.

The blended on-site and virtual strategy was instrumental in offering participation opportunities flexible enough for virtual participation of all practitioners, which allowed for organized work in smaller groups at on-site meetings.

Proposal to reflect on design of record tool shared by all institutions and practitioners helped show that information is crosscutting and also to envisage prospect of integrated health network favoring continuity of care.

Among conceptual agreements resulting from meetings, mention should be made of comprehensive view of health embodied in replacement of electronic medical record with comprehensive electronic health record; interdisciplinary approach and different types of care provision, reflected in substitution of care act for medical act; complex representation of multiple situations that may result in contact between practitioners and people in the community, captured by agreement to distinguish between reasons for consultation and problem situations; and significance of documenting practitioner activities that add quality to care act, such as education, team meetings, and activity planning.

The magnitude of these agreements is not only seen in change of terms used in the system, as with the EMR, a main tool in health information systems. Agreements also exhibit perspective which is not only medical/clinical and focused on health problems, but also comprehensive, interdisciplinary, and care-oriented.

Even though meetings aimed at defining terms for some of the record tool components, process complexity showed that, beyond terms, the way things are named bears highly symbolical and conceptual impact.

Based on achievement of conceptual agreements and term changes, in future, specific working groups will have to be organized for participatory design of functionalities representing primary care work flows.

As to limitations of these meetings, even when call was open and there was active participation, few practitioners were not challenged by the agenda and did not attend.

Conclusions

The experience derived from our meetings unambiguously stresses user participation in design of representative information tools as well as reaching conceptual agreements conducive to denominating health, disease, and health care processes recorded in the information system.

Implementation of these results will be carried out throughout 2018 and 2019, starting with change of term "Electronic Medical Record" to "Comprehensive Health Record," since other components demand software engineering.

References

[1] D.F. Sittig and H. Singh, A new socio-technical model for studying health information technology in complex adaptive healthcare systems, *Qual Saf Health Care* **19** (2010), 68-74. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3120130/

[2] F.G.B. de Quirós, M.V. Giussi Bordoni, F. Plazzotta, A. Baum, and C. Ilc, Preparation and implementation of a digital agenda in primary care in the Autonomous City of Buenos Aires, *J. Health Inform* **8** (2016), 787-798 . http://docs.bvsalud.org/biblioref/2018/07/906619/anais_cbis_2016_artigos_completos-787-798.pdf

[3] M.V. Giussi, A. Baum, F. Plazzotta, P. Muguerza, and F.G.B de Quirós, Change management strategies: transforming a difficult implementation into a successful one, *Stud Health Technol Inform* **245** (2017), 813-817. PMID: 29295211. http://ebooks.iospress.nl/publication/48264

[4] A. Kushniruk and C. Nøhr, Participatory design, user involvement and health IT evaluation, *Stud Health Technol Inform* **222** (2016), 139-151. https://www.ncbi.nlm.nih.gov/pubmed/27198099

[5] M.B. Schachner, F.J. Recondo, Z.A. González, J.A. Sommer, E. Stanziola, F.D. Gassino, M. Simón, G.E. López, and S.E. Benítez, User-centered design practices to redesign a nursing e-chart in line with the nursing process, *Stud Health Technol Inform* **225** (2016), 93-97. https://www.ncbi.nlm.nih.gov/pubmed/27332169

[6] P.R. Lawrence, How to deal with resistance to change, In: *Harvard Business Review*, 1969. https://hbr.org/1969/01/how-to-deal-with-resistance-to-change

[7] A.H. Krist, J.W. Beasley, J.C. Crosson, D.C. Kibbe, M.S. Klinkman, C.U. Lehmann, C.H. Fox, J.M. Mitchell, J.W. Mold, W.D. Pace, K.A. Peterson, R.L. Phillips, R. Post, J. Puro, M. Raddock, R. Simkus, and S.E. Waldren, Electronic health record functionality needed to better support primary care, *J Am Med Inform Assoc* **21** (2014), 764-771. https://www.ncbi.nlm.nih.gov/pubmed/24431335

[8] A.S. O'Malley, K. Draper, R. Gourevitch, D.A. Cross, and S.H. Scholle, Electronic health records and support for primary care teamwork, *J Am Med Inform Assoc* **22** (2015), 426-434. https://www.ncbi.nlm.nih.gov/pubmed/25627278

[9] R.M. Kamadjeu, E.M. Tapang, and R.N. Moluh, Designing and implementing an electronic health record system in primary care practice in sub-Saharan Africa: a case study from Cameroon, *Inform Prim Care* **13** (2005), 179-186. https://www.ncbi.nlm.nih.gov/pubmed/16259857

[10] E. Byrne and S. Sahay, Health information systems for primary health care: thinking about participation, In: *Proceedings of the IFIP* (Vol. 9). (2003), 237-250. https://www.researchgate.net/publication/232606970_Health_information_systems_for_primary_health_care_Thinking_about_participation

[11] Pan American Health Organization, *Primary Health Care-Based Health systems: Strategies for the Development of Primary Health Care Team*, PAHO, Washington, DC, 2009. http://iris.paho.org/xmlui/bitstream/handle/123456789/31221/9789275132623-eng.PDF?sequence=1&isAllowed=y

[12] D. Luna, C. Otero, M.L. Gambarte, and J. Frangella Terminology Services: Standard Terminologies to Control Medical Vocabulary. "Words are No What they Say but What they Mean", In: *Thomas F. Heston (Ed.) eHealth - Making Health Care Smarter*, IntechOpen, 2018. DOI: 10.5772/intechopen.75781.

[13] N.E. Adler and W.W. Stead, Patients in context – EHR capture of social and behavioral determinants of health, *N Engl J Med* **372** (2015), 698-701.

[14] G. Holeman, R. Wright, L. Jackson-Rossi, N.A. Dewan, and N.M. Lorenzi, Leading change in implementing technology, In: *Information Technology Essentials for Behavioral Health Clinicians*, Springer, London, 2010, 149-179.

[15] A. Baum and G. García, *Resisting: Change management in health information systems*, ebook (Spanish Edition) Kindle Edition (2018). Available on https://www.amazon.in/Resistiendo-Gesti%C3%B3n-Sistemas-Informaci%C3%B3n-Spanish-ebook/dp/B07B794LVX

Address for Correspondence

Lic. Florencia Faretta: florenciafaretta@gmail.com
Ministry of Health, Buenos Aires City, Argentina.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190277

Feasibility and Acceptability of Smart Augmented Reality Assisting Patients with Medication Pillbox Self-Management

Madeleine Blusi[ab], Juan Carlos Nieves[b]

[a] *Department of Community Medicine and Rehabilitation, Umeå University, Umeå, Sweden*
[b] *Department of Department of Computing Science, Umeå University Umeå, Sweden*

Abstract

Complex prescribed medicine regimens require extensive self-management. Handling multiple pills can be confusing; using a pillbox organiser is a common strategy. A smart Medication Coach Intelligent Agent (MCIA) can support patients in handling medicine. The aim of this research was to evaluate the feasibility and acceptability of the MCIA. A prototype was tested with 15 participants, age 17-76, filled a pillbox according to prescription assisted by the MCIA implemented in a Microsoft HoloLens. A quantitative method using questionnaires was applied. Results showed that using the MCIA implemented in an AR-headset, to assist people with prescribed polypharmacy regimen in filling a pillbox, was feasible and acceptable. There was a difference related to age regarding people's willingness to use an AR-headset for medication self-management. People older than 65 felt less comfortable using the technology and were also more hesitant to use the technology than those under 65.

Keywords:

Self-Management, Artificial Intelligence, Polypharmacy

Introduction

For people with chronic disease and multiple prescribed medications, self-management is considered an essential component [1]. Self-management requires an active role of the patient, including managing symptoms and medical treatment on a daily basis [2]. Using multiple medicines is commonly referred to as polypharmacy and is more common among older people [3]. On average persons older than 65 years have three chronic conditions, and more than 70% take five or more drugs every day. For patients with coexisting conditions, who take multiple prescribed medications distributed throughout the day adherence can be particularly challenging [4, 5].

Adherence to medication regimens is key to achieving better health outcomes for patients. Nonadherence to prescribed medication affects both the quality and length of life for patients [6]; it also places a significant cost burden on healthcare systems [7].

Common strategies used by older people to improve adherence to polypharmacy treatment is to associate medications with specific daily routines, such as meals, and using a pillbox [2, 8]. Instead of taking pills from multiple packages at different times throughout the day, the pillbox is filled up for a week at the time. A pillbox commonly has 28 compartments, distributed as seven horizontal compartments labeled Monday to Sunday, with four vertical slots for each day. The vertical compartments have labels suitable for the most common times to take pills during a day, such as morning, noon, evening, and bed.

Handling one's pills and filling the pillbox is part of medication self-management. Complexity in treatment regimens can lead to difficulties for people to continue self-management related to taking medicine and filling the pillbox. Medicine pills that look similar and use of generic medications can cause confusion and insecurity which in turn may be the trigger point for why people are unable to continue self-management and instead need to apply for health care services and help from a nurse to fill the pillbox for them. With this situation, patients lose their independence. It also puts a strain on society and health care systems, as there in many countries is a growing shortage of nurses and other health care professionals.

Based on the augmented reality (AR)-paradigm and intelligent coaching systems, a novel solution has been introduced. A smart mHealth application that can support patients with common problems related to the management of their medication makes it possible to provide patients with assistance and give advice when filling a pillbox, thus enabling continuous self-management and delay or minimize the need for nurse intervention [9, 10]. This smart mHealth application is designed and implemented as a Medication Coach Intelligent Agent (MCIA). By Intelligent Agent, we mean a software program able to be reactive, proactive, autonomous, and social. Autonomy plays a critical role in the behavior of an intelligent agent since it is expected to take decisions on its own. Hence, the MCIA has to manage different types of information such as medication plan (medication regime) of the patients, medication restrictions, as well as the patient's preferences and sensor input data from an AR-headset. It has proactive and reactive behavior in order to support patients in medication management. The MCIA is implemented in an AR headset and has autonomous reasoning capabilities that allows it to lead with long term goals in the settings of medication plans [11].

An augmented reality (AR)-headset makes it possible to have the mobility of a mobile device and still establish a social and friendly relationship with the user in hands-free interaction. Through holograms, it is possible to augment the users' field of view with an avatar. With a digital avatar, there are more possibilities regarding looks and appearances compared with a physical robot, which might have more effect on the intentions of the users. An AR headset, such as Microsoft HoloLens[1], makes it possible to be more aware of the environment and user activities because of its many sensors.

The main difference between classic smart pillboxes and the MCIA is that the technologies provide support in different stages of the medication self-management process. Classic

[1] https://www.microsoft.com/en-us/hololens

smart pillboxes and smartphone apps generally target patient groups who need help with reminders of when to take medicine, supporting the patient at the time when the patient is to put the pill in his or her mouth. The MCIA may be used for this purpose. However, it can also support individuals at an earlier stage of the medication process, when they are filling up weekly pillboxes with pills from prescribed cartons.

The aim of this research was to conduct a pilot study to evaluate the feasibility and acceptability of the proof-of-concept prototype of the Medication Coach Intelligent Agent (MCIA), implemented in an AR headset. The prototype was tested in a living lab context where participating individuals filled a pillbox according to a prescribed polypharmacy medicine regimen, assisted by the MCIA implemented in a Microsoft HoloLens. We aimed to answer the following questions: a) Is there a difference, related to age, regarding if people are willing to use an AR headset for medication self-management? and b) Is there a difference, related to the experience of using smart technology, regarding if people are willing to use an AR headset for medication self-management?

This paper represents a follow-up to our previous research. Details regarding the development process of the MCIA and technical specifications have been described in previous publications [9]. Focus in this present paper is on the user perspectives.

Methods

Quantitative methods were used for data collection and analysis. The material was explored through descriptive statistics. The research presented in this paper was conducted as a multidisciplinary collaboration between community medicine, computer science and, artificial intelligence researchers.

Setting and Participants

The prototype was tested in a living lab context. The setting was a quiet and home-like environment. Convenience sampling was used, where persons in a medium size town were asked to participate in a study and test a digital tool with the purpose to assist individuals in handling medicine and filling a pillbox. Description of participants is given in Table 1. The evaluation involved 15 participants who were selected based on the following criteria: 1) Different levels of management of medication on a regular basis, 2) Wide range of ages (medication management applies to people in all ages, not just elderly), 3) Different levels of experience from using smart technology in general and 4) A mix of men and women.

Table 1– Characteristics of Participants (n=15).

Parameter	Value (n)	Value %
Age (years)		
mean (range)	49 (17-76)	
median	58	
Gender		
women	9	60
men	6	40
Using medication	8	53
Helping others with medication	1	7
Have experience using smart technology	8	53
Familiar with AR	4	27
Persons older than 65 yrs	7	47

Procedure

Initially, participants were informed about the test procedure, which was divided into 4 sequences: 1) Fitting of the AR headset, including adjusting the size and adequate positioning on the head. 2) Becoming familiar with using the AR headset. Participants were guided by the test leader to explore features in the AR-headset for the purpose of becoming acquainted with using command functions and the feeling of receiving information through holograms. 3) Performing the actual test, handling prescribed medicine, and act on information given in holograms. 4) Responding to the evaluation questionnaire. None of the participants had used or tried on an AR-headset prior to the test.

Functionality of the MCIA

The functionality involved displaying information about the prescribed medicine, thereby helping the user to select the right medicine and dispense pills in a pillbox according to prescription. The participants were able to use voice, vision, and gestures to interact with the system and were presented with both visual and audible output. Visual outputs were in the form of holograms. Output information showed whether the person had selected the right medicine and gave information about how many pills of that particular medicine to put in each slot of the pillbox.

Instrument

Data was collected through a questionnaire which was developed at Nordic Telemedicine Center, NTC: The questionnaire was intended to be used for evaluation of digital innovations developed through the NTC [12]. The instrument comprised totally 20 items, distributed as seven questions, Q1-Q7, regarding personal background information and the following 13 questions, Q8-Q20, were formulated as statements relating to participants' experience of using the MCIA. Responses for Q8-Q20 were on a five-point Likert type scale graded from value 1 (strongly disagree) to value 5 (strongly agree). The lower bound to agree was made at value 4 (4 or 5 = agree). Examples of statements are: "It was easy to understand the presented information," "I felt comfortable using the technology," and "I am willing to use this technology if it was available right now."

Acting on Information Given in Holograms

This section gives a description of the above-mentioned test sequence number 3: Performing the actual test, handling prescribed medicine and act on information given in holograms. During the test, the participant sat at a table. On the table was a sheet of paper with a list of prescribed medicines, an empty pillbox, and five medicine boxes with labels for the prescribed medication (Figure 1).

Figure 1– Illustrations of How the MCIA is Used in a Homelike Environment for Distributing Prescribed Medicine in Pillbox.

All participants used the same mock-up medicine list, while medicine boxes were original cartons from a variety of pharmaceutical companies. When the patient looked at a medicine carton, the MCIA interpreted and acted on the information, supplying the participant with instructions on how many of each pill to put in each compartment of the pillbox. If the participant looked at a medicine which was not prescribed, the MCIA gave information that this was the wrong medicine (Figure 2). The images processed were of the cartons of the prescribed medicine. Typically a carton contains several blisters. Blisters were not scanned as the primary target group for the MCIA are patients with adequate cognitive functions, but who need support and assurance regarding variation in pharmaceutical brand names and generic drug substances.

Figure 2– Visual Information From the MCIA, Presented as Hologram. Examples: (left) instructions on how many pills to put in each compartment of the pillbox and (right) message to not take the medicine.

Results

The results are presented for the sample as a whole and also for subgroups 'Older than 65' and 'Have experience of using smart technology.' This is done in order to be able to answer the research questions regarding if there is a difference in willingness to use an AR-headset for medication management related to age and people's experiences of using smart technology (ST). Nearly half of the 15 participants (47%) were older than 65 years. A majority of participants (53%) had experience of using smart technology, and 27% said they were familiar with AR (Table 1). All participants with values 4 and 5 on Experience of ST were under 65 years old. Four participants older than 65 did have some experience of using ST, but they rated their experience lower than our cut off value (Figure 3).

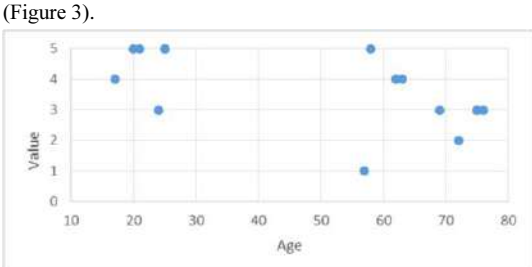


Figure 3 – Distribution of Response Values for Each Participant (n=15) for Variables 'Age' and 'Experience of using smart technology' (Values: 1=no experience and 5=very much experience).

Three of the 15 participants (20%) found it difficult to use the AR-headset, two of them were older than 65. The information was perceived as easy to understand (87%), and more than half of the participants (60%) felt comfortable while using the AR-headset to perform the medicine task. A difference related to age was discovered as all persons but one who said they felt comfortable using the technology, belonged to the younger age group. None of the participants, from either age group, felt that it was stressful to use the AR headset. Regarding the functionalities of the MCIA, 40% prefered talking over using gestures while interacting with the MCIA and a majority, 93%, appreciated receiving audio output along with the holograms. Preference for talking instead of using gestures with hands was more common in the younger age group.

Participants were asked if they, given the technology was available and affordable, would be willing to use the technology a) in the future or b) right now. Also on these questions there were differences between age groups, where all in the younger group said they would use the technology in both scenarios while 50% of those over 65 gave the same response. Variables where differences related to age were discovered are displayed in Table 2.

Table 2– Variables Where Differences in Response were Related to Age, Comparison Between Groups.

Variable	Age < 65 n=8	Age > 65 n=7
Experienced in using smart technology	87%	0%
Feel comfortable using the MCIA.	100%	15%
Prefer talking to interact with MCIA.	42%	28%
Willing to use the technology right now	100%	50%
Willing to use the technology in the future	100%	50%

Discussion

Even though none of the participants had used AR-headsets before they did not feel it was difficult to take instructions from the MCIA to perform the task of handling medicines, and most of them felt comfortable. One aim of the research was to explore if age may be a factor related to whether people are willing to use an AR-headset to support medication-related self-management. Four areas were identified, where differences in opinion could be seen between the different age groups. Participants older than 65 years were more sceptical than those who were younger when it came to feel comfortable while using the AR headset, being willing to use the technology in the future and being willing to use the MCIA now (if it was available and affordable). This result may be connected to the fact that older people in society are among those with least experience of using digital technologies and have more reluctance to change routines [13]. They generally have not used digital services throughout their working life which is an area where people in most professions today need to use digital technology.

One surprising result in the evaluation was that it was the younger group who preferred talking instead of using gestures when interacting with the MCIA. Considering that playing video games using motion sensors, as featured by Nintendo, Playstation and others, have been part of everyday life for families with children throughout the last 15 years, it could be expected that younger persons would find using motion sensors to be an ordinary way to communicate with technology. In previous research with digital interventions for seniors to use social platforms with web camera instead of the telephone, to enable visual contact during conversations with family members, participants stated that they initially felt more comfortable with the telephone because they were used to it [14]. The initiative to develop the MCIA and pursue the evaluation originated from clinical problems and challenges

experienced by nurses and patients in rural home health care environments in northern Sweden. Having difficulties self-managing their medicine and pillboxes is a common reason why patients apply for home health care. With the current shortage of nurses, it is challenging for the health care system to deliver the services and new solutions are needed to support self-management and safe handling of pills [15].

Available on the market today there are several devices and apps featuring smart medicine reminders. Based on input from the initiators to this research, nurses, and patients in rural home health care, the main target group for the MCIA is not patients with cognitive impairment or those who need assistance with reminders when it is time to take the medicine. Instead, the MCIA is aiming to support individuals who unwillingly are forced into dependence on healthcare services because they feel confused and insecure when handling the variety of generic medicines they come across based on their prescriptions. To our knowledge, there are not yet any supportive technologies for this stage of the medication process or targeting this level of self-management.

Limitations

We purposely recruited to include a wide range of ages as medication management applies to people of all ages. With the results at hand, it became apparent that age-distribution among participants was distorted. Most participants were either young adults or retired. Aiming to explore the influence of age related to willingness to use technology it would have been beneficial if the sample also included middle-aged individuals in their 40's and 50's. A majority of patients in the target group for the MCIA device is likely to be found among persons older than 65 years [4, 5]. This was the reason for placing the divider between age groups at 65 years. However, due to the ongoing demographic development, it would be relevant to repeat this evaluation with all participants older than 65 and compare relevance of age between those older/younger than 80 [16].

Another limitation was the cut off where values 4 and 5 were used to indicate "agree." This level was chosen to limit the risk of false positive answers. However, this made it impossible to address the research question concerning if the experience of using smart technology was related to people's willingness to use the MCIA. Several participants in the higher age group had chosen value 3 and with the applied cut-off level they were treated as 'not having experience' which instead gave a somewhat false negative result.

Future work

In our future work, we aim to test the MCIA for a more long term usability evaluation. The evaluation presented in this paper was conducted with test persons in a living lab using mock-up medicine list, a future evaluation is planned in collaboration with a home healthcare organisation, where patients with polypharmacy treatment will test the MCIA AR device in their own homes based on their personal prescriptions. However, based on the results in this paper, the previously mentioned evaluation with participants over 65 needs to be conducted prior to the in-home tests.

Let us point out that the MCIA belongs to a new generation of autonomous intelligent systems that aim to take decisions of their own in order to reach their designing goals. The autonomy of these systems for taking their decisions is also giving place to new research questions about responsibility. For instance, to which extent does an autonomous intelligent system such as the MCIA need to be certified by law enforcement institutions in order to be used by end-users? To answer this kind of questions is out of the scope of this paper. But, we highlight that law enforcement institutions will need to take an active role for certifying intelligent systems such as the MCIA. Let us observe that EU research projects, such as AI4EU, are appearing, and these research projects aim to develop new software methodologies that could guarantee transparency in the design and development of autonomous intelligent systems. In our future work, we will explore the social implications of the MCIA. In particular, to identify which kind of autonomous decisions by the MCIA that can put the end user in a risky situation will be part of our future work.

Conclusions

- We found that using the MCIA implemented in an AR-headset, to assist people with prescribed polypharmacy medicine regimen in filling a pillbox, was feasible and acceptable.
- There were differences related to age regarding people's willingness to use an AR headset for medication self management.
  - People older than 65 felt less comfortable using the technology.
  - People older than 65 were more hesitant to use the technology.

Acknowledgements

This research has been supported by Nordic Telemedic Center (EU Interreg Botnia Atlantica), SFO-V Strategic Research Area Health Care Science and CEDAR (Centre for Demographic and Aging Research).

References

[1] J. Barlow, C. Wright, J. Sheasby, A. Turner, and J. Hainsworth, Self-management approaches for people with chronic conditions: a review, *Patient Education and Counseling* **48** (2002), 177-187.

[2] M.W. Huygens, J. Vermeulen, I.C. Swinkels, R.D. Friele, O.C. Van Schayck, and L.P. De Witte, Expectations and needs of patients with a chronic disease toward self-management and eHealth for self-management purposes, *BMC Health Services Research* **16** (2016), 232.

[3] N. Masnoon, S. Shakib, L. Kalisch-Ellett, and G.E. Caughey, What is polypharmacy? A systematic review of definitions, BMC Geriatrics 17 (2017), 230.

[4] D. Koper, G. Kamenski, M. Flamm, B. Böhmdorfer, and A. Sönnichsen, Frequency of medication errors in primary care patients with polypharmacy, *Family Practice* **30** (2012), 313-319.

[5] T. Bodenheimer, and R. Berry-Millett, Follow the money - controlling expenditures by improving care for patients needing costly services, *New England Journal of Medicine* **361** (2009), 1521-1523.

[6] D.M. Cutler, and W. Everett, Thinking outside the pillbox — medication adherence as a priority for health care reform, *New England Journal of Medicine* **362** (2010), 1553-1555.

[7] R.L. Cutler, F. Fernandez-Llimos, M. Frommer, C. Benrimoj, and V. Garcia-Cardenas, Economic impact of medication non-adherence by disease groups: a systematic review, *BMJ Open* **8** (2018).

[8] J.J. Mira, I. Navarro, F. Botella, F. Borrás, R. Nuño-Solinís, D. Orozco, F. Iglesias-Alonso, P. Pérez-Pérez, S. Lorenzo, and N. Toro, A Spanish pillbox app for elderly patients taking multiple medications: randomized

controlled trial, *Journal of Medical Internet Research* **16** (2014), e99.
[9] M. Ingeson, M. Blusi, J.C. Nieves, Smart augmented reality mHealth for medication adherence, in : *The First Joint Workshop on AI in Health organized as part of the Federated AI Meeting (FAIM 2018), co-located with AAMAS 2018, ICML 2018, IJCAI 2018 and ICCBR 2018*, Stockholm, 2018.
[10] J.C. Nieves and H. Lindgren, Deliberative argumentation for service provision in smart environments, in: *European Conference on Multi-Agent Systems*, Springer, 2014, pp. 388-397.
[11] M. Ingeson, M. Blusi, and J.C. Nieves, Microsoft Hololens - a mHealth solution for medication adherence, in: *International Workshop on Artificial Intelligence in Health*, Springer, 2018, pp. 99-115.
[12] Nordic Telemedicine Center, https://www.nordictelemedicinecenter.eu/index.php/en/services-and-solutions/research, 2018. [accessed on March 28, 2019].
[13] P. Davidsson, M. Palm, and Å. Melin Mandre, *Svenskarna och internet 2018, Internetstiftelsen*, Svenska Internetsstiftelsen, IIS, Stockholm, Sweden, 2018.
[14] M. Blusi, M. Jong, and R. Dalin, Older People Using e-Health services - exploring frequency of use and associations with perceived benefits for spouse caregivers, *Inform*atics **3** (2016), 15.
[15] Socialstyrelsen, *Socialstyrelsens föreskrifter (SOSFS 2009:6) om bedömningen av om en hälso- och sjukvårdsåtgärd kan utföras som egenvård,* Swedish Board of Health and Welfare, Socialstyrelsen, Stockholm, 2017.
[16] United Nations, *World population prospects: the 2010 revision,* United Nations Population Division, New York, 2011.

Address for correspondence

Corresponding author: Madeleine Blusi, Department of Community Medicine and Rehabilitation, Umeå University, 901 87 Umeå, Sweden. E-mail: madeleine.blusi@umu.se Phone: +46 703551848

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190278

A Theory-Informed Digital Health Intervention in People with Severe Mental Health Problems

Sandra Bucci[ab], John Ainsworth[c], Christine Barrowclough[a], Shon Lewis[ab], Gillian Haddock[ab], Katherine Berry[ab], Richard Emsley[d], Dawn Edge[ab], Matthew Machin[c]

[a] *Division of Psychology and Mental Health, Manchester Academic Health Sciences, University of Manchester, Manchester, UK*
[b] *Greater Manchester Mental Health NHS Foundation Trust*
c Division of Informatics, Imaging & Data Sciences, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK
[d]*Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK*

Abstract

There is an important gap in scaling up psychosocial interventions for people with severe mental health problems so that these interventions are not only widely available but also delivered in a timely manner. We examined the feasability of adapting a psychological intervention traditionally delivered face-to-face onto a digital platform. We report both the clinical and technical processes used to adapt and develop the digital platform in a group of people in the early phase of psychosis. The digital platform prompts people to engage with the intervention multiple times a day over a 12-week period. Participants are also able to access a repository of multi-media content to support their mental health. The digital platform has been successfully validated by participants registered with early intervention for psychosis services in the Northwest of England, UK and is currently being tested in a powered efficacy trial.

Keywords:

Mental Health, Mobile Health, Smartphone

Introduction

Schizophrenia is considered a severe mental health problem affecting 24 million people worldwide, with the current cost to society estimated to be £11.8 billion per year in England alone [1]. The early phase of psychosis is a critical period, influencing the long-term course of difficulties. Up to 80% of people will relapse within five-years of initial episode; each relapse increases the risk of developing persistent psychotic symptoms and further disconnection from education, employment, social development and community connections, adversely affecting long-term psychosocial development [2]. Time-sensitive and accessible interventions are urgently needed. Psychological interventions such as cognitive behaviour therapy (CBT) are recommended in the treatment of psychosis. Current approaches to delivering CBT for psychosis involves a scheduled appointment with a qualified psychological therapist or CBT practitioner. Lengthy delays, fears about the stigma involved in accessing services, and a lack of trained staff are just some of the barriers that impact on timely delivery and receipt of psychological input. A high degree of smartphone ownership and use, even among people with severe mental health problems [3], means that it is possible to deliver real-time, ecologically-valid therapy that extends the reach of standard mental healthcare delivery. Smartphone technology offers an unprecedented opportunity to enhance health status by delivering real-time interventions which have the potential to extend the reach of treatment to a patient's own environment, thereby: i) transforming patient experience of, and how they engage with, services; ii) improving the efficiency and co-ordination of care; and iii) supporting people to self-manage their health and wellbeing.

Several self-management apps have either been developed and tested in clinical trials [4,5] or are currently in clinical trial [6], with researchers now turning to passive data collection including behavioural sensing and smartphone use [7] to monitor symptoms and support people with severe mental health problems. We developed a digital health intervention (DHI), Actissist, that delivers a theory-informed, in-the-moment intervention targeting distressing psychotic and psychosis-related experiences. Actissist is unconstrained by the limitations of existing treatment settings and serves as a conduit for interventions available anytime in almost any location. Leveraging technology reduces recall bias and generalisation of problems that often occurs when patients are asked to recall their symptoms, thereby improving validity of symptom reporting. Our intention is early detection of clinical deterioration while at the same time delivering timely, personalised care in one's own environment with a view to empowering people to make informed choices by way of opportunities to modify behaviour directly via a ubiquitous interface (smartphone). This paper describes the development of Actissist: a CBT-informed app for early psychosis. In particular, we describe the iterative technical and clinical process we adopted to developing the clinical content, protocol and technical architecture underlying the software. The Actissist app is currently being evaluated in a large-scale powered efficacy trial in the Northwest of England, UK.

Methods

Clinical Context

The users of the software application are people registered with an early intervention for psychosis service (EIS) in the North west of England and who are within five years of experiencing a first episode of psychosis. EIS exist worldwide and aim to

provide both pharmacological and psychosocial interventions. The Actissist app functions as a standalone app. Patients can choose to share the information in the app with their treating clinician /clinical care team, but they are not obliged to do so. Approval for this work was given by the relevant local NHS research ethics committee.

Developing the Clinical Content of the App

Actissist is grounded in the cognitive model of psychosis. We chose to adapt CBT methods because this is the recommended psychological treatment for psychosis but access to this therapy is significantly limited. The CBT approach proposes that cognitive appraisals contribute to the emergence of unhelpful beliefs and influence the interpretation of anomalous (unusual) experiences. Distress is largely linked to the meaning and interpretation of symptoms and beliefs regarding anticipated consequences. In developing the content, the clinical team distilled key theoretical elements from various academic texts and publications to ensure that the fundamental elements of CBT for psychosis are included in the architecture of the app. The elements identified include coping strategy enhancement, normalising, information, motivational interviewing techniques (e.g. decisional balance, change planning; cannabis misuse domain), psycho-education, activity scheduling, mindfulness, relaxation exercises, recovery videos, resilience building, behavioural experiments, safety behaviour work, cognitive restructuring, interactive fact sheets, to name a few.

Experience Sampling Methodology (ESM)

Psychotic experiences typically emerge in the realm of daily life and are influenced by contextual factors. Experience sampling methodology (ESM) is a method of assessment and intervention that is increasingly being used to understand psychopathology in daily life [8]. ESM uses a structured self-report diary technique to examine variables of interest as they occur in daily life. People are prompted to complete momentary questions multiple times a day over the course of a number of days [8]. As we are interested in delivering a real-time, ecologically-valid intervention in the context of one's own environment, we prompted people to engage with the Actissist app and to deliver in-the-moment CBT-informed tips and strategies to foster self-management of psychotic- and related experiences. Traditionally, ESM methods prompt people to answer a question set between 6-10 times per day over a 1-week period [8]. As the aim of this study was to not to conduct time-series analysis to assess psychopathology in daily life, we invited feedback from our user groups regarding the optimum number of alerts to facilitate engagement with the app. Our main reason for using a prompting schedule of alerts was to facilitate engagement with the app and to remind people to engage in the app, should they wish to do so.

Patient and Public Involvement (PPI) and Qualitative Work

We established an Expert Reference Group (ERG) comprising multidisciplinary group of service users, clinicians, software engineers, and clinical academics (including academic clinical psychologists and a clinical academic psychiatrist) to inform the design and development of the content, clinical and technical protocol in developing the Actissist platform. This was to ensure that the system developed remained meaningful to end-users and language and processes were appropriate for our target beneficiaries. A purposive sample of individual qualitative interviews with EIS patients and focus groups with staff who referred patients to the trial were conducted once the app was developed to understand what features both groups deemed necessary in an app of this nature and to gather ideas for improvement. All interviews were digitally-recorded, transcribed, checked for accuracy and analysed using qualitative methods.

Software Requirements

Software requirements were developed in an iterative cycle of discussing requirements, implementing those requirements deemed feasible by the software team, while incoprating feedback from end-users and stakeholders. The software team analysed the requirements provided by the clinical team in more detail to consider technical feasibility and to provide initial rough estimates of the amount of work required for each feature and the level of risk involved. These estimates were then fed back to the clinical team and resulted in two main related activities that were undertaken in a collaborative manner between software and clinical teams: (i) prioritisation of the requirements; and (ii) refinement of the requirements to align with what was considered feasible technically. The output of this initial phase was a list of the highest priority features for the app along with rough estimates of the amount of work required for each feature. The software team developed features in order of priority. Development followed the Scrum methodology. Scrum is a modern, agile, software development methodology that is widely accepted as an improvement on the more traditional Waterfall model. The iterative development approach within Scrum enables closer collaboration between the software development team and customer (in this case the clinical team and research participants) and allows changes in requirements to be incorporated throughout development. Scrum has the concept of a sprint, which is a timeboxed development effort usually between one and four weeks in duration. The software team used three-week sprints for this project as this presented a suitable balance between the need to be able to respond flexibly to changing requirements and the delivery of a sizeable chunk of developed software within each sprint. At the start of each sprint, a planning session took place attended by all of the software development and test team along with the technical project manager acting as the Product Owner and effectively acting as a proxy for the "customer" (clinical team).

Validation Study

We conducted beta-testing with 10 end-users over a 1-week period once a prototype version of the app was developed to find and eliminate defects and to ensure that the platform was accessible, clear and functional prior to commencing the clinical trial. Emphasis during testing was placed on usability and data security. We then carried out a proof-of-concept, single, blind, randomized controlled trial of Actissist plus treatment as usual (TAU) compared to a symptom-monitoring control condition plus TAU. Thirty-six early psychosis patients registered with EIS across the Northwest of England, UK were randomized on a 2:1 ratio to each arm of the trial. Engagement with the Actissist app was defined as participants completing >33% data entries over the intervention period. There was also a criterion applied to the data entry itself to determine whether it was a valid entry and could contribute to the overall adherence calculation; per participant, there was a maximum of three valid entries per day. Where the participant self-initiated access (or a combination of self- and prompt- initiated) which exceeded three entries per day, only the first three daily entries were contributed to the adherence figure. Satisfaction with the app was measured through qualitative interviews and via a satisfaction questionnaire [9].

Results

Clinical Protocol and Structure of App Content

The Actissist app was informed by CBT principles and, as far as we are aware, is the first CBT-informed app under clinical trial for early psychosis patients. Actissist is divided in two parts but functions as a single app. Actissist emits an alarm followed by a visual prompt at three pseudo-random points per day, six days per week between 10am and 10pm, for 12 weeks. The schedule of alerts used in this study was determined by user-feedback regarding the most acceptable structure of alerts over the intervention period. When a notification is accepted, or the user self-initiates use (when the user feels they would benefit from the support of the app, thereby supporting self-management), they are invited to choose a problem area. The developed app is split into two main areas: "What's bothering me?" and "My Toolkit", each of which are described below.

'What's bothering me?'

This is, what we believe, the core of the intervention. The problem areas targeted by the app are those that have been shown to predict relapse to psychosis [2]. Domains were agreed by the ERG, end-users and clinicians as areas of intervention priority (see Figure 1):

- Voices
- Suspicious thoughts
- Feeling Criticised
- Mixing with people
- Weed / Cannabis

Questions follow a branching system to mirror CBT-style questioning and are structured around a series of question-answer exchanges that focus on cognitive appraisals, rating conviction of beliefs, and identifying emotional and behavioural consequences of cognitive appraisals. Depending on the cognitive appraisal selected, the exchange is followed by rotating cognitive or behavioural tips/strategies for the user to engage in. If no problems are identified, the app simply offers a word of encouragement and is silenced until either the next alert or the user self-initiates use. For instance, for the domain 'mixing with people', questions included "How much have you been out and about?" and "How many people have you chatted to or spent time with?" Answers were selected from a radio button list. Participants were then asked how satisfied they were with this and able to choose from "Not at all", "A little", "Moderately" and "Extremely". If they responded "Extremely", then they were asked what's been good about it and received a positive reinforcement message. If the participant gave any other answer then they were asked some further questions and then received normalising information and helpful tips in managing difficult experiences (e.g. see Figure 2) and CBT-informed content. The messages they received were based on their answer to the previous questions but a number of different messages were available for each answer. This meant that participants received different intervention content each time they used the app even if they gave the same answers as the last time. This was intended to minimise boredom and repetition within the app.

'My Toolkit'

'My Toolkit' comprises a toolkit of multi-media options that operates in a standalone fashion and is split into the following main areas:

- Fact sheets (e.g. low mood, anxiety, self esteem, drug use, etc.)
- Mindfulness and relaxation audio
- Coping strategies
- Recovery videos

Should a user wish to access information in the app without going through the branching questions, they can enter the toolkit repository directly. That is, interacting with the app is not contingent on replying to the question-answer branching exchange.

In addition, a hamburger button directs the user to links including: i) help area, which includes an introductory video that provides a rationale to CBT, local service/support contacts, links to external websites (TED talks, blogs, etc), background information describing the development of the app, terms of use, acknowledgements; ii) daily diary to monitor thoughts and emotions; iii) settings / personalisation features (e.g. wallpaper, colour); and iv) a graphical readout and summary explanation of symptom fluctuations in order to self-monitor fluctuations in mood and psychotic symptoms.

In a 1-hour 'app set up session', a member of the clinical team meets with the participant and downloads the app on either the participant's own phone or on a loaned phone, sets the user a participant ID, username and password. Participants are also given the opportunity to further personalise the app (e.g. choose wallpaper, colour scheme) and orientate themselves to the different areas within the apps. The app is then activated. More specific details of this procedure are described elsewhere [9].

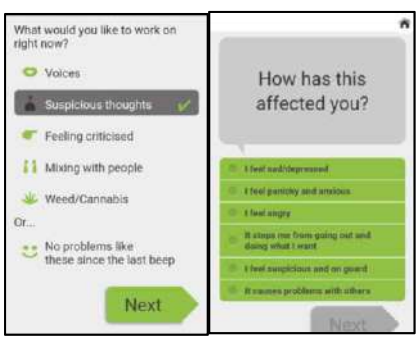


Figure 1 – Screenshot of Actissist Landing Page and Branching Questions

Figure 2 – Example of Normalising Content

Incorporation of User Feedback

User feedback was incorporated in two ways. First, our ERG met at study outset and every three months thereafter over the life of the project. An agenda was set before each meeting, with room for flexibility of topics covered, where system requirements, new app prototypes, and so on were reviewed. Second, qualitative interviews with staff and service users resulted in a number of suggestions for improvement. For example, we developed an initial pool of 79 bespoke sets of tips/strategies (socialisation: 15 tips/strategies; voices: 18 tips/strategies; suspiciousness: 13 tips/strategies; perceived criticism: 18 tips/strategies; cannabis: 15 tips/strategies). In our exit interviews, particiapnts said they wanted more tips and strategies throughout the app to minimise repetition and boredom particularly towards the end of the intervention period when users had selected the same domain repeatedly and starting receiving the same tips and strategies. Participants also said they wanted to personalise when the app sent an alert to engage with the app. Other suggestions included the ability to upload an image of a user's clinical formulation and the ability to set goals within the app to work towards a desired outcome. Where possible, this feedback will be incorporated into the next iteration of the app.

We developed a live document that stores suggestions and ideas for app refinements. This document is colour coded by the clinical team in terms of level of priority and feasiblity. This document is shared and discussed with the technical team when new versions of the app are being prepared.

App Implementation

The most important feature for the app was the branching question-answer exchange; therefore, this was prioritised. In the Scrum methodology, the aim is to provide working software at the end of each sprint. As this was a large segment of the app to build, it required several "sprints" for the development team to implement and it was therefore only possible to start providing working software after a few initial sprints. Thereafter, the implementation of other features were planned so that working software could be delivered at the end of each sprint.

As the app was developed, it was regularly reviewed by the clinical team and any improvements or changes needed were planned and incorporated into a future sprint. This also gave the clinical team an opportunity to re-prioritise requirements as the app took shape and they experienced how it worked in practice. As the development progressed, it became apparent that it would not be possible to implement the full set of features originally desired. The technical and clinical teams therefore re-prioritised the feature set, removing some of the items considered to be of lowest "value". Value was based on the amount of benefit a particular feature would provide compared with the effort and risk involved in implementing it, and whether it was deemed to be a core element of CBT-informed practice. The removed items were recorded for potential implementation in future iterations of the project; these items were compared with other feedback received from service users, clinicians and the Expert Reference Group (ERG).

Security

No identifiable data is captured in the app, instead a numeric id is used to pseudonymise it. Data captured by the app is transferred to a secure server over an encrypted HTTPS (TLS) connection. The data can then be accessed via a web interface by members of the research team. Each research team member has their own username and password to login to the web interface and access to the data is controlled based on a role-based permissions model. This means that some members of the research team can see all participants and some can only see their own participants. To improve security, login attempts are rate-limited to prevent against brute-force security attacks. The software team, who are all data protection trained, are also able to export the data collected during the study to enable the clinical team to carry out analysis. In this case, the data is secured inside an AES encrypted zip file. All of the data is still pseudonymised at this point and only the clinical team have access to patient identifiable information.

Trial Outcome

The Actissist app was successfully delivered and evaluated in a proof-of-concept trial with 24 people using the app (compared with 12 people randomised to a control symptom-monitoring app). In summary and in line with the CONSORT statement, we systematically demonstrated the safety, feasibility and acceptability of delivering a 12-week digital health intervention (DHI) using smartphone technology as follows: i) 24 patients used the app for 12-weeks; ii) the 'accept' milestone for compliance, pre-defined as 50% of the sample completing 33% or more of items was exceeded (actual achieved=75% of participants), as was the 'target' milestone for compliance pre-defined as 50% of the sample completing 50% or more of items (actual achieved=63% of sample); iii) the proportion of compliant participants was comparable to people completing either face-to-face CBT or computerised-CBT; iv) uptake was higher and drop-out lower than face-to-face and computerised-CBT; v) no reported safety issues or treatment withdrawals; vi) recruitment completed ahead of schedule, demonstrating engagement of all stakeholders to participate in, and willingness of participants to be randomised to, a trial of this nature; vii) significant improvements in psychotic and mood symptoms at the end of treatment; viii) high rates of acceptability and subjective benefit of the intervention to patients (90% participants said they would recommend the app to others in a similar position) and overall satisfaction with digital interventions expressed by staff. Full details of trial outcome are reported elsewhere [9].

Discussion

We developed a CBT-informed app for people in the early phase of psychosis. Safety, feasibility and acceptability was demonstrated in a proof-of-concept trial with 24 participants using the Actissist app compared against 12 people who used a symptom-monitoring app. The team has been successful in securing funding for a follow-on efficacy randomised controlled trial, Actissist 2.0, which has enabled a number of improvements to be made to the app functionality as well as to the look-and-feel of the app, which are summarised below. The list of potential requirements were based on feedback from participants during the proof-of-concept trial, feedback from the ERG, the list of items that did not get implemented previously, and some new ideas from the clinical team. These new requirements were prioritised in the same way as in the validation study and the following key changes in version two of the app were made:

1. Improvements to the application UI;
2. Addition of a goal setting feature so that the users can identify up to three goals s/he wishes to work towards during the intervention period;
3. Addition of an "Other" domain so that if a psychosis-specific domain is not relevant the user can be directed

to other potentially relevant mental health-related information to broaden the relevance of the app content to an individual's specific needs and concerns;

4. Option to store and access a clinical formulation image the user developed with their treating clinician.

Regarding point 1), UI design was undertaken by the software development team in the initial development phase of the app. This resulted in an app that had a reasonable look and feel and was sufficient for the proof-of-concept study. However, it became clear that specialist UI design skills were needed to meet user expectations regarding the look-and-feel of the app. A professional UI designer bought in to Actissist 2.0 will produce a series of functioning prototypes that were demonstrated to the clinical team and the ERG and were then subsequently amended to incorporate feedback and suggestions for improvements. We hope that version 2 of the app will further encourage engagement during our current efficacy trial. In response to user feedback in the validation study, version 3 of the app now comproses three parts: 'What's bothering me?', My Toolkit', and 'My goals'. The goal setting and formulations are examples of how user feedback was incorporated in the app: goal progress and uploading a formulation image were not part of the initial system requirements and were included only after users expressed a desire to monitor goals and have the ability to view an image of their clinical formulation.

Our current app has some limitations. At this stage, the Actissist app is only available for Android smartphones. This means that research participants who do not have an Android phone have to be loaned one from stocks held by the clinical team. In future, we intend to implement an iPhone version of the app. The app is limited in addressing only specific domains as opposed to delivering agile, formulation-driven CBT. Furthermore, a limited number of CBT-informed tips and strateiges are rotated within each domain of the app, also limiting the CBT content delivered within the context of the intervention period. The app has to date been trialed in an early psychosis group; we hope to extend trials of the efficacy of the intervention with later psychosis and other mental health problems.

Conclusions

We describe the development and testing of a CBT-informed app, Actissist, which targets priority domains in early psychosis. Actissist is delivered in the context of one's own environment and has been validated in a sample of 24 early psychosis users registered with EIS in the Northwest of England. Actissist was developed to close the gap to patients having access to mental health support and to provide time-sensitive intervention strategies. The app was shown to be safe and acceptable with promising signs regarding efficacy. Full scale efficacy is now being explored in a single, blind, randomised controlled. As a result of further funding, we have refined the app in line with user feedback from our validation study. We are using novel methodology to allow for three version updates of the app throughout the trial to ensure that the technology is not obsolete at the study end-point. Actissist supports self-management and can facilitate shared decision-making in treatment.

Acknowledgements

This work was supported by the Medical Research Council [grant number MR/LOO5301/1]. Thanks to the service users, clinical teams, and our ERG. S.B., J.A., S.L. are directors of Affigo CIC, a not-for-profit spun out of the University of Manchester.

References

[1] WHO Mental Health Gap Action Programme (MhGAP)," 2013. 1. WHO. (2013). Retrieved 12/03/2013 from: http://www.who.int/mental_health/mhgap/consultation_global_mh_action_plan_2013_2020/ en/index.html.

[2] M. Alvarez-Jimenez, A. Priede, S. Hetrick, S. Bendall, E. Killackey, A. Parker, P. McGorry, and J. Gleeson, Risk factors for relapse following treatment for first episode psychosis: a systematic review and meta-analysis of longitudinal studies, *Schizophrenia research* **139** (2012), 116-128.

[3] J. Firth and J. Torous, Smartphone apps for schizophrenia: a systematic review, *JMIR mHealth and uHealth* **3** (2015), e102.

[4] D. Ben-Zeev, S.M. Kaiser, C.J. Brenner, M. Begale, J. Duffecy, and D.C. Mohr, Development and usability testing of FOCUS: A smartphone system for self-management of schizophrenia, *Psychiatric rehabilitation journal* **36** (2013), 289.

[5] E. Granholm, D. Ben-Zeev, P.C. Link, K.R. Bradshaw, and J.L. Holden, Mobile Assessment and Treatment for Schizophrenia (MATS): a pilot trial of an interactive text-messaging intervention for medication adherence, socialization, and auditory hallucinations, *Schizophrenia bulletin* **38** (2011), 414-425.

[6] P.A. Garety, T. Ward, D. Freeman, D. Fowler, R. Emsley, G. Dunn, E. Kuipers, P. Bebbington, H. Waller, and K. Greenwood, SlowMo, a digital therapy targeting reasoning in paranoia, versus treatment as usual in the treatment of people who fear harm from others: study protocol for a randomised controlled trial, *Trials* **18** (2017), 510.

[7] D. Ben-Zeev, R. Brian, R. Wang, W. Wang, A.T. Campbell, M.S. Aung, M. Merrill, V.W. Tseng, T. Choudhury, and M. Hauser, CrossCheck: Integrating self-report, behavioral sensing, and smartphone use to identify digital indicators of psychotic relapse, *Psychiatric rehabilitation journal* **40** (2017), 266.

[8] I. Myin-Germeys, Z. Kasanova, T. Vaessen, H. Vachon, O. Kirtley, W. Viechtbauer, and U. Reininghaus, Experience sampling methodology in mental health research: new insights and technical developments, *World Psychiatry* **17** (2018), 123-132.

[9] S. Bucci, C. Barrowclough, J. Ainsworth, M. Machin, R. Morris, K. Berry, R. Emsley, S. Lewis, D. Edge, and I. Buchan, Actissist: proof-of-concept trial of a theory-driven digital intervention for psychosis, *Schizophrenia bulletin* **44** (2018), 1070-1080.

Address for correspondence

Dr Sandra Bucci
sandra.bucci@manchester.ac.uk
Ph: +44 161 3060422

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190279

Model-Driven Architecture Based Software Development for Epidemiological Surveillance Systems

Azanzi Jiomekong[a,b], Gaoussou Camara[b,c]

[a] *University of Yaounde I, Cameroon*
[b] *Sorbonne Université, IRD, UMMISCO, F-93143, Bondy, France*
[c] *LIMA, University Alioune Diop de Bambey, Bambey, Sénégal*

Abstract

Epidemiological surveillance systems enable collection, analysis and dissemination of information on the monitored disease to different stakeholders. It may be done manually or using a software. Given the poor performances of manual systems, the software approach is generally adopted. Epidemiological surveillance systems are based on existing softwares, softwares developed from scratch given the specifications or softwares provided by a vendor. These solutions are not always suitable because epidemiological surveillance systems evolve quickly (new drugs, new treatment protocols, etc.), leading to software updates, which can take time (while waiting for a new version) and be expensive. In this article, we present the use of the Model-Driven Architecture (MDA) approach to model and generate epidemiological surveillance systems. The result is a complete MDA based methodology and tool to develop epidemiological surveillance systems. The tool was used to model and generate softwares that are now used for epidemiological surveillance of tuberculosis in Cameroon.

Keywords:

Epidemiological Monitoring, Software Design, Tuberculosis.

Introduction

Epidemiological surveillance systems enable collection, analysis, interpretation, and dissemination of health information to different stakeholders. These information are essential to the planning, implementation and evaluation of public health practices [1]. To strengthen epidemiological surveillance, additional activities are generally integrated. For example, in Cameroon, to efficiently fight against tuberculosis (TB), the Cameroonian National Tuberculosis Control Program (NTCP) in addition to epidemiological surveillance, manages anti-TB drugs, follow-up appointments of patients, sensitize patients, etc. However, these activities are done manually, causing problems like: low rates of promptitude and completeness, the production of basic statistics are generally late, insufficient sensitization of the population, difficulties in the geolocalization of hospitals so as to help patients get to the nearest hospital, and difficulties in managing lost patients (who did not come to take their treatment). The poor performances of manual systems have oriented the adoption of the software approach [2; 3]. For example, Blaya et al. [2] have proven that a web-based system to transmit laboratories reports decrease delivery time of laboratory results, reduce redundancy in resource utilization, and provide faster and more complete notification for public health purposes, decrease the number of reporting errors to Hcs, and improved monitoring of patients because clinicians have greater access to their history and laboratory data.

The rapid advancements and availability of health Information and Communication Technologies offer remarkable enhancement opportunities for epidemiological surveillance systems. The softwares used may be developed from scratch or existing options (e.g., District Health Information Software-DHIS [4], OpenMRS [5]) may be used. These solutions are not always suitable because epidemiological surveillance systems evolve faster (new drugs, new treatment protocols, etc.), leading to software updates, which can take time (while waiting for new version) and be expensive. On the other hand, depending on the data gathered on the field, the epidemiologist may need to collect a new parameter in order to explain a phenomenon (for example, the height of the patients in order to calculate their body mass index). This task may be done by using supplementary materials such as paper form or spreadsheet software (which can lead to a problem of data integration) or new requirements can be introduced and the software updated (which can lead to the problem of software regression). These problems can be mitigated if the system is designed using a well-defined framework or architecture permitting the rapid development and refinement of the surveillance software by non-informatics experts such as health workers. MDA has proven to be one of the best choices [6-10].

MDA provides methods and tools that can be used by domain experts (institutional operators with a deep knowledge in the domain) with limited high-level IT skills to build visual models (composed of graphical notations) and generate source code. Consequently, non-technical users can safely and effectively make changes to their software to reflect their changing needs and understanding of their business [8; 9; 11]. In the context of health informatics, there is a significant number of examples adopting MDA approach [6-13]: tracking patient information, data collection, mobile-health, Crisis and Emergency Management, etc.

In this article, we present the use of an MDA based approach for rapid development/update of epidemiological surveillance systems.

We applied the approach to build and epidemiological surveillance system within the EPICAM (Epidemiology in Cameroon) project. The project aims at improving epidemiological surveillance systems in Cameroon and is conducted by the National Tuberculosis Control Program (NTCP) in Cameroon in collaboration with the Unit for Mathematical and Computer Modeling of Complex Systems (UMMISCO), Centre Pasteur of Cameroon (CPC), and MEDES in France.

The rest of the paper is organized as follows: section 2 presents the MDA based approach; section 3 presents the results of the

implementation of MDA for buiding the EPICAM platform; section 4 discusses the results; and section 5 concludes the paper.

Methods

Tuberculosis is a chronic infectious disease that kills almost two million people per year in the developing world [2]. To ensure high quality care is delivered in an efficient way, efficient information systems are essential. Then, the Cameroonian NTCP have planned a project called EPICAM with the goal to detect and treat TB patients and prevent disease from getting to people at risk (children who have been in contact with a patient or persons living with HIV). During this project, we have followed a set of principles, design activities and phases, based on agile software development and Model-Driven Architecture. In the next paragraphs, we will present the agile software development methodology, the MDA approach, and how we proceeded to develop epidemiological surveillance system in Cameroon.

Agile software development methodology

Agile software method relies on an iterative, incremental and adaptive development cycle which considers that the needs cannot be fixed and proposes to adapt the development to the changes. In a broader sense, agile method can be seen as a process consisting of an initialization step (the base version of the software is developed), an iterative step (new versions based on new specifications and users feedbacks are developed) and an incremental step (a set of new functionalities after an iteration is provided) [14; 15]. It's intended to support early and quick production of working code and allows fast deployment and adoption of health information [6].

MDA approach

MDA is an approach to software design, development and implementation which provides guidelines for structuring software specifications that are expressed as models. It focuses on forward engineering in which the executable source code is (semi)automatically generated from abstract, human-elaborated modeling diagrams such as a class diagram [6-8; 10; 12]. A class diagram describes the structure of a domain by identifying the domain classes (e.g., patient), their attributes (e.g., age, sex), their operations (e.g., calculate a body mass index) and the relationships amongst classes (e.g., the relation between a patient and his/her appointments at the hospital) [9; 13].

The process of building applications using the MDA approach can be summarized as follows: (1) the construction of a Computational Independent Model (CIM) which focuses on the environment in which the system will operate and its required features; (2) the construction of a Platform independent Model (PIM), which focuses on the aspects of the system features that are not likely to change from one platform to another; (3) the construction of a Platform Specific Model (PSM) which is obtained by integrating platform specific details to the PIM; (4) then, the PSM is converted into application code. It is the first and the second steps in the process that involves creativity and manual work; steps three and four are automated by the use of automated tools [10; 12]. The MDA approach is widely adopted to develop health information systems [6; 10].

MDA approach for epidemiological surveillance systems

Based on agile software development method and MDA, the methodology we propose for epidemiological systems development is composed of three main steps: pre-development, development and post-development.

Pre-development step: This step consists of the specification and the analysis of the system to be developed. During the pre-development step, the computer scientist works closely with the domain experts in order to make system specifications and analysis. The result is a first version of the software specification and analysis, containing sufficient information to develop the first version of the software.

Development step: This step is composed of two main phases:

- The first phase consists of the development of the first version of the software given the specifications and the analysis provided by the pre-development step. It proceeds as follows: **(1) Design:** The design consists to define the software architecture, choose the tools to be used to develop and deploy, and define the model (PIM) of the system in the form of a class diagram; **(2) Implementation:** The tools chosen in (1) will be used to transform the model into a platform specific model (PSM) which will be used to automatically generate the software.

  The result of this first phase is the baseline version of the software. The users will be trained on this basic version. Their use will generate feedback which will be very useful to continue the development.

- The second phase is an iterative and incremental phase in which each increment consists of the analysis of user feedback on the current version of software in order to add/remove functionalities. Each iteration proceeds as follows: **(1) Specifications and analysis:** the specifications and analysis are completed; **(2) Design:** the model (PIM) is updated according to new specifications; **(3) Implementation:** the model (PSM) is also updated and a new version of the software is generated. An increment is started after each iteration.

  The result of this second phase is a mature version of the software, tested and validated by the users.

- **Post-development step:** It consists of training the users on the use of the MDA tool to update the model and generate new versions of the epidemiological surveillance system. In fact, our goal is not necessarily to leave the software development to the domain expert, but to facilitate his/her task of the updating of the system to make it more close to his specifications without necessarily needing the help of an IT expert.

Development of the EPICAM platform

In this section, we present step by step how the MDA approach presented was used in order to develop the EPICAM platform for epidemiological surveillance of tuberculosis in Cameroon.

Specifications

To efficiently fight against tuberculosis, the Cameroonian National Tuberculosis Control Program has established a surveillance system through which it collects and shares data with the health professionals in health centers, health districts, health regions, the ministry of health, the general population and partners (Global fund, WHO, Centre Pasteur du Cameroun, etc.). However, this system is manually managed causing problems as previously introduced. Based on these problems, the specifications of an electronic system was developed.

To determine the specifications of the new system, we have collected information at each level (hospital, health district, health region and health ministry) of tuberculosis surveillance. These information were completed with the documents generally used for data collection and analysis. The desired system must permit users to: (1) collect, verify, synthesize data and make reports (weekly, monthly, quarterly) accessible at the district level, the regional level, and the central level; (2) follow

patients and make SMS recall for those who did not come to an appointment; (3) manage anti-tuberculosis drugs so as to prevent stockouts; (4) locate the closest hospitals with respect to the location of patients; (5) sensitize the population by SMS; (6) permit the users to work offline and update their data when they connect to the network; (7) update the system each year by adding/removing some information on data collection supports and reports.

Analysis and design

The software analysis and design helps define an architecture, a tool based on MDA, and a data dictionary. **(1) The definition of the architecture:** The architecture presented by Figure 1 is the architecture of the system which suits an environment like Cameroon where access to Internet is not always available. This architecture is composed of a user component and a server component. The user component may be a web browser connected to the server via Internet. In this case, the client shares data with the server in a synchronized manner. In the case of the offline use of the software, the user component will be a desktop/mobile application with a local database. This application will be used by the users to share data with the server in an asynchronous manner. The synchronization module works as a mediator, which permits the server get new data from the desktop application local database and the desktop application get new update from the server; **(2) Choice of tools:** to develop the software, we have made a survey of MDA based generation tools used in the health domain. This survey helps identify tools such as Open Data Kit [16], Magpi [17] and Imogene [18]. ODK and Magpi are not suitable for us because the softwares generated run on mobile phones. Hence, we have chosen Imogene. Imogene is an open source platform developed by MEDES in France and used for the generation of data collection softwares. It provides graphical tools for creating models and generating tools that help generate applications for different platforms (Android, Linux, Windows, and MacOS). Taking advantage of the MDA approach, Imogene is used to update an already deployed application by updating the model, regenerating the new application and redeploying. The applications generated may be used in a synchronous/asynchronous manner. Imogene was used to generate data collection forms for the prevention and the follow-up of diabetes in France in 2009 and for data collection for tuberculosis surveillance in Georgia. By studying Imogene, we remarked that Imogene cannot completely solve the problem because it just generates data collection applications; **(3) definition of the data dictionary:** A data dictionary, containing the entities, their attributes, their relationships and a clear definition of each entity given by domain experts was established. The definition given by the experts will permit us to present these entities in the user interfaces. From this data dictionary, a class diagram (PIM) was built. This class digram was used to represent the previous entities, their properties and relationships in a graphical manner.

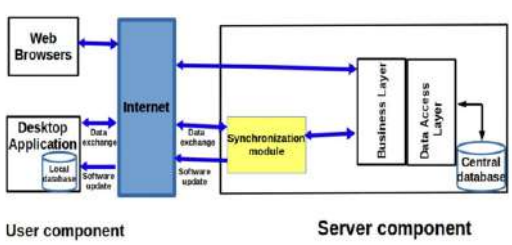


Figure 1– System architecture

Given the analysis and the *design* presented in the previous paragraph, we have developed EPICAM [19]. Based on Imogene, the EPICAM platform is an MDA based epidemiological surveillance system development. In addition to the data collection module provided by Imogene, EPICAM also integrates: **(1) reporting module:** for epidemiological report generation (in pdf format) using BIRT (Business Intelligence and Reporting Tool); **(2) geolocalization module** using OpenLayer, a library for creating interactive map on the Web; **(3) SMS module** for sensitization and patient recall; and **(4) managing drugs module** such as to prevent stockouts. EPICAM was developed to provide full software support for teams desiring to implement epidemiological surveillance system in their environment.

Implementation

Once the specifications, the analysis, and the design were completed, we have started the implementation. The EPICAM platform was therefore used to model (using visual notations) and generate a set of applications (presented in the results section). This first version was deployed in six hospitals in the two largest cities in Cameroon: Douala and Yaoundé. User feedback allowed us to complete the specifications (for example, integrate clinical radiology tests).

Iterations

After the development and the deployment of the software, the user feedback permitted us to complete the model and generate new versions of the applications. For users, everything is transparent, they simply discover new feature (for example, new field in a form).

After the application development, the NTCP selected twenty-five pilot hospitals for the deployment. These hospitals were selected given the quantity and the quality of the data they usually collect, and the ability of the health personnel to use computers. About fifty users were trained to use the softwares and three of them, from the central level were trained to use the EPICAM platform to update, generate and deploy a new version of the software.

Results

During the EPICAM project, we constructed a model, described in Figure 2, integrating the activities generally done by the NTCP. This model describes the entities of the domain, their attributes, relationships between entities and integrates the definition of each entities.

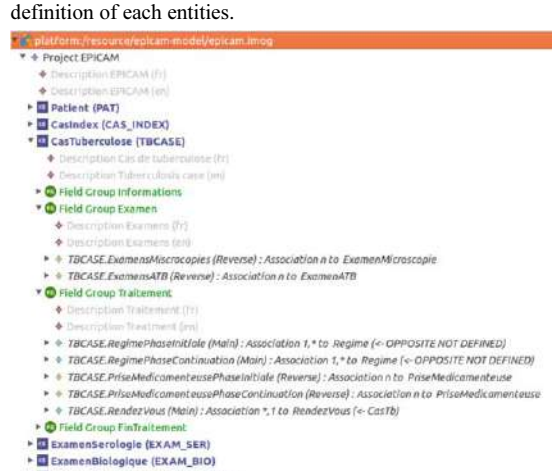


Figure 2– A model for tuberculosis surveillance software generation.

The model of Figure 2 permits us to generate a set of applications: **(1) Administration application:** used to manage health workers information, their roles, and their access rights on data; **(2) Web application:** used for epidemiological surveillance, manage anti-TB drugs, follow-up appointments of patients, sensitize patients, geolocalize the hospitals in order to orient patients. This application works in a synchronized manner. That is, when the health personnel using the application is connected to Internet; **(3) Desktop application** which has the same functionalities as the Web application, but is used in an asynchronous mode. That is, when Internet connection is not available, the system uses the local database as storage system and the local database is synchronized with the server when the Internet becomes availlable; and **(4) Synchronization application:** used to synchronize the client with the server (updating data or updating client applications).

Figure 3 presents an example of user interfaces completely generated using the model of Figure 2. This user interface is composed of the main user interface (entry point of the system) and a form used to register and follow patients (exams, drugs, appointments, etc.) during his/her treatment.

The system was deployed in twenty-five pilots hospitals, where the patients are generally treated for tuberculosis. We supervised the system in the course of the year 2015. Around 3900 (representing 15.6% of the annual number of TB cases in Cameroon) patients were registered and followed using the softwares. Given the success of this pilot phase, the NTCP have adopted the softwares generated as its electronic epidemiological surveillance softwares and extended it in ten regions, twenty new health centers in 2016 and 2017.

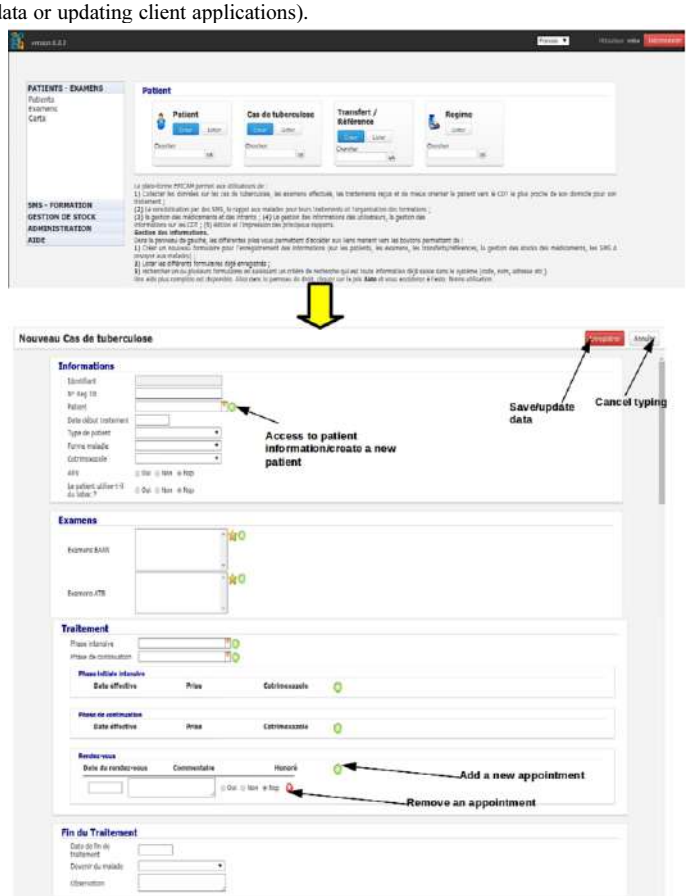


Figure 3– User interfaces completely generated using the model of Figure 2. It is composed of the software entry point and a form to register all the information during the patient treatment.

Discussion

During the EPICAM project, we produced a model for the generation of epidemiological surveillance system of tuberculosis. This model was used to generate a set of applications actually used for epidemiological surveillance of tuberculosis in Cameroon. In 2015, this platform proved its efficiency and has been adopted by the NTCP.

The use of agile software development methodologies to develop health information systems is not new. In fact, due to the complexity of the processes in health care, changes in the requirements may introduce a need for a correction of the implemented system, and can lead to regression [6]. This problem can be avoided by involving end users to the development process; and making an adapted methodology for iterative and gradual development. For example, Atanasovski et al. [6] have used an adapted agile methodology for the design and implementation of a health care information system used in Serbia and Macedonia. However, agile software development may have several drawbacks [15]. In the case of the EPICAM project, the experts were not available every time when needed. Then we were oriented to supplementary materials such as data collection forms, and statistics forms. These tools help us to advance in our job of modelling before user validation. During the validation, some elements not well modelled or defined

were corrected by the users, which took more time and energy. Atanasovski et al. [6] show that agile methodology permitted fast deployment and adoption of health information but they need another approach (e.g., MDA) to assure its extensibility, soundness, interoperability and standardization.

To adopt the MDA approach, several conditions must be considered [9; 12]: (1) the project team must be experienced in modelling. If we take the example of the EPICAM project, during the training of end users in the use of the tool in order to update the model and generate a new version of the system, just one user from the ones trained to model and generate the software has accepted to use the tool; (2) have advanced tooling with appropriate modelling formalisms. The MDA doesn't work if each project is completely different. The time spent to get the model right and the transformation will be higher than working only with code. In the case of the EPICAM project, we have presented that before the adoption of the MDA approach, we have made a survey of tools and one tool was selected and adapted by adding new functionalities in order to fit to the software to be developed. This tool can be used to model and generate epidemiological surveillance systems of other diseases than tuberculosis. Then, this will lead to a positive return on investment.

Conclusions

In this article, we have presented a methodology based on MDA and using an agile methodology to develop epidemiological surveillance systems. This methodology was used to develop EPICAM platform now used for epidemiological surveillance of tuberculosis in Cameroon. During the pilot project in the year 2015, the system showed its efficiency and has been adopted by the NTCP in Cameroon. It should be noted that the methodology presented in this work can be used for the development of any health informatics.

In the future, we plan to integrate the SMS module [20] into the platform in order to permit villages in which Internet access is unavailable or damaged by a disaster to make their data available in time.

Acknowledgements

The EPICAM project was partially supported by grants from PRODESO (Programme Franco-Camerounais pour un Développement Solidaire).

We would like to acknowledge all the health workers for their feedback and constructive comments while using the EPICAM software. Our gratitude goes also to Dr. Nolna Désiré and Dr. Abena Jean Louis from the National Tuberculosis Control Program in Cameroon; Dr. Gaetan Texier from Centre Pasteur du Cameroon; Prof. Maurice Tchuente and Dr. Hippolyte Tapamo from UMMISCO in Cameroon; Mr. Laurent Braak and Mr. Julien Dupouy from MEDES in France and finally, all those who contributed to the EPICAM project.

References

[1] R.R. Frerichs, Epidemiologic surveillance in developing countries, *Annual review of public health* **12** (1991), 257-280.

[2] J.A. Blaya, S.S. Shin, M. Yagui, C. Contreras, P. Cegielski, G. Yale, C. Suarez, L. Asencios, J. Bayona, and J. Kim, Reducing communication delays and improving quality of care with a tuberculosis laboratory information system in resource poor environments: a cluster randomized controlled trial, *PloS one* **9** (2014), e90110.

[3] M.A. Mapatano and L. Piripiri, Quelques erreurs courantes d'analyse d'un système d'information sanitaire (RD Congo), *Santé publique* **17** (2005), 551-558.

[4] R. Dehnavieh, A. Haghdoost, A. Khosravi, F. Hoseinabadi, H. Rahimi, A. Poursheikhali, N. Khajehpour, Z. Khajeh, N. Mirshekari, and M. Hasani, The District Health Information System (DHIS2): A literature review and meta-synthesis of its strengths and operational challenges based on the experiences of 11 countries, *Health Information Management Journal* (2018), 1833358318777713.

[5] W.M. Tierney, M. Achieng, E. Baker, A. Bell, P.G. Biondich, P. Braitstein, D. Kayiwa, S.N. Kimaiyo, B.W. Mamlin, and B. McKown, Experience implementing electronic health records in three East African countries, in: *MedInfo*, 2010, pp. 371-375.

[6] B. Atanasovski, M. Bogdanović, G. Velinov, L. Stoimenov, D. Sahpaski, I. Skrceska, M. Kon-Popovska, D. Janković, and B. Jakimovski, Transforming an Enterprise E-Health System from Process Oriented to Model Driven Architecture, in: *7th International Conference on Information Society and Technology ICIST*, 2017.

[7] B. Atanasovski, M. Bogdanovic, G. Velinov, L. Stoimenov, A.S. Dimovski, B. Koteska, D. Jankovic, I. Skrceska, M. Kon-Popovska, and B. Jakimovski, On defining a model driven architecture for an enterprise e-health system, *Enterprise Information Systems* **12** (2018), 915-941.

[8] H. Schlieter, M. Burwitz, O. Schönherr, and M. Benedict, Towards Model Driven Architecture in Health Care Information System Development, in: *12th International Conference on Wirtschaftsinformatik*, 2015, pp. 497-511.

[9] J.M. Kanyaru, M. Coles, S. Jeary, and K. Phalp, Using visualisation to elicit domain information as part of the Model Driven Architecture Approach, in: *First International Workshop on Business Support for MDA*, 2008.

[10] W. Raghupathi and A. Umar, Exploring a model-driven architecture (MDA) approach to health care information systems development, *International journal of medical informatics* **77** (2008), 305-314.

[11] V. Curcin, T. Woodcock, A.J. Poots, A. Majeed, and D. Bell, Model-driven approach to data collection and reporting for quality improvement, *Journal of biomedical informatics* **52** (2014), 151-162.

[12] R. France and B. Rumpe, Model-driven development of complex software: A research roadmap, in: *Future of Software Engineering*, IEEE Computer Society, 2007, pp. 37-54.

[13] K.S. Rubin, T. Beale, and B. Blobel, Modeling for Health Care, in: *Person-Centered Health Records*, Springer, 2005, pp. 125-146.

[14] C. Larman and V.R. Basili, Iterative and incremental developments. a brief history, *Computer* **36** (2003), 47-56.

[15] D. Turk, R. France, and B. Rumpe, Limitations of agile software processes, in: *Third International Conference on Extreme Programming and Flexible Processes in Software Engineering*, 2002.

[16] W. Brunette, M. Sundt, N. Dell, R. Chaudhri, N. Breit, and G. Borriello, Open data kit 2.0: expanding and refining information services for developing regions, in: *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications*, ACM, 2013, p. 10.

[17] Magpi, Home, https://home.magpi.com/, 2019. Accessed April 1, 2019.

[18] MEDES-IMPS, Welcome to imogene, https://github.com/medes-imps/imogene/wiki, 2019. Accessed April 1, 2019.

[19] UMMISCO, EPICAM Platform, http://github.com/UMMISCO/EPICAM, 2019. Accessed April 1, 2019.

[20] A. Jiomekong, L. Broto, D. Hagimont, S. Temate, and M. Tchuente, Data collection in a degraded network: case of developing countries or countries in crisis, in: *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance*, ACM, 2012, pp. 406-409.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190280

Do Medical Practitioners Trust Automated Interpretation of Electrocardiograms?

Cédric Delrot[a], Guillaume Bouzillé[b], Matthieu Calafiore[c], Michaël Rochoy[d], Bertrand Legrand[e], Grégoire Ficheur[f], Emmanuel Chazard[f]

[a] Univ. Lille, CERIM EA2694, F-59000 Lille, General practitioner, F-59179 Fenain, France
[b] Univ. Rennes, CHU Rennes, Inserm, LTSI – UMR 1099, F-35000 Rennes, France
[c] Univ. Lille, CERIM EA2694, F-59000 Lille, General practitioner, F-59150 Wattrelos, France
[d] Univ. Lille, Inserm U1171, F-59000 Lille, General practitioner, F-62230 Outreau, France
[e] Univ. Lille, CERIM EA2694, F-59000 Lille, General practitioner, F-59200 Tourcoing, France
[f] Univ. Lille, CHU Lille, CERIM EA2694, F-59000 Lille, France

Abstract

The objective is to study the way physicians use the ECG computerized interpretation (ECG-CI). Anonymous questionnaires were mailed to 282 primary care physicians (PCPs) and 140 cardiologists in France. 225 complete surveys were analyzed. PCPs performed a median of 5 ECGs per month, vs. 200 ECGs for cardiologists. Among PCPs with ECG, 57% felt confident about their skills in interpreting ECGs. Whereas 91.7% of cardiologists first interpreted the ECG by themselves, 27.9% of PCPs first read the computerized interpretation. PCPs found that ECG-CI was more reliable than cardiologists did for atrial or ventricular hypertrophy. PCPs and cardiologists agreed that ECG-CI was reliable for conduction troubles and "normal ECG" statement, but was not for other rhythm or repolarization troubles. PCPs are less experienced with ECG interpretation, but are also more likely to trust the computerized interpretation, whereas those interpreters are not fully reliable.

Keywords:

Electrocardiography, Computer interpretation

Introduction

Cardiovascular disease is the leading cause of death worldwide. According to the World Health Organizatio, 17.5 million people died from cardiovascular diseases in 2012 accounting for 31% of all deaths. In particular, 7.4 million were due to coronary heart disease and 6.7 million were due to stroke [1]. In 2014, cardiovascular disease was the second largest cause of death and the third cause of death before the age of 65 after cancer, accounting for roughly 150,000 deaths every year [2].

In France, cardiovascular risk factors are the second most common reason for consultation with primary care physicians (PCPs), accounting for 13% of primary care consulting grounds [3]. Those risk factors include uncomplicated arterial hypertension (7%), dyslipidemia (3.7%), and diabetes (2.4%) [3]. Cardiovascular diseases also represent 7.7% of emergency visits at the PCP's office. This proportion rises to 22.2% after the age of 70 [4].

Since the introduction of the string galvanometer by Willem Einthoven, the electrocardiogram (ECG) became the most widely used procedure for the diagnosis of cardiovascular disorders, and notably to confirm or exclude myocardial infarction. Since the first attempt of computerized ECG interpretation by Pipberger in the late 1950s [5], many advances in signal acquisition and diagnostic classification have been made. Currently, many electrocardiographs include a program that provides an ECG computerized interpretation (ECG-CI). Many studies have evaluated such computer programs. Those studies present the sensitivity, the specificity, or even the observed agreement rate (an intermediate computation for Cohen's Kappa coefficient, which should never be computed alone) [6–9]. Those metrics are not impacted by the low prevalence rates of diseases and may be too optimistic. Thus encouraging people to feel too confident with the ECG-CI.

The demographic atlas of the French Medical Association accounts 88,886 independent PCPs and 6,163 cardiologists in France, with a mean age of 52 and 51 years, respectively [10]. Among family physicians, 94.4% perform ECGs in the office in the USA [11], and only 62% in France [12].

In France, many studies at a local level have evaluated the use of ECG in general practice. Rural practice and group practice are the key predictors of owning an electrocardiograph. For PCPs, the limiting factors of owning an ECG device appear to be doubt about their own ability to interpret the exam, time consumption, and the fear of legal prosecutions in case of wrongful interpretation [13]. To our knowledge, no study investigated the way physicians consider and use the ECG-CI.

The aim of this study is to determine the attitude of French PCPs and cardiologists towards ECG-CI.

Methods

Participants and procedure

We obtained the 2016 French health professionals registry (RPPS) [14], which registers all physicians in France, and includes their names and specialty. We randomly draw 600 PCPs and 200 cardiologists. Retired physicians, military physicians, PCPs only practicing alternative medicine (e.g. homeopathy or acupuncture), salaried physicians, and physicians who were practicing in the French overseas departments and territories were excluded, as well as physicians whose postal address could not be retrieved. A questionnaire was then sent to 282 PCPs and 140 cardiologists, accounting for 0.32% of

all primary care physicians and 2.27% of all cardiologists in France (Figure 1). The survey was completed anonymously.

Paper questionnaire

A specific questionnaire was developed based on a review of the literature. It was designed to be self-administered, and to assess beliefs towards ECG computerized interpretation, and to evaluate the use of this diagnostic tool. It took approximately 3 minutes to complete. The questionnaire was then improved based on feedback from two academic public health physicians, and on pilot testing on two cardiologists and three PCPs. The final questionnaire consisted of 4 sections. The first section included questions regarding the participants' training in interpreting ECGs, self-perceived skills in ECG interpretation, distance from the nearest emergency department, practice type, and demographic information. The second section was for participants who did not have an electrocardiograph. It asked them to indicate if they formerly used to own an electrocardiograph, the limiting factors of owning an electrocardiograph, and how they handled the need to perform an ECG. The third section focused on participants' practices regarding the type of electrocardiograph they had, how often they used it if they used a support network for the interpretation of ECGs, and the medical causes to use it. The final section asked participants to indicate if they had an ECG-CI, how often they used it, and their confidence towards ECG-CI. The questionnaire included multiple-choice items and 5-point Likert scale items that ranged from "strongly disagree" to "strongly agree" or from "not helpful at all" to "very helpful," with an option either of "no opinion" or "does not apply." Each questionnaire was accompanied by a cover letter describing the study and a postage-paid reply envelope. The subjects were informed that all responses were anonymous and confidential. They had to return at the same time a waiver of informed consent.

Statistical analysis

Random drawing and statistical analyses were then performed using R statistical software, and R Studio [15,16]. Quantitative variables were described using the mean and standard deviation (SD) for symmetric distribution, and median and quartiles (Q1, Q3) for skewed distributions. 95% confidence intervals (CI95) of means were computed using a Student law. Categorical variables were described using number and percentage, and a binomial law was used to compute the CI95. In order to compare PCPs' and cardiologists' answers, we used χ^2 or Fisher's exact test for qualitative variables, Cochran–Armitage test for trend for ordered qualitative variables, and t-test for quantitative variables. All tests were double-sided, and a *p*-value lower than 0.05 was considered statistically significant.

Results

The response rate was 53.8% (Figure 1) and did not significantly differ between PCPs and cardiologists (p=0.46). The demographics and baseline characteristics of respondents are shown in Table 1. Most respondents were men (69.3%), the mean age was 52.6 years (SD=12.5). Respondents and non-respondents were similar in terms of gender and region of the country. Significant differences between PCPs and cardiologists existed for gender, region, practice location and setting, qualifications, distance from an emergency department, ECG-related training, self-perceived skills about ECG interpretation, and ownership of an ECG device (Table 1).

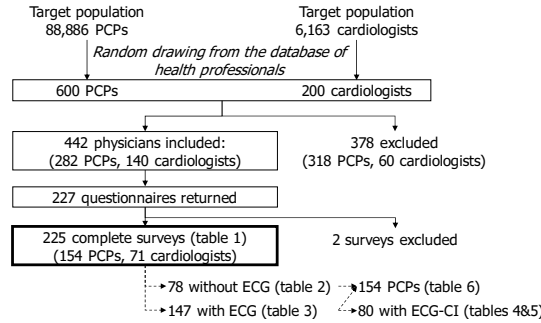


Figure 1. Participant flowchart (sampling fraction: 0.32% for PCPs & 2.27% for cardiologists; response rate: 53.8%)

Table 1. Descriptive characteristics of respondents. Percentages relate to columns.

Variable	PCPs (n=154)	Cardiologists (n=71)	*p* value
Age, mean (SD)	51.9 (10.8)	54.1 (11.3)	*0.17*
Men, n (%)	96 (62.3)	60 (84.5)	*<0.001*
Capital region, n (%)	25 (16.2)	16 (22.9)	*<0.001*
Practice location, n (%)			*0.003*
Rural	29 (19)	6 (8.5)	
Semi-rural	44 (28.8)	11 (15.5)	
Urban	80 (52.3)	54 (76.1)	
Main practice, n (%)			*<0.001*
Private practice	45 (29.2)	20 (28.6)	
Medical home	17 (11)	2 (2.9)	
Group practice	88 (57.1)	27 (38.6)	
Other	4 (2.6)	21 (30)	
Qualifications, n (%)			
ECG university degree	0 (0)	14 (30.4)	*<0.001*
Emergency specialty	11 (14.1)	0 (0)	*0.007*
Sports medicine	18 (23.1)	6 (13)	*0.24*
Other	49 (62.8)	26 (56.5)	*0.57*
On-call duty, n (%)	85 (55.6)	40 (56.3)	*1*
Distance from the nearest emergency department, km, median [Q1;Q3]	5 [2;15]	1 [0;5]	*<0.001*
Training in ECG interpretation, n (%)			
University degree	6 (4)	45 (68.1)	*<0.001*
Medicine studies	139 (90.9)	41 (62.1)	*<0.001*
Training seminar	18 (11.8)	5 (7.6)	*0.47*
Books	38 (25)	8 (12.1)	*0.046*
Internet	10 (6.6)	6 (9.1)	*0.57*
Other	13 (8.6)	5 (7.6)	*1*
"properly trained" to interpret ECGs, n (%)	55 (36.7)	65 (94.2)	*<0.001*
ECG device, n (%)	76 (49.4)	71 (100)	*<0.001*

The attitude of physicians who did not have an ECG device is reported in Table 2. It is worth noting that all cardiologists had an ECG device, therefore only PCPs are described in Table 2. The most important reason not to have an ECG device was to allow easy access to cardiologists or hospitals (71.2% [59.5;81.2]). Patients requiring an ECG in an emergency were first and foremost routed to an emergency room by 91% ([82.4;96.3]) of PCPs, and patients requiring a non-urgent ECG were first and foremost routed to a cardiologist by 93.5% ([85.5;97.9]) of PCPs.

Table 2. Reported attitude about ECG of physicians who do not own any ECG device. Percentages relate to columns.

Item, n (%)	PCPs (n=78)
Formerly had an ECG machine in their office	21 (27.3)
Reasons for not having an ECG machine:	
Easy access to a cardiologist or to a hospital	52 (71.2)
Difficulties in interpreting the ECG	32 (43.8)
Involvement of responsibility	31 (42.5)
Rare indications in current practice	29 (39.7)
Emergency services easily available	23 (31.5)
High purchase and maintenance costs	20 (27.4)
Insufficient reliability of the procedure	14 (19.2)
Time-consuming procedure	11 (15.1)
No diagnostic interest	2 (2.7)
Other	3 (4.1)
For urgent ECGs, patients are referred to:	
The emergency department	71 (91)
A cardiologist	6 (7.7)
A PCP who owns an ECG device	0 (0)
Other	1 (1.3)
For non-urgent ECGs, patients are referred to:	
A cardiologist	72 (93.5)
A PCP who owns an ECG device	2 (2.6)
The emergency department	1 (1.3)
Other	2 (2.6)

Table 3. Reported practices about ECG of physicians who own a device. Percentages relate to columns.

Variable	PCPs (n=63)	Cardiologists (n=65)	*p* value
Type of ECG machine, n (%)			*0.062*
Analog device	39 (61.9)	30 (46.2)	
AliveCor Mobile ECG	1 (1.6)	0 (0)	
Other digital ECG	23 (36.5)	35 (53.9)	
Number of leads, n (%)			*0.005*
<12 leads	17 (25.4)	4 (6)	
12 leads	50 (74.6)	59 (88)	
>12 leads	0 (0)	4 (6)	
Number of uses per month, median [Q1;Q3]	5 [2;10]	200 [150;300]	*<0.001*
Usual ECG indications, n (%)			
Diagnosis, symptoms	67 (91.8)	61 (87.1)	*0.42*
Disease monitoring	34 (46.6)	67 (95.7)	*<0.001*
Sports certificate	43 (58.9)	56 (80)	*0.007*
Before new treatment	29 (39.7)	57 (81.4)	*<0.001*
Baseline ECG	18 (24.7)	48 (68.6)	*<0.001*
Other	9 (12.3)	23 (32.9)	*0.005*
Use of the following aid for ECG interpretation, n (%)			
ECG ruler	37 (53.6)	60 (89.6)	*<0.001*
Automated measurements (intervals, freq.)	35 (50.7)	32 (47.8)	*0.73*
Book	14 (20.3)	2 (3)	*0.002*
Internet	6 (8.7)	1 (1.5)	*0.11*
Remote transmission	4 (5.8)	1 (1.5)	*0.37*
Device with available computerized interpreter, n (%)	44 (58.7)	36 (50.7)	*0.41*
Support network for ECG interpretation, n (%)			*<0.001*
Fax to a colleague	17 (23)	1 (1.5)	
Telemedicine network	4 (5.4)	0 (0)	
Paid service	0 (0)	0 (0)	
No support network	53 (71.6)	67 (98.5)	

Table 3 reports practices relating to ECG interpretation. The median number of ECGs performed per month was 5 for PCPs and 200 for cardiologists (p=0). PCPs and cardiologists had significantly different characteristics of the number of leads, reasons to perform the ECG, and support used for the interpretation (Table 3). In total, 54.8% of devices were equipped with ECG-CI.

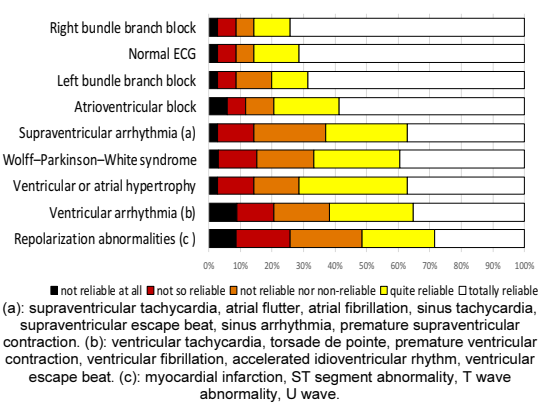
Table 4 shows the most important result of this study: PCPs and cardiologists did not use the ECG devices in the same way: 69.8% [54%; 83%] of PCPs first read the ECG and then the computerized interpretation, although this proportion fell to 52.8% [35%; 70%] for cardiologists, and 38.9% [23%; 57%] of cardiologists even never looked at the computerized interpretation.

Table 4. Usual attitude of physicians having ECG-CI (p <0.001). Percentages relate to columns.

Attitude, n (%)	PCPs (n=43)	Cardiologists (n=36)
Never read the ECG, only read the ECG-CI	0 (0)	0 (0)
First read the ECG-CI, then read the ECG	12 (27.9)	3 (8.3)
First read the ECG, then read the ECG-CI	30 (69.8)	19 (52.8)
Never read the ECG-CI, only read the ECG	1 (2.3)	14 (38.9)

Table 5. Perceived reliability of computerized interpretation (Confidence: 1=Not reliable at all; 2=Unreliable; 3=Neither reliable nor unreliable; 4=Reliable; 5=Very reliable).

Confidence (mean, SD) on the ECG-CI ability to detect some situations	PCPs (n=43)	Cardiologists (n=35)	*p* value
Normal ECG	4 (0.6)	3.9 (0.9)	*0.399*
Ventricular or atrial hypertrophy	3.5 (0.7)	3.1 (1.1)	*0.040*
Right bundle branch block	3.5 (0.8)	3.6 (1)	*0.558*
Left bundle branch block	3.6 (0.8)	3.7 (1)	*0.616*
Atrioventricular block	3.7 (0.8)	3.5 (1)	*0.262*
Wolff–Parkinson–White	3 (0.9)	2.7 (1)	*0.190*
Supraventricular arrhythmia	3.3 (0.9)	3 (1.1)	*0.174*
Ventricular arrhythmia	3.5 (0.8)	3.1 (1.2)	*0.095*
Repolarization abnormalities	3 (1)	2.7 (1.3)	*0.153*



(a): supraventricular tachycardia, atrial flutter, atrial fibrillation, sinus tachycardia, supraventricular escape beat, sinus arrhythmia, premature supraventricular contraction. (b): ventricular tachycardia, torsade de pointe, premature ventricular contraction, ventricular fibrillation, accelerated idioventricular rhythm, ventricular escape beat. (c): myocardial infarction, ST segment abnormality, T wave abnormality, U wave.

Figure 2. Perceived reliability of ECG-CI, by statement

Table 5 and Figure 2 display the perceived reliability of each category of statements of computerized ECG interpretation. The Cochrane-Armitage test reveals a significant difference between PCPs and cardiologists only for atrial or ventricular hypertrophy, where PCPs thought the interpretation was more reliable than cardiologists did (p=0.04). Globally, physicians thought the computerized interpretation was most reliable for conduction troubles and "normal ECG", and was less reliable for other rhythm troubles as well as repolarization troubles.

Among PCPs, Table 6 compares PCPs equipped with an ECG device or not. A physician who had an ECG device was most likely a man, in rural areas, with a private practice, ECG-related diplomas, who worked further away from emergency rooms or cardiologists, and was trained for ECG interpretation (Table 6). However, only 57% of them felt confident about their skills in interpreting ECGs.

Table 6. Descriptive characteristics of PCPs, equipped or not with ECG devices. Percentages relate to columns.

Variable	PCPs with ECG (n=76)	PCPs without ECG (n=78)	*p* value
Age, years, mean (SD)	51.1 (12.5)	52.7 (8.9)	*0.367*
Men, n (%)	56 (73.7)	40 (51.3)	*0.005*
Practice location, n (%)			*0.003*
Rural	22 (29)	7 (9.1)	
Semi-rural	23 (30.3)	21 (27.3)	
Urban	31 (40.8)	49 (63.6)	
Main practice setting, n (%)			*0.028*
Private practice	26 (34.2)	19 (24.4)	
Medical home	12 (15.8)	5 (6.4)	
Group practice	35 (46.1)	53 (68)	
Other	3 (4)	1 (1.3)	
Qualifications, n (%)			
ECG university degree	0 (0)	0 (0)	*0.651*
Emergency specialty	9 (22)	2 (5.4)	*0.051*
Sports medicine	10 (24.4)	8 (21.6)	*0.795*
Other	25 (61)	24 (64.9)	*0.816*
On-call duty, n (%)	50 (66.7)	35 (44.9)	*0.009*
Distance from the nearest, km, median [Q1;Q3]			
Emergency department	10 [3;20]	3 [2;10]	*<0.001*
Cardiology office	6 [1;18.5]	1 [1;5]	*<0.001*
Training in interpreting ECGs, n (%)			
University degree	2 (2.6)	4 (5.3)	*0.681*
Medicine studies	70 (92.1)	69 (89.6)	*0.780*
Training seminar	14 (18.4)	4 (5.3)	*0.022*
Books	23 (30.3)	15 (19.7)	*0.189*
Internet	6 (7.9)	4 (5.3)	*0.745*
Other	9 (11.8)	4 (5.3)	*0.245*
Self-perceived skills in ECG interpretation, n (%)	43 (57.3)	12 (16)	*<0.001*

Discussion

As more and more ECG devices embed software for computerized interpretation, and as the accuracy of such devices remains moderate, it was necessary to know whether physicians owned such devices, and how much they trusted the output of the program. We could analyze 225 questionnaires. An important result was that 49.4% of PCPs were equipped with ECGs. Among them, 58.7% had an embedded interpretation software, and 27.9% reported moderate confidence for its output. PCPs and cardiologists did not use the output the same way: whereas 91.7% of cardiologists first interpreted the ECG by themselves, and eventually read the ECG-CI output, 27.9% of PCPs first read the ECG-CI output, and then tried to interpret the ECG. PCPs and the cardiologists did not have the same qualification level: 68% of cardiologists had a university degree in ECG interpretation, versus 3.95% for PCPs. In addition, 63.3% of PCPs perceived a lack of skills in ECG interpretation. Furthermore, the physicians did not equally trust the computerized interpretation for all the diseases. This study also highlighted factors associated with ECG equipment: PCPs who owned an ECG device were more likely working in rural areas, far away from emergency departments. It is worth noting that, in France, general practitioners can bill ECGs, irrespectively from their post-graduate certificates. The nature of the devices also differed: 61.9% of PCPs were equipped with analog ECG devices whereas 53.9% of cardiologists used digital ECG devices. Only 28.4% of PCPs used support network for ECG interpretation, via fax for 23.0% of them. It also highlighted that having a personal device may have an impact on healthcare consumption: PCPs who did not own a device were more likely to address their patients to emergency rooms. Patients requiring an ECG in an emergency were first and foremost routed to an emergency room by 91% of PCPs, and patients requiring a non-urgent ECG were routed to a cardiologist by 93.5% of PCPs. The main limiting factors for non-equipped doctors was the easy access to cardiologists or hospitals,doubt in their ability to interpret ECGs, and the fear of legal proceedings for wrongful interpretation.

Those results are consistent with the literature. Previous works [13] showed that among PCPs in France, 6% had absolute confidence in computerized ECG interpretation, 65% had confidence when computer declared the ECG as normal, 26% trusted the computer for diagnosis of arrhythmias, and 21% for acute ischemia, whereas several analyses of computerized ECG interpretation accuracy concluded that frequent errors in the interpretation of cardiac rhythm occurred, with many false positives especially for the diagnosis of atrial fibrillation [17–24]. This study also highlighted that the median number of ECGs performed per month was 5 for PCPs and 200 for cardiologists, whereas the American College of Cardiology (ACC) and American Heart Association (AHA) recommended interpretation of 500 ECGs under supervision of an expert to attain initial competency, and the reading of 100 ECGs yearly to maintain competency in ECG interpretation [17].

To our knowledge, no study investigated specifically the way physicians considered and used the computerized ECG interpretation. The answers were of good quality. This work also presents some limitations. The questionnaires were self-administered, and collected declarations of the practitioners about their own practice instead of performing an objective evaluation. The response rate was moderate (53.8%).

Conclusion

The general practitioners are less experienced with ECG interpretation but are more likely to trust the results of ECG-CI. Paradoxically, insufficient knowledge may increase practitioners' trust in technologies that are not 100% reliable.

References

[1] World Health Organization, Global Health Estimates: Deaths by Cause, Age, Sex and Country, 2000-2012, WHO, Geneva, 2014.

[2] A.S. Go, D. Mozaffarian, V.L. Roger, E.J. Benjamin, J.D. Berry, M.J. Blaha, S. Dai, E.S. Ford, C.S. Fox, S. Franco, H.J. Fullerton, C. Gillespie, S.M. Hailpern, J.A. Heit, V.J. Howard, M.D. Huffman, S.E. Judd, B.M. Kissela, S.J. Kittner, D.T. Lackland, J.H. Lichtman, L.D. Lisabeth, R.H. Mackey, D.J. Magid, G.M. Marcus, A. Marelli, D.B. Matchar, D.K. McGuire, E.R. Mohler, C.S. Moy, M.E. Mussolino, R.W. Neumar, G. Nichol, D.K. Pandey, N.P. Paynter, M.J. Reeves, P.D. Sorlie, J. Stein, A. Towfighi, T.N. Turan, S.S. Virani, N.D. Wong, D. Woo, M.B. Turner, and American Heart Association Statistics Committee and Stroke Statistics Subcommittee, Heart disease and stroke statistics--2014 update: a report from the American Heart Association, *Circulation*. **129** (2014) e28–e292. doi:10.1161/01.cir.0000441139.02102.80.

[3] Letrilliart L, Supper I, Schuers M, Darmon D, Boulet P, Favre M, Guerin MH, Mercier A., ECOGEN : étude des Éléments de la COnsultation en médecine GENérale, *Exercer 2014;114:148-57.* http://www.exercer.fr/numero/114/page/148/.

[4] Gouyon M., Emergencies in General Medicine., *Direction de La Recherche, Des Études, de l'évaluation et Des Statistiques*. (2006). https://www.epsilon.insee.fr/jspui/bitstream/1/13662/1/seriestat94.pdf (accessed Nov 21, 2018).

[5] H.V. Pipberger, E. Simonson, E.A. Lopez, M.A. Araoye, and H.A. Pipberger, The electrocardiogram in epidemiologic investigations. A new classification system., *Circulation*. **65** (1982) 1456–1464. doi:10.1161/01.CIR.65.7.1456.

[6] D. Woolley, M. Henck, and J. Luck, Comparison of electrocardiogram interpretations by family physicians, a computer, and a cardiology service, *J Fam Pract*. **34** (1992) 428–432.

[7] J.L. Willems, C. Abreu-Lima, P. Arnaud, J.H. van Bemmel, C. Brohet, R. Degani, B. Denis, J. Gehring, I. Graham, G. van Herpen, H. Machado, P.W. Macfarlane, J. Michaelis, S.D. Moulopoulos, P. Rubel, and C. Zywietz, The Diagnostic Performance of Computer Programs for the Interpretation of Electrocardiograms, *New England Journal of Medicine*. **325** (1991) 1767–1773. doi:10.1056/NEJM199112193252503.

[8] T.-W. Ho, C.-W. Huang, C.-M. Lin, F. Lai, J.-J. Ding, Y.-L. Ho, and C.-S. Hung, A telesurveillance system with automatic electrocardiogram interpretation based on support vector machine and rule-based processing, *JMIR Med Inform*. **3** (2015) e21. doi:10.2196/medinform.4397.

[9] A.P. Shah, and S.A. Rubin, Errors in the computerized electrocardiogram interpretation of cardiac rhythm, *J Electrocardiol*. **40** (2007) 385–390. doi:10.1016/j.jelectrocard.2007.03.008.

[10] Conseil National de l'Ordre des Médecins, Interactive cartography of medical demographics, (2013). http://demographie.medecin.fr/ (accessed Nov 21, 2018).

[11] American Academy of Family Physicians (AAFP), Recommended Curriculum Guidelines for Family Medicine Residents, https://www.aafp.org/dam/AAFP/documents/medical_education_residency/program_directors/Reprint262_Cardio.pdf (accessed Nov 21, 2018).

[12] aafp, Electrocardiograms, Family Physician Interpretation of, (2013). http://www.aafp.org/about/policies/all/electrocardiograms.html (accessed Nov 21, 2018).

[13] Cretallaz Pierre, Factors limiting ECG devices acquisition in General Medicine, Thesis, Nice University, 2015. https://dumas.ccsd.cnrs.fr/dumas-01285060/document (accessed Nov 21, 2018).

[14] esante.gouv.fr, Accessing the RPPS data, *Le Portail de l'ASIP Santé*. http://esante.gouv.fr/services/referentiels/identification/acceder-aux-donnees-du-rpps-0 (accessed November 24, 2016).

[15] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2014. http://www.R-project.org/.

[16] RStudio Team (2015), RStudio: Integrated Development for R. RStudio, Inc. http://www.rstudio.com/.

[17] A.H. Kadish, A.E. Buxton, H.L. Kennedy, B.P. Knight, J.W. Mason, C.D. Schuger, C.M. Tracy, W.L. Winters, A.W. Boone, M. Elnicki, J.W. Hirshfeld, B.H. Lorell, G.P. Rodgers, C.M. Tracy, H.H. Weitz, ACC/AHA clinical competence statement on electrocardiography and ambulatory electrocardiography: A report of the ACC/AHA/ACP-ASIM task force on clinical competence endorsed by the International Society for Holter and noninvasive electrocardiology, *Circulation*. **104** (2001) 3169–3178.

[18] M. Hwan Bae, J. Hoon Lee, D. Heon Yang, H. Sik Park, Y. Cho, S. Chull Chae, and J.-E. Jun, Erroneous computer electrocardiogram interpretation of atrial fibrillation and its clinical consequences, *Clin Cardiol*. **35** (2012) 348–353. doi:10.1002/clc.22000.

[19] F. Bogun, D. Anh, G. Kalahasty, E. Wissner, C. Bou Serhal, R. Bazzi, W.D. Weaver, and C. Schuger, Misdiagnosis of atrial fibrillation and its clinical consequences, *Am. J. Med.* **117** (2004) 636–642. doi:10.1016/j.amjmed.2004.06.024.

[20] M.E. Guglin, and D. Thatai, Common errors in computer electrocardiogram interpretation, *Int. J. Cardiol.* **106** (2006) 232–237. doi:10.1016/j.ijcard.2005.02.007.

[21] S. Goodacre, A. Webster, and F. Morris, Do computer generated ECG reports improve interpretation by accident and emergency senior house officers?, *Postgrad Med J*. **77** (2001) 455–457.

[22] D.J. Brailer, E. Kroch, and M.V. Pauly, The impact of computer-assisted test interpretation on physician decision making: the case of electrocardiograms, *Med Decis Making*. **17** (1997) 80–86.

[23] A. Haeberlin, L. Roten, M. Schilling, F. Scarcia, T. Niederhauser, R. Vogel, J. Fuhrer, and H. Tanner, Software-based detection of atrial fibrillation in long-term ECGs, *Heart Rhythm*. **11** (2014) 933–938. doi:10.1016/j.hrthm.2014.03.014.

[24] J. Mant, D.A. Fitzmaurice, F.D.R. Hobbs, S. Jowett, E.T. Murray, R. Holder, M. Davies, and G.Y.H. Lip, Accuracy of diagnosing atrial fibrillation on electrocardiogram by primary care practitioners and interpretative diagnostic software: analysis of data from screening for atrial fibrillation in the elderly (SAFE) trial, *BMJ*. **335** (2007) 380. doi:10.1136/bmj.39227.551713.AE.

Address for correspondence

Emmanuel Chazard emmanuel.chazard@univ-lille.fr, CERIM EA2694, medicine faculty, 1 av Oscar Lambret, F-59035 Lille cedex, France.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190281

Project and Preliminary Evaluation of *SimHosp*, a Tool for Decision Making in Nursing

Gustavo Henrique Cervi[ac], Rute Merlo Somensi[a], Rita Catalina Aquino Caregnato[a], Ana Amélia Antunes Lima[a], Ana Respício[b] and Cecília Dias Flores[a]

[a] *Department of Health Sciences, Federal University of Health Sciences, Porto Alegre, Brazil*
[b] *Department of Informatics, Universidade de Lisboa, Lisboa, Portugal*
[c] *Faculdade IENH, Novo Hamburgo, Brazil*

Abstract

Computerized simulators are important tools that support teaching in many areas. The use of these instruments is often described as significant for students and teachers. In this paper, a software piece was developed to aid decision-making in nursing, allowing the simulation of real situations, addressed in the classroom. The simulator architecture corresponds to a multi-agent system supported by a state machine model. To build the knowledge base, a list of contents was selected, including the Nursing Intervention Classification (NIC). An experiment was carried out with the participation of eleven students of the third year of the course. A questionnaire was applied and, as a result, there were more than 90% of acceptance as a relevant educational tool. The simulation, through this tool, contributed to apply theoretical knowledge to the students, besides helping in the development of the nursing decision-making ability.

Keywords:

Health education, decision making, computer simulation.

Introduction

Decision-making results from a complex cognitive process that associates knowledge with practice [1]. Among management activities performed by the nurse, the management of people is often described as highly complex task [2]. In order to assist the teaching of people management in nursing, this paper proposes the use of a simulator as a tool to support the prediction of the amount of nursing personnel in a common hospital ward. To achieve its purpose, a computer simulator should be as realistic as possible in relation to the simulated scenario [3]. Several factors, including hospitalization unit, patients, relatives, doctors, nurses and other professionals should be considered [4]. The presence of an environment and several actors, interacting with each other, can characterize a multi-agent system where agents, supported by a knowledge base, acts by allowing a simulation of high verisimilitude [5]. In addition to the specialists, who contribute with their knowledge for the elaboration of rules, scripts and decision functions, this work uses the rationale of the Classification of Nursing Interventions (NIC) [6]. The NIC provides a knowledge base about interventions and nursing actions that can guide the simulator's decisions, such as the average times of each nursing activity and the level of knowledge required of the professional who will perform it [6]. Another source of knowledge used in this study is the Brazilian legislation related to the dimensioning of nursing professionals [7] in health services, determined by the level of complexity of care of patients assisted, by the managerial and care model adopted by the health institution, among other factors.

Methods

Technology

The simulator was initially developed as a prototype. After a series of tests, the code was translated to the .NET framework, written in C# language. The presentation layer is based on the WinForms (desktop application) for ease of support and low-level hardware performance (2D and 3D rendering). Since this software is available as free software, no proprietary library was used to build the simulator, resulting in a totally free software available online.

Knowledgebase

The knowledge base is composed of three main sources: 1) The Nursing Interventions Classification (NIC); 2) Experts of notorious knowledge and teachers of the discipline of Nursing Management and; 3) Resolution on the scaling of nursing staff addressed in the classroom. From the NIC taxonomy, the researchers mined a complete set of nursing activities, described through tables, and the average times related to each activity. With this information, it was possible to simulate how long each activity is performed. Other important information gained from the NIC, but with the support of specialists in the nursing area, is the level of knowledge of the professional needed to perform a given activity, whether a graduate or a nursing technician. Professors of the Nursing Management subject developed a set of clinical cases to challenge the student to answer questions during the simulation. These clinical cases were composed of a textual description containing clinical data of the patient allowing the definition of the complexity of the nursing care. The guiding legislation of nursing staffing in Brazil provided a basis for the construction of rules, such as the maximum time that a professional can work continuously and the minimum time of rest, between continuous workdays.

Models and architecture

The multiagent architecture of the simulator is composed of three types of agents: patients, nurses (bachelor degree) and technician nurses (secondary degree), who exhibit autonomous behavior and interact with each other [8]. These agents exhibit two fundamental characteristics: they are able to act autonomously by making decisions that lead to their goals and they are capable of interacting with other agents using human-inspired social interaction protocols and including coordination and co-operation functionalities.

Educational background

Many studies have confirmed the effectiveness of simulation in health teaching as well as in the assessment at the undergraduate and graduate health education levels [9]. Early studies showed the importance of technology in this area [10, 11]. The

computer simulator allows students to experience the work environment of an inpatient unit, including the calculation of the nursing staff, the analysis of the degree of inpatient complexity, clinical reasoning and decision making in nursing interventions. This is the main focus of the whole simulator: to provide a useful virtual environment for students and teachers to experience the day-to-day running of an inpatient unit.

Evaluation

The evaluation was carried out in a classroom with the participation of eleven students enrolled in the subject "Nursing Management" in the third year of the Nursing Bachelor. In this subject, the student learns how to manage a team of nursing professionals, how to deal with Law restrictions and other factors, such as the recommendations of nursing regulatory bodies. In the classroom, after a complete explanation of the use of the simulator, conducted by the research team, each student was connected to the simulator to perform all the tasks required by the virtual environment. After the tests, the students were invited to answer an anonymous, non-compulsory questionnaire. All students answered voluntarily. The questions were based on ten golden rules for the evaluation of health software [12]. The teacher was present during the experiment to inspect the researchers, without influence, control, acknowledgment or coercion on the students' (anonymous) responses.

Methods

A preliminary study was carried out to obtain the necessary data for the construction of the simulated hospitalization unit. A medium-size university-related hospital was used as a model. The adult clinical admission unit has, on average, 20 beds. On a typical day, occupancy is 80%. The work shift is 6 hours, so each day is divided into four shifts (morning, afternoon, early night and late night). Inpatient admission and discharge occur at the end of the shift. The simulator, in its first stage, does not allow the admission or discharge of patients, except in a simulated death, so each simulation process occurs in a period of six hours. Each patient presents a level of care complexity, being able to be classified in: minimum care, intermediate care, high dependence, semi-intensive care and intensive care. According to this classification, Brazilian legislation fixes how long it takes (in mean hours) of nursing care. To classify the complexity of the patient, the nurse performs a clinical evaluation based on predefined criteria presented in a table and the result is obtained from the score of these tables. In the initialization process, the simulator generates a random count of inpatients (up to 20), as well as the level of complexity of each patient. In a preliminary survey, it was concluded that it would be tiring and exhausting for the student interacting with too many simulated patients at the same time, so the strategy adopted was to restrict the student's access to direct care of only three simulated patients. These three patients will require student attention throughout the simulation, requesting medications, questioning and reporting his needs. However, for the management of people, the simulator considers an inpatient unit with a number of patients compatible with reality.

Agents

The "patient" agent is one of the two possible types: autonomous or programmed. The autonomous agent works in response to randomized events from the knowledge base (NIC and regulations) and based on the patient's level of complexity. For example, a patient with a minimal level of care requires random NIC activities of up to the maximum expected by their level of complexity. The scale for the arrangement of the complexity of the Patient used is defined by COFEN 543/2017 [7]. The other type of "patient" agent is the one programmed by the professor of the Nursing Management subject, which defines the actions required by the patient. These actions will cause the simulator to trigger a challenge to the student involving decision making regarding a nursing intervention. For example: the patient feels short of breath, what to do? The "nurse (bachelor)" and "nurse (secondary)" agents respond to patients' calls and perform nursing care. The main objective of these agents is to measure the workload of nursing professionals (nurses and technicians). When a nursing agent is assigned to care, he can't be allocated to another job until his work is done. Therefore, the complexity of care and the number of nursing actions to be performed in each patient indicate the number of nursing personnel and the distribution of these professionals among the patients in an inpatient unit. If the demand for care is too high, the available nursing staff will have to deal with the increased stress as well as with the impairment in quality of service. Figure 1 exemplifies the simulated "nurse" agent state machine. The network nodes are composed of rules that are used to define whether the professional is available or is interacting with another agent (table 1).

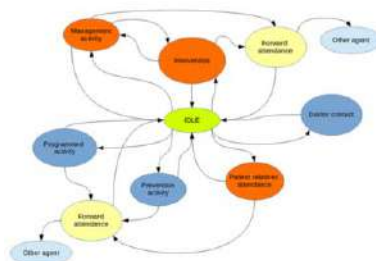


Figure 1 – Nurse state machine

To allow for multiple simulations during a teaching class, the simulated time has been accelerated to one-tenth of the actual time, it means that each simulated minute takes only six seconds to go, so each simulated session will last six minutes.

Table 1 – Agents

Name	Base	Interacts with
Patient	The specialist data; Teacher´s course	Nurses; Physicians; Students;
Nurse	Knowledge base; The specialist data; NIC; Laws and local rules;	Patients; Other nurses; Physicians;
Physician	The specialist data;	Patient; Nurses;

Communication between agents

Since the agents must interact with each other, a communication protocol was created to perform this function, this protocol was based on a data bus that runs on a thread accessible by instances of all agents, once an agent was created, it presents itself to the bus by registering in the main event manager. It is possible, over this bus, to interact with individual agents, one to one, one to many and many to many, allowing to simulate a wide range of communication types found in a hospital (phone call, internal sound system and personal conversation). This protocol allows another type of communication that is not based on messages, but based on events, since a patient can trigger some event like a heart rate monitor alarm and this event is directed to the nearest nurse, as a situation common in a hospital.

Figure 2 illustrates the communication design and the actors involved.

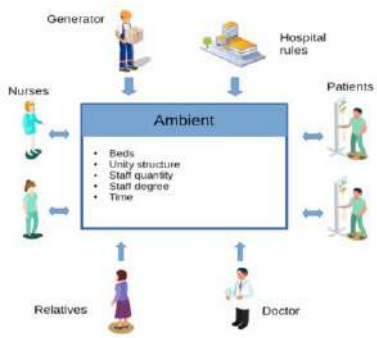


Figure 2 – Communication between agents

Ethics

This research was approved by the Research Ethics Committee under the title "Hospital Environment Simulator for Personnel Management and Dimensioning", CAAE: 57514916.3.0000.5345. Approved on: 08/18/2016. The project includes the classroom test, the use of the questionnaire and the participation of the specialists. All participating students signed a consent term and received full attendance during all phases. The participation was voluntary, and any refusal would not result in the reduction of grades at any level. Throughout the evaluation, students were free to leave participation, without penalties, however, no student left the activity. All images and artwork used in the simulator have been licensed by the corresponding author for use in this software and all software libraries and components are freely distributable.

Results

The results can be described in two different steps: the construction of the simulator and the validation of usability in the classroom.

Simulator

The first step, related to the construction of the simulator, resulted in the development of a discrete software, without complex dependencies, that can be executed with minimum efforts. After login (Figure 3), the student has access to the main screen (figure 4), where the simulation console is located. This screen is composed of inpatient controls, nurses and technicians, time controls, charts and a visual representation of the ward with beds, patients, physicians, nurses, technicians and family members. At this screen, the student has the time and complexity control of the simulation. The main console shows three graphs in real time while the simulation is running. The top chart represents the total of hospitalizations waiting for care. In case the team is dealing well with all internal calls, this graph may be empty otherwise the graph will show an increasing area meaning that the patients are waiting for attendance. The middle graph shows the hourly call amount, a nonlinear plot means a high entropy in the algorithm and shows a random set of calls during the simulation. The graph below represents the average time of care, the area of the chart is the total time of care applied to inpatients.

Figure 3 – Login screen

Although the simulator algorithm calculates the nursing team based on laws and constraints, it is possible to change the variables related to patient, nurses and technicians count through the sliders. By changing these controls, the student can suppress the current rules and try scenarios different from those expected by applying current regulations. At the bottom of the window, there are two more sliders, the top represents the time factor of care and the second is the complexity of care factor, these two sliders allow the student to experience a more flexible simulation.

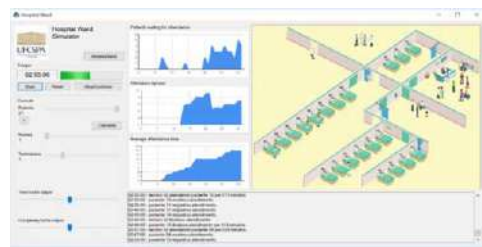


Figure 4 – Main simulation screen

In Figure 5, the student is challenged to evaluate the patient complexity, which results in a score used to classify the inpatient care, based on the recommendations of the COFEN [7]. The correct classification will contribute to the efficiency of care planning, allowing the nurse to distribute the amount of nursing personnel, according to the complexity of the care analyzed. An incorrect classification may cause stress to staff due to lack of people (inpatients labeled as below complexity that really it is), or may increase overall costs, should it require more people than necessary (inpatients labeled as above complexity that really it is).

Eventually, at random times, the student is challenged to solve situations to promote the development of clinical thinking and decision-making skills. In Figure 6, the student interacts with the inpatient due to a call followed by a problem description. Once answered, the simulator will report a feedback to the student. Based on the student's decision, the simulator can change the clinical picture of the hospitalized patient, which may be a better or a worse state

Figure 5 – Inpatient classification

Figure 6 – Inpatient attendance

Evaluation results

As a result of the classroom assessment, a table was compiled to illustrate the student's perception of the tool. The following table (table 2) represents the answers collected from the students, after the class where the simulator was used as a learning tool. The questionnaire was divided into 12 questions based on the ten golden rules for the evaluation of health software [12]. The questions are based on the Lickert style where "1" means "very bad" and "5" means "very good".

Table 2 – Questionnaire results

Question	Mean	SD
1) Was this exercise developed focused on the student improvement?	5	0
2) This type of exercise does stimulate student interactivity with technological tools to solve proposed problems?	5	0,302
3) How about the ability to motivate and focus on solving the problem proposed?	5	0,405
4) At some point during the development of the exercise did you feel pressured and forced to move forward without understanding the problem in its entirety?	0% Yes 90% No 10% Partially	
5) The exercise you have done, regarding your ability to generate autonomy in your learning, so that you can conduct your studies at your own pace?	5	0,302
6) The exercise, regarding the relevance and contextualization of the contents approached for your learning?	5	0,405
7) Do you believe that the exercise performed provides scientific depth on the subject addressed?	100% Yes 0% No 0% Partially	
8) As to the visual aspect of the exercise and how the content was presented, you judge the exercise:	5	0
9) How do you evaluate the ease of use of the exercise interface?	5	0
10) How do you evaluate the attractiveness of the exercise interface?	5	0
11) How do you evaluate the navigability of the exercise, i.e. the ease of finding the information and directions of the next steps?	5	0,522
12) Did you find the exercise too long?	0% Yes 100% No	

Discussion

The adoption of an agent-independent approach has brought important flexibility in the implementation of the simulator, first because it is appropriate to the context of the simulation of environments [4, 8] and second, because it is a simple implementation using parallel computing concepts [3]. In an empirical analysis, it was observed that the behavior of the agents in an inpatient unit would be analogous to the behavior of independent computational agents, and this was confirmed with the bibliographic studies found in [13, 14]. Another important factor for the realization of the project was the decision to implement the communication bus in three models: direct (p2p), group (multicast) and overall (broadcast) [15], making possible the simulation of three forms of communication between agents actors): direct communication (a telephone call), group communication, or overall communication (internal sound system) making compatible, in empirical analysis, with the behavior observed in a hospitalization unit of a medium/large hospital institution. The concept adopted to measure the states of a patient in discrete variables made viable for the use of state machines [3] to simplify the patient's assisted momentum, and the control of the machine states was filled by the professional who has the necessary knowledge, enabling the state transitions according to the information provided. This feature brought a special gain to the simulator because it allows interaction with the observer and the system, seeing not only simple passive analysis but an active approach, placing the user inside the unit, acting and interfering with the actors and the environment. The analysis of the results provided important reflections to the members of the study group that developed this work. All students considered that the exercise was focused on student growth, however, there was no unanimity about the perception that the simulated exercise stimulates the ability to solve the proposed problem (people management). Many hours of work were invested in the construction of the interface of the simulator, and this reflected in the perception of unanimous satisfaction about the ease of use and attractiveness of the simulator, but it is possible to identify a probable bias of confusion because the students do not have many references of simulators of this type for purposes of comparison, so this parameter, in spite of the absolute numbers in the answers, should depart from certainty and accuracy.

In general, it is possible to observe the major satisfaction with the tool, perhaps because the method provides an interactive way of experiencing the hospital ward and viewing staff calculations in a visually appealing manner that was denoted as "much better than numbers in the paper". Another important factor was hospital manipulation through simulation, based on the clinical case provided by the specialist (the subject professor in this case), students not only were able to choose a nursing care action, but also see the results of their decisions in dynamic show. On the other hand, apart from the educational environment, the divergence of the values recommended by the law and the reality found in the professional environments is evident. This divergence is related to a very important gap in the actual workload per nurse in a real hospital and the workload described in the regulations. First, the research group had been concerned about some misunderstanding or any other research bias, but after several analyzes in one unit and reference admission, the concern was resolved and this divergence between real and simulated resulted in a new study on this gap. This part is related to local (Brazilian) laws, but after some research [16] it was possible to determine that this problem affects other places also causing stress, low-quality care and turnover of nurses. Although this finding is very important for the quality and overall satisfaction of the workers, this was not the main objective of this study, but part of another study from the same research group at the same university.

Conclusions

Although the lack of formal validation, it is believed that simulators may play an important role in learning objects. Considering the increasing presence and necessity of virtual learning environments and the lag that traditional methods of evaluation have in such scenarios, more studies are needed to consolidate these tools as alternatives in health education. Simulators may play a role beyond the training environment. By explaining to the student the underlying mechanisms and the motivations that led him to make a decision, it can produce a deeper evolution in this same student, in this sense we believe in the importance of feedback during the execution of the simulation. The simulator may awake in the student the recognition of thought patterns and attitudes, motivating him to question established practices or acquire new skills and strategies for making decisions. We anticipate this, for example, in the still unexplored possibility of comparing log records and their respective motivators to the same student in different scenarios. Over time, the evolutionary analysis of these decision-making patterns could constitute a subsidy for conclusions about the entire process of student formation. We believe that it is necessary to carry out a validation of the method as an alternative to the formal evaluation, which can be obtained by comparing the traditional method (written tests and/or practices formulated by the teacher) and the simulator. It seems to us that an ideal methodology to be applied will consist of homogeneous classes, which receive the content in a usual way and then submitted to the traditional form of evaluation and to the simulator. The group will continue this form of work and intend in the next instance to correct the flaws pointed out in this research and then proceed to a new workshop for didactic-pedagogical validation of the simulator.

Acknowledgments

Acknowledgments: This work was partially supported by the Brazilian National Council for Scientific and Technological Development (CNPq), edital MCTI/CNPq/Universal 14/2014, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), edital Pró-Ensino na Saúde nº 024/2010, and the Portuguese Funding from FCT - Fundação para a Ciência e a Tecnologia, under the project: UID/MAT/04561/2013.

References

[1] B. Marquis and C. Huston, "Tomada de decisão, solução de problemas, raciocínio crítico e raciocínio clínico: requisi-tos para uma liderança e administração de sucesso," in *Administração e Liderança em Enfermagem: teoria e prática.*, 8 ed., Porto Alegre, Artmed, 2015.

[2] P. D. Wright and S. Mahar, "Centralized nurse scheduling to simultaneously improve schedule cost and nurse satisfaction," *Omega,* vol. 41, no. 6, pp. 1042-1052, 1 12 2013.

[3] S. Russel and P. Norvig, Artificial Intelligence, 3rd ed., Elsevier, Ed., 2013.

[4] F. D. L. Prieta, S. Rodrígues, J. Morchado and J. Bajo, "Infrastructure to Simulate Intelligent Agents," *IOS Press,* 2015.

[5] M. Wooldridge, An introduction to multiagent systems, John Wiley & Sons, 2009.

[6] T. H. Herdman, S. Kamitsuru and North American Nursing Diagnosis Association., NANDA International, Inc. Nursing diagnoses : definitions and classification : 2015-2017, Wiley-Blackwell, 2014, p. 483.

[7] "Resolução 543 de 18 de abril de 2017 - Atualiza e estabelece parâmetros para o Dimensionamento do Quadro de Profissionais de Enfer-magem nos serviços/locais em que são realizadas ativida-des de enfermagem," Brasília, 2017.

[8] G. Weiss, Multiagent systems, p. 867.

[9] Y. Okuda, E. O. Bryson, S. DeMaria, L. Jacobson, J. Quinones, B. Shen and A. I. Levine, "The Utility of Simulation in Medical Education: What Is the Evidence?," *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine,* vol. 76, no. 4, pp. 330-343, 1 8 2009.

[10] K. A. Hayes and C. U. Lehmann, "The interactive patient: a multimedia interactive educational tool on the World Wide Web.," *M.D. computing : computers in medical practice,* vol. 13, no. 4, pp. 330-4, 1996.

[11] A. Holzinger, M. D. Kickmeier-Rust, S. Wassertheurer and M. Hessinger, "Learning performance with interactive simulations in medical education: Lessons learned from results of learning complex physiological models with the HAEMOdynamics SIMulator," *Computers & Education,* vol. 52, no. 2, pp. 292-301, 2 2009.

[12] V. Jha and S. Duffy, "'Ten golden rules' for designing software in medical education: results from a formative evaluation of DIALOG," *Medical Teacher,* vol. 24, no. 4, pp. 417-421, 3 1 2002.

[13] J. A. Juanes and P. Ruisoto, "Computer Applications in Health Science Education," *Journal of Medical Systems,* vol. 39, no. 9, p. 97, 8 9 2015.

[14] L. Panait and S. Luke, "Cooperative Multi-Agent Learning: The State of the Art," *Autonomous Agents and Multi-Agent Systems,* vol. 11, no. 3, pp. 387-434, 11 2005.

[15] Andrew S. Tanenbaum., Computer Networks., Pearson India, 2013.

[16] E. H. Lee, S. J. Chang, H. O. Kim, J. Roh, E. J. Park and J. U. Won, "Relationship between Job Stress and Turnover of Registered Nurses in a University Hospital," *Korean J Occup Environ Med,* vol. 19, no. 2, pp. 93-104, 2007.

Address for correspondence

Gustavo Henrique Cervi

gustavohc@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190282

When Ants Take Care of Humans: ACO for Home-Care Services Planning Optimization

Yahia Chabane, Christophe Bortolaso and Mustapha Derras

Berger-Levrault, France

Abstract

Current planning approaches for home care services do not generally support the social and human dimension of planning. They focus on optimization criteria that are easily quantifiable, such as the cost. Whereas other criteria such as the quality of the relationship between caregivers and beneficiaries or the satisfaction of the latter are important too as they can highly influence the planning. To address this issue, we investigate in this work the problem of planning optimization for home care service. We propose an extension of the classical ant colony optimization algorithms. Optimization is carried out by several classic criteria such as the cost along with social-based criteria such as the relationship between caregivers and beneficiaries. We also propose a flexible and expressive language to represent the constraints in the form of predicates that can include variables, constants and functions of the problem. This allows each organisation to add its own constraints.

Keywords:

Home Care Services (HCS), Constraint Satisfaction Problem (CSP), Ant Colony Optimization (ACO)

Introduction

In France, according to the National Institute of Statistics and Economic Studies, one in three inhabitants would be 60 years or older in 2050. A lot of these people are in a situation of loss of autonomy and need to receive medical or social care. However, the specialized organisations lack of available places. Consequently, Home Care Services (HCS) are more and more emerging and several types of organisations were created recently. These HCS organisations reduce the overload and indebtedness of hospitals and other institutions and provide a better environment for beneficiaries.

The management of these organisations and more particularly the caregivers interventions planning is complicated and is still largely done manually. A large number of constraints must be considered when producing the plan such as the satisfaction of caregivers, of beneficiaries, skill requirements, working time regulations, variable duration of acts, etc. The production of a plan becomes a big challenge, as we often have to deal with hazards such as the integration of new beneficiaries, lack of staff, and various unanticiped changes.

Automatic plan generation constitutes a great solution to this problem and several approaches are proposed in the literature [10,15,18]. An automatic planning system must be flexible to deal with the hazards of planning, robust to satisfy the various constraints and be executed quickly to deal with emergencies, changes, or event crisis situations.

Current planning generation approaches generally address easily quantifiable criteria such as the cost, the travel distance and the number of caregivers, etc. We advocate that social and human dimensions of the care acts have a strong influence on the planning and must be considered. The satisfaction of the beneficiaries and their relationship with the caregivers are, for instance, important criteria to take into account while planing the caregivers interventions. In addition, constraints addressed traditionnally do not cover all needs of the HCS organisations.

To address this issue, we explore the problem of planning optimization of HCS and we study the potential of Ant Colony Optimization (ACO) algorithms to solve these problems. We chose to represent the problem as a Constraint Satisfaction Problem (CSP). Indeed, it allows us to reuse many tools, approaches and theoretical results about resolution and computational complexity of the problem instances. Moreover, CSP is well adapted to manage the constraints separately from the problem resolution which allows to easily customize or add social and human constraints.

Organisation of the paper. Preliminaries section recalls some basic notions regarding HCS planning problem, CSP and ACO. ACO for HCS optimization section presents our optimization approach, while Experimental Results section exposes the system implementation, experimental results and feedbacks from a qualitative study. Related Work section discusses related works. We conclude and draw future work in Conclusion and Future Work section.

Preliminaries

We introduce and recall some basic notions regarding HCS planning problem, CSP and ACO in this section.

Home Care Services Planning Problem

Home Care Services (HCS) are paid care services, proposed to people in their home. A beneficiary is a person, who is subscribed to a home care service. A caregiver is a competent person that carries out the set of care acts. An intervention is a set of acts performed at a time t by a caregiver c for a beneficiary b. A planning constraint is a condition for carrying out one or more interventions.

The HCS planning problem consists in the definition of a set of interventions to meet the demands of a set of beneficiaries taking into consideration the optimization of the satisfaction number of planning constraint.

Constraint Satisfaction Problem (CSP)

A CSP [19,22] is well adapted to represent several real problems with constraints. A CSP is defined by a triplet (X, D, C) where:

- $X = \{X_1, X_2, \ldots, X_N\}$ is a set of variables of the problem.
- D is a domain function that associates to each X_i its domain $D(X_i)$ (possible values of X_i).
- $C = \{C_1, C_2, \ldots, C_M\}$ is a set of constraints of the problem.

The associated optimization problem consists in the assignment of the variables and in the optimization/satisfaction of a set of constraints or/and an objective function. It is also called constrained optimization problem (COP).

Ant Colony Optimization (ACO)

ACO [7,9] is a meta-heuristic for solving combinatorial optimization problems. It is inspired by the behavior of real ants when foraging. The first ant colony algorithm was proposed by Dorigo [6] to solve the travelling salesman problem which consists of finding the shortest route that visits each city of a given set of cities. The basic principle behind ACO is to produce a collective intelligence from the interaction of individual behaviors to solve a complex problem. Indeed, the behavior of a single ant is not complex, but the result of the collaboration of multiple ants results in the emergence of a complex collective behavior.

Ants foraging consists in finding a source of food and the shortest path between this source and the anthill. Ants converge progressively to the shortest path through an indirect communication mechanism called stigmergy, achieved by the modification of the environment. This mechanism is realized by the deposit of a volatile hormone (i.e. pheromone) on the path taken by the ant. This pheromone has a direct impact on the behavior of the following ant, as they are attracted by the pheromone and more likely to move towards it. This results in ants following the same path as the first ant, but the decision remains stochastic, since other ants take other paths. An important characteristic of the pheromone lies in its ability to evaporate quickly. This makes the less traveled paths disappear (often the longest), and increases the amount of pheromone deposited on the shortest paths, because more and more ants take these paths. This phenomenon allows a short path to emerge (not necessarily the shortest possible path) that almost all ants follow even if some ants take other paths.

It is important for this type of approach to find a good balance between intensification (i.e. exploitation of collected information) and diversification (i.e. exploration of the research space). For ACO, intensification is realized by the deposit of pheromone and diversification through stochastic decision making of ants and the evaporation of pheromone.

Related Work

HCS planning problem is a problem with an exponential number of candidate solutions. Each candidate solution is evaluated, the goal is to find the best rated candidate solution. Intuitively, the resolution consists in listing all the candidate solutions and take the best rated one, but this is not possible in a reasonable time. The complexity [17] of this type of problem is in general *NP-complete* or *NP-hard*.

To design a HCS planning system, we need to answer three key questions: (i) What are the planning constraints ? (ii) How to evaluate a plan ? and (iii) how find the best plan ?

These questions are highly linked to each other, they represents what constitutes a HCS planning system. We investigate a brief overview of the possible answers to the above mentioned questions. The first question concerns the planning constraints of a HCS planning problem, the second question concerns the evaluation criteria of a plan and the used evaluatio function, and the third question concerns the used resolution approaches:

i. HCS planning constraints: they are conditions related to the validity or the quality of the plan. Some conditions must be verified, and others are related to the quality of the plan. The most used constraints in the literature are [5,10,15]: skill requirements, sectors, temporal dependency, time windows, continuity of care, workload balancing, breaks, etc.

ii. Evaluation function: an important factor in HCS planning system is the evaluation function. This function defines the criteria to optimize when generating a plan. Several criteria are used in the literature as [10,15,18]: number of caregivers, travel time, travel time, waiting time, preference, constraint violations, etc.

iii. Resolution approaches: Solving a HCS planning problem consists in finding the best solution optimizing one or more given evaluation criteria from an exponential set of candidate solutions. This set of candidate solutions is called the problem research space. Resolution approaches can be classified according to Completeness and Correctness [18,22]: a correct and complete approaches, a correct and incomplete approaches and an incorrect approaches. The approaches used are generally correct and incomplete, we can cite [5,10,15,18]: tabu search, genetic algorithm, greedy search, local search, adaptive large neighborhood search, etc.

As mentioned before, current planning approaches do not generally support the social and human dimension of planning, while the relationship and the satisfaction of the beneficiaries and the caregivers are very important. In addition, the constraints are very general and do not cover all needs of the HCS organisation.

ACO for HCS optimization

Our choice to represent the problem as a CSP enables us to separate the problem from the application field in order to better manage the representation of constraints and to simplify the adaptation of the solution to other applications. This information structure allow use to easily redefine constraints when required, to adapt the algorithm to different practices in different home-care organisations. About the resolution algorithm, ACO, has been used to solve several optimization problems [8,9,24], and also been successfully used to solve CSPs [13,21,22], which strengthens our choice. This naturally appeared as a relevant solution to our problem.

We present, next, our representation of the problem and BL-ANT-Planning, the proposed resolution mechanism.

Problem Representation

In the following, the representation of our problem in the form of a CSP. We have defined the variables of the problem, the function that associates each variable with its domain and the constraints of the problem.

Variables

We use a matrix $Assign[][]$ and an array $Planify[]$ to represent the variables of the problem, their size are $N \times M$ and N respectively, where N is the number of interventions and M is the maximum number of caregivers needed to perform an intervention.

- $Assign[1 \ldots N][1 \ldots M]$: it is the assignments of the interventions to the caregivers. $Assign[i][j]$ is the caregiver j who will perform the intervention i, such that $0 \leq i < N$ and $0 \leq j < M$.

- $Planify[1 \ldots N]$: it is the plan time of an intervention. $Planify[i]$ is the time slot identifier of the intervention i, such that $0 \leq i < N$. This identifier can be converted to a schedule.

Variables Domains

Let N be the number of interventions and let M be the maximum number of caregivers needed to perform an intervention, the function D that return the possible values of $Assign[][]$ and $Planify[]$ is defined by:

- $D(Assign[i][j]) = \{0, 1, 2, \ldots, K\}$ such that K is the number of caregivers and $0 \leq i < N$ and $0 \leq j < M$.
- $D(Planify[i]) = \{0, 1, 2, \ldots, T\}$ such that T is the number of time slots and $0 \leq i < N$.

Constraints

Let N be the number of interventions and let M be the maximum number of caregivers needed to perform an intervention. We define an expressive language to represent the constraints in the form of predicates that can include variables, constants and functions of the problem. For example:

- $\forall i \in [1; N], \forall j, k \in [1; M], if\ j \neq k$ $then\ Assign[i][j] \neq Assign[i][k]\ or\ Assign[i][j] = 0$ which means that an intervention cannot be assigned more than once to a caregiver.
- $\forall i \in [1; N], \forall j \in [1; M]$ $then\ Qualif\ (i) \subset Qualif\ (Assign\ [i]\ [j])$ where $Qualif\ (i)$ is a function that returns the skills of a caregiver (or necessary skills of acts), which means that a caregiver must have all necessary skills to perform care acts.

Based on that constraint representation, our test involve more than 17 constraints.

BL-ANT-Planning

Algorithm 1 describes the general principle of the approach. A first step (lines 2 and 3) for initializing the algorithm parameters and the pheromone. Then for each cycle (line 6), the ants generate solutions (one solution per ant) and each solution is improved by a local search (line 7). The solutions are then evaluated to update the best solution of the algorithm and the best solutions of the cycle (line 8). At the end of a cycle (line 9), the ants that have generated the best solutions of the cycle depose a pheromone. Finally, the best generated solution is then returned (line 10).

Initialization

In this step, the parameters of the algorithm are initialized and the pheromone is deposited on the possible solutions. Parameters include the number of cycles, the number of ants to use per cycle, the pheromone factor weight and the heuristic factor weight.

Algorithm 1: BL-ANT-Planning

```
   Input: problem ; the input data of the problem
   Output: bestSolution ; the best generated solution
1  begin
2      parameterInitialization(problem);
3      pheromoneInitialization(problem);
4      for i ← 0 to numberOfCycles do
5          for j ← 0 to numberOfAnts do
6              solution ← buildSolution(problem);
7              solution ← localSearch(problem, solution);
8              updateBestsSolution(solution, bestSolution, bestSolutionsOfCycle);
9          pheromoneUpdate(problem, bestSolutionsOfCycle);
10     return bestSolution;
```

Other parameters related to the pheromone are also initialized, the minimum pheromone factor and the maximum pheromone factor that can be found on a solution and the pheromone evaporation rate. Initially, a quantity equal to the maximum pheromone factor is deposited on all solutions.

Solution Generation

To generate a solution, the ant generates an assignment of the variables of the problem satisfying the constraints of the problem. What it generated is ignored if an assignment violates a constraint. The assignment is generated by assigning the variables one by one until there are no variables to assign. The order of the variable assignments is important, an assignment of a variable can restrict the domain of another variable or even make it empty.

1. *Variable selection:* the variable selection technique impacts the performance of the algorithm and the quality of the solution. Several techniques are proposed in the literature [3,19,23]. For performance reasons, we use for now a random selection of variables, where each variable has the same probability of selection. We plan to explore other selection techniques in future works.
2. *Value selection:* the value of the variable is selected from the subdomain that is compatible with the constraints of the problem. The value selection is stochastic and is based on the pheromone factor, deposited between the variable and its possible values, and on the heuristic of assignment evaluation. The probability of selecting the value v_i for the variable V_j is equal to: $p(v_i \rightarrow V_j) = \frac{[\tau(v_i \rightarrow V_j)]^{\alpha} \times [\eta(v_i \rightarrow V_j)]^{\beta}}{\sum_{v_k \in D_{V_j}} [\tau(v_i \rightarrow V_j)]^{\alpha} \times [\eta(v_i \rightarrow V_j)]^{\beta}}$

Where $\tau(v_i \rightarrow V_j)$ is the pheromone factor between v_i and V_j, $\eta(v_i \rightarrow V_j)$ is the heuristic factor of the assignment $v_i \rightarrow V_j$, α is the pheromone factor weight, β is the heuristic factor weight and D_{V_j} is the subdomain of V_j which is compatible with the constraints of the problem.

Improvement Solution

Local search is used to improve the solutions. It is performed before the solution evaluation in order to improve the ant generated solution. Caregivers satisfaction criterion is improved in this step by trying to eliminate unnecessary breaks.

Solution Evaluation

The solutions are evaluated in this step to keep the best solution of the algorithm and the best solutions of the cycle. Several criteria are used to evaluate a solution, including social and human dimension criteria. For example, the cost, the quality of the relationship between the caregivers and beneficiaries, the respect of the geographical area, the satisfaction of caregivers and beneficiaries, etc. Each criterion is evaluated on a scale of seven levels [1,16]. Then, an average of all criteria is calculated according to the weight of each criterion, this average will represent the evaluation of the solution.

Table 1. *List of identified planning constraints and optimization criteria.*

Related to	Planning constraint	Optimization criteria
Beneficiary	Time constraints and absences, state of the beneficiary and his entourage	Time preferences, regularity of the caregivers and plans
Caregiver	Qualification/skills, respect of employment contracts, lunch break, absences	Optimization of employment contracts, time preferences
Intervention	Synchronization with external services, caregiver/beneficiary incompatibility, replacement degree, equipment, travel time, mutual plans, implementation interval, intervention difficulty	Intervention priority, hardness
Organisation	Travel time, legislation, work rate, schedules of external structures	Compactness, costs, caregiver contract, sectors, substitute preferences

Pheromone update

At the end of a cycle, the pheromone factor is updated. A quantity evaporates and another quantity is deposited on the best solutions of the cycle. After each update, all the pheromone factors must be between the predefined minimum pheromone factor and maximum pheromone factor.

The evaporation of the pheromone is simulated at the end of each cycle by the multiplication of the pheromone factor by the pheromone evaporation rate. The pheromone deposit is not carried out as in nature, it is delayed until the end of the cycle and only performed on the best solutions of the cycle. The quantity of pheromone deposited depends on the quality of the solution. Pheromone deposition is only performed for the best solutions of the cycle unlike natural ants that deposit pheromone in all cases. We think this will help a faster emergence of solutions, we also plan to test other strategies.

Implementation

Our solution is implemented as a black-box service which takes as input a file formatted as JSON, specifying the list of caregivers, beneficiaries and the constraints to consider. The solution outputs a set of calendar files in iCal format. Each iCal file represents the plan of one caregiver.

The architecture of the application is modular, the definitions of the problem, of the constraints and of the resolution algorithm are decoupled. The software architecture make it also easy to integrate another resolution algorithm (i.e. other than ACO, such as genetic algorithm for instance), we can also change the representation language of constraints to cover a more expressive or less expressive language for better performance. This allows us to easily adapt our implementation to other applications and other problems.

Results

We test our approach in a twofold perspective. First, we did a quantitative study, focusing solely on algorithm performances. We then organised workshops with some HCS organisations to qualitatively test the algorithm on real cases. This enabled us to customise the solution evaluation function in order to improve the selection of solutions. This also allowed us to verify the accuracy and the usability of the proposed approach with practitioners.

Quantitative study

We started by studying for real case of planning, the influence of the algorithm parameters on the quality of the generated plans and the research quality. The quality of the generated plans is evaluated on the optimization criteria that we present in the following section and the research quality is measured by [4,11,20]: (i) a similarity ratio, in order to know if the collected information is well exploited and (ii) a resampling ratio, used to know if the research space is well explored by the algorithm. The best algorithm parameters found is (4, 2, 0.05, 0.1, 10) represents respectively the values of: 1) pheromone factor weight, 2) heuristic factor weight, 3) pheromone evaporation rate, 4) minimum pheromone factor and 5) maximum pheromone factor.

Regarding performance, we ran the algorithm on a machine with a 2.7Ghz i7-7500U processor and 16 GB of RAM. For an organisation with 268 beneficiaries, 37 caregivers, and 142 interventions to plan, the execution lasted 2 minutes for 10000 cycles with 15 ants by cycle.

We have a promising first results. They, nevertheless need to be consolidated and compared with other approaches. We have identified some benchmarks [2,12,14] for this comparison.

Qualitative study

We worked with five HCS organisations, initially to determine the planning constraints and optimization criteria they use. These HCS organisations are in different cities and are different size. The number of beneficiaries managed varies from 200 to 700 with 30 to 100 caregivers.

Table 1 lists the planning constraints and optimization criteria used by the five HCS organisations. Note that, a slight difference between the supported constraints list and the importance of each optimization criteria of the HCS organisations, this difference is related to the localization and the size of the HCS organisations.

We then tested the algorithm on the data of these organisations and we compared the generated plans with the plans they generate manually.

Table 2 presents a result of a comparison for an organisation with 268 beneficiaries, 37 caregivers, and 142 interventions to plan. This comparison is based on the number of caregivers needed for planning, the duration of all breaks and the regularity of interventions, such that an intervention is regular if the caregiver provided a car for the beneficiary at least twice.

Table 2. *Comparison of generated plans with organisations plans*

	Organisations plans	Result at 2min	Result at 30min	Result at 60min
Number of caregivers	21	19	16	15
Duration of all breaks	3120	3560	2660	2240
Number of regular intervention	140	142	142	142

The first results are promising, we improve the three criteria after 100000 cycles (30 min) and two criteria after 10000 cycles (2 min). Note that the number of regular interventions converges

quickly, we cover 100% of interventions directly after 10000 cycles. These results remain to be confirmed on benchmarks.

Our results indicates that our solution generates plannings which are alsmost as good as the one done manually after only 2 min of computation with a standard computer. Furthermore, it also shows that after 60 mins of computation it provides solutions involving potentially 28% less caregivers. This could results in better service allowing caregivers to spend more time with beneficiaries and/or cheaper home-care service.

Conclusion and Future Work

In this paper, we proposed and implemented a resolution algorithm for the planning problem in the context of Home Care Services, based on Ant Colony Optimization. We used a Constraint Satisfaction Problem to represent the problem. The implemented application is structured into modules to facilitate its extension and adaptation to other problems. Since the algorithm is non-deterministic with a stochastic solution construction, the correct parameterization that is at the core of the mechanism needs to be further validated.

The next objective of our project is to compare the performance of our approach with other approaches on benchmarks. We plan also to study the impact of different parameters and try other pheromone deposition techniques or variable selection strategy. We will also look at what is done on CSPs in terms of filtering variables domains, representation and constraints management.

References

[1] Duane F. Alwin and Jon A. Krosnick. 1991. The Reliability of Survey Attitude Measurement: The Influence of Question and Respondent Attributes. *Sociol. Methods Res.* (1991).

[2] Kris Braekers, Richard F. Hartl, Sophie N. Parragh, and Fabien Tricoire. 2016. A bi-objective home care scheduling problem: Analyzing the trade-off between costs and client inconvenience. *Eur. J. Oper. Res.* (2016).

[3] Sally C Brailsford, Chris N Potts, and Barbara M Smith. 1999. Constraint satisfaction problems: Algorithms and applications. *Eur. J. Oper. Res.* 119, 3 (1999), 557–581.

[4] E. Burke, S. Gustafson, and G. Kendall. 2002. A Survey And Analysis Of Diversity Measures In Genetic Programming. *GECCO '02 Proc. Genet. Evol. Comput. Conf.* (2002).

[5] Mohamed Cissé, Semih Yalçındağ, Yannick Kergosien, Evren Şahin, Christophe Lenté, and Andrea Matta. 2017. OR problems related to Home Health Care: A review of relevant routing and scheduling problems. *Operations Research for Health Care*.

[6] Marco Dorigo. 1992. Optimization, Learning and Natural Algorithms. *Ph.D. Thesis, Politec. di Milano* (1992).

[7] Marco Dorigo and Gianni Di Caro. 1999. New Ideas in Optimization. In David Corne, Marco Dorigo, Fred Glover, Dipankar Dasgupta, Pablo Moscato, Riccardo Poli and Kenneth V Price (eds.). McGraw-Hill Ltd., UK, Maidenhead, UK, England, 11–32.

[8] Marco Dorigo and Luca Maria Gambardella. 1997. Ant colony system: a cooperative learning approach to the traveling salesman problem. *{IEEE} Trans. Evol. Comput.* 1, 1 (1997), 53–66.

[9] Marco Dorigo and Thomas Stützle. 2004. *Ant colony optimization*. MIT Press.

[10] Christian Fikar and Patrick Hirsch. 2017. Home health care routing and scheduling: A review. *Computers and Operations Research*.

[11] J I van Hemert and T Bäck. 2002. Measuring the Searched Space to Guide Efficiency: The Principle and Evidence on Constraint Satisfaction. In *Parallel Problem Solving from Nature*.

[12] Gerhard Hiermann, Matthias Prandtstetter, Andrea Rendl, Jakob Puchinger, and Günther R. Raidl. 2013. Metaheuristics for solving a multimodal home-healthcare scheduling problem. *Cent. Eur. J. Oper. Res.* (2013).

[13] Madjid Khichane, Patrick Albert, and Christine Solnon. 2009. An ACO-Based Reactive Framework for Ant Colony Optimization: First Experiments on Constraint Satisfaction Problems. In *Learning and Intelligent Optimization* (Lecture Notes in Computer Science), 119–133.

[14] Dorota Slawa Mankowska, Frank Meisel, and Christian Bierwirth. 2014. The home health care routing and scheduling problem with interdependent services. *Health Care Manag. Sci.* (2014).

[15] Maria Di Mascolo, Marie Laure Espinouse, and Zied El Hajri. 2017. Planning in Home Health Care Structures: A literature review. *IFAC-PapersOnLine* (2017).

[16] Alberto Maydeu-Olivares, Amanda J. Fairchild, and Alexander G. Hall. 2017. Goodness of Fit in Item Factor Analysis: Effect of the Number of Response Alternatives. *Struct. Equ. Model.* (2017).

[17] Christos H Papadimitriou. 1994. *Computational complexity*. Addison-Wesley.

[18] Abdur Rais and Ana Vianaa. 2011. Operations research in healthcare: A survey. *Int. Trans. Oper. Res.* (2011).

[19] Francesca Rossi, Peter van Beek, and Toby Walsh (Eds.). 2006. *Handbook of Constraint Programming*. Elsevier.

[20] Alain Sidaner, Olivier Bailleux, and Jean Jacques Chabrier. 2002. Measuring the spatial dispersion of evolutionary search processes: Application to walksat. In *Lecture Notes in Computer Science*.

[21] Christine Solnon. 2002. Ants can solve constraint satisfaction problems. *IEEE Trans. Evol. Comput.* 6, 4 (2002), 347–357.

[22] Christine Solnon. 2010. *Ant Colony Optimization and Constraint Programming*. Wiley.

[23] Kostas Stergiou and Nikos Samaras. 2005. Binary Encodings of Non-binary Constraint Satisfaction Problems: Algorithms and Experimental Results. *J. Artif. Intell. Res.* 24, (2005), 641–684.

[24] Thomas Stützle and Holger H Hoos. 2000. MAX-MIN Ant System. *Futur. Gener. Comp. Syst.* 16, 8 (2000), 889–914.

Address for Correspondence

Yahia Chabane
yahia.chabane@berger-levrault.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190283

Automated Control of Codes Accuracy in Case-Mix Databases by Evaluating Coherence with Available Information in the Electronic Health Record

Robin Chaux[a], Isabelle Treussier[a], Bissan Audeh[b], Suzanne Pereira[c], Thierry Hengoat[d], Béatrice Trombert Paviot[ab], Cedric Bousquet[ab]

[a]Public Health and Medical Information Department, University Hospital Center of Saint Etienne, France,
[b]Sorbonne Université, Inserm, université Paris 13, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé, LIMICS, F-75006 Paris, France,
[c]Vidal, Issy-les-Moulineaux, France,
[d]Web100T, Mérignac, France

Abstract

Coding accuracy in case-mix databases enables efficient funding of health facilities and accurate epidemiological statistics based on patients' stays information. We assume that the data collected in the electronic health record, especially drug prescriptions and medical reports are relevant for checking the consistency of the coding of diagnoses. We evaluated a new coding control tool, "TOLBIAC control", embedded in the Web100T coding assistant. This tool interacts with the Vidal Application Programming Interface and the electronic health record of the University Hospital of Saint-Etienne. The micro-average F-measure was 0.76 for drug prescriptions and 0.55 for free text medical reports. This initial evaluation has revealed that drug prescriptions in EHRs can successfully be used to develop an automated ICD-10 code-control tool. Nevertheless the "TOLBIAC control" tool is not yet fully effective for widespread use because of its limited performance in text analysis, a feature that is currently undergoing improvements.

Keywords:

Drug Prescriptions, International Classification of Diseases, Diagnosis Related Groups, Clinical Coding, Electronic Health Records

Introduction

In many countries, funding of public and private hospitals is based on a prospective payment system. Hospitals are reimbursed according to Diagnosis Related Groups preset tariffs [1]. In France, this system relies on the French Hospital Activity Database which requires every inpatient and outpatient stay to be coded using a Standardized Discharge Summary (SDS) [2]. The SDS aggregates all diagnoses and medical procedures associated to patients' stays. Diagnoses are coded following the International Classification of Diseases, Injuries, and Causes of death (ICD-10) [3]. The process of coding has to follow mandatory national guidelines, published by the French Health Ministry, in order to ensure consistent practices among hospitals.

In addition to hospital reimbursement, the accuracy of SDSs is essential for nationwide hospital management, statistics, economy, research, and epidemiology [4]. Furthermore, physicians from the French National Insurance regularly ensure that clinical coding matches patients' medical records, if not, heavy fines may be applied [5].

Coding is mainly performed by physicians and specialized clinical coders. To support experts in the coding process, automated code prediction tools could be used. Such tools can suggest codes that are not manually detected, and thus can be acted on for completeness. Code prediction can be made based on the data from the EHRs using machine learning algorithms, rule-based systems, or their hybrid. An example of a rule is: "if serum potassium is more than 6.5 mmol/l, suggest E87.50 hyperkalemia" [6].

While many studies have shown interest in coding predictions [7–9] based on the comparison with elements available in the EHR, to our knowledge, no study has described an automated standardized code controlling tool that benefits from such data in addition to information collected in SDSs. To fill this gap, and to overcome the difficulties associated with handcrafted code harvesting [4,10], the TOLBIAC project (Terminologies and Ontologies for Linking Billing Information and Accurate Clinical data) aimed to define, model, implement, and evaluate the consistency between the information in the EHRs and the processed SDSs. As part of this project, the "Tolbiac-Control" tool was developed and integrated in the graphical user interface of Web100T® coding system to enable the use of ICD-10 codes for the hospital stays..The company managing the registered drug database VIDAL developed the tool.

To control the validity of existing codes, "Tolbiac-Control" is based on three hypotheses. First, hospital drug prescriptions are a good indicator of the presence of various pathologies [11]. Some studies successfully predicted codes using information on drugs prescribed in EHRs [9], and other studies showed that prescriptions may be a valid proxy of prevalence of diseases [12,13]. Therefore, even if there are some known limitations [14], data concerning drugs prescribed in hospitals could be used to assess whether an ICD-10 code rightfully belongs to the patient's SDS or not.

The second hypothesis is that the use of unstructured data from EHRs, such as text from clinical reports may be used for code control. Unstructured data has been widely used for code prediction and suggestion [8], and may likewise be useful for code control.

Abhyankar et al. [15] and Halfon et al. [14] showed that the utilization of multiple data sources improves the prediction of codes. Moreover, some pathologies are complex and not represented in a single data source [15]. Consequently, our third hypothesis is that combining the usage of prescribed drugs and clinical text altogether may improve code control.

In this pilot study, we present and evaluate the "Tolbiac-Control" tool in its prototype phase. The evaluation is achieved for each of the control features proposed by the tool: the control based on drugs prescribed in EHR and the control based on unstructured text from medical reports associated to patients' stays. We also performed an exploratory evaluation using the combination of both the features.

Methods

Implementation

The "Tolbiac-Control" tool takes as input the drug codes, free text from medical reports, and ICD-10 codes selected by physicians or clinical coders. With this input, the tool proposes two checking features, one based on the drug prescription structured data, and the other on free text reports (Figure 1). A colored button is automatically generated for every single code for both the controls under verification feature to indicate if the code is justified (green), unjustified (red) or should be verified (orange). The lack of a button for the drug verification means that there is no drug indication for a specific diagnosis. The tool uses the VIDAL database and the ICD dictionaries for this verification (Figure 2).

For drug verification, the UCD identifiers (*Drug Common Unit of Dispensing*) of the drugs prescribed during the stay are used to extract drug indication codes from the VIDAL database. Free-text verification is achieved using a REST web service proposed by the text mining engine "VIDAL" [16]. The service takes the free text of stays' report encoded in UTF8 as input. The text-mining engine uses the multi-terminology approach F-MTI to propose ICD-10 codes based on the input text.

Figure 1- Tolbiac-Control Interface in a Patient's SDS Integrated in Web100T® Encoding Software

The codes extracted from drug and free-text verification methods are compared separately to the codes assigned by health professionals to associate a "Justified" marker for matching codes, "Unjustified" for mismatch, and "To verify" if the tool is not able to decide based on the available data. In addition to these three markers, a fourth marker "No UCD available" is assigned when a coded diagnosis has no corresponding specific drug indication (e.g. physical traumas, congenital conditions).

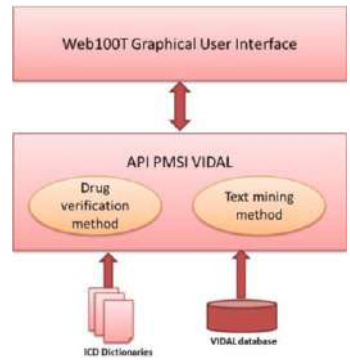


Figure 2- Structure of Tolbiac-Control Tool

Evaluation

We randomly selected a corpus of 32 inpatient stays in 17 medical specialties at the Public and University Hospital of Saint-Etienne during June 2017. The stays had previously been coded as usual by clinical coders or physicians, and utilized 282 ICD-10 codes. The average length of stay was 8 days and an average of 10 drugs were administered per stay.

To evaluate the "Tolbiac-Control" tool, we compared its automated code-control verifications with those of two assessors who had access to the same documents. The first was a trained clinical coder (IT) who looked for elements of EHRs justifying ICD-10 codes in evaluated SDSs. The second was a medical resident (RC) who ensured that the markers "no UCD available" were appropriately reported and assisted the clinical coder in case of uncertain situations. Verifications achieved by the assessors were considered as the "ground truth".

Quantitative approach

The results were tabulated in a confusion matrix, with the output from the "Tolbiac-Control" in rows and the cross-checking by the assessors in columns. As the verified codes were not uniformly distributed over the markers, we had to look at the effectiveness of the classifier one class at a time before averaging the entries. We estimated the one-versus-all confusion matrix for each marker, as seen in figure 3 (3 entries for text control with 3 markers, and four entries for UCD control with 4 markers).

		Clinical coder verification, ground truth	
		Justified	**Other markers**
TOLBIAC code control markers	**Justified**	True Positives	False Positives
	Other markers	False Negatives	True Negatives

Figure 3- Example of a One-versus-all Confusion Matrix for the "Justified" Marker

For example, for the text-control element, we used 3 binary classification tasks where one class (e.g. "Justified") was considered as the positive class while the combination of all the other classes ("Unjustified" and "To verify") constituted the negative class.

The main statistical measure used to evaluate the "Tolbiac-Control" tool was the micro-averaged F-measure, which is the harmonic mean of precision and recall. It is a synthetic indicator that allows to perceive the overall preciseness and robustness of the classifier [9]. Its formula is described in equation (1):

$$f - measure_{micro} = 2\frac{precision*recall}{precision+recall} \quad (1)$$

Micro-averages were preferred over macro-averages. As some marker modalities such as "To verify" or "No UCD" were vastly underrepresented, we decided to take into account the respective weight of each modality for a better representation of the results.

Precision, also called positive predictive value, can be defined as the fraction of relevant marker modalities among the retrieved markers. For an information retrieval system, high precision means that it retrieves more relevant markers than irrelevant ones. The formula is described in equation (2) with "k" being the modality marker, TP for true positives and FP for false positives:

$$Precision_{micro} = \frac{\sum_k TP_k}{\sum_k TP_k + \sum_k FP_k} \quad (2)$$

Recall, also called true positive rate or sensitivity, is the fraction of relevant marker modalities that have been retrieved over the total amount of relevant markers. An information retrieval system with high recall means that it retrieves most of the relevant results. The formula is provided in equation (3) with FN for false negatives:

$$Recall_{micro} = \frac{\sum_k TP_k}{\sum_k TP_k + \sum_k FN_k} \quad (3)$$

Accuracy is the proportion of true results, both positives and negatives, among the total number of examined cases. It measures how well a classification system correctly identifies or excludes a situation. With TN for true negatives, the formula for quantifying accuracy in a one-versus-all confusion matrix is described in equation (4):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Classification error is the percentage of incorrect markers to the total number of markers analyzed, and was calculated along with its 95% confidence interval.

To analyze the combined effect of the two code-control features (drug-based and text-based), we built the decision tree described in figure 4 on the basis of the following rules:

1. The code is labeled "Justified": if the information that justifies the code is compliant with a drug identifier and/or present in free text.
2. The code is labeled "Unjustified": if both sources (drug-based and text-based) state that the information is "Unjustified" or "To verify".

We dichotomized the decisional outcome with only "Justified" and "Unjustified" markers, assuming that the clinical coder has to check back the validity of the code if it is "Unjustified" or "To verify".

The decision result was assessed against the clinical coder's ground truth using the previously described metrix. Statistical analyses were performed with the R language and environment for statistical computing and graphics version 3.4.0 (with the packages "ggplot2" version 2.2.1, "lattice" version 0.20-35 and "caret" version 6.0-8.0).

Qualitative approach

A clinical coder (IT) from the Medical Information Unit of the University Hospital Center of Saint-Etienne who is familiar with several coding tools evaluated the graphical interface. The evaluation included reviewing usability of the module according to different criteria: ease of use, clarity and precision of the display. Text analysis was assessed by evaluating elements of free text from EHRs that led to the chosen marker of a code. The consistency of the detected terms and their use by the TOLBIAC text control tool were also verified.

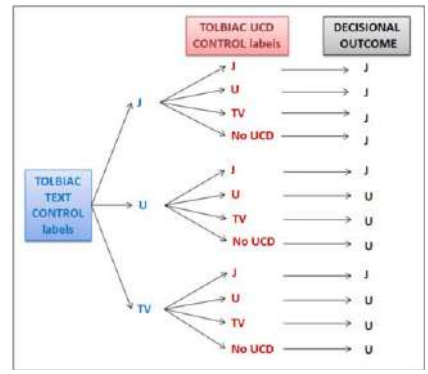


Figure 4- Decision Tree for Combined UCD and Free-text control. J= "Justified" Marker, U= "Unjustified" Marker, TV= "To verify" Marker, "No UCD"= No UCD

Results

Quantitative approach

Among 282 codes, the UCD code control feature of the "Tolbiac-Control" tool found 112 (39.7%) codes "Justified", 97 (34.4%) "Unjustified", 15 (5.3%) "To verify", and 58 (20.6%) codes with no UCD indication. Table 1 shows the results of the evaluation of the UCD code control.

Table 1 - VIDAL UCD Control Feature (n= 282)

Accuracy, n (%)	212 (0.752)
Classification Error, % [95% CI]	0.248 [0.198-0.298]
Recall	0.752
Precision	0.762
F-measure	0.757

Among the same 282 codes, the text code control feature of the VIDAL module found 66 (23.4%) codes "Justified", 172 (61.0%) "Unjustified", and 44 (15.6%) codes "To verify". Table 2 presents the results of the evaluation of the text code control.

Table 2 - VIDAL Text Control Feature (n= 282)

Accuracy, n (%)	146 (0.517)
Classification Error, % [95% CI]	0.483 [0.425-0.541]
Recall	0.519
Precision	0.673
F-measure	0.555

Among 282 codes, the text and UCD combined code control found 146 (51.8%) codes "Justified" and 136 (48.2%) "Unjustified". Table 3 shows the results of the texts and UCD code control combined.

Table 3 - VIDAL UCD and Text Control Features Combined (n= 282)

Accuracy, n (%)	187 (0.663)
Classification Error, % [95% CI]	0.337 [0.282-0.392]
Recall	0.626
Precision	0.869
F-measure	0.728

Qualitative approach

The clinical coder was able to easily use the interface of "Tolbiac-Control" within the encoding software without any prior training. She found that the information was displayed clearly with immediate visual access to the code-control markers for each ICD-10 code. In addition, the interface regrouped all drug prescriptions as well as medical text retrieved from EHRs, which can be conveniently accessed directly from the coding tool. No issues were reported with the graphical interface, and the usability of the tool was described as convenient.

Some issues related to text analysis were observed. The "Tolbiac-Control" tool did not identify negation. For example, "there is no appendicitis" written in a medical report is acknowledged as the ICD-10 code K35 "acute appendicitis". In addition, the tool did not always detect pathology localization. For example, for "fractured radius" in an operative report, the tool identified "fractured" but not "radius".

Furthermore, the tool might wrongly identify word fragments as individual terms. For example, in "transverse", the word fragment "ver" (which means "worm" in English) was recognized within the word, and a coded Helminthiase (an intestinal parasitic worm) was wrongly acknowledged as "Justified" in a patient's discharge summary.

In some cases, the same code could be assigned several times for the same stay, which often happens with handcrafted coding. In our sample, this was the case with seven codes. We noticed that for some the codes, the tool generated a different marker for each occurrence from the same EHR.

Discussion

We recognize that code control using drug prescriptions is promising for this pilot study of the "Tolbiac-Control" tool. The drug control feature retrieved 75% of the relevant information and classified them with the right marker 76% of the time. This result is consistent with other studies showing the reliability of this approach for code prediction or as proxy of disease prevalence [13]. We showed in this study that using drug prescriptions is also useful for code control. In addition, this data source is reliable as drug prescriptions are systematically encoded in EHRs with no missing values.

The text control feature has less conclusive results. It retrieved about half of the relevant markers, and classified them correctly in 67% of the cases. As with numerous other studies [17,18], using free-text unstructured data to improve coding is challenging. In particular, hospitalization reports and clinician notes, called "notational text", often contain abbreviations, lack of punctuation, poor grammatical structures [17] as well as spelling mistakes and missing values [7]. These inherent obstacles added to the limitation of the text analysis feature of the "Tolbiac-Control" tool and might explain its lower performance. However, the F-measures are within the boundaries of the results found in the literature for natural language processing and machine learning methods currently applied to ICD-9 or ICD-10 coding in different languages (about 0.50-0.90) [17,18,20].

With drugs and text features combined, we observed a 10% improvement in precision as compared to the use of each data source separately. However, the recall and F-measure decreased when using the combined features compared to the use of the drug-based feature alone. Thus, contrary to other studies, classification was not improved by combining two data sources, mostly due to the deficiencies of the text-control feature.

As clinical coding is growing in complexity and importance in hospitals, there is an ever increasing need for automated computerized assistance to increase quality, productivity, and efficiency [8]. The "Tolbiac-Control" tool gathers drug identifiers, medical reports files, and code control markers in a simple window in already existing encoding software. Coupled with its easy-to-use visual interface, the tool might be a suitable choice for the current needs of code control professionals.

Although "Tolbiac-Control" was tested in one hospital, we were able to achieve a broad range evaluation with different medical specialties and large sets of codes unlike previous studies that focused on limited types of specific diagnosis [5]. Knowing that some pathologies are easier to code using an automated tool than others [14], our study showed the overall effectiveness of the proposed control tool and the potential of its use in a wide variety of clinical situations. While our study focused on a specific national institution, accurate coding in Diagnosis Related Groups-based information systems is an international concern. Our work, especially related to the use of drug identifiers for code-control, could be used in other institutions and countries.

The main limitation of the current tool certainly is the low performance of the text-based control feature. This feature needs to better identify the link between medical entities in clinical free text. Another practical enhancement is to automate free-text extraction from available reports instead of the required manual upload of unstructured data in the current version. The VIDAL is currently working on improving its API to overcome these limitations. The generalization of the tool with other encoding software is also in progress.

Another limitation of the study is the limited sample size. As the "Tolbiac-Control" tool is still a prototype, the analysis was made on 282 ICD-10 codes, imported from 32 patient stays. Further experiments on a larger sample size are required to confirm the research hypothesis with higher statistical power. Also, we relied on a single medical coder to evaluate the interface and usability of the tool. Considering the subjective nature of qualitative approaches, it is an important limitation to bear in mind at this stage of the project.

Once an improved version of the "Tolbiac-Control" is available, we aim to perform further analysis with larger samples to verify if code control is more efficient with some specialties than with others, as revealed by previous studies regarding code prediction [14]. Furthermore, it would be interesting to assess the utility of using the results of biological tests and implantable medical devices as features for code control along with drug-based and text-based features. The integration of biological tests analysis is already scheduled for the next version of "Tolbiac-Control", and it is expected to improve the quality of the tool.

Conclusions

This work shows that information contained in drug prescriptions in EHRs can successfully be used to develop an automated and systematized ICD-10 code-control tool. The "Tolbiac-Control" tool could become an asset for clinical coders, and might improve code precision, which would enhance the quality of SDSs data.

The proposed tool evaluation was done in a prototype phase. Although this tool is not yet entirely functional for a general use, it is currently undergoing improvements based on the observed issues, mainly about the usage of unstructured data.

Acknowledgements

This work benefited from a grant from the French agency for research (ANR – *Agence nationale pour la recherche*) ANR-13-TECS-0010 for the TOLBIAC project. The VIDAzL API server is developed by the VIDAL Company, which employs Suzanne Pereira. WebPIMs® and WebCAD® are owned by the Web100T Company, which employs Thierry Hengoat. The other authors have no conflicts of interest to declare.

References

[1] F.A. Aiello, and S.P. Roddy, Inpatient coding and the diagnosis-related group, *Journal of Vascular Surgery*. **66** (2017) 1621–1623. doi:10.1016/j.jvs.2017.08.030.

[2] T. Boudemaghe, and I. Belhadj, Data Resource Profile: The French National Uniform Hospital Discharge Data Set Database (PMSI), *International Journal of Epidemiology*. **46** (2017) 392-392d. doi:10.1093/ije/dyw359.

[3] 2011 - ICD-10 International statistical classification o.pdf, (n.d.). http://www.who.int/classifications/icd/ICD10Volume2_en_2010.pdf (accessed November 12, 2018).

[4] S.A.R. Nouraei, S. O'Hanlon, C.R. Butler, A. Hadovsky, E. Donald, E. Benjamin, and G.S. Sandhu, A multidisciplinary audit of clinical coding accuracy in otolaryngology: financial, managerial and clinical governance considerations under payment-by-results, *Clinical Otolaryngology*. **34** (2009) 43–51. doi:10.1111/j.1749-4486.2008.01863.x.

[5] J. Lardon, H. Asfari, J. Souvignet, B. Trombert-Paviot, and C. Bousquet, Improvement of Diagnosis Coding by Analysing EHR and Using Rule Engine: Application to the Chronic Kidney Disease, *Stud Health Technol Inform*. **210** (2015) 120–124.

[6] A. Guardia, R. Meyer, and P. Rohner, The CASSANDRE Project: automated alerts for optimal coding of diagnosis and interventions, *World Hosp Health Serv*. **49** (2013) 21–24.

[7] E. Scheurwegs, K. Luyckx, L. Luyten, W. Daelemans, and T. Van den Bulcke, Data integration of structured and unstructured sources for assigning clinical codes to patient stays, *Journal of the American Medical Informatics Association*. **23** (2016) e11–e19. doi:10.1093/jamia/ocv115.

[8] E. Scheurwegs, B. Cule, K. Luyckx, L. Luyten, and W. Daelemans, Selecting relevant features from the electronic health record for clinical code prediction, *Journal of Biomedical Informatics*. **74** (2017) 92–103. doi:10.1016/j.jbi.2017.09.004.

[9] N. Triki, C. Bousquet, J. Lardon, H. Asfari, R. Spiga, and B. Trombert-Paviot, Inconsistencies Between Antiparkinsonian Drugs and ICD-10 Codes in Inpatients: A TOLBIAC Project Case Study, *Stud Health Technol Inform*. **228** (2016) 364–368.

[10] K. Furu, B. Wettermark, M. Andersen, J.E. Martikainen, A.B. Almarsdottir, and H.T. SÃ¸rensen, The Nordic Countries as a Cohort for Pharmacoepidemiological Research, *Basic & Clinical Pharmacology & Toxicology*. **106** (2010) 86–94. doi:10.1111/j.1742-7843.2009.00494.x.

[11] K.L. Sloan, A.E. Sales, C.-F. Liu, P. Fishman, P. Nichol, N.T. Suzuki, and N.D. Sharp, Construction and Characteristics of the RxRisk-V: A VA-Adapted Pharmacy-Based Case-mix Instrument, *Medical Care*. **41** (2003) 761–774. doi:10.1097/01.MLR.0000064641.84967.B7.

[12] K. Furu, S. Skurtveit, A. Langhammer, and P. Nafstad, Use of anti-asthmatic medications as a proxy for prevalence of asthma in children and adolescents in Norway:a nationwide prescription database analysis, *European Journal of Clinical Pharmacology*. **63** (2007) 693–698. doi:10.1007/s00228-007-0301-9.

[13] R.E. Cossman, J.S. Cossman, W.L. James, T. Blanchard, R. Thomas, L.G. Pol, and A.G. Cosby, Correlating pharmaceutical data with a national health survey as a proxy for estimating rural population health, *Population Health Metrics*. **8** (2010). doi:10.1186/1478-7954-8-25.

[14] P. Halfon, Y. Eggli, A. Decollogny, and E. Seker, Disease identification based on ambulatory drugs dispensation and in-hospital ICD-10 diagnoses: a comparison, *BMC Health Services Research*. **13** (2013). doi:10.1186/1472-6963-13-453.

[15] S. Abhyankar, D. Demner-Fushman, F.M. Callaghan, and C.J. McDonald, Combining structured and unstructured data to identify a cohort of ICU patients who received dialysis, *Journal of the American Medical Informatics Association*. **21** (2014) 801–807. doi:10.1136/amiajnl-2013-001915.

[16] D. Ragain, E. Pommiers, J.-F. Forget, and D. Blum, « Optimiser » n'est pas un gros mot : préconisations automatiques de diagnostics du Programme de médicalisation des systèmes d'information grâce à la gestion du médicament dans un système d'informations hospitalier optimisé, *Revue d'Épidémiologie et de Santé Publique*. **59** (2011) S52. doi:10.1016/j.respe.2011.03.029.

[17] JP. Pestian, C. Brew, P. Matykiewicz, D. J. Hovermale, N. Johnson, K. Bretonnel Cohen, and W. Duch. A Shared Task Involving Multi-label Classification of Clinical Free Text. *Proceedings of the Workshop on BioNLP 2007 Biological, Translational, and Clinical Language Processing.* (2007) pp 97–104.

[18] T. Baumel, J. Nassour-Kassis, R. Cohen, M. Elhadad, and N. Elhadad. Multi-Label Classification of Patient Notes: Case Study on ICD Code Assignment. *Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence. (2018)* pp 410-416.

[19] R.C. Barrows Jr, M. Busuioc, and C. Friedman, Limited parsing of notational text visit notes: ad-hoc vs. NLP approaches, *Proc AMIA Symp*. (2000) 51–55.

[20] C. Lin, CJ. Hsu, YS. Lou, SJ. Yeh, CC. Lee, SL. Su, and HC. Chen. Artificial Intelligence Learning Semantics via External Resources for Classifying Diagnosis Codes in Discharge Notes, *J Med Internet Res*. **19** (2017) e380 p.1.

Address for correspondence

Corresponding author: Robin Chaux.

Email: robinvchaux@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190284

Characteristics and Hospital Activity of Elderly Patients Receiving Admission Avoidance Home Visits: A Population-Level Record Linkage Study

Maria Cristina Martin[a], Matt-Mouley Bouamrane[b], Paul Woolman[c], Kimberley Kavanagh[d], David Young[d]

[a] *Department of Computer Science, University of Strathclyde, Glasgow, Scotland, UK*
[b] *Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Scotland, UK*
[c] *Information Services, NHS Forth Valley, Falkirk, Scotland, UK*
[d] *Department of Mathematics and Statistics, University of Strathclyde, Glasgow, Scotland, UK*

Abstract

As pressures on healthcare systems increase, due to an ageing population, hospital admission avoidance interventions have been emphasised. These interventions can be difficult to objectively evaluate due to non-randomised roll-out, requiring observational methods with carefully selected control groups. This study aims to identify the defining characteristics of elderly patients receiving admission avoidance home visits. We conducted a record linkage study using routinely collected data to compare characteristics and outcomes of the general elderly population and a subset of high-risk patients. Intervention patients were found to have significantly different demographics and admission rates compared to the general population, having four times higher admission rates at baseline. However, they share similarities with high-risk patients, particularly in that after a period of increased admissions, both groups experienced a reduction in the following year. Identifying defining characteristics of the target intervention population can guide the careful selection of a control group for evaluation.

Keywords:

Home Care Services; Evaluation Research; Informatics

Introduction

Within the last decade, there has been an increased emphasis on reducing unscheduled admissions to the hospital, reflecting the need to better manage the increased pressures faced by health systems due to the changing demographic profile of European populations [1]. This has led to the development of alternative care models, focusing on proactive rather than reactive care. Alternative care models include proactive chronic condition management, intermediate care, community-based care interventions and using telemedicine, telehealth and digital health solutions. Young states that intermediate care "is conceived as a range of service models aimed at 'care closer to home' by expansion and development of community health and social services," including hospital-at-home services [2]. There are currently two main hospital-at-home models: *early discharge models* where hospital-level care is provided at home following early discharge from hospital, and *admission avoidance models*, in which multidisciplinary rapid response teams provide treatment, assessment and support for a short period of time [1,2].

The latter models are concerned with unscheduled care, differentiating them from models such as health promotion visits or preventive home visits, which involve assessments primarily aimed at preventing new problems for patients living independently in the community [3].

Hospital-at-home models are complex interventions with several interconnecting parts. Their evaluation can prove challenging, often because the interventions have not been fully defined or developed at roll-out [4]. Furthermore, they are also prone to evolve over time, particularly within a community setting [5]. Randomisation may be unfeasible or inappropriate, with the decision to evaluate often being made in hindsight [5].

Observational studies can provide an alternative evaluation approach where randomisation is unfeasible; however, these evaluation studies require a robust design and methodology. Often evaluations may use the general non-intervention population as a control group and use standardisation of likely confounders, such as age and sex. However, the inclusion criteria of hospital avoidance interventions will usually be linked to the outcomes being measured for evaluation [5].

Hence, trends in hospital activity for the intervention population and the characteristics that define them will differ greatly from that of a general elderly population. This is particularly the case in people over 65 years of age with a history of emergency admissions (also known as high-risk patients), whose levels of hospital use have been shown to naturally reduce over time compared to the general elderly population, due to both mortality and regression to the mean [6]. In their study, Roland et al warranted further research for defining high-risk patient groups for interventions to reduce admissions [6]. To evaluate the effect of a healthcare intervention on hospital admissions, a carefully selected control group is essential and, in the case of interventions for high-risk patients over 65 years of age, must match the intervention group according to their defining characteristics [6].

The aim of this paper is to identify the defining characteristics of patients receiving admission avoidance home visits, for which the decision to evaluate was made in retrospect and referral criteria has been loosely defined (as is common in complex community interventions).

This was done by comparing patients receiving the intervention to two groups: the general population over 65 years of age and a subset of high-risk patients in the area of interest. We first compared patient characteristics in these three groups, obtained through healthcare record linkage of several datasets, some of which have never previously been used in research.

We then made a temporal comparison of hospital admission rates for the three groups. Identifying characteristics that define the intervention patients will enable the appropriate selection of a comparison group for evaluation, which is essential for a robust evaluation in this setting, and may prove useful to others evaluating similar services and interventions.

Background

'Closer to Home' programme

In 2011, the Scottish Government rolled out the Reshaping Care for Older People (RCOP) strategy with the vision that older people should live full and positive lives at home or in a homely setting [7]. The allocated RCOP strategy Change Fund resulted in a number of local initiatives across Scotland aiming towards promoting home and community care. The Forth Valley (FV) health-board includes a central area of Scotland with an estimated population of 57,317 residents aged 65 and over in 2017 [8]. 'Closer to Home' is a coordinated programme, set up in December 2015, to achieve the RCOP aims within the NHS Forth Valley health board. We have developed an evaluation framework for the 'Closer to Home' programme which has been previously described elsewhere [9].

'Enhanced Community Teams'

The 'Closer to Home' programme includes Enhanced Community Teams (ECTs) which are multidisciplinary rapid response teams of nurses, physiotherapists, occupational therapists, social care staff and, more recently, specialty-trained general practitioners (GPs) (from January 2017). In this paper, ECT is the intervention of interest.

The ECTs provide both early discharge and admission avoidance home visits, with a focus on the latter. Care is provided 24 hours a day, 7 days a week. Referrals are made mainly through the patient's GP, however, referrals can also be made by discharge coordinators, emergency department clinicians, the ambulance service, and other community services. The criteria for referral are loosely defined but state that the patient must require immediate hospital-level support that can be provided at home and must be registered with a FV GP. Treated patients are predominantly elderly, frail patients over 65 with long-term conditions. Acutely ill elderly patients who did not receive the intervention received usual care which includes hospital admission along with community services such as district nursing which may support a patient's recovery but are not aimed at admission avoidance.

Implementation of the ECTs was staggered in the first year in a non-randomised fashion, area by area through contact with GP practices and other potential referral sources to promote the teams. ECTs use NHS FV's Multidisciplinary Information System (MiDIS) for the recording of community activity data, including assessments and contacts with patients such as home visits.

Methods

Constructing a study cohort

A cohort of patients was defined to include any patient within FV who was aged 65 years or older at any point between a year before the time the ECTs were started (December 2014) and the time at which the cohort was constructed (April 2018). To construct this cohort, two main datasets were combined: one consisting of a list of the patients registered with a GP in FV, and one consisting of records of deaths registered in FV by the National Records of Scotland. These datasets are used nationally and go through several quality checks. These were combined to create a cohort of 65,189 patients.

Record linkage methods and ethics

Linkage of patient datasets was conducted using deterministic record linkage in SQL Server (Management Studio 2008), which allows full control of the linkage process. Pre-merge data cleansing to resolve typographical differences was required to enable linking keys to be matched [10]. In Scotland, the Community Health Index (CHI) is used to uniquely identify patients, often used in linking patient datasets [11]. Using the CHI and postcode as linking keys, demographic, hospital activity and prescribing records were linked to the compiled cohort described previously.

Research ethics approval was not required for this study as it is for the purpose of a service evaluation based on retrospective analysis of routinely collected data. Caldicott approval within FV and the Information Services Division (ISD) for Scotland was obtained for the request for prescribing data. All analysis was conducted on pseudonymised data.

Identification of intervention patients

Intervention patients were identified from datasets collected from MiDIS. These datasets are not used for national reporting and prior to this study, have never been used in research, hence they required multiple linkages and data cleansing in consultation with the ECTs to understand each variable held. The main linked MiDIS datasets were a master patient dataset and datasets of episodes of care, individual contacts, and episode registrations. This linkage combined with consultation with ECT members enabled the compilation of a validated dataset containing episode details, including number of contacts, type of patient and discharge reason, from which intervention patients were identifed (1,294 records).

Main linked datasets

The linked datasets which are locally held include emergency department attendance data, community health visit data and a master patient dataset for any patient having received inpatient, day case or outpatient care in FV. The linked datasets which are nationally held include outpatient attendance data (Scottish Morbidity Record for outpatients (SMR00)) and hospital inpatient stay data (Scottish Morbidity Record for general acute inpatients and day cases (SMR01)) [11]. Data for prescribed items dispensed in the community were obtained for each patient in the cohort through a request for data through ISD.

Demographic data linkage

The linked demographic variables include age, gender, GP practice, locality, deprivation, ethnicity, marital status, smoking status, living alone, and a care home stay indicator. These primary variables were compiled in an analytical dataset and additional secondary variables were created from these. Due to missing values for the primary variable living alone, marital status was incorporated to create a secondary variable indicating if the patient has been recorded as living alone or not at any point, with not married being classified as living alone and married/cohabiting as not living alone. This reduced the missing values by 34.0%. An additional variable was created combining nursing home residency and having had a stay in a nursing home, to characterise patients who have had a nursing home stay (used as a proxy for functional status with high dependency needs).

Hospital inpatient stay records were linked to obtain a Charlson comorbidity score for each patient, which identifies and gives weights to each of 17 comorbidities according to the relative risk of one-year mortality [12].

The comorbidities were identified from International Classification of Diseases codes (10^{th} revision) (ICD-10) recorded as hospital diagnoses [12] (923,465 records). An algorithm within R package "icd" was then used to generate the Charlson scores for each patient using ICD-10 codes recorded in the past 5 years [13].

Hospital activity and prescribing data linkage

Hospital activity data linked for each patient included the number of emergency and elective inpatient hospitalisations (from SMR01), two years before and after the implementation of ECT (2014-17) (100,071 records). Prescribing data linked for each patient included prescription items dispensed and reimbursed by the NHS in the community (94.6% were prescribed by a GP). This data included the number of items prescribed and the number of British National Formulary (BNF) classes (paragraphs) covered by the prescribed items (1,510,018 records). The average monthly number of BNF classes was selected for use as it reduces variation and reduces the effect of exaggerated polypharmacy for patients who have multiple medications in the same BNF paragraph (class), hence it was deemed a better proxy measure of multimorbidity [14].

Population selected for analysis

The activity data collection period was set as data registered between the 1st January 2016 to the 1st May 2018. From the 65,189 patients in the compiled cohort, exclusion criteria were applied before analysis. Patients receiving the intervention outside of the collection period were excluded (including only episodes occuring between 1st January 2016 to the 1st May 2017, to allow for outcomes to be collected for one year post-intervention). Patients were included if they were over 65 years of age at the start of the collection period, registered with a GP in FV, had not transferred to another health board and had not died at the start of the collection period. Intervention patients who were receiving palliative care or whose episodes of care were recorded as failed (due to inappropriate referrals or inability to contact the patient) were excluded. This left a total of 566 intervention patients and 60,901 patients from the general population.

Data analysis

A subset of high-risk patients, defined as those who had two or more emergency admissions in the same year that the intervention patients experienced a deterioration in health (2016), were grouped separately from the general population in order to compare their characteristics and identify their hospital activity patterns, following Roland et al's comparison [6]. To test for differences between demographic variables in the intervention group, the high-risk group and the remaining population, Welch two sample *t*-tests were used for continuous variables, chi-squared tests were used for categorical variables and Fisher's exact test was used for categorical variables with small observed values in contingency tables.

A temporal comparison was made in admission rates per person between the intervention patients, high-risk patients and the general population (all patients) using the denominator equal to the number alive in each year of the comparison, which Roland et al were unable to do because of a lack of data on deaths. They have previously shown, however, that including mortality has little impact on overall conclusions [6]. Data analysis was conducted in R Studio (R v3.4.1).

Results

Demographic differences

The ECT intervention group and the general population were shown to be very different in the analysis of demographic data. The differences between the general population and both the intervention and high-risk groups were all found to be significant at the 0.1% level for each of the variables shown in *Table 1*. Differences in ethnicity and the average number of prescriptions were not found to be significant between the intervention and high-risk group, while other variables showed significant differences (at the 5% level for care home stay and deprivation, and at the 0.1% level for all other variables).

The proportion of patients aged 75 and older was found to be higher in the intervention group than the other groups (83.5% of intervention patients, 62.4% of high-risk patients, and 38.9% of general population) (*Figure 1*). A greater proportion of intervention patients were female and had a care home stay compared to the other groups. Another major difference between the groups was that 38.5% of the general population were found to have no hospital inpatient stay records in the past 5 years compared to 4.6% in the intervention group, hence had no Charlson score. Greater proportions of high-risk' patients had moderate to severe Charlson comorbidity scores than intervention patients, and only 3.2% of the general population had severe scores. The intervention group had an average of 6.6 monthly prescriptions (BNF classes) with 21.7% of the group having more than ten prescriptions, while the general population had 4.2 on average with only 3.0% having more than ten prescriptions.

Other differences were observed in ethnicity, deprivation, smoking status, and having lived alone, however one of the greatest differences is in the missing values for these variables. There are much lower numbers of missing values for the intervention and high-risk groups, indicating that they have greater interaction with the healthcare systems where these data are recorded. Excluding patients with missing values, 41.5% of intervention patients were recorded as having lived alone, compared to 19.6% of high-risk patients and 12.5% of the general population. Among ethnicity, smoking status and deprivation, when excluding patients with missing values, such great differences were not observed between all three groups.

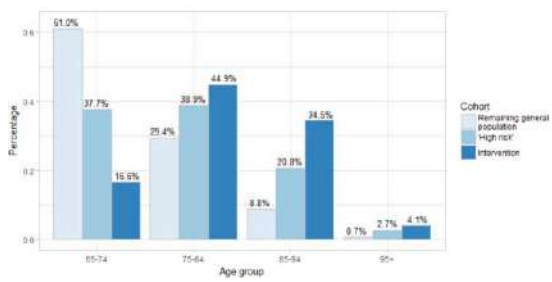


Figure 1– Age distribution of population aged 65 and over

Hospital activity differences between groups

All reported admission rates account for deaths as previously described. In the year prior to ECT implementation, the general population (all patients) had an admission rate of 0.17 emergency admissions, while the other groups had much higher admissions (0.68 in the intervention group and 0.81 in the high-risk group).

This means that prior to implementation of ECT the admission rate in the intervention group was 4.0 times higher than that of the general population but 0.8 times lower than that of the high-risk sub-group.

In the year following the implementation of the intervention, the rate of admission in the full general population increased by 27.5%, while both the high-risk and intervention groups had multiple times higher rates (see *Table 2* and *Figure 2*), reflecting a deterioration in health for intervention patients at the time of being admitted to the ECTs.

Two years on from the implementation of ECT (2017), the admission rate for intervention and high-risk patients decreased drastically from the previous year (by 32.0% and 60.8%

respectively), while in the general population a small increase was observed (5.0%).

Table 1 – Characteristics of the intervention group, high-risk and remaining population over 65 years in Forth Valley

Variable	ECT group (n=566) n (%)	High-risk group (n=2,467) n (%)	Remaining population (n=58,434) n (%)
Age, mean (SD)	82 (7.5)	78 (8.4)	73 (7.5)
Female sex	351 (62.0)	1,267(51.4)	26,458 (45.3)
Ethnicity			
White	546 (96.5)	2,407 (97.6)	47,859 (81.9)
Other	3 (0.5)	14 (0.6)	270 (0.5)
Unspecified	17 (3.0)	46 (1.9)	9,053 (15.5)
Null	0 (0.0)	0 (0.0)	1,252 (2.1)
Deprivation quintile[a]			
1	91 (16.1)	415 (16.8)	7,616 (13.0)
2	153 (27.0)	687 (27.8)	13,551 (23.2)
3	98 (17.3)	527 (21.4)	11,787 (20.2)
4	114 (20.1)	493 (20.0)	12,974 (22.2)
5	108 (19.1)	345 (14.0)	12,505 (21.4)
Null	2 (0.4)	0 (0.0)	1 (0.0)
Has lived alone			
Yes	235 (41.5)	483 (19.6)	5,849 (10.0)
No	331 (58.5)	1,978 (80.2)	42,485 (72.7)
Null	0 (0.0)	6 (0.2)	10,100 (17.3)
Care home stay			
Yes	132 (23.3)	481 (19.5)	2,956 (5.1)
No	434 (76.7)	1,986 (80.5)	55,478 (94.9)
Smoking status			
Yes	54 (9.5)	131 (5.3)	760 (1.3)
Ex-smoker	95 (16.8)	323 (13.1)	1,381 (2.4)
No	222 (39.2)	439 (17.8)	2,900 (5.0)
Null	195 (34.5)	1,574 (63.8)	53,393 (91.4)
Charlson comorbidity score group[b]			
No ICD-10 codes[b]	22 (3.9)	0 (0.0)	22,477 (38.5)
0	87 (15.4)	251 (10.2)	18994 (32.5)
1-2	196 (34.6)	840 (34.0)	11704 (20.0)
3-4	129 (22.8)	720 (29.2)	3386 (5.8)
≥5	132 (23.3)	656 (26.6)	1873 (3.2)
Average prescriptions[c], mean (SD)	6.6 (3.3)	6.4 (3.4)	4.1 (2.8)
<5	161 (28.4)	762 (30.9)	36,210 (63.7)
5-10	333 (58.8)	1405 (57.0)	20,677 (36.3)
>10	72 (12.7)	300 (12.2)	1,547 (2.7)

[a]*Scottish Index of Multiple Deprivation (SIMD) quintile (1=within most deprived fifth of population, 5=within least deprived fifth)*

[b]*Charlson comorbidity score groups: 0 (No comorbidities identified) 1-2 (Mild), 3-4 (Moderate), ≥5 (Severe), from ICD-10 codes in past 5 years*

[c]*Average monthly number of prescription classes (BNF paragraphs) in the year prior to ECT implementation (2015)*

Table 2 – Emergency inpatient admission rate by population sub-group by with number alive each year as denominator

Population sub-group	2014	2015	2016	2017
Intervention patients[a] (n=566)	0.408	0.678	1.977	1.345
High-risk patients[b] (n=2,467)	0.557	0.808	3.561	1.396
Remaining population (n=58,434)	0.123	0.139	0.095	0.191
All patients (n=60,901)	0.143	0.171	0.218	0.229

[a]*Receiving ECT intervention between Jan 2016-May 2017*
[b]*With ≥ 2 emergency admissions in 2016*

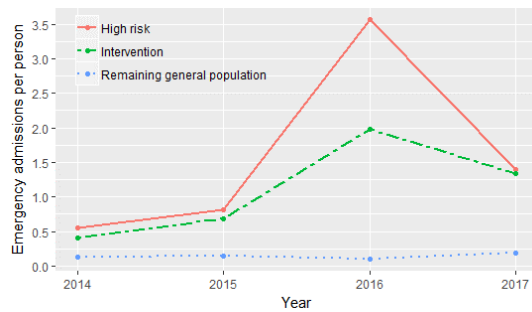


Figure 2– Trend in number of emergency admissions for population aged 65 and over in Forth Valley by sub-group

Discussion

The results of the study confirm that the characteristics of intervention patients differ significantly from those of the general elderly population in NHS Forth Valley. In comparison, the intervention group are older, a greater proportion were female, have higher comorbidity scores, higher prescriptions, four times higher hospital admission rates in the year prior to the intervention and a much higher proportion of intervention patients have lived alone or had a care home stay (proxy for high functional dependency). These characteristics, among others, have previously been identified as being associated with the need for home care [15]. A previous systematic review has also identified older age and Charlson score as risk factors for hospitalisations in community-dwelling elderly patients [16].

Compared to the high-risk patients, intervention patients were older, a greater proportion were female, a higher proportion had lived alone or had a care home stay but had a lower proportion of moderate to severe comorbidity scores. The difference that was most evident in variables with missing values was that the intervention and high-risk groups have much lower frequencies of missing values, indicating greater interaction with the healthcare system.

Overall, ntervention patients had lower admission rates than the high-risk group, however, their patterns of hospital activity were similar. In the year after ECT implementation (2016), both intervention and high-risk patients experienced multiple times higher emergency admission rates, followed by a dramatic reduction (rates including mortality) in the subsequent year; in the general population, a small decrease was observed. These results have been observed in a similar comparison of high-risk patients as previously noted [6]. Similarly, Roland et al. found a drastic reduction in emergency admissions per person in the high-risk patient group (a 75.0% decrease compared to 69.5% decrease in our study when including mortality) [6].

Hence, as Roland et al also found, admissions after the intervention should be compared with a control group satisfying the same criteria that define the intervention group (which can be approximated by their characteristics) [6].

Limitations

Due to the retrospective nature of this study, it is subject to a number of limitations. Retrospective analysis is always limited by the fact that the data may not have been collected for research purposes. For example, input options in MiDIS for discharge reasons can differ according to the user profile, hence this required standardisation through the consultation process with ECT. This study is also limited by data availability. The intervention patients are often described as frail and elderly. Frailty has been difficult to define but has been characterised by physical function, gait speed and cognition [17], which are not routinely collected for the full FV population. Hence, these variables may define the intervention patients, but we have been unable to investigate them. The Charlson comorbidity measure used in this study also has its limitations due to its reliance on ICD-10 codes which in FV are held in hospital records. Hence comorbidity information was not available for patients with no hospital records. In addition, the deterministic data linkage by personal identifiers used in this study is limited in that identifiers are subject to recording errors, hence links can be missed [11]. For prescribing data, ISD was able to confirm that in NHS FV personal identifiers were captured correctly for 96.3% of prescribed items between 2015-17.

Conclusions

In conclusion, this analysis has enabled the identification of defining characteristics of the intervention patients compared to both high-risk patients and the general elderly population, and has highlighted the differences in hospital activity between them. The intervention group are significantly different to the general elderly population, but share some similarities with high-risk patients. The identified characteristics will aid the selection of an appropriate control group, in the evaluatation of the effectiveness of the intervention, which currently ongoing.

The results also highlighted that high-risk patients experience reduced hospital admissions after a period of increased admissions, which has also been shown in other studies. Overall, the analysis indicates that a carefully selected control group is required for the evaluation of the intervention.

Acknowledgements

This research is supported by a University of Strathclyde / Capita Strategic Technology Partnership PhD scholarship (2016-2019). We would like to acknowledge practical support provided by NHS FV, the ‘Closer to Home’ team and ISD.

References

[1] S. Shepperd, S. Iliffe, H.A. Doll, M.J. Clarke, L. Kalra, A.D. Wilson, and D.C. Gonçalves-Bradley, Admission avoidance hospital at home, *Cochrane Database Syst. Rev.* (2016).

[2] J. Young, The development of intermediate care services in England., *Arch. Gerontol. Geriatr.* **49 Suppl 2** (2009) S21-5.

[3] J.C. van Haastregt, J.P. Diederiks, E. van Rossum, L.P. de Witte, and H.F. Crebolder, Effects of preventive home visits to elderly people living in the community: systematic review., *BMJ.* **320** (2000) 754–758.

[4] M. Campbell, R. Fitzpatrick, A. Haines, A.L. Kinmonth, P. Sandercock, D. Spiegelhalter, and P. Tyrer, Framework for design and evaluation of complex interventions to improve health., *BMJ.* **321** (2000) 694–6.

[5] A. Steventon, M. Bardsley, J. Billings, T. Georghiou, and G.H. Lewis, The Role of Matched Controls in Building an Evidence Base for Hospital-Avoidance Schemes: A Retrospective Evaluation, *Health Serv. Res.* **47** (2012) 1679–1698.

[6] M. Roland, M. Dusheiko, H. Gravelle, and S. Parker, Follow up of people aged 65 and over with a history of emergency admissions: analysis of routine admission data., *BMJ.* **330** (2005) 289–92. doi:10.1136/bmj.330.7486.289.

[7] NHS Scotland, The Scottish Government, and COSLA, Reshaping Care for Older People: A Programme for Change 2011-2021, 2011.

[8] National Records of Scotland, Mid-year population estimates: Scotland and its NHS Board areas by single year of age and sex: 1981 to 2017, *Natl. Rec. Scotl.* (2017). https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates/population-estimates-time-series-data (accessed September 24, 2018).

[9] M.C. Martin, M.-M. Bouamrane, K. Kavanagh, and P. Woolman, Preventing frail and elderly hospital admissions: developing an evaluation framework for the “Closer to Home” quality improvement programme in NHS Forth Valley, in: Proc. 2017 Int. Conf. Digit. Heal. - DH ’17, 2017: pp. 231–232.

[10] A. Sayers, Y. Ben-Shlomo, A.W. Blom, and F. Steele, Probabilistic record linkage., *Int. J. Epidemiol.* **45** (2016) 954–64.

[11] M. Fleming, B. Kirby, and K.I. Penny, Record linkage in Scotland and its applications to health research, *J. Clin. Nurs.* **21** (2012) 2711–2721.

[12] V. Sundararajan, T. Henderson, C. Perry, A. Muggivan, H. Quan, and W.A. Ghali, New ICD-10 version of the Charlson Comorbidity Index predicted in-hospital mortality, *J. Clin. Epidemiol.* **57** (2004) 1288–1294.

[13] J.O. Wasey, icd: Comorbidity Calculations and Tools for ICD-9 and ICD-10 Codes. R package version 3.2.1., (2018). https://cran.r-project.org/package=icd.

[14] S.L. Brilleman, and C. Salisbury, Comparing measures of multimorbidity to predict outcomes in primary care: a cross sectional study., *Fam. Pract.* **30** (2013) 172–8.

[15] K.H. Bowles, M.D. Naylor, and J.B. Foust, Patient Characteristics at Hospital Discharge and a Comparison of Home Care Referral Decisions, *J. Am. Geriatr. Soc.* **50** (2002) 336–342.

[16] S.K. Inouye, Y. Zhang, R.N. Jones, P. Shi, L.A. Cupples, H.N. Calderon, and E.R. Marcantonio, Risk factors for hospitalization among community-dwelling primary care older patients: development and validation of a predictive model., *Med. Care*. **46** (2008) 726–31.

[17] S.A. Sternberg, A.W. Schwartz, S. Karunananthan, H. Bergman, and A. Mark Clarfield, The Identification of Frailty: A Systematic Literature Review, *J. Am. Geriatr. Soc.* **59** (2011) 2129–2138.

Address for correspondence

Maria Cristina Martin, E-mail: cristina.martin@strath.ac.uk

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190285

Improving Adherence to Clinical Pathways Through Natural Language Processing on Electronic Medical Records

Noa P. Cruz[a], Lea Canales[a], Javier García Muñoz[b], Bernardino Pérez[b], Ignacio Arnott[c]

[a] *Medsavana S.L., Madrid, Spain*
[b] *Servicio de Salud de Castilla-La Mancha, Castilla-La Mancha, Spain*
[c] *Accenture, Madrid, Spain*

Abstract

This paper presents a pioneering and practical experience in the development and implementation of a clinical decision support system (CDSS) based on natural language processing (NLP) and artificial intelligence (AI) techniques. Our CDSS notifies primary care physicians in real time about recommendations regarding the healthcare process. This is, to the best of our knowledge, the first real-time CDSS implemented in the Spanish National Health System. We achieved adherence rate improvements in eight out of 18 practices. Moreover, the provider's feedback was very positive, describing the solution as fast, useful, and unintrusive. Our CDSS reduced clinical variability and revealed the usefulness of NLP and AI techniques for the evaluation and improvement of health care.

Keywords:

Critical pathways; decision support systems, clinical; natural language processing

Introduction

A clinical pathway is a structured multidisciplinary care plan tool built on evidence-based medicine, which states essential goals and key elements in the management of a well-defined group of patients [1]. Although there is no single and widely accepted definition of a clinical pathway, the aim of these tools is to improve quality of care, reduce risks, improve patient satisfaction, and promote efficient use of healthcare resources [2].

The benefits of implementing clinical pathways in routine clinical practice are well documented and include increased hospital efficiency [3; 4], decreased length of hospital stay [5], and lowered costs [6]. In order to obtain these benefits, implementations of clinical pathways should comply with a minimum set of basic principles [7]; the lack of implementation of these principles by healthcare professionals negatively impacts patient care [8-10].

Computerized approaches are considered promising tools for increasing adherence to clinical pathways [11]. For example, adherence to clinical guidelines for acute decompensated heart failure [12] and glucose regulation in an intensive care unit [13] were successfully improved using computerized pathways in user-friendly formats. Other studies implemented a CDSS that sends computerized clinical reminders to professionals [14; 15]. Yet another evaluated adherence to acute bacterial rhinosinusitis guidelines in three phases: 1) inform each participant about personal adherence rate, 2) perform an educational intervention about clinical pathways, and 3) introduce a CDSS [16]. However, a study that investigated whether the redesign of a CDSS leads to a more appropriate prescription of antimicrobials was not able to detect the expected improvement [17]. Obviously, many factors, such as poor usability or lack of follow-up, can have a negative impact on the desired outcome. A successful CDSS implementation needs to detect and eliminate these factors.

Clinical pathway adherence can be measured from the patients' perspective (adherence to treatments) [18-20] or the professionals' perspective (adherence to clinical pathways) [21-24]. However, what these studies have in common is that the adherence was evaluated using questionnaires, a process which requires significant amounts of time and resources. A thoughtful implementation of healthcare information technology (HIT) can greatly improve the efficiency of clinical pathways [25].

Inspired by the discussion about the use of clinical practice guidelines in primary care in Spain [26], we present a real-time CDSS to improve adherence rates to clinical pathways. The real-time CDSS proposed was developed by Accenture[1] and integrates Savana's[2] EHRead Technology which automatically extracts valuable medical information from unstructured free text information contained in electronic health records (EHRs). Our results demonstrate improvements in clinical adherence due to the use of the real-time CDSS. The real-time CDSS reduces the variability of healthcare and establishes a method for measuring adherence. It also helps to understand the difficulties in implementing clinical pathways and how to overcome them.

Methods

Savana System

Savana [27] was founded in 2014 with the goal of using AI to improve the quality of health services. Savana has developed EHRead, a powerful technology that applies the latest NLP, machine learning and deep learning techniques to analyze the unstructured free text information written in millions of EHRs and automatically extract highly valuable medical information. The pipeline consists of different modules performing tasks such as sentence segmentation, tokenization, text normalization, acronym disambiguation, negation detection, and a multi-dimensional ranking scheme. To execute these tasks, the pipeline combines linguistic knowledge, statistical evidence and state-of-the-art continuous vector representations of words and documents in the clinical domain learned via shallow neural networks. The Savana Manager software is the first medical linguistic engine that applies this concept to the Spanish language, converting the

[1] https://www.accenture.com/
[2] https://www.savanamed.com/

valuable information contained in EHR free text into an interpretable and user-friendly format. A general overview of the CDSS interface is shown in Figure 1.

Figure 1 – Savana Manager Interface

Implementation and Evaluation Process

The real-time CDSS was implemented and evaluated in the primary care area of Castilla-La Mancha's Health Service (SESCAM)[3], serving a region in Spain with more than two million inhabitants. 24 healthcare centers with a total of 86 physicians (general practitioners and pediatricians) participated in this pioneering experience.

The CDSS proposed consists of a real-time tool developed by Accenture which continuously processes the notes taken by professionals during their medical consultation. This technology integrates Savana's EHRead Technology whose objective is to extract and identify the most important clinical information from EHRs. The CDSS warns professionals when the case meets a criteria for a recommendation.

In order to assess the ability of this CDSS to improve clinical pathway adherence, a well-defined methodology was carried out. This methodology consists of the following steps:

1. Translation of clinical pathways into rules by the Savana medical team. These rules define which clinical terms are relevant for each pathway and how they relate to each other.
2. With these rules and the information automatically extracted by Savana Manager, the professionals' adherence to the different clinical pathways is evaluated. To achieve this goal, a dataset of more than 2.5 million documents from emergency and primary care services, archived during a time period ranging from 01/01/2016 to 31/05/2017, was analyzed. We refer to this as the first measurement period.

   From this evaluation, those clinical pathways that had less than 90% compliance were translated into recommendations by our medical team. These recommendations were implemented in a real-time alert system in an attempt to improve adherence rates. Table 1 shows the 18 recommendations defined and their degree of compliance in clinical practice in the first measurement period.
3. Once the proposed CDSS was implemented, an additional adherence evaluation was performed. In this second measurement period, a dataset of more than 345,000 clinical documents, archived during a time period ranging from 27/11/2017 to 31/12/2017, was analyzed in order to assess the benefits of our CDSS.

To quantify the changes in adherence rates to clinical pathways between the first and second measurement periods, we compared the number of documents in which the patients' cases followed the correct clinical pathway before implementation with the number correct after the implementation.

We tested whether the difference between the two proportions (adherence rates before and after implementation) was significant using a two-proportion z-test, which is calculated as:

$$Z = \frac{P_2 - P_1}{\sqrt{\bar{P} \cdot (1 - \bar{P}) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $\bar{P}$ is the proportion of successes over all of our samples combined:

$$\bar{P} = \frac{n_1 \cdot P_1 + n_2 \cdot P_2}{n_1 + n_2}$$

P_1 and P_2 are the adherence rates of samples 1 (before implementation) and 2 (after implementation); n_1 and n_2 are the respective sample sizes.

The null hypothesis H_0 is that there is no improvement in the adherence after implementation:

$$H_0: P_2 <= P_1$$
$$H_1: P_2 > P_1$$

The p-value is equal to 1−Z, assuming the standard normal cumulative distribution. H_0 is rejected if the p-value is lower than 0.05.

Results

Adherence rates to clinical pathways improved in eight out of 18 recommendations when our real-time CDSS was employed, achieving 100% compliance in some cases (Table 2, Figure 2). The improvement was statistically significant in three cases ($p < 0.05$). Considering that this was a preliminary study, these results are promising. The greatest increase in degree of improvement (76%) occurred in recommendation #10 ("SSRIs should not be prescribed to patients with chronic lower back pain unless they are also suffering from depression"). This means that before implementation of our CDSS, professionals often prescribed the use of SSRIs in cases for which it was not clinically recommended; afterwards, the number of these cases decreased dramatically.

The remaining ten recommendations could not be evaluated due to one of two main reasons: 1) these were recommendations for referrals to another service or requests for evidence which require a long period of time to check whether the referral or follow-up in question was requested, or 2) the volume of data was insufficient to make an analysis and draw conclusions. Although one month is a short period of time to evaluate the impact of a CDSS, subsequent evaluations will enrich these results. Importantly, this preliminary assessment made it possible to come up with an initial measure of the CDSS's effectiveness.

[3] See http://sescam.castillalamancha.es/

Table 1– Set of recommendations implemented and their degree of compliance before CDSS implementation

	Recommendation	Degree of compliance (%)
Osteoarthritis of the Hip and Knee	1. Glucosamine and chondroitin sulphate should not be routinely given to patients with hip or knee OA, as there are no relevant clinical benefits	84
	2. The routine use of NSAIDs is not recommended in patients with OA of the hip or knee who have coexisting cardiovascular pathologies such as heart failure or ACS	88
Anxiety	3. Patients with post-traumatic stress disorder should be referred to the mental health team	83
	4. Patients with anxiety should be prescribed pregabalin rather than benzodiazepines due to improved tolerance, fewer side effects, and the possibility of long-term use	9
Headache	5. Fluoxetine doses should not be divided due to a half-life of up to 72h	86
	6. All patients over 50 with headache and temporal arthritis should be referred to a specialist	83
	7. Patients suffering from cluster headaches should be advised to avoid alcohol	2
Dyspepsia	8. Patients over 55 years old with dyspepsia should be referred for a gastroscopy	61
Lower Back Pain	9. Physiotherapy is the treatment of choice for acute lower back pain	42
	10. SSRIs should not be prescribed to patients with chronic lower back pain unless they are also suffering from depression	24
Heart Failure (HF)	11. The use of antialdosterones is recommended for all patients with NYHA stage II-IV HF and a LVEF of < 35%	75
	12. The follow-up for HF patients should include a chest X-ray at least once a year in order to evaluate the cardiothoracic ratio	71
Urinary Tract Infection (UTI)	13. Patients with prostatitis should have a urine sample sent for microscopy and culture	44
	14. Fosfomycin and co-amoxiclav should be avoided when treating UTIs in young men (14-45 years old)	82
	15. Pregnant women diagnosed with a UTI should have a urine sample sent for microscopy and culture	50
Heart Murmur	16. Patients with a heart murmur and a personal or family history of cardiac illness should be referred to a specialist	72
Osteoporosis	17. Calcium should not be routinely given to patients under 70 years old who are diagnosed with osteoporosis	68
	18. Pharmacological treatment (biphosphonates) should not be routinely given to patients over 80 years diagnosed with osteoporosis	84

OA=Osteoarthrosis; NSAID=Non-Steroid Anti-Inflammatory Drug; ACS=Acute Coronary Syndrome; SSRI=Selective Serotonin Reuptake Inhibiting Drug; NYHA=New York Heart Association; LVEF=Left Ventricular Ejection Fraction.

In order to evaluate how our CDSS affected physicians, we monitored the number of alerts/recommendations on a daily basis. In the second measurement period, a total of 534 recommendations were provided, which translated into 15.7 alerts per day. This means that, on average, each of the 86 physicians that participated in this experiment received 1.8 alerts per day. This shows that our real-time CDSS is non-intrusive and is therefore highly unlikely to produce alert fatigue.

Table 2 – Comparison of the degree of compliance with recommendations before and after the implementation of the CDSS

Rec #	Compliance % (Before)	Compliance % (After)	Improvement %
4	9	16	7 (p=0.145)
10	24	100	76* (p=0.01)
12	71	87	16* (p=0.01)
13	44	71	27 (p=0.053)
14	82	100	18 (p=0.108)
15	50	100	50* (p=0.013)
17	68	73	5 (p=0.314)
18	84	100	16 (p=0.096)

Statistically significant improvements (*t*-test, $p < 0.05$) are starred (*).
The term "before" refers to the first measurement period (01/01/2016 - 31/05/2017).
The term "after" refers to the second measurement period (27/11/2017 - 31/12/2017).
The period between 01/06/2017 and 26/11/2017 was used to extract the data provided by the healthcare centers and to train and configure the CDSS.

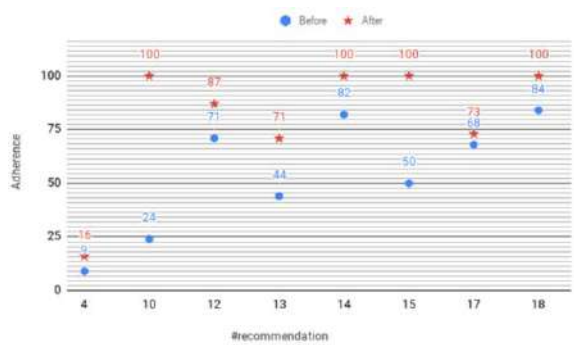


Figure 2 – Graphical representation of the improvement in adherence rates between the two measurement periods

Finally, a presidential session with our team and the participating physicians was carried out to directly obtain feedback from the users. The general response was very positive; the solution was described as fast, very useful and unintrusive, since it did not affect the dynamics of their work.

Discussion

The adoption of AI and big data in healthcare is increasing since numerous studies have shown that these techniques help to solve a variety of problems for patients, healthcare centers, and the healthcare industry in general. These new applications

are more than just technological and analytical tools; they transform healthcare ecosystems by interactively connecting medical data, personalized medicine, and AI with clinical staff to improve patient care.

A new example of this ongoing transformation is the real-time CDSS presented in this paper. In fact, the results are very promising, although their preliminary nature means they must be interpreted with caution. An improvement in adherence rates was achieved in eight of the 18 established recommendations, this improvement being statistically significant in three of them. An important feature of our CDSS is its ability to adapt recommendations over time, thereby achieving proactive management of best practices.

One limitation of our project – indeed, any project working with EHRs – is EHR incompleteness. This can sometimes result in an insufficient amount of information for meaningful processing and measurement.

Our CDSS is currently in operation, and a more complete analysis will be provided in the near future. We will extend the project to the rest of the Healthcare Centers of Castilla-La Mancha and add recommendations for nurses. We will also add the capability to deactivate recommendations that achieve a degree of compliance of greater than 90%. Finally, we will enable professionals to add their own recommendations and follow their evolution in their adherence to clinical pathways.

Conclusions

This paper presents the first real-time CDSS based on NLP and AI techniques implemented in the Spanish National Health System. It allows measurement and improvement of the adherence of primary care physicians to the clinical pathways established by the Health Service. This CDSS helps professionals in their decision-making process and reduces the variability of clinical practice in the healthcare centers, having a very positive effect on overall quality of care.

Acknowledgements

We thank medical writer Ana Lopez-Ballesteros for her help in writing the paper.

References

[1] L. Kinsman, T. Rotter, E. James, P. Snow, and J. Willis, What is a clinical pathway? Development of a definition to inform the debate, *BMC Med* **8** (2010), 31.

[2] A.K. Lawal, T. Rotter, L. Kinsman, A. Machotta, U. Ronellenfitsch, S.D. Scott, D. Goodridge, C. Plishka, and G. Groot, What is a clinical pathway? Refinement of an operational definition to identify clinical pathway studies for a Cochrane systematic review, *BMC Med* **14** (2016), 35.

[3] J.M. Villar del Moral, V. Soria Aledo, A. Colina Alonso, B. Flores Pastor, M.T. Gutierrez Rodriguez, J. Ortega Serrano, P. Parra Hidalgo, and S. Ros Lopez, Clinical Pathway for Thyroidectomy, *Cir Esp* **93** (2015), 283-299.

[4] J.G. Yetzer, P. Pirgousis, Z. Li, and R. Fernandes, Clinical Pathway Implementation Improves Efficiency of Care in a Maxillofacial Head and Neck Surgery Unit, *J Oral Maxillofac Surg* **75** (2017), 190-196.

[5] S.D. Pearson, S.F. Kleefield, J.R. Soukop, E.F. Cook, and T.H. Lee, Critical pathways intervention to reduce length of hospital stay, *Am J Med* **110** (2001), 175-180.

[6] M.A. Bryan, A.D. Desai, L. Wilson, D.R. Wright, and R. Mangione-Smith, Association of Bronchiolitis Clinical Pathway Adherence With Length of Stay and Costs, *Pediatrics* **139** (2017).

[7] A.E.J. Ruiz López P, Ferrándiz Santos J., El diseño de la calidad: la gestión de vías clínicas en el contexto de planes de calidad, *JANO* **65** (2004), 75-80.

[8] J.J. Gomez-Doblas, [Implementation of clinical guidelines], *Rev Esp Cardiol* **59 Suppl 3** (2006), 29-35.

[9] M.D. Cabana, C.S. Rand, N.R. Powe, A.W. Wu, M.H. Wilson, P.A. Abboud, and H.R. Rubin, Why don't physicians follow clinical practice guidelines? A framework for improvement, *Jama* **282** (1999), 1458-1465.

[10] J. van de Klundert, P. Gorissen, and S. Zeemering, Measuring clinical pathway adherence, *J Biomed Inform* **43** (2010), 861-872.

[11] S. Ferrante, S. Bonacina, G. Pozzi, F. Pinciroli, and S. Marceglia, A Design Methodology for Medical Processes, *Applied clinical informatics* **7** (2016), 191-210.

[12] N.J. Gardetto, K. Greaney, L. Arai, A. Brenner, K.C. Carroll, N.M. Howerton, M. Lee, L. Pada, M. Tseng, and A.S. Maisel, Critical pathway for the management of acute heart failure at the Veterans Affairs San Diego Healthcare System: transforming performance measures into cardiac care, *Crit Pathw Cardiol* **7** (2008), 153-172.

[13] E. Rood, R.J. Bosman, J.I. van der Spoel, P. Taylor, and D.F. Zandstra, Use of a computerized guideline for glucose regulation in the intensive care unit improved both guideline adherence and glucose regulation, *J Am Med Inform Assoc* **12** (2005), 172-180.

[14] F.O. Kooij, T. Klok, B. Preckel, M.W. Hollmann, and J.E. Kal, The effect of requesting a reason for non-adherence to a guideline in a long running automated reminder system for PONV prophylaxis, *Applied clinical informatics* **8** (2017), 313-321.

[15] S.J. Wu, M. Lehto, Y. Yih, J.J. Saleem, and B.N. Doebbeling, Relationship of estimated resolution time and computerized clinical reminder adherence, *AMIA Annu Symp Proc* (2007), 334-338.

[16] K.L. Smith, D. Tran, and B.L. Westra, Sinusitis Treatment Guideline Adherence in the E-Visit Setting: A Performance Improvement Project, *Appl Clin Inform* **7** (2016), 299-307.

[17] M.T. Baysari, J. Del Gigante, M. Moran, I. Sandaradura, L. Li, K.L. Richardson, A. Sandhu, E.C. Lehnbom, J.I. Westbrook, and R.O. Day, Redesign of computerized decision support to improve antimicrobial prescribing. A controlled before-and-after study, *Appl Clin Inform* **8** (2017), 949-963.

[18] J. González-Bueno, M.D. Vega-Coca, A. Rodríguez-Pérez, M.D. Toscano-Guzmán, C. Pérez-Guerrero, and B. Santos-Ramos, Intervenciones para la mejora de la adherencia al tratamiento en pacientes pluripatológicos: resumen de revisiones sistemáticas, *Atención Primaria* **48** (2016), 121-130.

[19] L.d.l.Á. Martín Alfonso, J.A. Grau Ábalo, and A.D. Espinosa Brito, Marco conceptual para la evaluación y mejora de la adherencia a los tratamientos médicos en enfermedades crónicas %J Revista Cubana de Salud Pública, **40** (2014), 222-235.

[20] V.V.M.-H.G.G.G.J.G. Roales-Nieto, Aplicación de un programa de mejora de la adherencia en pacientes hipertensos debutantes, *Atención Primaria* **47** (2015), 83-89.

[21] M. Abd Rahman, R. Ahmad Zaki, R. Sarimin, M.I. Ariff, Z. Suli, M. Mahmud, K. Hong Bee, C. Anthonysamy, A. Abdul Rahim, B. Singh Gill, S. Rudra Deva, A.F.

Abdullah Sani, E.Z. Romli, I.M. Mohamed Ghazali, M.A. Mohd Yusof, N. Ahmad Lutfi, S.E. Shuib, N. Mohd Darus, R. Bakri, and Yahya, Adherence to Clinical Practice Guidelines (CPG) management of dengue infection in adults (revised 2nd edition), *PLoS One* **12** (2017), e0184559.

[22] N.S. Bauer, A.E. Carroll, C. Saha, and S.M. Downs, Computer Decision Support Changes Physician Practice But Not Knowledge Regarding Autism Spectrum Disorders, *Appl Clin Inform* **6** (2015), 454-465.

[23] M. Fragoso Marchante, A. Espinosa Brito, G. Álvarez Amador, I. González Morales, J. Bernal Muñoz, and M. Mosquera Fernández, Adherencia a las guías de prácticas clínicas sobre neumonía adquirida en la comunidad y su relación con la mortalidad: un problema sociomédico %J MediSur, **8** (2010), 49-56.

[24] A.F.H. I.S. Carballosa, A.M.H. Cardoso, J.J.N. López, and R.T. Peña., ADHERENCIAS A LAS GUÍAS DE BUENAS PRÁCTICAS CLÍNICAS DE ANGINA INESTABLE. SERVICIO DE CARDIOLOGÍA.CIENFUEGOS 2004, in, 2004.

[25] W. Dong and Z. Huang, A Method to Evaluate Critical Factors for Successful Implementation of Clinical Pathways, *Applied clinical informatics* **6** (2015), 650-668.

[26] J. Gene-Badia, P. Gallo, J. Cais, E. Sanchez, C. Carrion, L. Arroyo, and M. Aymerich, The use of clinical practice guidelines in primary care: professional mindlines and control mechanisms, *Gac Sanit* **30** (2016), 345-351.

[27] L.T. Espinosa-Anke, J.; Pardo, A.; Medrano, I.; Ureña, A.; Salcedo, I.; Saggion, H. , Savana: A Global Information Extraction and Terminology Expansion Framework in the Medical Domain *Procesamiento del Lenguaje Natural* **57** (2016), 23-30.

Address for correspondence

Noa P. Cruz, Medsavana S.L.

C/ Jiloca 4 - 5 Derecha, 28016 Madrid

+34-91 069 69 03 contact@noacruz.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190286

The Digitization of the ICU: An Evaluation of Usability and Hospital-Wide Acceptability

Racha Dabliz[a], Simon K. Poon[b] , Angus Ritchie[cd], Kevin Kuan[b], Jonathan Penm[a]

[a] The University of Sydney, Faculty of Medicine and Health, School of Pharmacy, Sydney, New South Wales, Australia.
[b] School of Computer Science, University of Sydney, Sydney, NSW, Australia
[c] Concord Clinical School, University of Sydney, Sydney, NSW, Australia
[d] Health Informatics Unit, Sydney Local Health District, Camperdown, NSW, Australia

Abstract

This paper explores the impact of an electronic medication management system (EMMS) on users in an intensive care unit using the Unified Theory and Use of Technology constructs. It also explores the impact of having a consistent EMMS hospital wide, as it is the first Australian hospital to implement the same EMMS hospital wide. The research model was evaluated using survey data from 100 nurses, doctors and pharmacists both within the ICU and externally, to assess the usability and acceptability of the system. Results showed that performance expectancy, effort expectancy, social influence and facilitating condition all correlate with overall user satisfaction. Overall, teams external to the ICU are in strong favor of its implementation whist user acceptance from within the ICU itself is poor.

Keywords:

Electronic Health Records, Intensive Care Units, Medication Systems

Introduction

The Intensive Care unit (ICU) is a complex environment that requires healthcare providers to balance competing tasks and responsibilities in their care for patients. Caring for complex patients requires communication and coordination of multiple healthcare team members and changes in work routines could affect their ability to provide safe, high-quality care. The ICU team consists of a range of staff including nurses, doctors and pharmacists. It involves coordinated patient management by internal intensivists and doctors, as well as external teams to the ICU such as surgeons, infectious diseases specialists, anesthetists, geriatricians and several others. All members require ease of communication, visibility of the medication charts and changes to medication regimens.

The patient record remains the principal instrument for ensuring continuity of care. The ICU is a data-rich environment and there is discordance between the mass of data and the capacity of paper-based documentation, which can lead to major defects in information processing [1]. There are many barriers to efficiency using a paper-based documentation system. Traditionally, on admission to, and discharge from, the ICU, a patient's paper-chart would be ceased, and a new chart would be written up by the ICU doctor. Transcription errors on admission to and from the ICU, as well as illegible handwriting [2,3], time spent on manual data entry [4], low quality of, and frequency of, medication errors [5] are just a few of the challenges that can result in inefficient workflow, medication errors and poor productivity [6]. Electronic medication management systems (EMMS) aim to improve work processes for all end users by presenting medication information that can be easily accessed during a patient's hospital stay.

However, EMMS that are unable to support clinical work-flow efficacy [7] can generate unintentional consequences that can harm patients [8]. Unintended consequences resulting in errors have been variously labeled 'system-related', 'technology-induced,' and 'computer-related' [9,10]. A frequent subgroup is system-related errors arising from the use and functionality of an EMMS which would be unlikely or unable to occur in paper-based medication ordering systems [10]. These are typically caused by the inability of the EMMS to match healthcare work patterns and settings, creating user acceptance barriers. It is therefore necessary to evaluate the claim that an EMMS will enhance the quality of patient care and increase documentation efficiency.

A Case Study

In Australia, the NSW state government eHealth Strategy for 2016–2026 is to develop a digitally-enabled and integrated health system delivering patient-centered health experiences and quality health outcomes [11]. A key component of this strategy is the roll-out of a commercial electronic medication management system referred to locally as- 'eMeds' (Millennium®, current code level 2015.01.25, Cerner Corporation, Kansas City, MO) to replace paper medication charts in general wards in 178 NSW state hospitals. eMeds is a physician order entry system. Prescribing using eMeds involves selecting items from a drop-down menu of predefined order sentences triggered on drug selection. When selecting a medication, it includes order sentences which contain details of the drug, strength, dose, and form, with the option to edit details in the order sentence. All orders are subjected to series of checks including drug allergies and interactions. EMeds is incorporated into the patient electronic medical record which contains all other aspects of patient care such as pathology, imagining etc.

In NSW hospitals, eMeds is being rolled out to general wards, whilst the ICUs are generally implementing an alternate commercial EMMS, known as the electronic record of intensive care or 'eRIC' (MetaVision ICU, iMDsoft®, Tel Aviv, Israel) or are remaining paper-based. This is creating hybrid or dual prescribing system environments. Studies have shown stakeholders believe this practice to have negative impacts on communication, with some users reporting missing patient information [12]. To ensure crucial patient information was not missed, it was proposed that eMeds EMMS be implemented in a 13-bed general ICU of a tertiary hospital in NSW, Australia, to align with the general wards which include surgical, acute and aged care wards. Prior to its implementation the remainder of the 750-bed general hospital had already been using this system for 3 years while the ICU during that time was out of scope, using the paper National Inpatient Medication Chart (NIMC) [13]. Across NSW there are 81 public hospital ICUs [14], 15[11] of these have implemented the alternative eRIC system [15] in the ICU and eMeds throughout the remainder of

their hospital. This is one of the first hospitals in Australia to implement the eMeds EMMS system in the ICU, creating a consistent prescribing system across its facility. However, unlike the remainder of the hospital that is using computerized physician order entry (CPOE) and electronic clinical documentation across all aspects of care, the ICU are only using CPOE for medications and diagnostics tests. Additionally, some orders, such as continuous infusions and blood products, remained on a large-format daily ICU flowsheet at the request of the ICU medical staff.

This provided a unique opportunity to evaluate the impact of this EMMS in the ICU setting, as well as the effect of having the same system facility wide. Despite the coordinated care required amongst teams within and external to the ICU, most studies have focused primarily on a single group rather than the impact across all teams involved in the medication management process. Our research goal is to apply an existing technology acceptance model to evaluate the usability and acceptability of nurses, doctors and pharmacists within the hospital.

Methods

We chose a formative evaluation for this study. This paper focused on the results of a survey that was part of a larger case-based mixed-methods approach, as recent review of evaluation of health care IT recommends methodological pluralism, including both qualitative and quantitative methods [16].

Participants

The evaluators focused on the three main clinical groups (nurses, doctors and pharmacists) in both the ICU and the remainder of the general hospital. Usability evaluation studies need to provide a comprehensive image of usability by focusing on more than one single end-user perspective[17]. The survey link was sent to 70 ICU nurses, 15 ICU doctors and 2 ICU pharmacists. The possible maximum of non-ICU staff was difficult to obtain due to the nature of rotating shifts between nurses and doctors for a single patient's hospital admission. Staff were anonymous and of mixed age, gender, experience and seniority. These groups were chosen as they are central to the operation of the EMMS. Staff external to the ICU were included to determine the hospital wide effect of transitioning from a hybrid prescribing hospital environment of paper-based and an EMMS to a homogenous one. It also allowed for the comparison of the impact on teams involved in the direct use of the system with those on the receiving end of patients transitioning in and out of the ICU. Staff completing the survey were expected to have worked in the ICU and other relevant wards three months prior and during the implementation of eMeds in the ICU.

Survey Design

The survey was developed by the project team and based on the Unified Theory on Acceptance and Use of Technology (UTAUT) to evaluate the acceptance and use of eMeds. The UTAUT [18], was chosen as the framework for the development of the survey as it has been widely applied and empirically tested to investigate factors that could influence individuals to adopt and use technology in various environments [19].

UTAUT integrates eight theories on technology adoption and provides a comprehensive view of the factors related to users' adoption behavior [20]. The main UTAUT constructs are [18]:

- Performance expectancy (PE): ''The degree to which an individual believes that using the system will help him or her attain gains in job performance.''
- Effort expectancy (EE): "The degree of ease associated with the use of the system."
- Facilitating conditions (FC): ''the degree to which an individual believes that an organizational and technical infrastructure exists to support use of the system."
- Social influence (SI): "the degree to which an individual perceives that important others believe he or she should use the new system."

According to the UTAUT, PE, EE and SI are theorized to influence behavioral intention to use a technology, while behavioral intention and FC determine technology use [20]. The UTAUT does not specify the methods or parameters to be used, as it is a case by case basis. In this paper we report on the survey outcomes designed against the UTUAT framework.

The survey used a 7-point Likert scale (1 – strongly disagree; 7 – strongly agree). It consisted of 69 questions capturing a range of user feedback questions, however this paper focuses only on data from 33 questions relevant to the UTAUT framework. The survey also had two open ended questions that asked, '*why do you think eMeds in the ICU is or isn't sustainable?*' and *'general comments regarding eMeds'*. Surveys sent to the ICU and non-ICU staff were identical, except for the questions under the effort expectancy construct. Effort expectancy between ICU and non-ICU were not compared directly as the questions under this construct were not in both surveys. Due to the nature of the questions, i.e. *'I find it easy to get the eMEDs system to do what I want it to do,'* which is not specific to the impact of eMeds in the ICU on non-ICU wards, they were removed from the non-ICU staff survey. The surveys were pilot tested on four participants. The phrasing and selection of question involved a methodological trade-off between following established standards and adapting these to adequately fit the case at hand.

Survey Distribution

Three months after the implementation of eMeds in the ICU, the survey was distributed to ICU nurses, doctors and pharmacists. One hundred ICU staff were emailed an online survey link or given physical copies. The second survey was sent to the nurses, doctors and pharmacists on the wards that ICU patients are commonly transferred to. Two reminder emails were sent out within 1-month after the initial email. Non-ICU staff completing this survey were expected to have been involved in the care of a patient transferred from the ICU. Participants were provided a coffee voucher upon survey completion.

Analysis

Data from the survey was collected and stored using a Research Electronic Data Capture Tool (REDCap), 8.3.1 (Vanderbilt University, Nashville). It was subsequently analyzed for descriptive, correlation and Cronbach's alpha statistics using the Statistical Package for the Social Sciences 22 (SPSS) (IBM Corp. Released 2016, Version 24.0. Armonk, NY: IBM Corp).

Results

A total of 100 surveys were received. Table 1 gives the response rates for the individual wards and professions.

Overall User Satisfaction

In technology-acceptance research, factors that may influence people's acceptance of systems are typically correlated with (self-reported) usage of systems. Because use of this system was mandatory, the items included in this study were correlated with user's overall assessment of the system. We first examined

the user's general satisfaction with the EMMS by examining responses to the survey item, 'overall, I am satisfied with the way of working with eMeds.' Overall, ICU doctors were dissatisfied, (mean = 1.78, SD =1.72) particularly in comparison to non-ICU doctors (mean = 5.71, SD =1.14). Similarly, ICU pharmacists were less satisfied than non-ICU pharmacists (mean = 3.00, SD=1.41 and mean = 5.22, SD= 1.20 respectively) and ICU nurses less than non-ICU nurses (mean = 4.83, SD= 1.37 and 5.72, SD = 1.49 respectively. Of the 6 groups, non-ICU nurses were found to be the most satisfied. When examining the correlation between overall satisfaction and the UTAUT constructs, overall satisfaction was significantly correlated with all four constructs ($p < 0.01$). Overall satisfaction was positively moderately correlated with performance expectancy (r= 0.499), facilitating condition (r= 0.455) and effort expectancy (ICU, r= 0.361, non-ICU, r=0.463). It was most strongly correlated with social influence (r= 0.510). Suggesting that all 4 constructs play a role in the overall assessment of the system.

Table 1 – Survey Response Rates

	ICU staff	Non-ICU Staff
Profession	Respondents N, (response rate %)	Respondents * (N)
Doctors		
Senior doctors	5, (45%)	3
Junior doctors	4, (80%)	5
Unspecified	0, (0%)	7
Nurses	29, (41%)	36
Pharmacists	2, (100%)	9
Total:	40, (45%)	60

**Unable to obtain response rate % as the possible maximum of non-ICU staff was difficult to obtain due to the nature of rotating shifts between nurses and doctors for a single patient's hospital admission.*

The UTAUT Constructs Reliability

Cronbach's alpha was calculated for each of the constructs and found to be, for performance expectancy, $\alpha = 0.94$ consisting of 3 items, for facilitating condition, $\alpha = 0.93$ consisting of 3 items and for social influence, $\alpha = 0.89$ which consisted of 2 items. For the ICU staff survey, the effort expectancy construct consisted of 3 items with $\alpha = 0.91$. All the alpha values are above the recommended threshold of 0.7 [21] . The reliability of all 4 constructs are deemed satisfactory.

Performance Expectancy (PE)

The non-ICUs overall assessment of performance was satisfaction with the system compared to ICU staff, with a mean difference of -1.70 ($p<0.001$) (Table 2). Statistically significant mean differences between groups within the ICU were unable to be calculated due to the small sample size of individual professions. Of the three groups examined within the ICU, doctors perceived the system as supporting them the least in attaining gains in their job performance (Table 3). However, perception of performance expectancy varied across the three ICU groups. ICU nurses and pharmacists were slightly in favor of the EMMS increasing their job performance, with an average mean response rate of more than 4. Interestingly, all professional groups external to the ICU staff, rated performance expectancy higher than ICU groups and all groups perceived the EMMS useful to some degree in achieving a greater job performance. The highest mean response was from the nurses (mean = 5.81, SD= 1.58), followed closely by the doctors (mean = 5.60, SD= 1.60) and pharmacists (mean = 5.44, SD= 1.13).

Effort Expectancy (EE)

Overall, the ICU staff were neutral in their view on the degree of ease associated with the system (mean = 4.28, SD= 1.71). However, when broken down to the individual groups, table 3 shows that ICU doctors did not believe there was a degree of ease associated with use of the system (mean= 2.0, SD= 1.58). Similarly, to the trend in PE, ICU nurses and pharmacists (table 3) believed there to be a degree of ease associated with the system, with an average mean response rate of more than 4. Externally to the ICU, mean response to the question, 'the features of the eMEDs system meet the needs of my work tasks,' was favourable and had a mean value of 5.33 (1.69) and were positively supported by all three groups (Table 3).

Facilitating Condition (FC)

Unlike the other two constructs, there was consistency in the perception of FC across all groups both within and external to the ICU. Table 2 shows that both the ICU and non-ICU wards believed the hospital provided them with the required implementation and ongoing support for the system. All sub-groups both internal and external to the ICU had a mean average of more than 4. When comparing which group felt they had the most support, of all the groups Table 3 shows that both the ICU and non-ICU pharmacists agreed that the support was greatest. Nursing groups internal and external to the ICU rankings were the next followed closely by externals doctors and internal ICU doctors. The impact of hospital support is consistent across all three groups internal and external to the ICU.

Table 2 – Comparison of the Overall Mean for Each Construct of ICU and Non-ICU staff

	ICU Mean (sd), N	Non-ICU Mean (sd) N= 60 (All items)	Mean difference
PE	4.00 (1.66) N= 40	5.70 (1.51)	-1.70 (t_{78}= 5.08 , P <0.01)
EE	4.28 (1.71) N= 40	5.0 (1.65)	*
FC	5.16 (1.39) N= 37	5.73 (1.54)	-0.56 (t_{81} =1.69, P <0.10)
SI	4.35 (1.73) N= 37	5.73 (1.59)	-1.374 (t_{73} = 3.93, P < 0.01)

** Unable to compare means of EE as questions under the constructs differed between ICU and non-ICU ward. The N varies between each construct for ICU staff due to removal of non-response data. Missing data occurred throughout the survey, but did not exceed 10%.*

Social Influence (SI)

Both the ICU and non-ICU collectively perceived their seniors and colleagues as being in favor of the implementation of the EMMS in the ICU. The mean average of both the ICU and non-ICU group was more than 4 (Table 2). This suggests that department heads have provided managerial support throughout the implementation of the system. However, when broken down to individual group levels, Table 3 shows that the greatest managerial support was held by nurses within and externally to the ICU. Alternatively, whilst overall the ICU believed their seniors to be in support of the system, Table 3 shows that the ICU doctors alone had a contrasting view (mean = 2.44, SD= 1.67)

Across all 4 constructs, Table 2 shows that for the ICU staff, there were overall lower means and higher variations across all 4 constructs. However, non-ICU staff have higher means and lower variability across constructs, indicating there is greater consistency in the support for EMMS from external groups to the ICU, whilst there is less support and greater variability in opinion amongst groups within the ICU.

Table 3 – Comparison of mean responses of individual groups for ICU and non-ICU staff under the four UTAUT constructs.

	ICU STAFF mean (sd)			NON-ICU STAFF mean (sd)		
	Nurses	Doctors N=9	Pharmacists N= 2	Nurses N= 36	Doctors N= 15	Pharmacists N= 9
PE	4.59 (1.30) N=29	2.0 (1.32)	4.50 (0.7)	5.81 (1.58)	5.60 (1.60)	5.44 (1.13)
EE**	4.89 (1.08) N= 29	2.0 (1.58)	5.5 (0.71)	5.69 (1.53)	4.80 (1.90)	4.78 (1.79)
FC	5.31 (1.32) N= 26	4.44 (1.42)	6.50 (0.71)	5.75 (1.80)	5.50 (0.97)	5.89 (1.51)
SI	5.00 (1.30) N=26	2.44 (1.67)	4.50 (0.71)	5.94 (1.56)	5.90 (1.85)	4.78 (1.09)

** *Effort expectancy between ICU groups and NON-ICU groups not directly comparable due to the difference in which asked between each group due to the nature of the questions and their relevance to the ward. The N varies between each construct for ICU nurses due to the removal of non-response data. Missing data occurred throughout the survey, however this did not exceed 10%*

Comments Section

The survey responses to the open-end questions for ICU staff were more commonly negatively swayed. The most common responses given by ICU-doctors focused on the increased time it takes to prescribe medication and the negative impact it has on workflow. Similarly, all three ICU groups commented that whilst eMeds is appropriate for a general ward, this benefit did not extend to the fast-paced ICU environment. Across the three groups within the ICU, nurses had the most positive comments, outlining that it improved workflow and transition of care. Groups external to the ICU also commented that having eMeds within the ICU positively impacted on their own workflow and the transition of care of patients.

Discussion

The first major objective of the introduction of eMeds in the ICU was to optimize workflow and workload. Usability is an intrinsic characteristic of a technology that impacts end-users' interaction with the technology; it leads to higher work efficiency in case of good usability, but in case of poor usability it may also slow down user performance, decrease users' satisfaction, and expose users to use errors [17].

This study reveals a clear difference in overall satisfaction between ICU and non-ICU staff. The ICU believed the eMeds EMMS not to be an appropriate fit for their setting. Task Technology Fit (TTF) focuses on the degree to which systems characteristics match user task needs [12]. Studies have suggested that in the absence of customized interfaces and tailored workflow support, the fit between the information entry and review needs of doctors, and the features offered by EMMS is likely to differ [22]. Previous studies show that the system and its users should be studied together and considered, as both are vital for implementation in order for the process of system adoption to be met with less resistance [23]. As revealed in the comments, the ICU staff believe the EMMS is appropriate for general wards but does not align with the ICU workflow. Suggesting that an EMMS fit for the general hospital, does not necessarily transcend to the ICU setting.

An EMMS should facilitate the aggregation and synthesis of multiple data elements for physicians [22]. Traditional paper-charts involves the extraction of data from various sources, over several pages in the file, and then collate the information. An EMR system, in contrast, makes such data easy to retrieve and, review [22]. As the ICU has continued to adopt the hybrid method of CPOE and paper documentation as well as selected medications, this scattered display of information and various sources continues. Thereby potentially impacting the perception of the benefits of an EMMS by the ICU staff.

General wards external to the ICU are at the stage of optimization of the system to align with workflow, thereby have overcome initial user resistance issues that may arise during the shakedown phase and may have facilitated use. Whereas, currently within the ICU there are continued efforts being employed to make improvements. It is currently unknown if these issues would remain after 6-12 months of using the system in ICU. Evidence suggests that it may take up to 2 years post-implementation until the unit returns to complete stability [24]. The greater satisfaction of ICU pharmacists in comparison to doctors could also explain this as ICU pharmacists were using this same system in the rest of the hospital prior to its implementation in the ICU, whilst the other two groups (doctors and nurses) were not.

Similar to previous studies, these results confirm that groups of professionals react differently when EMMS are implemented, making it difficult to implement a one-EMMS-fits-all across professions and departments. The finding that within the ICU, doctors were least satisfied compared to the nurses and pharmacists in areas related to PE and EE could be attributed to doctors being responsible for the entering of information into the EMMS. Physicians are at the frontline and perform not only knowledge work, such as making decisions and crafting treatment regimen based on patient information, but also data entry. Accordingly, the influence of EMRs on different groups may differ if the technology provides disparate impacts with respect to information review versus information entry [22]. Previous studies found that a low level of usability plays an important role in unsatisfactory implementation of an EMMS which led to disruptions of workflows and accordingly negative impacts on job performance [25,26]. A literature review that looked at the impact of a critical care information system on time spent documenting by nurses and physicians revealed 25% of studies found an increase in time spent charting, 42% found no difference, and 33% of studies reported a decrease [25]. The benefits in effort are more likely reaped by those downstream who are not required to manually enter data.

The study also showed that users put a stronger emphasis on PE and EE, rather than facilitating condition or social influence . This could be attributed to its use being mandatory and had hospital management support. Resulting in the necessary resources and support for its successful implementation. This study also aimed to determine the effect of having the same EMMS hospital-wide rather than the popular dual prescribing environments. It was assumed that introduction of eMeds would favor workflow, workload and be beneficial hospital-wide. The results showed that whilst the general wards perceived having eMeds in the ICU as supporting them in their productivity, and effort, the ICU itself did not believe the system supported their work environment. This is the first Australian study that investigates the user perception from all three key groups within the ICU involved in the use of eMeds. It is also the first to explore the benefits of having the same EMMS facility wide. A key outcome is that external ICU users believe it to create a safer workflow and eases the patient transition throughout the hospital. The significant difference in impact on performance expectancy and effort expectancy between groups within and externally to the ICU could be attributed to the benefits associated with the electronic

environment such as remote access and transparency of the patient medical journey.

Limitations

This paper focuses on the quantitative results of a survey which was part of a larger case-based mixed-methods approach. Concepts and results found from the survey will be triangulated with qualitative and medication safety data. Furthermore, the 45% response rate from ICU staff, the difference in sample size between ICU and non-ICU staff and within sub-groups calls for caution when interpreting the results.

Conclusions

This study aimed to determine the acceptance and usability of an EMMS system in the ICU setting, and the impact of having a homogenous EMMS hospital wide. It found that performance expectancy, effort expectancy, facilitating condition and social influence were all moderately correlated with user's overall satisfaction with the EMMS. It demonstrated that implementation entailed changes in work processes which were challenging for all ICU groups. Whilst user acceptance from within the ICU itself is poor, teams external to the ICU were in strong favor of its implementation. As this is part of a mixed-method case study design which includes triangulation of data, further investigations are being made into the core user-resistance concepts through qualitative evaluation methods.

Acknowledgements

Acknowledgements to the Sydney Local Health District Health Informatics Scholarship for the funding and ethics approval (approval number: LNR/18/CRGH/92) to complete the study.

References

[1] G.D. Martich, C.S. Waldmann, and M. Imhoff, Clinical informatics in critical care, *J Intensive Care Med* **19** (2004), 154-163.
[2] A. Bobb, K. Gleason, M. Husch, J. Feinglass, P.R. Yarnold, and G.A. Noskin, The epidemiology of prescribing errors: the potential impact of computerized prescriber order entry, *Archives of Internal Medicine* **164** (2004), 785-792.
[3] M.J. Hartel, L.P. Staub, C. Röder, and S. Eggli, High incidence of medication documentation errors in a Swiss university hospital due to the handwritten prescription process, *BMC Health Services Research* **11** (2011), 199.
[4] J. Allan and J. Englebright, Patient-Centered Documentation: An Effective and Efficient Use of Clinical Information Systems, *JONA: The Journal of Nursing Administration* **30** (2000), 90-95.
[5] C.A. Bond, C.L. Raehl, and T. Franke, Medication Errors in United States Hospitals, *Pharmacotherapy:The Journal of Human Pharmacology & Drug Therapy* **21** (2001), 1023-1036.
[6] M. Lising and C. Kennedy, A Multimethod Approach to Evaluating Critical Care Information Systems, *CIN: Computers, Informatics, Nursing* **23** (2005), 27-37.
[7] J.J. Saleem, A.L. Russ, A. Neddo, P.T. Blades, B.N. Doebbeling, and B.H. Foresman, Paper persistence, workarounds, and communication breakdowns in computerized consultation management, *International Journal of Medical Informatics* **80** (2011), 466-479.
[8] D. Chou, Health it and patient safety: Building safer systems for better care, *JAMA* **308** (2012), 2282-2282.
[9] P. Rahadhan, S.K. Poon, and L. Land, Understanding unintended consequences for EMR: a literature review, *Studies in health technology and informatics* **178** (2012), 192.
[10] J.I. Westbrook, M.T. Baysari, L. Li, R. Burke, K.L. Richardson, and R.O. Day, The safety of electronic prescribing: manifestations, mechanisms, and rates of system-related errors associated with two commercial systems in hospitals, *Journal of the American Medical Informatics Association : JAMIA* **20** (2013), 1159-1167.
[11] eHealth, ehealth-strategy-for-nsw-health-2016-2026.pdf, in: N.G. Health, ed., 2016.
[12] M. Baysari, L. Richardson, W.Y. Zheng, and J. Westbrook, Implementation of electronic medication management systems in hospitals-A literature scan, in, Australian Commission on Safety and Quality in Health Care, Sydney, 2016.
[13] National Inpatient Medication Chart User Guide, in: ASCQHC, ed., 2016.
[14] Australian hospital statistics, in: A. Government, ed., 2017, p. 27.
[15] Electronic Record for Intensive Care (eRIC), in, NSW Ministry of Health, 2018.
[16] R. Lilford, J. Foster, and M. Pringle, Evaluating eHealth: How to Make Evaluation More Methodologically Robust, *PLoS Medicine U6* **6** (2009).
[17] E. Ammenwerth and M. Rigby, *Evidence based Health Informatics* IOS Press BV. Amsterdam, Netherlands, 2016.
[18] V. Venkatesh, M.G. Morris, G.B. Davis, and F.D. Davis, User Acceptance of Information Technology: Toward a Unified View, *MIS Quarterly* **27** (2003), 425-478.
[19] H. Khechine, S. Lakhal, and P. Ndjambou, A Meta-analysis of the UTAUT Model: Eleven Years Later, *Revue Canadienne des Sciences de l'Administration/Canadian Journal of Administrative Sciences* **33** (2016), 138-152.
[20] M.D. Williams, N.P. Rana, and Y.K. Dwivedi, The unified theory of acceptance and use of technology (UTAUT): a literature review, *Journal of Enterprise Information Management* **28** (2015), 443-488.
[21] J.C. Nunnally, *Psychometric theory McGraw-Hill series in psychology*, McGraw-Hill, The University of Michigan, 1978.
[22] H.K. Bhargava and A.N. Mishra, Electronic Medical Records and Physician Productivity: Evidence from Panel Data Analysis, *Management Science* **60** (2014), 2543-2562.
[23] S.Y. Park, S.Y. Lee, and Y. Chen, The effects of EMR deployment on doctors' work practices: a qualitative study in the emergency department of a teaching hospital, *Int J Med Inform* **81** (2012), 204-217.
[24] A. Donati, V. Gabbanelli, S. Pantanetti, P. Carletti, T. Principi, B. Marini, S. Nataloni, G. Sambo, and P. Pelaia, The Impact of a Clinical Information System in an Intensive Care Unit, *Journal of Clinical Monitoring and Computing* **22** (2008), 31-36.
[25] R.L. Mador and N.T. Shaw, The impact of a Critical Care Information System (CCIS) on time spent charting and in direct patient care by staff in the ICU: A review of the literature, *International Journal of Medical Informatics* **78** (2009), 435-445.
[26] K. Saarinen and M. Aho, Does the implementation of a clinical information system decrease the time intensive care nurses spend on documentation of care?, *Acta Anaesthesiologica Scandinavica* **49** (2005), 62-65.

Address for correspondence
Racha Dabliz
E : rdab8832@uni.sydney.edu.au

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190287

A Digital Health Platform to Deliver Tailored Early Stimulation Programs for Children with Developmental Delays

Raquel da Luz Dias[a], Marcela de Oliveira Lima[b], João Guilherme Bezerra Alves[b], William Van Woensel[a], Asil Naqvi[a], Zahra Take[a], Syed Sibte Raza Abidi[a]

[a] Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia, Canada
[b] Rehabilitation Centre Prof. Fernando Figueira Integral Medicine Institute, Recife, Pernambuco, Brazil

Abstract

Developmental delay is a deviation from the regular development of normative milestones during childhood. Early stimulation is a standardized and straightforward technique to support children with developmental delays (aged 0-3 years) in reaching basic motor skills, which are essential for the execution of everyday activities, such as playing, feeding and locomotion. In doing so, early stimulation reduces the chances of permanent motor impairment, thus allowing the child to live a more functional life. However, outcomes of this treatment depend heavily on the involvement of the family, who are required to continue the early stimulation activities at home on a daily basis. To empower and educate families to administer standardized early stimulation programs at home, we developed an electronic early stimulation program, which provides personalized guidance to parents to administer early stimulation; together with evidence-based clinical decision support to therapists in tailoring ESP to observed needs.

Keywords:

Child development, Mobile health, Patient education

Introduction

Child developmental delay is defined as a deviation from the regular development of normative milestones in the areas of cognitive, language, social, emotional and motor functioning [1]. Globally, developmental delays are the most common chronic condition during early childhood. With the recent outbreak of the Zika virus, which causes microcephaly in newborns limiting their development, delays in child development have become a major public health concern in Brazil [2]. Treatment for developmental delays in children between 0-3 years comprises a standardized exercise program—known as the Early Stimulation Program (ESP) [3] that targets cognitive and motor function development. ESP therapy is administered by a therapist, who determines the extent of the developmental delay, with respect to age-specific normative milestones. From an intervention standpoint, the therapist formulates a child-specific ESP to help the child attain normative, age-specific developmental milestones.

Given the large number of Brazilian children with developmental delays, there exists an *imminent and urgent need* to provide them with comprehensive ESP support. However, ESP therapy is only available at a limited number of specialized therapeutic centers, meaning that many of the affected children do not get appropriate treatment in a timely manner. To address the threat to the child development, in 2016 the Brazilian government launched the Brazilian Early Stimulation Guidelines [3] which aim to guide health professionals about how to provide standardized treatments for developmental delays. Additionally, the government, international organizations, health care centers and communities to uphold children's right to treatment for developmental delays through parent education programs about how to perform ESP to their child. However, these initiatives were not enough to reach the high volume of children who experience developmental delays. Therefore, there is an imminent need to implement a feasible, scalable and accessible solution to addresses the lack of healthcare resources for children with developmental delays, especially in large country like Brazil.

To overcome the issue of timely access to ESP resources, we are pursuing a *home-based ESP delivery* approach that remotely engages and trains the parents of affected children on how to administer early stimulation exercises at home. To train parents on how to perform specific early stimulation exercises, we will deliver short videos (max. 2 minutes) illustrating how to perform the early stimulation exercises through mobile phones. Video-based patient education programs have been shown to be an effective method to improve the patient's knowledge, decision-making ability and self-management efficacy [4]. Based on the video-based educational material, parents will be able to perform the early stimulation exercise at home and also record the child's treatment progress through a mobile app for follow-up consultation.

In this paper, we present our digital health based ESP therapy intervention called the BraziLian Early Stimulation System (BLESS) that is designed to provide early stimulation therapy to children in the age group 0-3 years so that they achieve the five most important age-specific motor development milestones—i.e. head control, rolling, sitting, standing and walking. To engage and educate parents to consistently and effectively pursue their child's ESP, we use the "prepared-informed-motivated" approach, proposed by the WHO Innovative Care for Chronic Conditions (ICCC) framework [5] which aims to empower the child's family to self-manage their child's chronic problems in partnership with professional healthcare providers. We use the developmental motor milestones scale proposed by World Health Organization (WHO) [6] and the International Classification of Functioning and Disabilities (ICF) [7] to assess a child's developmental delay and in turn to provide a ESP for the child. We develop a personalized ESP for the child that takes into account the parents' educational and efficacy levels to administer ESP with different levels of self-management of the child's treatment (i.e. acquisition, fluency, maintenance and

generalization) [8]. To personalize the ESP to the child's needs and parents' efficacy level, we have developed an intelligent ESP personalization model to assist therapists in designing ESP that is tailored to *the child's developmental needs* as well as the *parents' observed relevant skills*.

BLESS incorporates **(a)** a mobile health app that serves as an *accessible and personalized ESP self-management tool* for the child's family. It allows parents to receive the prescribed early stimulation videos and associated educational material and it supports them in administering the prescribed ESP in a home-based setting, as well as tracking parent's adherence to ESP; and (**b**) an *ESP therapy planning and decision support tool* to support therapists to design an evidence-based ESP. The desktop-based therapy planning tool is suitable for fast-paced clinical routines as it assists the therapist to assess the child developmental status using standardized assessment tools as per clinical guidelines; design a personalized ESP for a child based on this assessment; conduct follow-up examinations to monitor the child's developmental progress; overview compliance and progress of their patients with their ESP; and, if adjustment is needed communicate with the child's parents.

BLESS is being deployed at a reference center for Maternal and Child healthcare in the Northeast of Brazil, where there is a high incidence of developmental delays in children, due to the Zika virus—but only limited ESP therapy is available.

In this paper, we discuss BLESS in terms of its ESP personalization decision model, the formulation of the parent training content, and technical architecture and functionalities. A pilot evaluation of BLESS is currently underway and will be reported in a follow-up publication.

BLESS ESP Personalization Decision Model

To generate an ESP that will maximize the child's motor and cognitive development, our approach involves tailoring the relevant ESP based on the child's developmental needs. For that purpose, we have developed an ESP personalization model that based on an assessment of a child's impairment for normative milestones selects the appropriate early stimulation exercises. To computerize the ESP personalization decision model, we applied knowledge translation strategies [9,10] to collect knowledge from (a) ESP clinical guidelines, local protocols and standardized child development assessment tools; and (b) local experts, therapists and families whose children have followed ESP therapy. BLESS decision model is shown in Fig. 1 and is based on the criteria listed in Table 1.

For effective child development therapy, it is important to prioritize one developmental milestone at a time as it engages parents to focus on the child's most relevant need at the time, improving their skills in a step-wise way, and prevents them from being overwhelmed if the child has multiple development needs. This treatment delivery method has been shown to be effective in improving parents' skills to administer home-based therapies to their child [8,11].

In operation, considering the milestones expected for the child's age group, the therapist assesses each successive milestone and inputs the diagnosed impairment value (ICF value) into the decision model shown in Fig. 1. In line with the criteria from Table 1, the decision model proceeds as follows:

- In case of major impairment (ICF value = 3/4), select the current milestone for ESP. Else, continue processing the successive milestones until attaining a severe/complete diagnosis, or until all milestones are processed. Health experts consider tackling major impairments more important than dealing with prior milestones in the achievement pattern, even if these have mild/moderate impairment (ICF value = 1/2).
- Once all milestones are processed and no major issues were found, select the first milestone in the achievement pattern with any issues (ICF value > 0) for ESP (if any exist). Since no severe/complete impairment (ICF value >= 3) was found for any age-relevant milestone, the achievement pattern can be safely followed to improve the child's development.

Table 1 – Criteria for BLESS personalization decision model

Criteria	Especification
Age	Five age-groups, related to the average age of motor milestones achievement (0-4 months; 5-6 months; 7-9 months; 10-12 months; + 13 months).
Expected milestones for age	Five motor milestones expected for each age-group (head control, rolling, sitting, standing, walking) [6].
Impairment classification	Impairment level classification for each milestone (no impairment; mild/moderate impairment; and severe/complete impairment) [7].
Impairment severity	Achieving the milestone classified as a severe / complete impairment level is prioritized for obvious reasons—also, families tend to better engage in the child's most visible needs.
Order of milestone achievement	The milestone achievement pattern, respecting milestones which are prerequisite for subsequent milestones (e.g., head control before sitting, etc.).

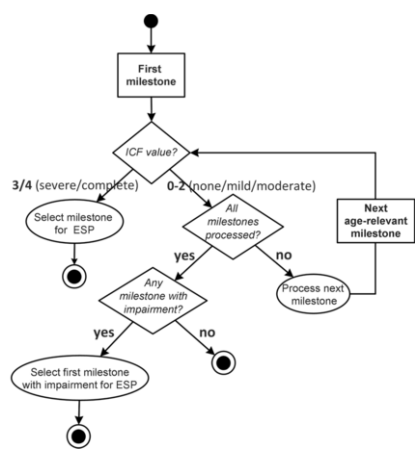


Figure 1 – BLESS personalization decision model

Once a suitable developmental milestone is selected, the personalization decision model proceeds by selecting ESA pertaining to the milestone and the child's specific impairment level. At the end, this process generates an ESP that is personalized to the child's current needs.

BLESS Patient Education Material

BLESS provides patient education content to improve the parent's knowledge and skills to effectively perform early stimulation exercises to their child at home. Our education approach is guided by chronic care frameworks [5,11] and other that stipulates strategies and content to engage and

empower individuals so that they can achieve effective self-management. We gathered and incorporated ESP-related educational content from a variety of local resources, including the Brazilian Early Stimulation Guidelines [3], official documents of the Brazilian Ministry of Health [12], as well as tacit knowledge and practical expertise of therapists.

For our approach, videos are a key medium to educate parents to perform early stimulation exercises with their child [4]. The primary educational medium are videos that illustrate how to perform a specific early stimulation exercise. We have prepared 60 early stimulation exercise videos, where each video is 90-120 seconds long and demonstrates in a step-by-step manner how to perform an exercise with the child, with a commentary of instructions narrated by a trained therapist. Additional educational content was organized in terms of 12 orientation packages, which comprise written educational material about handling their child's everyday activities, such as promoting extra stimulation during feeding, waking up, bathing, playing, posturing, communicating, etc. These orientation packages are selectively provided to parents based on their informational needs and skill levels as observed by the therapist. The videos and educational content are in Portuguese in plain language [13] at a readability level compatible with 6-8 years of educational level. All educational material was reviewed and validated by therapists and experts in ESP.

Digital Health Components of BLESS

BLESS comprises of two main components: (1) an *ESP therapy planning and decision support tool*, which allows therapists to generate evidence-based, personalized ESP; and (2) a *mobile BLESS family app*, which provides early stimulation videos to guide children's parents to perform early stimulation with their child; together with orientation packages, to better support their child during everyday activities at home.

The ESP therapy planning and decision support tool was developed using a 3-tier architecture, where the user-interface tier was programmed using ReactJS and Javascript d3 libraries; the application tier was developed using Java SpringBoot; and the database tier deploys MySQL. The mobile BLESS app was developed for Android, using Android Room for persistence and retrofit library for network communication.

ESP Therapy Planning and Decision Support Tool

The BLESS *ESP therapy planning and decision support tool* is a web-based tool to assist therapists in adminstering ESP therapy, based on clinical guidelines and local clinical protocols. The tool is designed as a shared decision making and planning environment, engaging both the therapist and the child's parents to determine the course of ESP therapy for the child. The tool provides the following functionality: (**a**) assess the child developmental status, using standard assessment tools, (**b**) design a personalized ESP for the child, in reponse to the assessed developmental deficiencies; (**c**) perform follow-up examinations, to monitor developmental progress and prescribe corresponding ESP; and (**d**) provide an overview of the therapist's patients in terms of compliance and progression with their assigned ESP.

The therapy planning workflow involves three 3 main tasks: (**i**) patient registration, (**ii**) family registration, and (**iii**) ESP consultation. Steps (**i**) and (**ii**) record demographic data about the child and the family needed for patient / family follow up and outcome analysis. Following the registration steps, the tool assists the therapist to perform the consultation in 3 steps:

Assessment Step

We computerized standard patient and family assessment tools, to (**a**) assess the child's impairment of normative development milestones; and (**b**) assess the family's performance in daily living activities, including observations on postures and positions, playing, and communication; as well as questions about feeding, playing and other activities; and (**c**) registering other impairments, needs and therapies, accomapnied by optional comments. Fig. 2 shows the screenshot of part (**a**) of the assessment step.

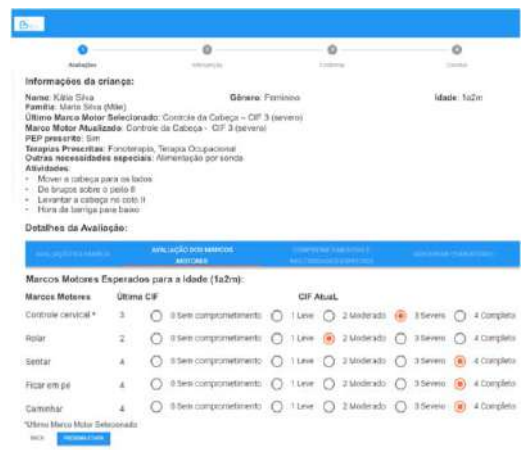


Figure 2 – Assessment step of BLESS therapist platform

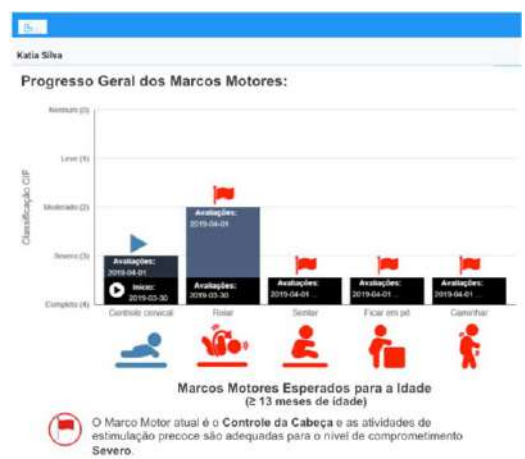


Figure 3 – Overall patient progress of BLESS therapist platform

ESP Planning Step

Our strategy for therapy planning involves designing ESP that focus only on the most critical developmental milestone to avoid putting a cognitive overload on parents and hence compromise their ability to comply with the ESP therapy schedule [14]. An ESP spans 8 weeks, and comprises 4 different early stimulation exercises that should be performed at least 3 times a week for 2 weeks.

The ESP therapy planning tool operationalizes the ESP personalization decision model (Fig. 1), and based on the assessment results supports the therapist and child's parents to jointly design an ESP that is tailored to the child's development needs. To educate the child's parents on how to perform activities of daily living the therapist further selects appropriate ESP orientation packages.

Confirmation Step

At the end of the consultation, the tool generates a report summarizing the child's and family information, current developmental stage of the child and the prescribed ESP withorientation packages. This confirmation step concludes the in-clinic consultation process. After the prescribed ESP is activated and the constituent early stimulation exercise videos with the educational material are delivered to the BLESS mobile app on the parents' mobile phone. The patient's profile page summarizes their progress in the form of an overall progress chart detailing the achieved milestones, and indicates the extent of family engagement by tracking the ESA execution via the BLESS mobile app (Fig. 3).

BLESS Family Mobile App

The BLESS mobile app is designed to guide parents to administering the prescribed early stimulation activities to their child. It offers the following features:

Guidance for ESP Exercises

In line with the prescribed ESP, the BLESS app's home screen indicates the stimulation activity that should currently be performed (Fig. 4a). After selecting an exercise, parents are shown the educational video together with associated info (Fig. 4b). Realizing that consistent internet connection may be an issue in some areas, therefore the mobile app downloads the ESP-relevant videos once after which they can be viewed as many times as needed regardless of the Internet connection. As new ESP are assigned, the stored videos are automatically deleted from the mobile phone (to save storage space). Synchronization with the server occurs in an opportunistic way, i.e., whenever Internet connectivity is available.

Monitoring of ESP Completion

The BLESS mobile app allows parents to monitor their progress in completing the ESP, as shown in Fig. 4c. This progress is informed by the completion of exercises as indicated by the parents (Fig. 4b). Further, the recorded progress is communicated to the ESP therapy planning and support tool, and used to update patients' status.

ESP Orientation Packages

The BLESS mobile app provides orientation packages on daily living activities for parents to effectively perform the ESP. The orientation packages are structured into separate pages that are suitable for the form factor of a smartphone. The mobile app supports navigation between pages and bookmarking them for repeated use (Fig. 4d).

BLESS in Action

During operation BLESS runs in concert with the therapist, whereby in a shared decision making environment the family and health professional tailor an ESP towards the child's development needs. To support the parents perform the ESP relevant educational material is also delivered to the parents.

We demonstrate the working of BLESS using a case-study of a 13-months-old child who has multiple levels of impairment for 3 milestones—i.e., severe impairment for head control, moderate impairment for rolling and complete impairment for sitting, standing and walking. Moreover, the family demonstrates not knowing how to stimulate the child at feeding, bathing and playing time. Based on the therapist's assessment, the decision logic model (Fig. 1) will create a personalized ESP by (a) selecting a suitable milestone to be targeted, based on the decision criteria (Table 1), and (b) selecting a group of ESA based on the impairment level. In this case, BLESS will select an ESP focused on head control due to its higher impairment level (severe) and its precedence in the milestone achievement pattern—note that it will not be possible for the child to achieve *rolling* or *sitting* without proper head control. Further, based on the therapist's observations of their DLA needs, orientation packages on feeding, bathing and playing issues can be selected and deleivered to the child's parent.

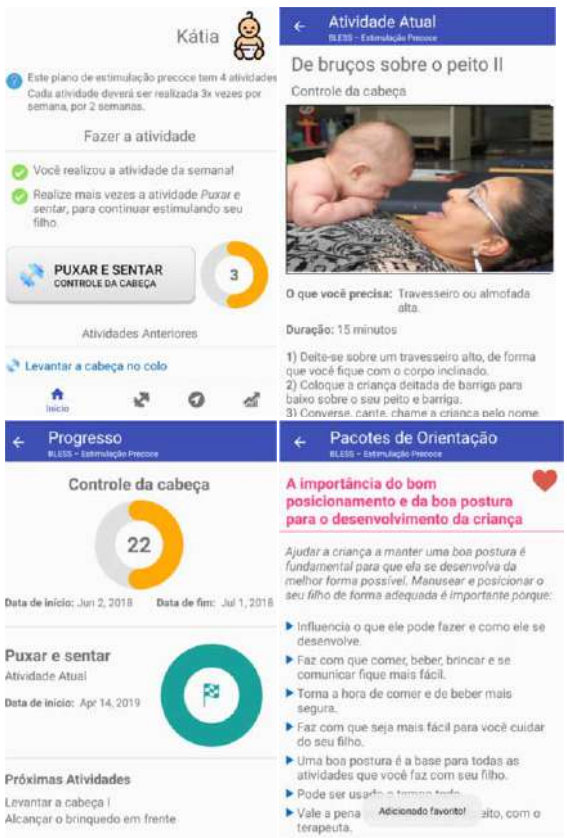


Figure 4 –BLESS mobile app screenshots: Guidance for ESP exercises (4a); ESA video screen (4b); Overview of ESP progress (4c); and orientation packages (4d)

As a follow-up, once the head control milestone has improved and the family demands addressed, a subsequent milestone is selected and new orientation packages can be chosen.

BLESS Implementation and Evaluation

BLESS is to be deployed at a reference centre for Maternal and Child healthcare in the Northeast of Brazil where we are gearing for a 6 month pilot study, involving 30-35 patients with their families and 4 therapists, to assess the (a) usability

of the BLESS system components; (b) impact on child development, parent knowledge and therapist practice routine; and (c) impact on service delivery costs. The pilot involves three stages as follows [15]: (i) Engaging and training the therapists to uptake BLESS. The thereapists will have one month to familiarize themselves with BLESS, where they will run case studies to build familiarity and confidence in the decision model operationalized by BLESS; (ii) Enrolling patients and their families into the pilot study, which includes installing BLESS on their mobile phones and training them to use it; (iii) Collecting study data over a 4-month period to evaluate BLESS. Both the therapist and family components of BLESS will track information about usage and compliance.

Discussion

We presented a Digital Healthapproach to translate knowledge on the achievement of normative motor milestones and early stimulation guidelines into a point-of-care ESP platform for therapists and patients' families. The key contribution of our solution is the integration of two recognized assessment instruments (i.e., gross motor milestones and ICF classification) to generate evidence-based ESP the educational content. An additional innovative aspect of BLESS is the personalization of the ESP based on standard child development assessment tools—the child development systems suggest relevant developmental activities but they are not tailored towards the child's developmental needs [3,12,16].

From a clinical standpoint, the computerization of knowledge on child development assessment and early stimulation enables the therapists to make better decisions regarding therapeutic planning and in turn offer quality health education for parents. The use of the mobile app allows us to reach a larger population of children with developmental delays which currently do not receive timely ESP. From the parents' perspective, the contribution of our work is a self-management program which engages and empowers parents to manage the child's condition in home-based setting, whilst reducing the frequency of in-person consultations which lessen the family burden and financial costs for families living in remote areas.

Conclusions

BLESS presents a unique solution to provide standardzied, tailored and timely treatment to children affected by developmental delays. Our solution takes into account the child's current therapy needs to achieve the necessary normative milestones. Our approach engages and educates the child's parents to administer ESP at home, whereas the therapist is remotely monitoring the child's progress and coordinating the selection and delivery of specific ESA to the parents. Families now have the opportunity of receiving cultural compliant and comprehensive educational content, with the goal of increasing the adherance to the child's treatment and ultimately leading to better health outcomes. Our solution is innovative and relevant to current Brazilian public health needs, and provides a low-cost treatment easily accessible to Brazilians, even in remote areas; and may be scalable to other aspects of child development.

Acknowledgements

This research is supported by the Brazilian Federal Agency for Support and Evaluation of Graduate Education – CAPES.

References

[1] United Nations for Child and Education Funding; and World Health Organization. Care for Child Development: Guide for Clinical Practice, (2012).

[2] Brasil, Ministério da Saúde, Secretaria de Vigilância em Saúde. *BoletimEpidemiológico*: *Monitoramento integrado de alterações no crescimento e desenvolvimento relacionadas à infecção pelo vírus Zika e outras etiologias infecciosas, até a Semana Epidemiológica 28/2017*, Brasilia, 2017.

[3] Brasil, Ministério da Saúde, Secretaria de Atenção à Saúde, *Diretrizes de estimulação precoce: crianças de zero a 3 anos com atraso no desenvolvimento neuropsicomotor decorrente de microcefalia.* Ministério da Saúde, Secretaria de Atenção à Saúde, Brasília, 2015.

[4] N.B. Giuse, T.Y. Koonce, A.B. Storrow, Kusnoor, S. V., & Ye, F. (2012). Using health literacy and learning style preferences to optimize the delivery of health information. *Journal of health communication*, *17*(sup3), 122-140.

[5] World Health Organization. *Innovative Care for Chronic Conditions: Building Blocks for Action.* World Health Organization, Geneva, 2002.

[6] World Health Organization, Multicentre Growth Reference Study Group, & de Onis, M. WHO Motor Development Study: windows of achievement for six gross motor development milestones. *Acta Paediatrica*, *95* (2006), 86-95.

[7] World Health Organization, *International classification of functioning, disability and health:* ICF, World Health Organization, Geneva, 2008.

[8] R.J. Palisano, M. Intensity of therapy services: what are the considerations? *Physical & occupational therapy in pediatrics* 29 (2009), 107-112.

[9] S.S.R. Abidi, S. Shayegani. Modelling the form and function of clinical practice guidelines: An ontological model to computerize clinical practice guidelines. In: Workshop on Knowledge Management for Health Care Procedures. *Springer*, Berlin Heidelberg (2008), 81-91.

[10] S.B Wasif, S. Abidi, S.S.R. Abidi. An Ontological Model of Behaviour Theory to Generate Personalized Action Plans to Modify Behaviours. In: *Exploring Complexity in Health: And Interdisciplinary Systems Approach. 26th European Medical Informatics Conference, IOS Press, Munich* (2016). 399-403.

[11] W.V. Woensel, P.C. Roy, S.R.Abidi, S.S.R.Abidi. A Mobile and Intelligent Patient Diary for Chronic Disease Self-Management. In: *MedInfo 2015: e-Health-enabled Health-Proceedings of the 15th World Congress on Health and Biomedical Informatics*, *IOS Press,* São Paulo (2015). 118-122.

[12] Brasil, Ministério da Saúde, Secretaria de Atenção à Saúde, Departamento de Atenção Básica. *Saúde da criança: crescimento e desenvolvimento*, Ministério da Saúde, Brasília, 2012.

[13] M. Wilson. Readability and patient education materials used for low-income populations." *Clinical Nurse Specialist* 23 (2009), 33-40.

[14] B. Jafarpour, S.R. Abidi, A.M.Ahmad, S.S.R. Abidi. INITIATE: An Intelligent Adaptive Alert Environment. In: *MedInfo 2015: e-Health-enabled Health-Proceedings of the 15th World Congress on Health and Biomedical Informatics*, *IOS Press,* São Paulo (2015). 285-289.

[15] S. Abidi, M. Vallis, H. Piccinini-Vallis, A.S. Imram, S.S.R. Abidi. Diabetes-Related Behavior Change Knowledge Transfer to Primary Care Practitioners and Patients: Implamentation and Evaluation of a Digital Health Platform. *JMIR Med Inform* (2018), 25-30.

[16] K. Verch. *Primeira Infância Melhor: Transforming the attention towards the first years of life in Latin America: challenges and achievements of a public policy in southern Brazil*, Banco Interamericano de Desenvolvimento, Divisão de Proteção Social e Saúde, IV (2017).

Address for correspondence

Raquel da Luz Dias, Faculty of Computer Science Dalhousie University 6050 University Avenue Halifax NS B3H 4R2 Canada. Email: raquel.dias@dal.ca

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190288

A Generic Rapid Evaluation Support Tool (GREST) for Clinical and Commissioning Decisions

Jack Dowie[ab], Vije Rajput[c], Mette Kjer Kaltoft[b]

[a] *London School of Hygiene and Tropical Medicine, London, UK*
[b] *University of Southern Denmark, Odense, Denmark*
[c] *Stonydelph Health Centre, Tamworth, UK*

Abstract

A fast and frugal generic tool can provide decision support to those making decisions about individual cases, particularly clinicians and clinical commissioners operating within the budget and time constraints of their practices. The multi-national Generic Rapid Evaluation Support Tool (GREST) is a standard preference-sensitive Multi-Criteria Decision Analysis-based tool, but innovatory insofar as an equity criterion is introduced as one of six. Equity impact reflects the number of population QALYs lost or gained in moving from Old (current intervention) to New (contemplated intervention). In the exemplar UK implementation Claxton's NHS Willingness to Pay per QALY is the numeraire. Any weight from 0 to 100% may be assigned to the equity criterion but its presence affirms that it is persons-as-citizens who experience any opportunity harms or benefits arising from actions within the health service commons. A fully-operational but demonstration-only version is available on open access, as proof of concept and method.

Keywords:
Clinical, Decision Support Systems, Health Equity

Introduction

Clinicians and clinical commissioning groups are routinely making decisions about the use of new and/or expensive interventions for individual cases, including patients with multiple morbidities. They may be facing an increasing number of such decisions in the UK, if the recent High Court decision regarding off-label/licence drugs is upheld

(https://www.bbc.co.uk/news/health-45588983).

We perceive a major gap in the support available for such decisions. NICE-type evaluations can cover only a tiny minority of interventions and their remit is currently restricted to on-label use. Decision aids developed according to normative standards such as IPDASi [1] are very limited in coverage, constitute information aids rather than decision support tools that produce an opinion, and rarely introduce budget considerations. Guideline recommendations, such as those produced by GRADE, cannot, by definition, reflect individual preferences, only group or sub-group averages, and leave the decision maker analytically unsupported in the task of processing the extensive summary of evidence at the point of decision.

Whenever a proposed NEW intervention is to be comparatively evaluated against a current OLD one within time and resource constraints, a flexible, rapid, generic and inexpensive decision support tool is needed.

The required tool must be practical and useful. It must meet the SMART criteria - Specific, Measurable, Achievable, Relevant, and Timely - to the extent each of these are reasonable in the given decision context. More simply, the decision support must be 'fast and frugal'. It must be pitched at the most appropriate point (trade-off) on the 'rigour-relevance' continuum. It will therefore be a long way from a highly analytical NICE-type evaluation in one direction, but also far removed from an expert deliberation-based guideline in the other. While endorsing long term efforts to develop normatively superior methods of linking evidence and clinical practice [2-6] or dealing with the complexities of value-based care [7-8] we have the simple but limited ambition of providing a rapid and practical method of improving on the present decision making process, whatever it is. The tool will not seek to replace or deter the development of superior tools, but reflect the belief that the *normatively* best, or even normatively *better*, may be the enemy of the *empirically* better – a point well-accepted in relation to drugs and devices, not yet for decisions. Evaluations should reflect this and use the actual process they would replace as comparator.

Method

Multi-Criteria Decision Analysis (MCDA) can provide the basis for such fast and frugal decision support tools. Given their basis in this technique they can provide an opinion on the worth of each Option (one of the actions that can be taken) by combining its Ratings (how well it performs on relevant Criteria) with the criteria Weightings (how important each criterion is relevant to the others), making the opinion a preference-sensitive one. Introductory materials and numerous examples of the implementations of MCDA in the decision support context are available at http://cafeannalisa.org.uk including the short video 'Powtoons'. In the space available and to avoid duplication we leave further details of Method to the following Results section and the online tool introduced there.

In Multi-Criteria Decision Analysis-based tools, it is entirely feasible to include an equity outcome criterion and our proposed tool, implemented in the Annalisa template [9], does this. However, we emphasize that this innovation occurs in the context of a more extended, multicriterial analysis, otherwise

the tool will not be appropriate in person/citizen-centred healthcare decision making.

Result

The Generic Rapid Evaluation Support Tool (GREST) is presented here as proof of concept and method, based on internal testing, and as the basis for feedback in relation to future development and implementation.

The two options in GREST are OLD (the current intervention, such as usual/standard/current care) and NEW (the contemplated replacement). The six criteria in the default prototype are:

- A condition/decision-specific Biomarker (e.g. Bone Mineral Density)
- A condition/decision-specific Function(al) Index (e.g. Six Minute Walk Test)
- Option (test/treatment) Side effects
- Option (test/treatment) Burden (e.g. arising from frequency and mode of delivery of medication)
- Health-Related Quality of Life (HRQOL)
- Equity (reflecting the Harm/Foregone Benefits to Others in 'North-East' cases and Benefit/Foregone Harm to Others in 'South-West' ones – see below)

It will be noted that Life Expectancy is not a criterion in the current GREST, so if there is any effect on this from the new intervention, it would need to be discussed separately in the light of the GREST opinion. It will also be noted that monetary cost is not a separate criterion, being introduced only through the sixth, equity, criterion.

To engage with the tool and understand the method underlying it, go to https://ale.rsyd.dk (enter 1513 as survey ID). This version is in English and uses the EQ-5D-5L tariff for 11 countries to establish Health-Related Quality of Life (HRQOL) values used in deriving the ratings for both the HRQOL and Equity criteria. Nine tariffs are from https://euroqol.org/eq-5d-instruments/eq-5d-about/valuation-standard-value-sets/crosswalk-index-value-calculator[10]; others tariffs included are for Poland [11] and China [12],

This is a demo version provided as a proof of concept and method on open access. Only anonymous, non-confidential data should be entered. No responsibility is taken for data security.

The performance ratings of OLD and NEW on the first five criteria in GREST are elicited in whatever way is compatible with the clinician's or clinical commissioning group's practice timescale and resources. The Ratings for both OLD and NEW should be the BEANs (Best *Estimates* Available Now) and hence as evidence-based as is possible, and as expertise-based as is necessary, within the *actual* resource and time constraints of the decision makers. The person/patient is to be regarded as the expert on Treatment Burden. The derivation of the ratings for the equity criterion is explained in a separate sub-section below.

The relative importance Weighting of each criterion, including equity, is elicited on a 0-100% scale, where 0 indicates of no importance and 100% of extreme importance. The six responses are summed and percentaged to give and display the set of provisional criterion weights that add to 100%. In this interactive tool they may be changed by cursor on inspection of their graphical display. More sophisticated weight elicitation procedures, such as swing weights or discrete choice experiments, may have greater normative appeal than Visual Analog Scales, but lack either individual applicability or practicality in the typical time and resource scale contemplated.

Figure 1: North-East Case with 5.9% Equity Weight

Figure 2: North-East Case with 50% Equity Weight

The GREST output is in the form of a 'deciographic' - a single screen showing all Ratings, Weightings, and the evaluation Scores for OLD and NEW. As in all standard MCDA applications, the Score combines the relative importance criterion Weightings with the evidence- and expertise-informed option performance Ratings, by the Expected Value algorithm.

Giving 50% weight to equity in Figure 2 (bottom) rather than the 5.9% in Figure 1 (top) flips the opinion to OLD.

The ratings in the above example are purely illustrative. We imagine NEW to be superior on Biomarker, Function, and HRQOL, but inferior on Side Effects and Treatment Burden – and, by the definition of a 'North-East' GREST, it produces net 'Harm to Others'. The 50% Equity rating for OLD reflects complete uncertainty about its value. In the Ratings panel a longer bar always means better and in the Weightings panel a longer bar always means more important.

The Equity Criterion

Person-centred care – and value-based healthcare - is not all about the rights of the person. Persons are also citizens who have responsibilities and duties within a resource-constrained public health service. Only by applying a *generic outcome measure* as a criterion within a personalised *multi-criterial* decision analysis can we move to the coherent involvement of equity in decision making about clinical cases. In GREST, the equity outcome criterion is defined and measured as the Harms or Benefits to Others created in moving from OLD to NEW in a resource-constrained service. In measuring these,

any *condition-specific* outcome such as *cancer* mortality/morbidity, as contrasted with *all-cause* mortality/morbidity, is ruled out as equity-irrelevant.

We have chosen Health-Related Quality of Life (HRQOL) as our generic equity outcome measure, the current GREST explicitly eschewing evaluations where NEW alters life expectancy. We use EQ-5D-5L as the HRQOL metric, though the method is not tied to any specific instrument.

Where the NEW intervention is *more expensive* than the OLD, but also *more effective*, we are in the North-East quadrant of the Cost-Effectiveness plane. There will be 'opportunity harms' (*foregone benefits*) to others, usually anonymous and unidentifiable, as a result of substituting NEW for OLD.

Where a NEW intervention is *less expensive* than the OLD, but also *less effective*, we are in the South-West quadrant of the Cost-Effectiveness plane, the only one recognized in NICE, where lower effectiveness is disallowed as a source of improved cost-effectiveness. This restriction of cost-effectiveness to *incremental* cost-effectiveness constitutes political interference with the implementation of a neutral technique [13]. In this quadrant there will be 'opportunity benefits' (*foregone harms*) to others, usually anonymous and unidentifiable, as a result of substituting NEW for OLD.

In measuring the foregone benefits and harms in the UK context, we use Karl Claxton's estimate of the NHS's revealed willingness to pay (WTP) for a Quality-Adjusted Life Year (QALY) of £12,000. 'The central or mean estimate of £12,936 is likely, if anything, to be an overestimate', hence our use of £12,000. [14] (p4).

Equity impact is calculated as the number of QALYs that would be moved from Others to the recipient of the NEW intervention. So, if the NEW intervention (assuming no life expectancy effect) costs £24,000 per year compared with £1,600 for the current OLD treatment (Extra cost = £22,400), and it improves HRQOL from 0.23 to 0.59 (QALY gain = 0.36), the Equity impact (Harm to Others) created by adopting NEW is calculated as:

£22,400 / 0.36 = £62,222 per QALY
£62,222 / £12,000 = 5.2.

The generation of the 1 QALY involved in substituting NEW for OLD for a single (identifiable) person creates a QALY loss of 5.2 to anonymous others elsewhere in the service. At the population level there is a net loss of 4.2 QALYs from the shift. This loss may all be borne by one anonymous person, or be distributed as lost quality-adjusted weeks or days across small or large numbers of people. The fact is that this distribution is unknown. (The issue of 'statistical compassion' is raised in the final section.)

If we reverse the OLD and NEW data and move into the South-West quadrant of *decremental* cost-effectiveness, the arithmetic is the same but the effect the opposite. Substituting NEW for OLD now creates a QALY gain of 5.2 to anonymous others elsewhere in the service. At the population level there is a net gain of 4.2 QALYs.

To map the QALY movement on to the required 0-1 ratio scale for the MCDA-based tool, we take its reciprocal, subtracting this from 1 in the South-West case. Moving 5.2 QALYs becomes 0.193 in the NE case (Figures 1 and 2) and 0.807 in the SW (Figures 3 and 4).

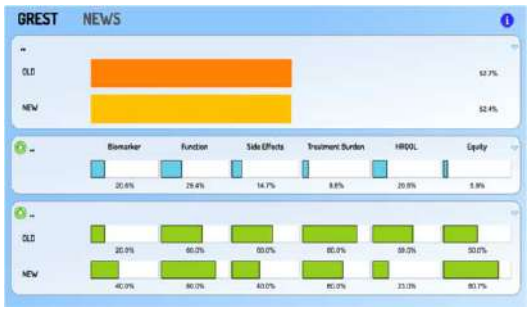


Figure 3: South-West case with 5.9% equity weight

Figure 4: South-West case with 50% equity weight

The above screens result from reversing the data in the earlier example, making NEW a South-West intervention, where it produces lower HRQOL but at greatly reduced cost, in contrast to the North-East situation where it produces the same amount of higher HRQOL but at greatly increased cost.

Giving about 6% weight to equity in Figure 3 (top) is insufficient to shift the verdict in favour of NEW, but assigning 50% weight in Figure 4 (bottom) clearly does so.

What can we conclude about the change in equity in these two situations? The most reasonable assumption is that the anonymous others who are either gainers or losers from NEW will be randomly distributed. Under this assumption and assuming an equity rating for OLD of 50% (reflecting complete uncertainty), the net effect of NEW will be to *increase inequity* if it is a North-East intervention (replacing the less costly and less effective OLD) and *reduce inequity* if it is a South-West one (where it is less costly and less effective than OLD).

The magnitude of the effects will be a function of the numbers in the particular clinical case. But we can note that at a service level, the potentially very large numbers of opportunity harms from NE interventions may add up to a massive increase in inequity while the failure to consider SW interventions can add up to a huge potential failure to reduce inequity. The same applies in the individual screening context.

Discussion

Claxton makes our basic point in relation to NICE decisions about new drugs, where the threshold being used is far above the average observed WTP for a QALY of £12.000, but it is equally pertinent at the clinical/clinical commissioning level.

"The evidence suggests that more harm than good is being done, but it is the unidentified and unrepresented NHS patients who bear the true (health) opportunity costs. Although finding reasons to approve new drugs is undoubtedly politically expedient, this cannot be ethically literate, because the interests of NHS patients, whether they are identifiable or not, are just as real and equally deserving of the type of care and compassion that can be offered by a collectively funded health care system. It is to be hoped that NICE will begin to place the unidentified NHS patients who bear the real opportunity costs at the heart of its deliberative process." [15] (p6).

The GREST tool does not mandate any amount of concern with the impact on others, but it does mandate a statement of the weight to be assigned to it, in the 0 to 100% range, in generating the opinion of the decision support tool. As persons-as-citizens, we incur both a right and responsibility to be informed of the impact on others of our use of the health service commons in self-producing and co-creating health [16].

Objections to introducing a societal element into clinical decisions can be expected from clinicians as well as patients, but rejected on the ground that equity is a social matter for all citizens, and it is not for a profession to decide as they should to be able to 'protect their patients' from concern with it. In any case, a healthcare budget-holder cannot escape this wider responsibility for all those within their aegis.

Treating any individual person as a means to achieve some population policy end– to reduce the incidence of this, increase the uptake of that – has to be rejected even if it is in their *perceived* individual best interests. (Infectious diseases are an exception where the law rightly takes away the right of an individual to directly jeopardise the health of others.) But neither is it ethical for the person as citizen to be left in – or allowed to opt into - complete ignorance as to the impact on others, in whatever direction. An empirical resolution of the clash between individual and community [17] is brought no nearer by simple repeated noting of its existence, or by attempts to solve it solely at the collective level, leaving the individual uninvolved in any direct way.

The person is not only a citizen, but usually a member of a family or other group of significant others. Nothing said here is meant to imply that the person will not wish to take the feelings and consequences for these near others into their decision making. Nor to imply that the benefits of a 'family focus' in provider care should be ignored or underestimated. We merely note that if relatives are driving demands to 'do everything possible' against the perceived implicit, or explicitly expressed, wishes of the patient, this transparent support will be helpful in empowering the patient.

Both the ethical requirements of 'patient-centred' care and the legal requirements of 'reasonable patient' care suggest that the introduction of 'equity' into clinical practice guidelines, recently advocated [18-21], is not an appropriate way to go. The GRADE subgroup propose that guidelines panels recommend that clinical practitioners 'consider' the ways in which an individual patient may be affected through being a member of a 'disadvantaged' group. However, this seems to be ethically suspect, insofar as it could potentially distort the personalized option ratings for, and personal criterion weightings of, the specific individual. The case for a personalized decision analytic approach using individualised ratings and weightings is well made by Wasfy and colleagues [22] and not satisfactorily refuted by Lightner [23]. GREST is on offer whether or not guidelines panels were to abdicate from making a recommendation in the face of uncertainty [24].

Everyone seeking to participate in the equity debate is confronted by the sociopsychological phenomenon, emphasised by Williams and Cookson, among others, of the difficulty individuals have in relating to the thousands of anonymous others who have equal rights in the health service commons.

"For many people [the] notion of 'statistical compassion' seems to create both intellectual and psychological difficulties. It is as if personal empathy with one or two individuals is possible, but, paradoxically, if many individuals are involved, this capacity to empathise diminishes. This difference between focusing on groups and focusing on individuals also distinguishes economists (and managers) from clinicians and others dealing with people at an individual level. The latter often claim that they are under an ethical duty to do everything possible for the person in front of them no matter what the consequences might be for everybody else. If this assertion is taken at its face value, it would imply that clinicians should ignore their responsibilities for the welfare of their other patients except when that patient is in front of them. It seems most unlikely that any clinician would actually behave in that way, so perhaps the statement should not be taken at its face value, but regarded instead as part of the rhetoric of medical practice, designed to bolster the doctor-patient relationship. But whatever may be the role of such statements, it is clear that in a public policy context, where distributive justice is an explicit objective, it is clearly not ethical for a clinician to ignore the consequences of his or her actions concerning the treatment of one patient for the health of other patients for whom the system is also responsible." [25] (p1866).

Conclusion

The individual can only self-produce their health, and co-create it with healthcare professionals and significant others, within the wider environment and socioeconomic constraints in which they live and work. Most of the recent advances in key health indicators (e.g. life expectancy) are attributable to sectors other than healthcare, so that the creation of health and health equity needs to be approached within a much wider and comprehensive framework [26]. Equitably, cost-effective, cross-sectoral public health policies are an essential complement to any changes in clinical decision making [27].

However, the transparent introduction of equity considerations into decision making in relation to individuals can play an important role in affirming the social nature of a public health service. A public health service is not a private health service. It is ultimately the person-as-citizen who is experiencing any opportunity harm or benefit that arises from an action within the health service commons [28]. All health service decision makers should therefore be showing the 'statistical compassion' appropriate to their level of responsibility. GREST provides a simple way to allow the opportunity implications of a decision to be highlighted, but then set at whatever level is desired - including zero.

Author Contributions: JD developed the generic GREST tool in response to a challenging clinical decision, incorporating the insights and suggestions of VR and MKK. He wrote the paper, incorporating their observations, revisions and additions.

Funding: This research received no external funding. The software used for GREST was installed at https:// ale.rsyd.dk as part of a nationally funded project to develop decision support tools. (Danish National Board of Health/SATS J.nr. 1-1010/116/27.)

Conflicts of Interest: Jack Dowie has a financial interest in the Annalisa software implementation of MCDA, but does not benefit from its non-commercial use.

References

[1] G. Elwyn, A.M. O'Connor, C. Bennett, R.G. Newcombe, M. Politi, M-A. Durand, et al., Assessing the quality of decision support technologies using the International Patient Decision Aid Standards instrument (IPDASi), *PLoS One* **4** (2009), e4705.

[2] C.F. Manski, Reasonable patient care under uncertainty, *Health Econ* **27** (2018), 1397–1421.

[3] A. Basu, Manski's views on patient care under uncertainty: Comment. *Health Econ* **27** (2018), 1422–4.

[4] K. Claxton, Positive tails and normative dogs: Comment on Manski 2018, *Health Econ* **27** (2018), 1425–7.

[5] E. McIntosh, Decentralized decision making through adaptive minimax regret — Complex yet intuitively appealing: Comment on Manski 2018, *Health Econ* **27** (2018), 1428–30.

[6] C.F. Manski, Response to commentaries on "Reasonable patient care under uncertainty*", Health Econ* **27** (2018), 1431–4.

[7] L.P. Garrison, M.V. Pauly, R.J. Willke, and P.J. Neumann, An Overview of Value, Perspective, and Decision Context—A Health Economics Approach: An ISPOR Special Task Force Report 2, *Value Heal* **21**, 2018, 124–30.

[8] P.J. Neumann, R.J. Willke, and L.P. Garrison, A Health Economics Approach to US Value Assessment Frameworks—Introduction: An ISPOR Special Task Force Report 1, *Value Heal* **21**, 2018, 119–23.

[9] J. Dowie, M.K. Kaltoft, G. Salkeld, and M. Cunich, Towards generic online multicriteria decision support in patient-centred health care, *Health Expect* **18** (2013), 689–702.

[10] B. van Hout, M.F. Janssen, Y.S. Feng, T. Kohlmann, J. Busschbach, D. Golicki, et al., Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets, *Value in Health* **15** (2012), 708-715

[11] D. Golicki, M. Niewada, B. van Hout, M.F. Janssen, and A.S. Pickard, Interim EQ-5D-5L Value Set for Poland: First Crosswalk Value Set in Central and Eastern Europe, *Value Heal Reg Issues* **4** (2014), 19-23.

[12] N. Luo, G. Liu, M. Li, H. Guan, X. Jin, and K. Rand-Hendriksen, Estimating an EQ-5D-5L Value Set for China, *Value Heal* **20** (2017), 662–9.

[13] J. Dowie, M.K. Kaltoft, J.B. Nielsen, and G. Salkeld, *Caveat emptor* NICE: biased use of cost-effectiveness is inefficient and inequitable. *F1000Res* **4** (2015), 1078.

[14] K. Claxton, S. Martin, M. Soares, N. Rice, E. Spackman, S. Hinde, et al., Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold, *Health Technol Assess* **19** (2015), 1–504.

[15] K. Claxton, M. Sculpher, S. Palmer, and A.J. Culyer, Causes for concern: is NICE failing to uphold its responsibilties to all NHS patients? *Health Econ* **7** (2015), 1–7.

[16] J. Dowie and M.K. Kaltoft, The future of health is self-production and co-creation based on apomediative decision support, *Med Sci* **6** (2018), 66.

[17] A. Granata and A. Hillman, Competing practice guidelines: using cost-effectiveness analysis to make optimal decisions, *Ann Intern Med* **128** (1998), 56–63.

[18] V.A.Welch, E.A. Akl, G. Guyatt, K. Pottie, J. Eslava-Schmalbach, M.T. Ansari, et al. GRADE equity guidelines 1: considering health equity in GRADE guideline development: introduction and rationale*, J Clin Epidemiol* **90** (2017), 59–67.

[19] E.A. Akl, V. Welch, K. Pottie, J. Eslava-Schmalbach, A. Darzi, I. Sola, et al. GRADE equity guidelines 2: considering health equity in GRADE guideline development: equity extension of the guideline development checklist, *J Clin Epidemiol* **90** (2017), 68–75.

[20] V.A.Welch, E.A. Akl, K. Pottie, M.T. Ansari, M. Briel, R. Christensen, et al., GRADE equity guidelines 3: considering health equity in GRADE guideline development: rating the certainty of synthesized evidence. *J Clin Epidemiol* **90** (2017), 76–83.

[21] K. Pottie, V. Welch, R. Morton, E.A. Akl, J.H. Eslava-Schmalbach, V. Katikireddi, et al., GRADE equity guidelines 4: considering health equity in GRADE guideline development: evidence to decision process, *J Clin Epidemiol* **90** (2017), 84–91.

[22] J.H. Wasfy, K. Armstrong, C.E. Milford, and T.M. Sundt, Bicuspid aortic disease and decision making under uncertainty - The limitations of clinical guidelines*, Int J Cardiol*. **181** (2015), 169–71.

[23] D.J. Lightner, Re: Wasfy et al., Guidelines versus decision analysis is the wrong conversation, *Int J Cardiol* **186** (2015), 60.

[24] M. Bretthauer and M. Kalager, When no guideline recommendation is the best recommendation, *Lancet* **392** (2018), 898–9.

[25] A. Williams and R. Cookson, Equity in health, *Handbook of Health Economics* **1B** (2000*),*1863–910.

[26] J. Bircher and S. Kuruvilla, Defining health by addressing individual, social, and environmental determinants : New opportunities for health care and public health, *J Public Health Policy* **35** (2014), 363–86.

[27] G. Mooney, *The Health of Nations: Towards a New Political Economy*, Zed Books, London, 2012.

[28] K. Hassanally, Overgrazing in general practice: the new *Tragedy of the Commons*, *Br J Gen Pract* **65** (2015), 81.

Address for correspondence
Jack Dowie. Email: jack.dowie@lshtm.ac.uk.

Queries about survey on https://ale.rsyd.dk
Mette Kjer Kaltoft. Email: mkaltoft@health.sdu.dk

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190289

Design and Implementation of a Tool for Pharmacists to Register Potential Errors in Prescribed Medication

Santiago Frid[a], Valeria Zapico[a], Adriana Mansilla[b], Matías Álvarez[a], Gastón López[a], Carlos Otero[a], Daniel Luna[a]

[a] *Department of Health Informatics, Hospital Italiano de Buenos Aires, Ciudad de Buenos Aires, Argentina*
[b] *Department of Pharmacy, Hospital Italiano de Buenos Aires, Ciudad de Buenos Aires, Argentina*

Abstract

Adverse drug events are frequent and may be mitigated with the implementation of functionalities within Health Information Systems. We developed a tool that allows Pharmacists to register and communicate to providers potential errors in prescribed drugs in terms of medication omission, unjustified stop of medication or other reasons.

We included all interventions performed by Pharmacists for admitted patients between July, 31st 2018 and October, 23rd 2018.

During the study period, 193 interventions were carried out by Pharmacists. 117 (60%) were intended for registering medication omission, 7 (4%) for unjustified stop of medication and 69 (36%) for other reasons. 112 interventions lead to the provider performing the suggested action (58%), 77 (40%) were rejected and 4 (2%) required no action.

Although there were errors in the use of the tool, a great amount of interventions were accepted, thus representing a better quality of care for patients.

Keywords:
Electronic prescribing, patient safety, medication errors

Introduction

Medications are widely used therapies associated with multiple benefits for patients, but they can also cause adverse events, leading to thousands of hospital admissions every year [1]. Adverse Drug Events (ADE) are defined as harm caused by medical interventions related to drug administration [2]. Most of the preventable ADEs are due to errors in the prescription of medication [3].

The Institute of Medicine (IOM) strongly recommends the development of automated information systems with clinical provider order entries (CPOE), either with or without clinical decision support systems (CDSS) [4], in order to enhance reliability, quality and safety in the use of medication [5]. There is evidence that the implementation of CPOE decreases many of the problems associated with handwriting issues (illegible writing, incomplete orders, incorrect doses) [6,7]. Furthermore, when CPOE are complemented with CDSS it can offer predetermined values of dose, route of administration and frequency for commonly prescribed drugs [8], verify medication allergies, interactions, dose adjustment and therapeutic duplication [9], leading to a further reduction in ADEs [10].

However, either because alerts are ignored due to "alert fatigue" [11] or because the patient's clinical condition can vary (e.g., alterations in the renal function), changing the appropriateness of a prescription, asynchronous safety nets are needed besides CDS systems [12]. This is why it is important to combine technology with clinical judgement.

Clinical pharmacists provide high value for patients, physicians and additional members of the healthcare team in the management of medication [13], and the IOM considers them an essential asset for the safe use of drugs [14]. One of their missions is to formulate recommendations to providers with the intention to improve the management of patient medication (called pharmaceutical interventions or PIs). PIs can be related to the adequacy of a prescribed medication (wrong or ineffective medication) or a dose (overdose, low dose), the lack of a needed medication (omission of a usual medication or need of a new one) or the need to obtain a new laboratory test for monitoring [15,16].

An indirect way to evaluate the contribution of clinical pharmacists to the optimization of pharmacological therapy is to determine the amount of medication-related problems that are addressed or prevented with their interventions [17]. Rates of acceptance of such interventions by providers of up to 90% have been reported [18,19]

Hospital Italiano de Buenos Aires (HIBA) has an Electronic Health Record (EHR) with CPOE and CDSS, and pharmacists perform and register on average 2,500 interventions monthly, out of the 165,600 prescriptions they check (1.51 interventions/100 prescriptions). Given the need of the staff of the Pharmacy Department to perform and register interventions related to the omission of a medication, and in order to comply with the standards of the Joint Commission International (JCI) to optimize medication reconciliation, we decided to develop and implement a tool for that purpose.

Our general objective is to describe the development and implementation of a tool that allows pharmacists to perform and register interventions on medication that is not currently prescribed (omission, unjustified stop or other reasons).

Our specific objective is to analyze preliminary results of the use of this tool by physicians and pharmacists during the first 12 weeks of implementation, considering the actions taken within the tool and the actual prescription of drugs.

Methods

Setting

HIBA is a non-profit organization founded over 165 years ago in Buenos Aires, Argentina. Its network includes a highly

specialized university hospital which provides healthcare for ambulatory care, emergencies, admissions, clinical and surgical specialties, critical care, home care and chronic patient care. It has its own health maintenance organization (HMO), called Plan de Salud, with over 165,000 affiliates, and it provides healthcare services to 1,500,000 patients insured by other companies. Every year, over 45,000 patients are admitted, 45,000 surgical procedures are performed and 3,000,000 ambulatory consultations take place in our network. In 2015 HIBA achieved accreditation from the JCI, and became reaccredited in 2018.

Since 1998 we gradually implemented a Health Information System developed in-house to manage clinical and administrative information. Our Electronic Health Record (EHR), named ITALICA, is web-based, modular and problem-oriented. This EHR covers all the spectrum of healthcare: outpatient, inpatient, emergencies and home care. ITALICA includes functionalities such as CPOE for drug prescription and laboratory and imaging studies and visualization of study results.

The pharmacy service (PS) consists of a head of service, two sub-managers, 23 pharmacists and 11 pharmacy residents, who add up to a total of 34 pharmacists who perform medication validation. There are also 40 people dedicated to the preparation of medication.

Prior to drug dispensation, pharmaceutical validation is carried out. In the 1990s, the American Society of Health-System Pharmacists (ASHP) set the minimum standards for Hospital pharmacy services. The JCI defined that a pharmacist must validate all prescribed drugs before dispensation [20,21].

Through medication validation, pharmacists verify the prescription, evaluating if the prescribed medication is in accordance with the patient's clinical condition, if there are any contraindications or interactions and if dose-adjustment is required according to age, weight and laboratory results. This confers a greater value to the pharmaceutical process, allowing crossed control and providing greater security to the cycle.

However, pharmacists could only validate existing prescriptions, and they were unable to register in the EHR an intervention on a medication that was not currently prescribed. Therefore, communication between pharmacists and providers was carried out informally (through e-mails and telephone calls), thus elevating the chances of occurrence of potential errors and not registering important information in the EHR. Additionally, there was no chance of finding patterns that could help improve the prescription process.

Study design

This is an observational, descriptive and analytical study of a retrospective cohort in the HIBA. We included all interventions performed by pharmacists with the new tool on admitted patients of all ages from July, 31st 2018 through October, 23rd 2018.

Categorical variables are presented as absolute frequency and relative frequency (percentage). The rate is presented as prevalence, with its respective 95% confidence interval.

Features of the tool

We developed a tool that allows pharmacists to register interventions related to one or more of the following:

- Medication omission: lack of prescription of a necessary medication
- Unjustified stop of medication: prescription of a medication stopped without providing a reason for doing so
- Other: specification of type of intervention and reason.

When using it, the pharmacist must also explicate whether it is a miss (defined as an event that caused harm to the patient) or a near miss (defined as a potential error that was prevented before it caused harm to the patient), and describe the intervention in a free-text field. Lastly, the user must choose which providers to communicate the intervention to (the treating team is set by default) (Figure 1).

Figure 1– Capture of the developed tool

Afterwards, these physicians receive an e-mail with all the aforementioned information and a telephone number of the PS in case doubts arise regarding the intervention.

These interventions are visualized by providers in the EHR in the Medical prescriptions menu. They can choose to accept the intervention (which redirects to the drug prescription module) or reject it (providing a reason for rejection).

All medical and pharmaceutical users were informed about this tool through institutional e-mails and tutorials posted in the EHR prior to implementation.

Results

Types of intervention

Throughout the 85 days of the study period, from July, 31st 2018 to October, 23rd 2018, 193 pharmaceutical interventions were carried out, of which 184 were near misses and 9 were misses. 60% (n=117) of such interventions were related to medication omission, 36% (n=69) to other reasons and 4% (n=7) to unjustified stop of medication (Figure 2).

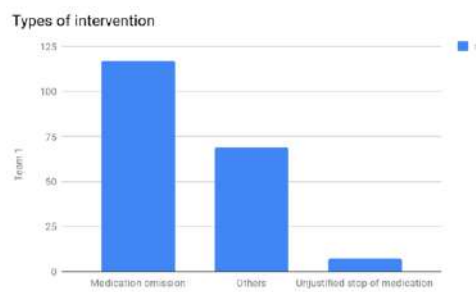


Figure 2– Types of intervention

The rate of interventions was 2.27/day (193 interventions/85 days). The mean number of weekly admissions at HIBA during 2018 was 697 (99.59 admissions/day), so the estimate ratio of interventions to admissions is 0.02. If we consider the number of prescriptions (165,600/month, approximately 5,520/day), the estimate ratio of interventions to prescriptions is 0.0004.

Use of the tool by providers

As to the response of physicians, 29.02% (n=56) of interventions were accepted, 9.84% (n=19) were rejected and in 41.45% of the cases (n=80) no action was taken. Pharmacists cancelled 17.62% (n=34) of the interventions they performed themselves, and 2.07% (n=4) did not require acceptance nor rejection, since they were merely for informative purposes (Figure 3).

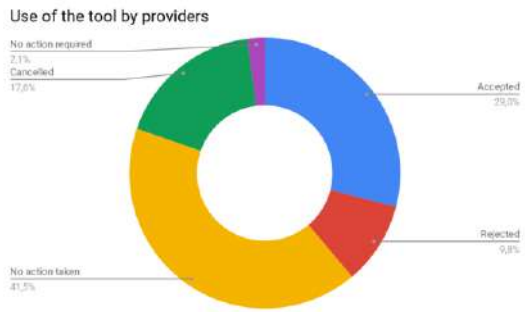


Figure 3– Use of the tool by providers

Action taken by providers

Although the proportion of interventions accepted by providers was low, these rates vary significantly when taking into account the course of action actually carried out by physicians (Figure 4).

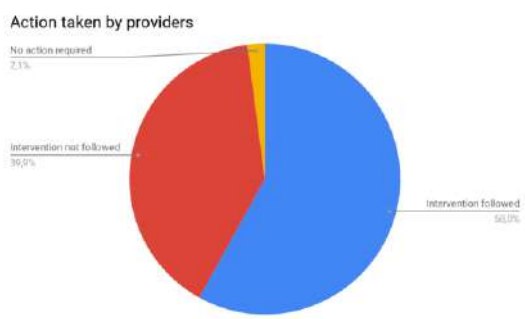


Figure 4– Action taken by providers

There were 112 interventions in which providers followed the pharmacist's suggestion, which means an actual acceptance rate of 58.03% (95% CI 50.73-65.08%), including 4 that were duplicated. Of those, 3 prescriptions had already been performed by physicians minutes before the intervention, which could be due to the simultaneous log-in to the EHR of both users or a parallel communication channel between them. Furthermore, in 8 cases the provider prescribed some of the suggested drugs but not all of them. Lastly, there were 4 interventions in which the pharmacist suggested to change the medication due to lack of stock in the PS, and 2 in which this communication channel was used to correct the registration of prescriptions in the EHR that had already been performed, either verbally or in an incorrect section of the EHR.

On the other hand, there were 77 interventions (39.90%) that lead to no action taken by the provider (that is, that the intervention was not followed). This includes three patients that were discharged on the same day, three interventions regarding lack of stock of a medication in which the prescription was not modified and one intervention that lacked the name of the drug omitted.

The remaining 4 interventions (2.07%) include the following: 1 to communicate that the PS would send a different medication because of lack of stock, 1 to remind the need for authorization by the financial entity prior to drug dispensation, 1 suggestion for clinical evaluation and 1 to register the communication with the nurse staff about the correct rate of infusion for a drug.

Interventions by Medical Department

As to the medical departments that received the interventions, Internal Medicine was the main recipient (38.86%, n=75), followed by Surgery (10.88%, n=21), Obstetrics (8.81%, n=17), Cardiology (6.22%, n=12), Pediatrics (5.18%, n=10) and Traumatology (5.18%, n=10) (Figure 5). The difference between the number of interventions by medical departments is probably related to the number of inpatients treated by each.

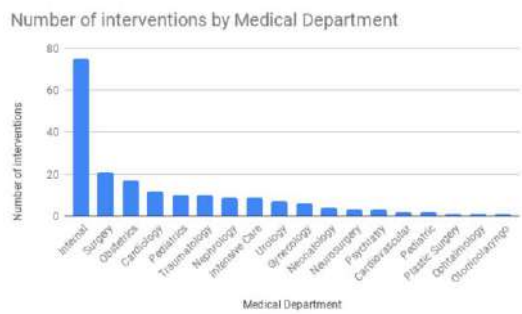


Figure 5– Action taken by providers

Interventions for medication omission

When analyzing all interventions for medication omission (117 out of all 193 interventions), 73 (62.39%) recorded the lack of prescription of a medication due to errors in the medication reconciliation process, with a total of 98 medications (1.34/intervention). The rest (44/117, 37.61%) were about suggestions of medication that was relevant to the current clinical context of the patient (e.g. potassium for hypokalemia).

Discussion

In this paper we describe the evaluation of a tool that allows pharmacists to register in the EHR and communicate to providers about potential errors in the medical prescriptions, either by omission of relevant medication, unjustified stop of medication or other reasons.

This tool was created mainly to optimize the medication reconciliation at admission, given the extensive evidence about the importance of this process [22,23]. However, when analyzing the results we observed that 37.61% of the interventions for medication omission were performed to register the lack of a prescription of a medication that was relevant to the current clinical context of the patient. In our

institution, the PS has a leading role in the medication cycle, having achieved validation rates close to 100%, which is why we find it very important to involve pharmacists in the improvement of the patient safety related to the pharmacological treatment.

On the other hand, the high use of the type of intervention *Other* of the tool provides evidence of the need of a formal communication channel between pharmacists and physicians. This type of intervention was mainly used to provide administrative information, to suggest improvements in the registry of the clinical information, to indicate the need of a clinical evaluation (e.g., corroborate known allergies) and, in some occasions, to indicate medication omission.

Also, it is worth mentioning that in one case this tool was used to register a telephonic communication carried out between the pharmacist and a nurse to provide information about the rate of infusion of a chemotherapy drug. This led us to reflect on the lack of existence of a formal communication channel between the PS and the Nursing Service, and the possibility to explore the need of such channel through inquiries to the stakeholders.

Another issue that is evident when evaluating the results is that the tool was not used properly (or at least not in the way it was expected according to how it was designed). Either by an error in the type of intervention selected by the pharmacists or by an incorrect selection of the action taken by the providers, the raw data shows a discrepancy between the action registered and the action taken. This could be improved by optimizing communication and training of the correct use of the tool. In any case, when we focus on the impact that the tool had on the actions taken by physicians, 112 interventions were accepted throughout the study period, representing 58% of all interventions.

Conclusions

Despite the discrepancy between the registry of acceptance of interventions through the tool and the amount of interventions eventually followed by providers (29.02% and 58.03%, respectively), it is clear that the functionalities of the tool are yet to be further exploited. The final number of interventions by the PS that were carried out by physicians during the study period is significant (193 in 85 days), and the rate of acceptance was high. It is worth mentioning that the existence of rejected interventions does not imply per se an error in the design and implementation of the tool or in the actions taken by pharmacists. They mainly represent a decision taken by providers based on their clinical judgement, which lead them to override the suggestions.

Furthermore, it is clear to us that the correct training of the staff involved in the process (namely, physicians and pharmacists) is critical in achieving our goals for improving quality of care and patient safety through the use of this tool.

References

[1] D.C. Classen, S. Phansalkar, and D.W. Bates, Critical drug-drug interactions for use in electronic health records systems with computerized physician order entry: review of leading approaches, *J. Patient Saf.* **7** (2011) 61–65.

[2] L.L. Leape, T.A. Brennan, N. Laird, A.G. Lawthers, A.R. Localio, B.A. Barnes, L. Hebert, J.P. Newhouse, P.C. Weiler, and H. Hiatt, The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II, *N. Engl. J. Med.* **324** (1991) 377–384.

[3] H.M. Seidling, S.P.W. Schmitt, T. Bruckner, J. Kaltschmidt, M.G. Pruszydlo, C. Senger, T. Bertsche, I. Walter-Sack, and W.E. Haefeli, Patient-specific electronic decision support reduces prescription of excessive doses, *BMJ Qual. Saf.* **19** (2010) e15–e15.

[4] Committee on Data Standards for Patient Safety, Board on Health Care Services, and Institute of Medicine, Patient Safety: Achieving a New Standard for Care, National Academies Press, 2004.

[5] G.J. Kuperman, A. Bobb, T.H. Payne, A.J. Avery, T.K. Gandhi, G. Burns, D.C. Classen, and D.W. Bates, Medication-related clinical decision support in computerized provider order entry systems: a review, *J. Am. Med. Inform. Assoc.* **14** (2007) 29–40.

[6] C. Senholzi, and J. Gottlieb, Pharmacist interventions after implementation of computerized prescriber order entry, *Am. J. Health. Syst. Pharm.* **60** (2003) 1880–1882.

[7] C. Meyer-Massetti, V. Hofstetter, B. Hedinger-Grogg, C.R. Meier, and B.J. Guglielmo, Medication-related problems during transfer from hospital to home care: baseline data from Switzerland, *Int. J. Clin. Pharm.* (2018). doi:10.1007/s11096-018-0728-3.

[8] S.R. Ranji, S. Rennke, and R.M. Wachter, Computerised provider order entry combined with clinical decision support systems to improve medication safety: a narrative review, *BMJ Qual. Saf.* **23** (2014) 773–780.

[9] N. Page, M.T. Baysari, and J.I. Westbrook, A systematic review of the effectiveness of interruptive medication prescribing alerts in hospital CPOE systems to change prescriber behavior and improve patient safety, *Int. J. Med. Inform.* **105** (2017) 22–30.

[10] D.M. Ashcroft, P.J. Lewis, M.P. Tully, T.M. Farragher, D. Taylor, V. Wass, S.D. Williams, and T. Dornan, Prevalence, Nature, Severity and Risk Factors for Prescribing Errors in Hospital Inpatients: Prospective Study in 20 UK Hospitals, *Drug Saf.* **38** (2015) 833–843.

[11] S. Phansalkar, A. Desai, A. Choksi, E. Yoshida, J. Doole, M. Czochanski, A.D. Tucker, B. Middleton, D. Bell, and D.W. Bates, Criteria for assessing high-priority drug-drug interactions for clinical decision support in electronic health records, *BMC Med. Inform. Decis. Mak.* **13** (2013) 65.

[12] R.M. Reichley, T.L. Seaton, E. Resetar, S.T. Micek, K.L. Scott, V.J. Fraser, W.C. Dunagan, and T.C. Bailey, Implementing a commercial rule base as a medication order safety net, *J. Am. Med. Inform. Assoc.* **12** (2005) 383–389.

[13] P.J. Kaboli, A.B. Hoth, B.J. McClimon, and J.L. Schnipper, Clinical pharmacists and inpatient medical care: a systematic review, *Arch. Intern. Med.* **166** (2006) 955–964.

[14] Institute of Medicine, Academic health centers: leading change in the 21st century, *Acad. Emerg. Med.* **11** (2004) 802–806.

[15] L. Cornuault, V. Mouchel, T.-T. Phan Thi, H. Beaussier, Y. Bézie, and J. Corny, Identification of variables influencing pharmaceutical interventions to improve medication review efficiency, *Int. J. Clin. Pharm.* **40** (2018) 1175–1179.

[16] B. Pfister, J. Jonsson, and M. Gustafsson, Drug-related problems and medication reviews among old people with dementia, *BMC Pharmacol. Toxicol.* **18** (2017) 52.

[17] K.K. Viktil, and H.S. Blix, The impact of clinical pharmacists on drug-related problems and clinical outcomes, *Basic Clin. Pharmacol. Toxicol.* **102** (2008) 275–280.

[18] S.N. Kucukarslan, M. Peters, M. Mlynarek, and D.A. Nafziger, Pharmacists on rounding teams reduce preventable adverse drug events in hospital general medicine units, *Arch. Intern. Med.* **163** (2003) 2014–2018.

[19] K.M. O'Dell, and S.N. Kucukarslan, Impact of the clinical pharmacist on readmission in patients with acute coronary syndrome, *Ann. Pharmacother.* **39** (2005) 1423–1427.
[20] M.K. Riaz, M. Riaz, and A. Latif, Review - Medication errors and strategies for their prevention, *Pak. J. Pharm. Sci.* **30** (2017) 921–928.
[21] American Society of Hospital Pharmacists, ASHP guidelines: minimum standard for pharmacies in hospitals, *Am. J. Health. Syst. Pharm.* **70** (2013) 1619–1630.
[22] O. Fernandes, Medication reconciliation in the hospital: what, why, where, when, who and how?, *Healthc. Q.* **15 Spec No** (2012) 42–49.
[23] K.M. Gleason, M.R. McDaniel, J. Feinglass, D.W. Baker, L. Lindquist, D. Liss, and G.A. Noskin, Results of the Medications at Transitions and Clinical Handoffs (MATCH) study: an analysis of medication reconciliation errors and risk factors at hospital admission, *J. Gen. Intern. Med.* **25** (2010) 441–447.

Address for correspondence

Santiago Andrés Frid MD, Department of Health Informatics, Hospital Italiano de Buenos Aires. Juan Domingo Perón 4190, C1181ACH. Ciudad de Buenos Aires, Argentina. email: santiago.frid@hospitalitaliano.org.ar

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190290

Telemedicine for Upper Respiratory Tract Infections During 2018 Epidemiological Outbreak in South America

Santiago Andrés Frid[a], María Florencia Grande Ratti[b], Ana Pedretti[b], Javier Pollan[b], Bernardo Martínez[b], Alejandro López Abreu[a], Gloria Diodati[a], Gastón López[a], Janine Sommer[a], Daniel Luna[a], Fernando Plazzotta[a]

[a] *Department of Health Informatics, Hospital Italiano de Buenos Aires, Ciudad de Buenos Aires, Argentina*
[b] *Department of Internal Medicine, Hospital Italiano de Buenos Aires, Ciudad de Buenos Aires, Argentina*

Abstract

Telemedicine is an increasingly used strategy for providing care to patients. The prevention and treatment of Upper Respiratory Tract Infections (URTIs) during outbreaks still require new management approaches.

We aimed to describe patients' characteristics and the care process after the creation and implementation of a virtual care program for patients with URTI during the epidemiological outbreak.

We studied all consultations that took place between May, 21st, and September, 14th 2018 at Hospital Italiano de Buenos Aires (HIBA). After applying exclusion criteria 218 consultations were left for the analysis. Most patients did not need a referral to a care center for a face-to-face assessment. The consultation rate to the Emergency Department (ED) within 7 days was 11.92% (26/218) with a 95% CI of 7.94-16.99%.

This new approach in patient care has a great potential for relieving the overcrowding in EDs, decreasing waiting times and preventing the infection spread in waiting rooms.

Keywords:
upper respiratory tract infections; telemedicine

Introduction

Technological advances have a direct influence on the provision of health services, replacing in many cases the traditional ways in which Medicine has been exercised by offering remote medical services. Many doctors and health institutions are increasingly adopting the use of technology to provide care for their patients, either through websites, e-mails, chats, video calls, and/or mobile applications. The World Health Organization (WHO) defines Telemedicine as "the provision of health services by health professionals using information and communication technologies for the diagnosis, treatment, prevention of diseases, injuries, research, evaluation and continuing education, with the interest of taking care of the health of individuals and communities where distance is a critical factor". In many institutions in several countries, valuable experience in this field is progressively being acquired [1–5].

On the other hand, Upper Respiratory Tract Infections (URTI is defined as a term that includes several infectious diseases of the upper respiratory tract: rhinosinusitis, the common cold, pharyngitis/adenitis, laryngitis, and otitis media) are a reason for frequent consultation in hospitals throughout the world, both in adults and in the pediatric population [6], especially during the winter time.

Non-urgent consultations to the ED during the epidemiological outbreak contribute to a phenomenon known as *overcrowding* [7], and they overload the health system, negatively affecting patient care. According to a study performed in our institution, the overall prevalence of consultations for URTI at HIBA during 2015-2016 in our walk-in area was 12.01% with 95%CI 11.86%-12.16% (21,581/179,597). Management strategies for the prevention and treatment of influenza during the seasonal epidemiological outbreak are necessary to redirect patients to the outpatient setting [8].

In addition, the healthcare sector faces two apparently contradictory demands: first, to ensure equitable access to quality healthcare services, and second, to reduce, or at least control, the rising costs of these services.

Telemedicine could help meet these two demands by optimizing the use of existing resources (expertise and equipment) by means of telecommunications. Despite this, the use of Telemedicine for URTIs has not been thoroughly investigated. A tool that informs patients about the level of care required was reported to be widely accepted among users[9], but there is also evidence that the implementation of Telemedicine for the management of acute respiratory illnesses may increase health costs[10]. With this in mind, in this study, we describe the characteristics of the patients and the care process implemented for patients insured by HIBA's own Health Maintenance Organization (HMO) that consulted through the virtual URTI care program implemented in 2018.

Methods

Setting

Hospital Italiano de Buenos Aires (HIBA) is a non-profit organization with 165 years of history in Argentina. Its healthcare network includes a university hospital of high complexity that covers health care for outpatient, inpatients, emergencies, critical care, home care, chronic care, and medical and surgical specialties. It has its own medical insurance service (health maintenance organization), with more than 160,000 affiliates, and provides health services to 1,500,000 people with other health insurances. Annually, more than 45,000 patients are admitted to their hospitals, and 45,000 surgical procedures and 3,000,000 outpatient visits take place.

Since 1998, the HIBA has its own health information system (in-house) that includes the management of clinical and administrative information. It's Electronic Health Record (EHR) is an integrated, modular, problem-oriented and patient-centered system, used in the different clinical scenarios (ambulatory, hospitalization, emergency center, and home care) [11,12].

As part of the information system, an integrated Personal Health Records (PHR) called POPES is available to all outpatients since 2007. PHR allows patients receiving medical care in the hospital network to access and verify clinical and administrative information, and to interact with the health system. Among its main functionalities, the PHR allows users to update their personal information, share information, manage scheduled appointments, view test results, check, order and buy prescribed medication, and the possibility to consult with the healthcare team through Telemedicine tools, and it has a messaging service for communication with the general practitioner" At present, POPES has approximately 400,000 registered users [13].

Study Design

This study is an observational, descriptive, and analytical study of a retrospective cohort in the HIBA.

All patients included in this study were adults (defined as >= 18 years of age at the time of consultation) insured by HIBA's HMO, who consulted via the URTI virtual attention program between May 21st, 2018 and September, 14th, 2018.

Patients were followed-up for 7 days from the date of the virtual consultation for consultations to the ED and/or hospitalization.

The URTI program was publicized from May, 2nd 2018 through different institutional communication channels: HIBA TV (an internal television channel that displays short videos in waiting rooms), e-mails, Health Portal, HIBA's Intranet, and brochures distributed in the hospital (Figure 1).

It was initially designed to provide virtual attention to adults aged 18 to 65, insured by HIBA's HMO (a restriction due to administrative issue), with a Patient's Portal account, without risk factors (defined as: chronic pulmonary conditions -COPD, asthma , cystic fibrosis, bronchiectasis, tracheostomy-, chronic heart failure, chronic kidney disease in dialysis, chemotherapy treatment in the prior month, transplant, hospitalization in the prior month, pregnancy, and / or immunosuppression) that self-registered for attention for URTI symptoms. Providers could choose to refer patients older than 65 years old and/or with any of the stated risk factors to the ED, or decide to go through with the consultation if they judged the patient to be fit for it –e.g: mild asthma.

The service was implemented on May, 21st 2018, and was made available on business days from Mondays to Fridays, between 9 A.M. and 9 P.M. The consultation was registered in the Electronic Health Record (EHR) using a predetermined structured form designed for this project. The variables included in this form are used as secondary datasets for the description of the care process implemented.

The virtual attention system consisted of a video conference with video, audio and chats as functionalities to optimize effective communication (see Figure 2). The attending doctors could create and send to the patient's health portal a PDF certificate of care, indicating the prescribed hours of rest.

Figure 1– Advertising leaflet

The teleconsultation service implemented was designed in-house and embedded in the EHR and the PHR. Its development was based on the premise of not needing the installation of any software, besides Google Chrome. With that in mind, we used the Web-RTC (Real-Time-Communication) framework for the transmission of high quality and performance audio and video, a signaling service built with WebSocket and HTML-5 for the front-end and access to peripherals. This project required a server, developed in NodeJS with socket.io for establishing connections, and a gateway TURN/STUN in order to offer this service without compromising the security of the hospital network. We carried out iterative cycles of development and testing until we obtained a satisfactory final version for both desktop and mobile devices.

Figure 2 - Teleconsultation service embedded in the EHR and the PHR

The research project was approved by the Institutional Research Protocols Ethics Committee. The anonymity and confidentiality of the information were guaranteed. There were no potential risks for the patients involved in this study.

The quantitative variables are presented according to the distribution as means and standard deviation (SD), median and interquartile range (RIC) or 25-75 percentiles. The categorical variables are presented as absolute frequency and relative frequency (percentage). The rate is presented as prevalence, with its respective 95% confidence interval.

Results

During the study period, 434 virtual consultations were performed. We excluded those that were not effective (due to interruptions or other technical issues that precluded effective connection), patients that did not meet inclusion criteria and those with incomplete records of the consultation in the EHR (due to missing data on the care process). After the exclusions there were a total of 218 consultations for the analysis (Figure 3).

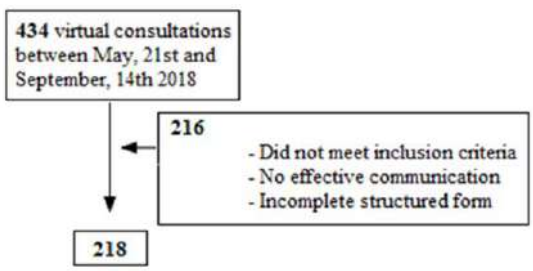


Figure 3– Flow chart of included patients

The most frequent causes of communication failures were:

- *Other reasons for consultation, not related to URTI symptoms*: request for test results, electronic order request for laboratory testing and urine culture, request for proof of blood type, information about flu vaccination, information about preparation for fecal occult blood, peripheral facial paralysis (one patient, who was referred to the ED), skin rash, dermatitis, request for a medical certificate of good health for work.
- *Technical failures in communication:* the patient initiated the communication but never answered the doctor's video call, technical issues with audio and/or video, interruptions in communication, patients with hearing loss, logistical problems when doctors changed shifts.
- *Non-compliance of age:* adult patients registered requesting attention to their children < 18 years old, patients under 18 (referred to a pediatrician), caregivers of patients > 65 years old registered to request attention to their patients.

It should also be noted that, within the subgroup with effective communication, we accomplished to overcome distance barriers for 1 patient from Bariloche (1,000 miles away from Buenos Aires) and 1 patient abroad in Chile.

The frequency of consultations had little variation during each month of the study (see Figure 4), and they were distributed as follows: 6 in May, 57 in June, 59 in July, 65 in August and 31 in September. The service provided was limited to business days between 9 A.M. and 9 P.M., and the daily distribution of consultations was as follows: 23.4% (51) Tuesdays, 21.6% (47) Wednesdays, 20.2% (44) Thursdays, 18.8% (41) Mondays and 16.1% (35) Fridays.

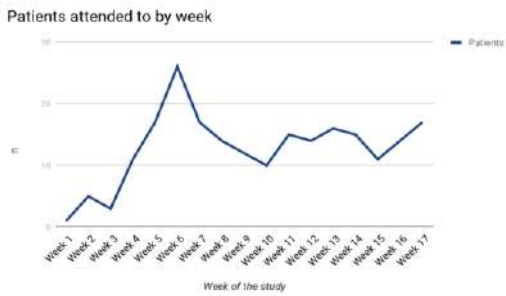


Figure 4– Distribution of frequency of patients attended to, according to date

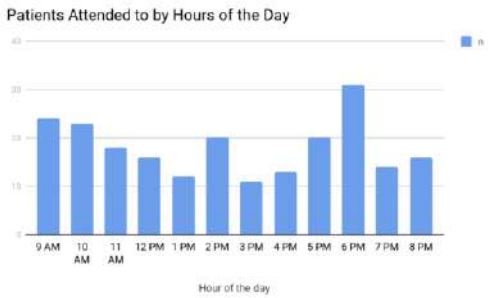


Figure 5– Distribution of frequency of patients attended to, according to time of day

The busiest range of hours was from 9 A.M. to 12 P.M., with 65 out of 218 consultations (global percentage 29.82%, see Figure 5), although the period of highest usage was between 6 pm and 7 pm.

The patients had a mean age of 46.6 years (SD 16.1), and 72.0% (157) were female. Table 1 shows patients' characteristics, including age, gender and other relevant clinical information recorded prospectively in the structured form. 206 patients carried out 218 consultations (10 patients consulted twice, 1 did so 3 times and 1 consulted 4 times).

Table 1– Patients characteristics

	n=218
Age, in years *	46.6 (16.1)
Female	72.0% (157)
Chronic diseases	
Chronic pulmonary conditions	4.6% (10)
Congestive Heart Failure	0.5% (1)
Immunosuppression	1.4% (3)
Dialysis	0% (0)
Chemotherapy in the last month	0% (0)
Transplant	0% (0)
Pregnancy	0.9% (2)
Non-planned admissions in the prior month	0.9% (2)
* mean (standard deviation)	

The data on the care process registered in the EHR are detailed in Table 2. The most frequent presumptive diagnoses were influenza and pharyngitis, with a combined 66.1%. Most patients (86.7%) were prescribed with symptomatic treatment. The vast majority of patients solved their consultation by this means, without the need for referral to a care center for a face-to-face assessment.

Table 2– Consultations and patient attention process

	n: 218
Dyspnea	5.1% (11)
Fever	39.0% (85)
Cough	72.9% (159)
Presumptive diagnosis	
Bronchitis	12.4% (27)
Pharyngitis	20.2% (44)
Influenza	45.8% (100)
Rhinitis	12.8% (28)
Sinusitis	2.8% (6)
Tracheitis	6.0% (13)
Symptomatic treatment	86.7% (189)
New virtual consultation suggested	8.7% (19)
Consultation in walk-in care center suggested	10.6% (23)
Referral to the walk-in care center	5.5% (12)
Referral to ED	6.4% (14)
Ambulance dispatch	-
Doctor home visit scheduled	0.5% (1)

The patients were followed up for 7 days from the time of the virtual consultation in order to estimate the consultation rate in the ED. Only 14 patients were referred by physicians to the ED, but a total of 26 patients consulted the ED, representing a rate of 11.9% (95% CI 7.9-17.0%). 23 of them did so for reasons related to the virtual consultation –e.g.: pharyngitis, laryngitis–, and 3 for unrelated reasons –diverticulitis, back pain.

In the follow-up period, there were no hospitalizations for any of the participants.

Discussion

Based on our work, we deem this new care strategy to have great potential during the seasonal outbreak: it could help prevent hospital *overcrowding*, reduce long delays inattention (or at least offer the possibility of waiting at home), avoid unnecessary referrals, and limit infections in waiting rooms. In accordance with previous studies, we consider that Telemedicine and e-health systems will have a growing role in the care process, bringing information and services to patients wherever they are [14] and allowing professionals to have greater degrees of mobility.

Users of this system were mostly young patients, which could be related to the fact that they are more frequently active and busy during work hours, or that they are more keen on choosing this specific channel for addressing low complexity health issues. This also explains why the levels of comorbidity burden were low. Older population may use this system with less frequency due to technological barriers (smartphone, computer or tablet required) and/or their preference for personal contact with caregivers.

It is also interesting that 72% of patients were female. This is clearly in relation to the population targeted in this project: patients insured with HIBA's HMO, which consists of an inverted population pyramid with over 70% of women.

It should be noted that some of the users displayed a certain resistance, since there were very few registered patients in the first weeks (perhaps due to insufficient advertising), and so did some of the providers, who expressed doubts or fears on Telemedicine during training sessions. It is clear that technological advances are shaping new paradigms in the relationships between individuals. Some authors mention advantages over the usual practice of Medicine, such as avoiding the need for transportation to a health institution, cost savings for the health system and time-saving for doctors (since video consultations are much shorter than the conventional ones [10]).

The elevated number of consultations during this first experience with a teleconsultation system for instant treatment suggests that healthcare users are willing to use new alternative strategies to conventional ones, even considering the selection bias inherent in the study design, since it was directed to adults over 18 years old, insured with HIBA's HMO, with an active account in the patient's health portal. It is also noteworthy that there were many consultations aborted due to technical inconvenience. Even though we were unable to account for all of these issues, many patients that were contacted by telephone afterward reported connectivity problems, which is very usual in-home connections in our country.

We must highlight three findings derived from our study. In the first place, there is a need for Telemedicine for pediatric patients with this pathology, which could guide the population targeted for the use of this tool in the future. Furthermore, no patients with a history of transplant, chronic renal failure or chemotherapy in the last month have attempted to use the service, probably as a result of education as an important tool in the management of patients with these chronic diseases [15]. Third, there were many patients with no respiratory symptoms seeking medical attention through this channel, including reasons for consultation that could be solved or managed by a video call. On the one hand, this leads us to reflect on the effectiveness of existing communication channels that were designed to solve these issues [16]. On the other hand, it allowed us to consider that Telemedicine could play a bigger role in the health system, prompting us to begin a new project for implementing this tool as a medical triage, which will be published in due time.

As for presumptive diagnoses, the most frequent were influenza and pharyngitis. Accordingly, 86.7% of patients were prescribed symptomatic treatment, and the vast majority of patients solved their consultation by the means of Telemedicine, without the need for a referral to a care center. This is probably related to the fact that the most frequently reported symptom was a cough, without the presence of fever. This shows that URTIs are low complexity pathologies, which contribute to the unnecessary *overcrowding* of EDs, requiring a strategy to decentralize medical attention and educate the population on the right channels for consultation. The shift of the healthcare model towards preventive services is essential to successfully accomplish a person-centered healthcare system, leading to patient satisfaction as an essential component and an important indicator of the quality of care [17].

Although it would have been interesting to explore user satisfaction, adoption, and resistance, we lacked the time to analyze this aspect prior to implementation, which constitutes a clear limitation to our study. However, we plan on evaluating these aspects in future lines of work, along with many lessons learned from this experience regarding technical issues and usability.

Conclusions

We can conclude that the use of a Telemedicine tool with video conference for URTI symptoms was attractive and practical, both for patients and professionals. This new care strategy could help prevent *overcrowding* in high complexity health institutions, reduce delays in attention and avoid unnecessary referrals, all of which could have a significant impact on healthcare costs and patient and provider satisfaction.

References

[1] A. Carbo, M. Gupta, L. Tamariz, A. Palacio, S. Levis, Z. Nemeth, and S. Dang, Mobile Technologies for Managing Heart Failure: A Systematic Review and Meta-analysis, *Telemed. J. E. Health.* (2018). doi:10.1089/tmj.2017.0269.

[2] C. Rat, S. Hild, J. Rault Sérandour, A. Gaultier, G. Quereux, B. Dreno, and J.-M. Nguyen, Use of Smartphones for Early Detection of Melanoma: Systematic Review, *J. Med. Internet Res.* **20** (2018) e135.

[3] B.C. Helsel, J.E. Williams, K. Lawson, J. Liang, and J. Markowitz, Telemedicine and Mobile Health Technology Are Effective in the Management of Digestive Diseases: A Systematic Review, *Dig. Dis. Sci.* **63** (2018) 1392–1408.

[4] G. Labiris, E.-K. Panagiotopoulou, and V.P. Kozobolis, A systematic review of teleophthalmological studies in Europe, *Int. J. Ophthalmol.* **11** (2018) 314–325.

[5] T. Alanzi, Role of Social Media in Diabetes Management in the Middle East Region: Systematic Review, *J. Med. Internet Res.* **20** (2018) e58.

[6] R. Cutrera, E. Baraldi, L. Indinnimeo, M. Miraglia Del Giudice, G. Piacentini, F. Scaglione, N. Ullmann, L. Moschino, F. Galdo, and M. Duse, Management of acute respiratory diseases in the pediatric population: the role of oral corticosteroids, *Ital. J. Pediatr.* **43** (2017) 31.

[7] E. al Giunta DH, [Analysis of Crowding in an Adult Emergency Department of a tertiary university hospital]. - PubMed - NCBI, (n.d.). https://www.ncbi.nlm.nih.gov/pubmed/28898330 (accessed November 23, 2018).

[8] J.A. Bentley, S. Thakore, W. Morrison, and W. Wang, Emergency Department redirection to primary care: a prospective evaluation of practice, *Scott. Med. J.* **62** (2017) 2–10.

[9] S.T. Rosenbloom, T.L. Daniels, T.R. Talbot, T. McClain, R. Hennes, S. Stenner, S. Muse, J. Jirjis, and G. Purcell Jackson, Triaging patients at risk of influenza using a patient portal, *J. Am. Med. Inform. Assoc.* **19** (2012) 549–554.

[10] J.S. Ashwood, J. Scott Ashwood, A. Mehrotra, D. Cowling, and L. Uscher-Pines, Direct-To-Consumer Telehealth May Increase Access To Care But Does Not Decrease Spending, *Health Affairs*. **36** (2017) 485–491. doi:10.1377/hlthaff.2016.1130.

[11] F. Plazzotta, D. Luna, and F. González Bernaldo de Quirós, [Health information systems: integrating clinical data in different scenarios and users], *Rev. Peru. Med. Exp. Salud Publica*. **32** (2015) 343–351.

[12] D. Luna, F. Plazzotta, C. Otero, F. González Bernaldo de Quirós, A. Baum, and S. Benítez, Incorporación de tecnologías de la información y de las comunicaciones en el Hospital Italiano de Buenos Aires, (2012). http://hdl.handle.net/11362/3959 (accessed November 23, 2018).

[13] M.I. Smith, G. Garcia, M. Simon, L. Bruchanski, J. Frangella, J.A. Sommer, M.V. Giussi, and D.R. Luna, Lessons Learned After Redesigning a Personal Health Record, *Stud. Health Technol. Inform.* **245** (2017) 216–220.

[14] J.L. Monteagudo, L. Serrano, and C. Hernández Salvador, La telemedicina: ¿ciencia o ficción?, *An. Sist. Sanit. Navar.* **28** (2005). doi:10.4321/s1137-66272005000500002.

[15] O. Cristina Herzog, Educación al paciente y la familia en un programa de trasplante, experiencia en CLC, *Revista Médica Clínica Las Condes*. **21** (2010) 293–299.

[16] J.G. Anderson, M.R. Rainey, and G. Eysenbach, The impact of CyberHealthcare on the physician-patient relationship, *J. Med. Syst.* **27** (2003) 67–84.

[17] R. Hemadeh, R. Hammoud, O. Kdouh, T. Jaber, and L. Ammar, Patient satisfaction with primary healthcare services in Lebanon, *Int. J. Health Plann. Manage.* (2018). doi:10.1002/hpm.2659.

Address for correspondence

Santiago Andrés Frid MD, Department of Health Informatics, Hospital Italiano de Buenos Aires. Juan Domingo Perón 4190, C1181ACH. Ciudad de Buenos Aires, Argentina. email: santiago.frid@hospitalitaliano.org.ar

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190291

Diagnostic Informatics: Its Role in Enhancing Clinical Excellence, Patient Safety and the Value of Care

Andrew Georgiou, Rae-Anne Hardie, Maria R. Dahm, Julie Li, Judith Thomas, Gorkem Sezgin, Ling Li, Johanna I Westbrook

Centre for Health Systems and Safety Research, Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

Abstract

Diagnostic informatics encompasses the role of information technology in key areas of the diagnostic testing (pathology and medical imaging) process, including the selection of appropriate tests and interpretation and follow-up of test results. We present three case studies employing diagnostic informatics methodologies to demonstrate their potential use and value in health services research: (1) Data analytics applied to diagnostic data linked with patient outcome data as a means of enhanancing the monitoring of the quality and appropriateness of diagnostic test choices; (2) Business process modelling which can help to highlight healthcare processes in the diagnostic pathway as a means of improving safety and performance, and (3) Consumer involvement in the diagnostic research process to assist in the establishment of person-centred test result management systems. The case studies provide evidence of the role that diagnostic informatics can have in improving the quality and safety of patient care.

Keywords:

Diagnosis; Quality; Safety.

Introduction

Diagnostic testing, (laboratory medicine, anatomic pathology and medical imaging), is an essential part of healthcare systems. Diagnostic testing generates information that is crucial to the prevention, diagnosis, prognosis, stratification of risk, and treatment of disease [19]. Whilst diagnostic testing may account for a small (less than 5%) proportion of most hospital budgets, there is evidence to show that laboratory services influence 66% of clinical decision making [8]. Despite the importance of laboratory and medical imaging in clinical care, diagnostic testing and its impact on healthcare processes has been an overlooked area of health services research [2].

Diagnostic error involves the failure to either establish an accurate and timely explanation of a patient's diagnosis, or communicate the explanation to the patient [15]. Diagnostic error poses a serious risk to patient safety, with major studies showing that it contributes to approximately 10% of patient deaths and 6-17% of hospital adverse events [15].

Factors which can contribute to diagnostic error include: problems with collaboration and communication among clinicians, patients and their families; lack of infrastructure to support the diagnostic process; and inadequate attention to understanding the problem and its causes. [15]. Failures can occur across a number of areas in the diagnostic process beginning with a failure to engage with a patient, to order the correct or appropriate test, follow up test results with the patient, or gather, integrate or interpret the necessary information, which can result in diagnostic errors [5]. Existing evidence has shown that over 40% of patients leave the hospital before all test results are finalised, 9.4% of which were deemed actionable by independent review [18]. The failure to follow-up test results can lead to missed diagnoses, delayed treatments, unnecessary healthcare utilisation and preventable harm [4]. Effective solutions must engage all stakeholders to arrive at decisions about who needs to receive the test results, how and when the results are communicated, and how they are acknowledged and acted upon [10].

We use the term *diagnostic informatics* to define the role that information technology plays in generating, gathering, integrating, interpreting, and communicating clinical test data and information. The diagnostic informatics landscape encompasses key areas of the diagnostic process, starting with the selection of the appropriate test/referral to address a clinical question, the quality and efficiency of the analytical process, and finally the interpretation, communication and follow-up of test results (including engagement with patients) and their impact on enhancing the value of care and patient outcomes (see Figure 1).

Figure 1. Diagnostic informatics landscape

Major areas of diagnostic informatics research can thus include the study of the choice of the appropriate laboratory/medical imaging request, the quality and efficiency of the analytical process, and the interpretation and follow-up of test results and their impact on patient care outcomes. Information and communication are a key component of the whole diagnostic process. The aim of this paper is to demonstrate how diagnostic informatics research can be used to inform decision-making and improve health outcomes. We outline three case studies to illustrate the application of diagnostic informatics research to the effectiveness of healthcare:

1. **Data analysis and linkage**, the impact of electronic ordering on the rate of potentially unnecessary repeat tests for older hospital inpatients.
2. **Organisational workflows**, the use of Business Process Model and Notation (BPMN) techniques to model and simulate crucial test result management and communication workflows in the diagnostic process.
3. **Patient/consumer involvement in the co-development of research approaches**, where consumers are partners in the development of the research process.

Case studies

Case study 1: Using data linkage to study repeat testing amongst older hospital patients

Aim

The aim of this study was to determine whether electronic provider ordering of laboratory could contribute to a reduction in potentially unnecessary repeat tests.

Methods

A retrospective study investigated 1,367,015 laboratory tests from 55,979 admissions of inpatients aged 80 years and over across three metropolitan hospitals from New South Wales, Australia, between 2014 and 2016. Data from the pathology service laboratory information system (Omnilab v9.4.2 SR10 updated to v11.1.1 SR23 in 2016), containing laboratory test information, and the patient administration system (Cerner PowerChart v2010.02.16), containing admission information, were linked by matching de-identified patient medical record number, gender, date of birth, hospital, and date-times of laboratory tests and admissions-discharges. The linked dataset was used in the analyses. The five most frequently utilised laboratory tests were identified and used to investigate repeat testing. The proportion of repeats were reported for electronic and paper-based test orders.

Results

The most frequently utilised laboratory tests for older inpatients aged 80 years and over were Electrolytes-Urea-Creatinine (EUC), Full Bood Count (FBC), Calcium-Magnesium-Phosphate (CaMgPhos), C-Reactive Protein (CRP), and Liver Function Test (LFT). The total number of tests by the ordering method (paper or electronic) are shown in Table 1. Among EUC, FBC, and LFT tests, there were more repeat tests among paper-based orders compared to electronic orders. The reverse was true for CaMgPhos and CRP tests. (Table 1).

Table 1. Proportion of tests which are repeats of a previously conducted test during a patient's admission, shown by paper-based or electronic ordering of tests

Test		Repeat Tests (%)	Total (n)
EUC	Paper	82.1	7,474
	Electronic	78.3	236,883
FBC	Paper	80.6	6,662
	Electronic	77.6	231,170
CaMgPhos	Paper	68.5	5,070
	Electronic	70.6	121,125
CRP	Paper	67.2	4,460
	Electronic	72.0	85,915
LFT	Paper	63.5	4,085

The repeat test interval (i.e., the time between successive repeat tests) was shorter for paper-based orders (examples of two of these tests, FBC and CRP, are shown in Figure 2). The difference in the cumulative proportions of repeat-tests between electronic and paper ordering became progressively smaller with increasing time between the repeat tests. Similar time-trends were observed for all of the top five tests.

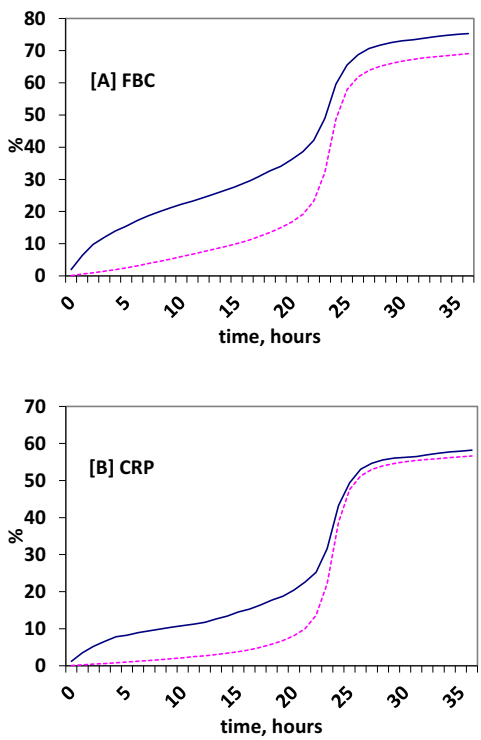


Figure 2. Cumulative proportions of repeat testing (horizontal axis) as a proportion of all test orders, by time (vertical axis; hours) for two analytes. Time intervals up to 36 hours shown. (A) Full blood count (FBC) (B) C-reactive protein (CRP) (Solid line=paper, dashed line=electronic)

Discussion

This study showed that paper-based orders were repeated with shorter time between tests when compared to electronic orders. This observation is similar to the study by Li *et al.* [14], which also found shorter time between repeat tests in paper-based test orders among paediatric intensive-care unit patients. Guidelines outline the minimum repeat testing intervals for a meaningful change to be observed as: 12 hours for EUC and FBC tests, 36 hours for LFT and CaMgPhos tests, and 24 hours for CRP tests [13]. Although repeats before the suggested minimum retest interval were observed for all tests, paper-based orders had a higher proportion of repeats with shorter intervals between tests when compared to electronic orders. Potential reasons for this difference as proposed by Li *et al* included: (1) duplicate order alerts were provided in the electronic ordering system, notifying clinicians that an identical test had already been ordered within 24 hours for the same patient; and (2) clinicians could more easily see existing test orders and results in the electronic ordering system, thus leading to better self-regulation of ordering decisions [14]. This study provides valuable information on the differences between electronic and paper-based laboratory test orders on repeat tests. The findings suggest that electronic ordering can potentially reduce potentially unnecessary short-interval repetitions of tests.

Case study 2: Using Business Process Model and Notation (BPMN) techniques to model communication workflows in the diagnostic process

Aim

The aim of this study was to model communication workflows in the diagnostic process.

Methods

In order to form an in-depth understanding of the context in which a system (e.g. electronic test result management) operates, we created a generic high-level process model for test-results management. The model was developed using the Business Process Model and Notation (BPMN) standard [16], a technique which uses structured analysis to diagrammatically represent business processes using standard symbols to provide insights into the complexity of healthcare processes. Models were created with the Bizagi Modeller freeware application.

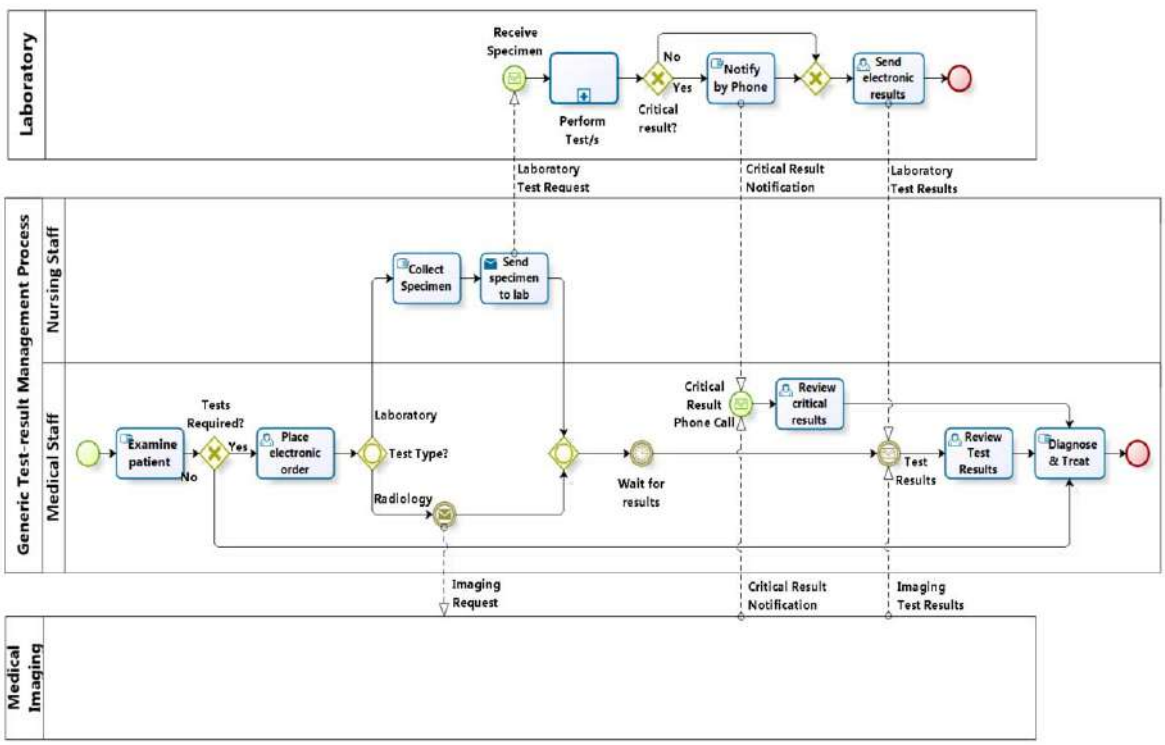


Figure 3. High-level process model for test results management developed using the Business Process Model and Notation (BPMN) standard

Results

The BPMN 2.0 standard allows processes to be represented using '*pools'* (which can be divided into multiple *'lanes'*) to clearly identify which participants perform each '*activity'*. This is demonstrated in Figure 3 as a high level process model. Three 'pools' are represented, namely the laboratory, medical imaging and a clinical department (represented by a medical staff 'lane' and a nursing staff 'lane'). Medical imaging is represented as a 'black box' i.e. only the message flows into and out of the pool are represented. In contrast, the laboratory and clinical department are modelled as 'white boxes' where all activities related to the process are visible. Following the arrows from the start (green circle event) of the process shows the sequence flow of activities (blue rectangles). The process pathway deviates based on the decision outcomes of questions posed at each gateway (yellow diamond).

The model in Figure 3 depicts the interaction between the pool containing the process of test result management by medical and nursing staff in a hospital test setting, with the pools of laboratory and medical imaging processes. Diamond shaped *'gateways'* depict decisions in a process where one or more

alternative paths may be taken (e.g. is a test required, or is a result critical). Circular shaped '*events'* (e.g. receiving a specimen, waiting on results, or calling through a critical result) can be used to indicate pertinent events in the process including (but not limited to) the start or end of a process, time related events or instances where messages are sent/received.

Discussion

The BPMN model highlights the value of using structured analysis to visualise healthcare processes. The technique allows proposed changes to processes to be modelled to determine their impact on existing workflows. As more data are collected, it will be possible to use BPMN to run theoretical process simulations.

BPMN enables business processes to be visually represented for ease of understanding and analysis. It can thus aid in enhancing the design and evaluation of evidence-based interventions, such as test result management tools/interventions (including electronic decision support aids) to consider how they impact on existing communication processes. This is particularly important for the laboratory notification of high-risk test results to the appropriate and responsible clinician, for immediate clinical action. Other clinical scenarios that have applied BPMN techniques include modelling clinical pathways [17; 20], for process improvement [1; 6; 20], for pre/post implementation workflows [9] and in genetic testing processes [6].

Case study 3: Engaging consumers in the analysis of diagnostic studies

Aim

The aim of this study is to involve consumers in the diagnostic informatics research process to assist in the establishment of person-centred test result management systems.

Methods

Consumer-focused research is a critical element in the development of person-focused care and shared decision making, which can contribute to improvements in the safety and quality of care. Health consumers can be defined as previous, current, or potential patients and their carers accessing healthcare services [11].

A core component of our research strategy was the establishment of a Consumer Reference Group which engaged consumers as partners in the research process. The value of research increases when consumer participants are familiar with how research is designed and undertaken [12]. Meaningful engagement of patients as partners in the research team requires the provision of adequate preparation and training for their research roles [3], which are achieved through capacity building activities and workshops. These hands-on learning activities ensure that the patient/consumer is involved in all phases of the research project.

Consumer representative organisations provided input during the development of our research proposal. We held a stakeholder forum to launch the project, where key recommendations were identified to address threats to patient safety resulting from failure to follow up test-results.

Informed by the stakeholder forum, we conducted semi-structured interviews with a diverse purposive sample of clinicians, radiology and laboratory staff, and patients within three NSW Emergency Departments in Australia. Consumer (including through interviews with patients) and clinician perspectives positively shaped the direction of the research study [7]. Interview results were used to compare current work processes and gauge patients experience of the test-result management cycle. Re-iterative qualitative thematic analysis was conducted.

Results

Ten consumers formed a Consumer Reference Group (CRG) and ranked themes (previously identified in interview analysis) according to their chosen order of importance. The CRG then analysed (in an interactive qualitative data analysis workshop with members of our research team) the two topics with the highest ranking: 1) 'Transitions of care' (how and if results move from one healthcare setting to another including procedures related to patient discharge); and 2) 'Access' (whether and how a person can have access to their results).

Discussion

This innovative, consumer-driven approach engaged consumers/patients in inclusive research which provided rich insights into consumer experience and expectations in test-result management. Moving forward, consumers will be invited to contribute to the preparation of research publications, including contributions of intellectual content and critical revisions of the manuscript. Dissemination of the study findings via academic and policy publications will inform future directions for research in this area.

Figure 4. Consumer reference group and research team

Enhancing consumer contribution through inclusive research has required, and continues to necessitate, a rigorous, open, negotiated, and interactive process. Ensuring consumer/patient involvement during all phases of the project, including the inclusion of consumers in data analysis, represents a step towards research that is co-produced with consumers. Our consumer engagement strategy helps build capacity by assisting consumers in developing relevant research skills which will enable them to effectively navigate and inform outcomes-based research.

Conclusions

Diagnostic informatics requires an innovative and inter-disciplinary approach to enhancing clinical excellence and quality care through the incorporation of outcomes-based approaches to monitoring and measuring diagnostic quality. These approaches include the building of robust communication workflows and evidence-based electronic decision support systems, and the establishment of shared decision-making through consumer/patient involvment.

The three case studies outlined above demonstrate how diagnostic informatics can be used to answer key research questions. In Case 1, linking diagnostic data with patient outcomes uncovered valuable information on differences

between electronic and paper-based laboratory test orders and the impact of electronic ordering on repeat tests. This evidence can contribute to the development of decision support tools to promote appropriate and safe test ordering. Case 2 demonstrated how business process modelling can be used to visualise healthcare processes in the diagnostic pathway, helping to optimise design through the identification of potential problem areas and the modelling (and simulation) of interventions. The ability to identify and explicate health care processes provides a valuable means of enhancing the design and implementation of digital systems, thus providing a valuable connection for the development of research tools identified in Case 1. Finally, Case 3 showcased the importance of consumer-driven analysis and interpretation of study findings, which allowed us to identify consumer-perceived barriers, leading to the development of person-centred, safe and effective test result management systems. A better understanding of diagnostic informatics as an integral part of the diagnostic process is crucial to the future quality and safety of patient care.

Acknowledgements

This work was supported by National Health and Medical Research Council, Partnership Project Grant number 1111925 with partners including NSW Health Pathology and the Australian Commission on Safety and Quality in Healthcare. We also acknowledge and thank all participants who participated.

Ethics approval was granted by the Human Research Ethics Committee of the relevant Local Health District, ratified by Macquarie University. All relevant site-specific assessments have been approved.

References

[1] G. Antonacci, A. Calabrese, A. D'Ambrogio, A. Giglio, B. Intrigila, and N.L. Ghiron, A BPMN-based automated approach for the analysis of healthcare processes, *Proceedings - 25th IEEE International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises, WETICE 2016* (2016), 124-129.

[2] D.W. Bates and H. Singh, Two Decades Since To Err Is Human: An Assessment Of Progress And Emerging Priorities In Patient Safety, *Health Affairs (Millwood)* **37** (2018), 1736-1743.

[3] A. Black, K. Strain, C. Wallsworth, S.-G. Charlton, W. Chang, K. McNamee, and C. Hamilton, What constitutes meaningful engagement for patients and families as partners on research teams?, *Journal of Health Services Research & Policy* (2018), 1355819618762960.

[4] J. Callen, A. Georgiou, J. Li, and J. Westbrook, The safety implications of missed test results for hospitalized patients: a systematic review, *BMJ Quality and Safety* **20** (2011), 194-199.

[5] K. Cosby, Diagnostic error, in: *Diagnosis - Interpreting the Shadows,* P. Croskerry, K. Cosby, M. Graber, and H. Singh, eds., CRC Press, Florida, 2017, pp. 161-186.

[6] E.M. Cutting, C.L. Overby, M. Banchero, T. Pollin, M. Kelemen, A.R. Shuldiner, and A.L. Beitelshees, Using Workflow Modeling to Identify Areas to Improve Genetic Test Processes in the University of Maryland Translational Pharmacogenomics Project, *AMIA Annual Symposium Procedures* **2015** (2015), 466-474.

[7] M. Dahm, A. Georgiou, R. Herkes, A. Brown, J. Li, R. Lindeman, A. Horvath, G. Jones, M. Legg, L. Li, D. Greenfield, and J. Westbrook, Patient groups, clinicians and healthcare professionals agree - all test results need to be seen, understood and followed up. *Diagnosis* 2018;(5):215-222.

[8] R.W. Forsman, Why is the laboratory an afterthought for managed care organizations?, *Clinical Chemistry* **42** (1996), 813-816.

[9] A. Georgiou, E.J. McCaughey, A. Tariq, S.R. Walter, J. Li, J. Callen, R. Paoloni, W.B. Runciman, and J.I. Westbrook, What is the impact of an electronic test result acknowledgement system on Emergency Department physicians' work processes? A mixed-method pre-post observational study, *International Journal of Medical Informatics* **99** (2017), 29-36.

[10] D. Hanna, P. Griswold, L. Leape, and D. Bates, Communicating critical test results: safe practice recommendations, *Joint Commission Journal on Quality and Patient Safety* **31** (2005), 68-80.

[11] Health Consumers NSW, Consumer and Community Engagement Model- An outcome of the WentWest Health Consumers NSW Joint Consumer Engagement Project, in, Sydney, 2015.

[12] J.P. Ioannidis, Why Most Clinical Research Is Not Useful, *PLoS Med* **13** (2016), e1002049

[13] Leeds Pathology, Responsible Pathology Requesting, Available at: www.pathology.leedsth.nhs.uk/pathology/Departments/BloodSciences/MinimumReTestIntervals.aspx Accessed 25 November 2018

[14] L. Li, A. Georgiou, E. Vecellio, A. Eigenstetter, G. Toouli, R. Wilson, and J.I. Westbrook, What is the effect of electronic pathology ordering on test re-ordering patterns for paediatric patients?, *Studies in Health Technology Informatics* **204** (2014), 74-79.

[15] National Academies of Science Engineering and Medicine, *Improving diagnosis in health care*, The National Acadamies Press, Washington DC, 2015.

[16] BPMN 2.0 Object Management Group, About the Business Process Model and Notation Specification Version 2.0, Available at: www.omg.org/spec/BPMN/2.0/About-BPMN/ Accessed 30 March 2019.

[17] S. Rodriguez-Loya, A. Aziz, and C. Chatwin, A service oriented approach for guidelines-based clinical decision support using BPMN, *Studies in Health Technology Informatics*. **205** (2014), 43-47.

[18] C.L. Roy, E.G. Poon, A.S. Karson, Z. Ladak-Merchant, R.E. Johnson, S.M. Maviglia, and T.K. Gandhi, Patient safety concerns arising from test results that return after hospital discharge, *Annals of Internal Medicine* **143** (2005), 121-128.

[19] M. Rubinstein, R. Hirsch, K. Bandyopadhyay, B. Madison, T. Taylor, A. Ranne, M. Linville, K. Donaldson, F. Lacbawan, and N. Cornish, Effectiveness of Practices to Support Appropriate Laboratory Test Utilization: A Laboratory Medicine Best Practices Systematic Review and Meta-Analysis, *American Journal of Clinical Patholology* **149** (2018), 197-221.

[20] F. Zerbato, B. Oliboni, C. Combi, M. Campos, and J.M. Juarez, BPMN-based Representation and Comparison of Clinical Pathways for Catheter-related Bloodstream Infections, *2015 IEEE International Conference on Healthcare Informatics (ICHI 2015)* (2015), 346-355.

Address for correspondence

Professor Andrew Georgiou: andrew.georgiou@mq.edu.au

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190292

Possible Usages of Smart Contracts (Blockchain) in Healthcare and Why No One Is Using Them

Alain Giordanengo[a,b]

[a] *Department of Computer Science, UiT The Arctic University of Norway, Norway,*
[b] *Norwegian Centre for E-health Research, University Hospital of North Norway (UNN), Tromsø, Norway*

Abstract

Security, privacy, transparency, consent, and data sharing are major challenges that healthcare institutions must address today. The explosion of the Internet of Things (IoT), the enactment of the General Data Protection Regulation (GDPR), the growing trend of patients self-managing their diseases, and the eagerness of patients to share their self-collected health data with primary and secondary health organisations further increase the complexity of these challenges. Smart contracts, based on blockchain technology, can be a legitimate approach for addressing these challenges. Smart contracts define rules and penalties in an agreement, enforce those rules, and render them irrevocable. This paper presents a state-of-the-art review (as of May 2018) of the possible usages of smart contracts in healthcare and focuses on data sharing between patients, doctors, and institutions.

Keywords:

smart contracts, healthcare, blockchain, trust

Introduction

Since the enactment of the General Data Protection Regulation (GDPR) in May 2018, the security, privacy, transparency, and consent for patient-owned medical data have been at the forefront of the concerns of healthcare institutions. The explicit consent of patients for processing health data and the transparency notice explaining what data will be collected, how it will be collected and patients' rights to full access to their health data [1] have greatly affected healthcare information systems.

In addition to the data generated by healthcare institutions, patients are increasingly active in managing their diseases by collecting health data using mobile devices and sensors [2]. Sharing patients' self-collected data with medical systems has a positive effect on disease management [3], and patients are eager to participate [4].

Blockchain technology is receiving extensive publicity in healthcare and has promised great improvements, such as smart healthcare management and patient empowerment [5]. Smart contracts implemented using blockchains, sometimes referred to as Blockchain 2.0, are protocols permitting the verification and enforcement of legal agreements between two or more parties and rendering them irrevocable. Interest in smart contracts has been growing ever since the creation of Ethereum, the first blockchain-based solution that integrated smart contracts, which was publicly released in 2015. Smart contracts can allow patients to manage access to their health records, secure data exchange, and ensure the privacy of those exchanges [6].

This paper presents a state-of-the-art review of the possible usage of smart contracts in healthcare, their objectives, and their limitations, with a focus on data sharing, and discusses why no one is using them in a real situation today.

Methods

Scientific and grey literature search

The author followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology to perform a scientific and grey literature search. Figure 1 shows the keywords and the search query selected by the author. Three peer-reviewed online databases were searched: PubMed, IEEE Xplore and Web of Science, together with Google Scholar. The author tailored the search query for each online database according to its specific functionalities. The search query was limited to the metadata fields: title, abstract and keywords.

---

smart contracts[Title/Abstract/Keywords] AND
(clinical[Title/Abstract/Keywords] OR
healthcare[Title/Abstract/Keywords])

Figure 1: Search query and keywords used for the scientific literature review.

The author imported all the results from PubMed, IEEE Xplore and Web of Science, as well as the results displayed on the first page of Google Scholar, to Rayyan [7], an online tool that facilitates the review process. The author chose Rayyan based on its lack of cost and flexibility compared to other tools [8]. The author first excluded results based on their metadata fields (title, abstract and keywords) using criteria listed in the next section. The author then reviewed the remaining results for inclusion based on the full texts.

Inclusion and exclusion criteria

The papers needed to meet several criteria to be included in this review. The papers needed to do one of the following:

- Describe a model or an implementation using smart contracts in a healthcare-related situation;
- Illustrate an idea for, or the potential effects of, smart contracts in healthcare systems or medical workflow.

Systematic or literature reviews that provided sufficient information regarding smart contract usages in healthcare were also included.

Papers focusing on the blockchain technology stack or smart contract algorithms, but without illustrating their uses in a clinical setting, were excluded.

Studies reported in languages other than English were excluded.

Data categorisation and data collection

The content of the papers has been organised according to a taxonomy defined by the author for presenting an overview of the usage of smart contracts in healthcare. The categories comprise the following:

1. *State of the presented work*: the part of the life cycle in which the described work is positioned (e.g. proof-of-concept [POC], prototype, production);
2. *Objective*: the situations in which the smart contracts can be used and what their goals are, or what challenges they are addressing;
3. *Content of the smart contracts*: the data or information that the smart contracts contain;
4. *Technology stack*: the frameworks, components, software, or standards on which the smart contracts rely on;
5. *Concerned Actors*: the actors affected by the introduction of the presented work in healthcare (e.g. electronic health records vendors, clinicians, patients);

The author used these categories to evaluate and analyse the included papers. Each included paper was expected to address at least one of these categories.

Results

Reviews on literature

Figure 2 shows the selection of articles. In total, forty-three articles were identified from the literature search: thirty-three from peer-reviewed literature and ten from Google Scholar. Eight duplicates were identified and removed. The author reviewed titles, keywords and abstracts of thirty-five papers, and fifteen were excluded based on the criteria specified in the previous section, leaving twenty articles for full-text assessment. Ten further articles were identified for exclusion for the following reasons:

- Out-of-scope papers (8): five papers cited healthcare settings as potential examples but did not include them at any stage of their studies, while two others limited their trials to blockchain technology that did not involve smart contracts. One paper focused on metrics for assessing blockchain-based healthcare apps instead of describing a model or an idea.
- Inappropriate description (1): the description or testing of an idea included insufficient details that would permit solid reproduction of the claims made.
- Full article innaccessible for review (1).

Ten papers were included in the final collection and analysis.

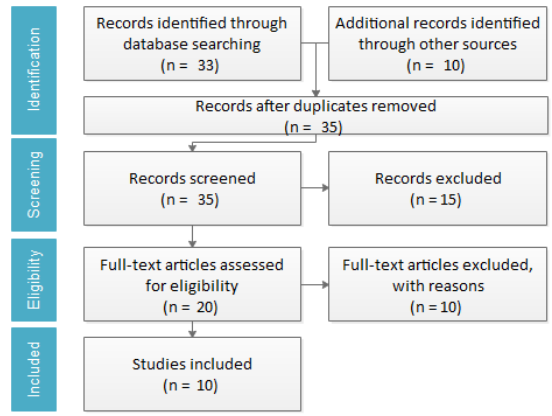


Figure 2: PRISMA flow diagram

Data extraction from included articles

The evaluation and analysis of the included articles (Table 1) are based on the categorizations described previously in the methods section.

State of the presented work

It is important to note that none of the nine studies (omitting one literature review) presented have reached the production stage. Of the nine presented studies, one is a model (i.e. a solution that is not entirely functional), three are POCs (i.e. demonstrating the feasibility of a concept) and five are prototypes (i.e. providing almost all features of an end product). None have been tested in a real-life situation.

Objectives of smart contracts, their contents and concerned actors

Six studies used smart contracts for managing data sharing. These studies concerned patients, medical workers and healthcare institutions.

- Dubovitskaya et al. [9] defined a prototype using smart contracts for exchanging data between patients and doctors and to manage access permissions. The smart contracts contain three types of blocks: 1) patient-defined permissions for allowing doctors to access or share patient- or healthcare-generated health data. The permissions can specify a data category, particular rights (read, write, and share) and a timeframe. They can also force the anonymisation of data. 2) clinical metadata, which contains all required information for accessing the corresponding data files stored off-chain (i.e. in a classic cloud solution). The clinical metadata also contains a hash of the data files to ensure the unforgeability of the data stored in the cloud. 3) patient private data directly attached to the chain by the patient, such as self-collected health data. This is the only prototype system that allows patients to exchange their data actively, without relying solely on data generated by healthcare institutions.
- Dagher et al. [10] proposed using six smart contracts as access controls for sharing medical records between healthcare and insurance providers. The first contract records the users and the mining operations. The second classifies users as patients, providers or third-parties. The third defines the relationships between users. The fourth defines the ownership of medical records, the fifth specifies the access permissions for those records and the last shares symmetric encryption

keys (SEK). Patients interact with the blockchain by changing the access permissions, while the providers use the SEKs to encrypt or decrypt medical records before sending or after receiving them via an off-chain communication channel.

- Azaria et al. [11] used several smart contracts in their data-sharing prototype for different purposes: 1) registrar contracts, which map participant (patient) identification strings (e.g. social security numbers) to their public signing keys to be used in a blockchain. These contracts also contain policies regarding the creation or updating of identities, and only certified institutions can generate them. 2) patient–provider relationship contracts, which allow patients to fine tune the access rights of their providers regarding any portion of their medical data. These contracts also contain data pointers and can be used between providers. 3) summary contracts, which contain the history of all contracts signed by all parties. For instance, they include all the patient–provider relationship contracts of a patient, who can consult them. They also act as a backup.
- Xia et al. [12] used smart contracts for sharing medical data between cloud providers and medical and research organisations. The smart contracts are used for three main purposes: 1) encrypting medical reports, 2) identifying actions performed on sent data, and 3) revoking access to violated data. The smart contracts contain a data sensitivity level, IDs of the owner and requestor (i.e. who is requesting the data), data IDs, permissions and the cryptographic keys.
- The POC defined by Ahram et al. [13] used smart contracts but for limited purposes compared to the previous studies. First, a smart contract ensures that a patient and only a patient is creating the initial version of their medical records during the first visit to a clinic. A second type of smart contract then ensures the update or transfer of the medical record by or to a provider.
- The POC by Saravanan et al. [14] used smart contracts for sharing health data with clinicians that has been self-collected using sensors. The contracts contain access logs and the shared health data. This solution requires patients to share their private keys with their clinicians off-chain before starting to use the solution.

Two other studies rely on smart contracts for improving medical trials. These studies concern researchers, participants in medical trials and research institutions.

- Benchoufi and Ravaud [15] proposed using two smart contracts to ensure the integrity and transparency of medical trials. The first ensures the irrevocability of the trial protocol by containing the protocol of the study and the statistical analysis plan and by defining the data monitoring committee. The second smart contract contains patient enrolment data (consent and information forms), data collection, trial monitoring, data management and data analysis. Using this approach, the authors claim that the reproduciability is improved and study reports and dissemination of results are impartial. Any public institution can monitor the flow and progress of a study and verify its validity.
- Nugent et al. [16] proposed similar usage of two smart contracts for improving the data transparency in clinical trials. The first is a regulator contract, containing clinical trial authorisation details, which is managed by public regulators (e.g. US Food and Drug Administration). The second is a trial contract, managed by the research organisations, which is used for storing trial protocols, consent forms and anonymised participant information.

The final study, by McFarlane et al. [17], focused on the adjudication of medical billing and the provision of medical access in case of emergency. In the first situation, a smart contract containing patient identification, institution denomination, and the debt owed would be issued. The smart contract would be auto-updated once the patient has paid the debt. In the second situation, a smart contract containing a secondary private key (derived from the original private key) could be issued by the patient to allow emergency services to access medical records, should the patient be unresponsive, have their mobile phone present and have configured emergency access to that phone by bypassing the lock screen. The second situation is only an early model, and no more details are given.

Technology

While a comparison of the different blockchain technologies is outside the scope of this article, it is interesting to note that none of the studies are interoperable, even if they use the same blockchain "family". This is due to the use of proprietary data types, with different types of rules and custom codes for managing the automatic execution of smart contracts. In addition, only one addresses interoperability issues by proposing the use of the Fast Healthcare Interoperability Resources (FHIR) specification to represent medical data.

Five types of technologies are used in these studies. Ethereum, a permission-less blockchain (i.e. any user can create and run code, and its execution relies on miners), was the most used (6 studies of 9), together with specialised libraries or languages that targetthis blockchain, such as Solidity (a contract-oriented high level language targetting the Ethereum Virtual Machine). One of the studies relies on Hyperledger, which is a permissioned blockchain. The authors of that study claimed that Hyperledger is more suited for sharing data than Ethereum [9]. It is permission-limited, and the impersonalisation and risk of data misuse due to the anonymisation of permission-less-typed blockchains both increase the likelihood of a Hyperledger system being used and remove the need to pay for transaction execution (mining). Another study relies on IBM blockchain, and one proposes the usage of ErisDB (renamed Monax in 2017 - https://monax.io/2016/11/08/eris-0120-release/) as well as Ethereum.

Conclusion and Discussion

This paper shows that smart contracts could be used in healthcare in different situations, from data sharing to the improvement of clinical trials. Two studies presented allow patients to upload their self-collected data into a blockchain and share it with their clinicians.

However, the small numbers of studies included (n=9, omitting a literature review) and the fact that none of them were at a commercialisation or production stage raise questions about the usability of this technology in real-life situations. A wider systematic review of blockchain technology, conducted in 2016, showed the same limited results, with only three articles examining smart contracts and no production-ready services [18]. Several possibilities could explain this situation in healthcare:

1. Blockchain will not change how medical records are stored. Blockchain is usable as a registry only, because inserting vast amounts of medical data, such as computed tomography (CT) scans, would render the Blockchain bloated and difficult to manage. The challenges of medical data storageing are the same, whether blockchain is used or not.
2. Blockchain technology is not necessary when trusted parties or regulators control the decision-making processes (e.g. creating smart contracts or mining), as in healthcare. Moreover, private blockchains are arguably only a shared database with at best a journaling of the data, which has existed since the seventies [19]. However, blockchain has proved its usefulness in decentralized situations in which parties cannot be trusted, even if some security issues remain unaddressed today [18].
3. Accessing encrypted patient data in the blockchain requires the healthcare institutions to use the patients' private keys (the public keys being used for encrypting the data). The sharing of a private key renders it public, and therefore not secure. In addition, this raises the question of trusted parties, described in point 2.
4. The GDPR states that patients have full access to their data [1], meaning that they have the opportunity to both manage the access rights and to move any portion or all of their data from one provider to another. Moving data between providers implies the deletion of data held by the old provider. However, it is not possible to delete anything from a blockchain without voiding its integrity and recalculating all the hashes.
5. Some of the actors cited are vapourware. For instance, ErisDB (or Monax, as it is now branded) provides no documentation nor access to a single piece of code, but still advertises its products. These practices increase doubts about the usefulness of the technology.
6. There are contradictions regarding the potential impacts of the costs of using a blockchain-based solution by healthcare institutions; some suggest that cost savings could be made [20] while others point out probable cost increases due to the nature of blockchain itself (e.g. computational power and storage increase due to data replication) [21].

Based on these considerations, the author believes blockchain-based technologies are not adapted and not ready yet for usage in healthcare, at the time this study was conducted (May 2018). Moreover, another study has suggested that the usage of these technologies is extremely immature and lacks public or expert knowledge, making it hard to form a clear strategic vision of its true future potential [22].

References

[1] V. Hordern, Data Protection Compliance in the Age of Digital Health, *European Journal of Health Law* **23** (2016), 248-264.

[2] M. Haghi, K. Thurow, and R. Stoll, Wearable Devices in Medical Internet of Things: Scientific Research and Commercially Available Devices, *Healthcare Informatics Research* **23** (2017), 4-15.

[3] M. Peleg, Y. Shahar, S. Quaglini, T. Broens, R. Budasu, N. Fung, A. Fux, G. Garcia-Saez, A. Goldstein, A. Gonzalez-Ferrer, H. Hermens, M.E. Hernando, V. Jones, G. Klebanov, D. Klimov, D. Knoppel, N. Larburu, C. Marcos, I. Martinez-Sarriegui, C. Napolitano, A. Pallas, A. Palomares, E. Parimbelli, B. Pons, M. Rigla, L. Sacchi, E. Shalom, P. Soffer, and B. van Schooten, Assessment of a personalized and distributed patient guidance system, *Int J Med Inform* **101** (2017), 108-130.

[4] M.S. Thomas Pickard, Big Desire to Share Big Health Data: A Shift in Consumer Attitudes toward Personal Health Information, *AAAI Publications, AAAI Spring Symposium Series* (2014).

[5] M. Mettler, Blockchain technology in healthcare: The revolution starts here, *IEEE 18th International Conference on e-Health Networking, Applications and Services* (2016), pp. 1-3.

[6] K. Peterson, R. Deeduvanu, P. Kanjamala, and K.B. Mayo, A Blockchain-Based Approach to Health Information Exchange Networks, (2016).

[7] O. Mourad, H. Hossam, F. Zbys, and E. Ahmed, Rayyan---a web and mobile app for systematic reviews, *Systematic Reviews* **5** (2016), 210.

[8] L. Kellermeyer, B. Harnke, and S. Knight, Covidence and Rayyan, *Journal of the Medical Library Association : JMLA* **106** (2018), 580-583.

[9] A. Dubovitskaya, Z. Xu, S. Ryu, M. Schumacher, and F. Wang, Secure and Trustable Electronic Medical Records Sharing using Blockchain, *CoRR* (2017).

[10] G.G. Dagher, J. Mohler, M. Milojkovic, and P.B. Marella, Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology, *Sustainable Cities and Society* **39** (2018), 283-297.

[11] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, MedRec: Using Blockchain for Medical Data Access and Permission Management, *2nd International Conference on Open and Big Data* (2016), pp. 25-30.

[12] Q. Xia, E.B. Sifah, K.O. Asamoah, J. Gao, X. Du, and M. Guizani, MeDShare: Trust-Less Medical Data Sharing Among Cloud Service Providers via Blockchain, *IEEE Access* **5** (2017), 14757-14767.

[13] T. Ahram, A. Sargolzaei, S. Sargolzaei, J. Daniels, and B. Amaba, Blockchain technology innovations, *IEEE Technology & Engineering Management Conference* (2017), pp. 137-141.

[14] M. Saravanan, R. Shubha, A.M. Marks, and V. Iyer, SMEAD: A secured mobile enabled assisting device for diabetics monitoring, in: *IEEE International Conference on Advanced Networks and Telecommunications Systems* (2017), pp. 1-6.

[15] M. Benchoufi and P. Ravaud, Blockchain technology for improving clinical research quality, *Trials* (2017),335.

[16] T. Nugent, D. Upton, and M. Cimpoesu, Improving data transparency in clinical trials using blockchain smart contracts, *F1000Research* **5** (2016), 2541.

[17] C. McFarlane, M. Beer, J.J. Brown, and N. Prendergast, Patientory: A Healthcare Peer-to-Peer EMR Storage Network v1.1, (2017).

[18] J. Yli-Huumo, D. Ko, S. Choi, S. Park, and K. Smolander, Where Is Current Research on Blockchain Technology?-A Systematic Review, *PLoS ONE* **11** (2016), e0163477.

[19] A. NARAYANAN, "Private blockchain" is just a confusing name for a shared database, URL:https://freedom-to-tinker.com/2015/09/18/private-blockchain-is-just-a-confusing-name-for-a-shared-database/. Accessed: 2019-03-22. (Archived by WebCite® at http://www.webcitation.org/773slK5tg), (2015).

[20] Credit Suisse, Blockchain. URL:https://www.finextra.com/finextra-downloads/newsdocs/document-1063851711.pdf.

Accessed: 2019-03-17. (Archived by WebCite® at http://www.webcitation.org/76wJ7KQJD), (2016).
[21] D. Wood, ETHEREUM: A SECURE DECENTRALISED GENERALISED TRANSACTION LEDGER, (2014).
[22] I. Radanović and R. Likić, Opportunities for Use of Blockchain Technology in Medicine, *Applied Health Economics and Health Policy* **16** (2018), 583-590.
[23] E. Karafiloski and A. Mishev, Blockchain solutions for big data challenges: A literature review, *IEEE EUROCON 2017 -17th International Conference on Smart Technologies* (2017), pp. 763-768.

Address for correspondence

Alain Giordanengo
email: alain.giordanengo@ehealthresearch.no

Table 1 – Papers included in the review

Ref.	State	Objective of smart-contracts	Content	Technology	Actors
[17]	Model	1. Adjudication of medical billing 2. Emergency access of health records allowance	Patient (1) Institution (1) Debt (1) Secondary private key (2)	Ethereum FHIR Amazon Web Services ErisDB	Patients Institutions Debt collectors Insurances Emergency Services
[23]	Literature Review	---	Reference Paper [11]	---	---
[9]	Prototype	Data sharing between patients and doctors, with data generated from both sides	Permissions (patients to doctors) Clinical Metadata Patients' private data	Hyperledger Chaincode ARIA Varian Cloud Go	Doctors Patients
[15]	POC	Improving medical trials by managing consent and ensuring integrity and transparency of the trials	Trial protocol and setup (1) Patients enrolment (2) Data Collection (2) Trial Monitoring (2) Data Management (2) Data Analysis (2)	Ethereum Solidity Chainscript	Trial participants Researchers
[13]	POC	Consent of the patients Record transfer between healthcare networks	Any Protected Health Information Involved health networks	IBM Blockchain Bluemix NodeJS	Patients Doctors
[16]	Prototype	1. Capturing clinical trial authorization 2. Storing clinical trial protocols and collected data	Clinical trial authorization Protocols Collected data	Ethereum Javascript Solidity	Regulators Research Organizations Researchers Doctors Patients
[11]	Prototype	1. Mapping patients ID to their public keys 2. Logging patient-providers relationships, access rights and data retrieval pointers 3. Managing Medical Record history	Patients ID Patients Ethereum address Provider ID Patients-Providers relationships Access permissions Data pointers	Ethereum PyEthereum PyEthApp SQLite Flask	Patients Providers
[10]	Prototype	1. User registration and mining 2. Classify users as patients/providers/third party 3. Relationships of nodes 4. Ownerships of medical records 5. Permission access to medical records 6. Proxy re-encryption	Ethereum address Users ID Relationship status Access Conditions Hashes Symmetric Encryption Key	Ethereum QuorumChain Ethereum-Go	Patients Providers Healthcare Insurance
[12]	Prototype	1. Encrypt reports 2. Identify actions performed on sent data 3. Revoke access to violated data	Cryptographic keys Reports Permissions Data sensitivity level IDs	Undisclosed	Cloud providers Research organizations Medical organizations
[14]	POC	Share self-collected health data	Medical data Access logs	Ethereum	Patients Clinicians

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190293

Developing a Taxonomy of Online Medical Calculators for Assessing Automatability and Clinical Efficiency Improvements

Tim A Green [a], Chi-Ren Shyu [a, b, c]

[a] *Informatics Institute, University of Missouri, Columbia, MO, USA*
[b] *Electrical Engineering and Computer Science Department, University of Missouri, Columbia, MO, USA*
[c] *School of Medicine, University of Missouri, Columbia, MO, USA*

Abstract

Medical calculators play an important role as a component of specific clinical decision support systems that synthesize measurable evidence and can introduce new medical guidelines and standards. Understanding the features of calculators is important for calculator adoption and clinical acceptance. This paper presents a novel classification system for medical calculators. Metadata on 766 medical calculators implemented online were collected, analyzed, and categorized by their input types, method of presenting results, and advisory nature of those results. Reference rate, publication year, and availability of references were collected. We found the majority of calculators are likely not automatable. 16% of medical calculators present advisory results to clinicians. 83% of medical calculators provide references. We show a 9-year lag from publication to implementation of calculators. New medical calculators should be developed with EHR integration and the advisory nature of results in mind so that calculators may become integral to clinical workflow.

Keywords:
Medical Calculator; Clinical Decision Support Systems.

Introduction

Electronic health records (EHR) are becoming highly prevalent in hospital systems [1]. Clinical decision support (CDS) within EHRs is also ubiquitous. Technologies such as the SMART platform [2], the HL7 FHIR data interface [3], and CDS Hooks [4] are helping drive the development of CDS that can be used in any EHR system. At the same time, studies have shown quality, workflow, and efficiency benefits for users of decision support systems [5,6]; however, these benefits are not universal for all CDS [7,8].

Some CDS systems have medical calculators as a major component; therefore, it is important to understand medical calculator attributes. Medical calculators embody evidence-based medicine and are typically based on scientific literature [9]. Some medical calculators are embedded into EHRs and can be considered ubiquitous such as the automatic BMI calculation. The proliferation of technologies, such as the internet and EHRs, have obvious implications on the accessibility of patient data and access to medical calculators. While the majority of medical calculators are simple and straightforward, there exist many online, web-based medical calculators that may be provisioned within an EHR.

Workflow integration and dissemination techniques are common themes in literature examining CDS. Previous broad studies on CDS have identified workflow, adoption, effectiveness, and dissemination of knowledge as top challenges [10,11]. Appropriate integration of CDS has been problematic, with alert fatigue being well studied [12,13]. Recent studies have investigated the potential for automating calculation of medical calculators, highlighting the opportunities and challenges of doing so [14,15].

The appropriate provisioning of CDS was characterized by the "five rights": making the right information available to the right person, in the right format, through the right channel, at the right time [16]. The automatic provisioning of CDS can have a positive impact on important healthcare issues, such as patient safety [17], racial and gender disparities [18], and process adherence [19]. In addition, prior studies show that factors such as automatic provisioning of CDS tools [6,20] can impact the adoption and success of CDS. Moreover, the Kawamoto study [21] identified several important relevant factors driving CDS adoption that are applicable to medical calculators: a) automatic provision of decision support as part of clinician workflow, b) provision of recommendation rather than just an assessment, and c) computer-based generation of decision support.

Classification of medical calculators is an important topic that impacts provisioning techniques. There is no widely accepted standard classification of CDS, and no comprehensive taxonomy for medical calculators. Osheroff et al., proposed a generic CDS taxonomy based on user interface [16], while Berlin et al. developed a framework for the classification of CDS (the CDSS Taxonomy framework) [22]. Calculator inputs and outputs have not been well studied. Dziadzko et al. [23] classified a subset of online calculators by their specialty, calculation methods, and goal, but did not further describe the output modes of a calculator. Aakre et al. [15] studied the specific availability of the inputs of 168 clinical calculators within the EHR and classified them as easily extractable, extractable with advanced techniques, or not extractable, but did not provide a taxonomy to describe different input types and the impact those types have on automatic calculation. Of the existing literature, the Berlin et al. framework provides the most broadly applicable framework for assessing a CDS like medical calculators. Their Reasoning Method, Recommendation Explicitness, and Explanation Availability attributes are particularly pertinent to calculators due to their simple nature.

The importance of workflow integration and automation on CDS adoption is clearly defined in literature; however, current research does not address the specific contributions that the structure of a medical calculator may have on the ability to automate and integrate these types of CDS into EHR workflow. These currently unknown attributes of calculators may have a direct impact on medical calculator adoption. We expand the current state of CDS classification by identifying attributes that are unique to medical calculators. Their potential for automatic

calculation and delivery of advisory information to clinicians and calculator input and output modalities are important factors for clinical acceptance and workflow integration. We also examine literature references of calculators to determine availability and lag between publication and implementation of online calculators.

Methods

We performed an assessment of three currently available online services that provide access to medical calculators, consisting of two free services and one commercial service. These services are anonymously referred to as Service 1, Service 2, and Service 3, respectively. The two free services were the first two non-medical-specialty specific web-based services appearing in the top 10 "organic" search results using the term "medical calculator" through a Google search. The commercial service was selected due to its availability in the University of Missouri Health System (UMHS). In total, these three online medical calculator services contained 766 implemented medical calculator algorithms.

Input types were determined by performing HTML data scraping of the HTML input tag from Service 3. Each input was classified into a type by examining the HTML input type (radio, checkbox, number, or text), and whether the data was a discrete value, a logical computation of a discrete value, required interpretation or the opinion of a clinician, or were worded in such a way as to require data from a patient and be unlikely to be stored in the EHR. The resulting types were checked for completeness during classification of the entire set of calculators.

Calculator output types were determined by examining all calculators in the study. Each calculator page was opened and classified into one or more of the output type categories. Categories were added as new output types were encountered. The calculators' targeted user (physician or patient) was captured and their references were collected where available. The calculator type was also assessed by examining the input and output modalities and targeted user to arrive at a classification. Calculators that did not fall into an already encountered type were assigned to a new type.

Results

Using the CDSS Taxonomy framework, we accounted for Reasoning Method, Recommendation Explicitness, and Explanation Availability during our data collection. Data inputs, calculator outputs, and calculator references were documented for each calculator in the three services. Calculators were then categorized based on these factors.

Calculator Inputs

To provide a generalized guide for future calculator development, we examined the inputs necessary for medical calculators and generally classified them as follows:

1. Discrete Data Elements – these are atomic pieces of data stored in an EHR. For example, the rate of creatinine clearance.
2. Non-discrete Data Elements – inputs of a non-discrete nature can ask for medical opinions of providers, for example, the likelihood of a diagnosis.
3. Logical Computation on discrete data elements – a calculator that asks if a value is over or under a certain threshold, or within a specified range, requires logical computation to determine an input value. For instance, in a point-based calculator, assigning points based on age ranges falls into this category.
4. Obscure Data Elements – Data elements unlikely to be contained as structured data within an EHR. For example, the NIH Stroke Score requires the patient to identify the current month and his or her own age.

Calculator Output

For demand-driven calculators, the way in which calculator results are delivered (Recommendation Explicitness [22]) were considered germane in our review as they are related to the advisory nature of the calculator output. Advisory calculators suggest a diagnosis or recommendation, and non-advisory are assessment only, providing a probability, score, or discrete information result. We identified five different types of results display, classified as either non-advisory (types 3, 4, and 5) or advisory (types 1 and 2), with Table I showing the distribution of these.

1. Diagnosis - Calculator presents a potential diagnosis, for example the Duke Criteria for Infective Endocarditis [24] provides a definite, probable, or rejected diagnosis for infective endocarditis
2. Advice/Recommendation - Calculator suggests or recommends a specific course of action, such as the HEMORR2HAGES Score for Major Bleeding Risk [25] which suggests initiating therapy based on calculator results.
3. Probability - Calculator provides a probability of patient having or developing a condition. The APACHE II Score [26] provides a probability of mortality
4. Classification - Calculator classifies patient in one or more categories. For example, the Apgar Score [27] classifies infants as normal or requiring intervention.
5. Discrete Information - Calculator provides a discrete data value for provider to use. The BMI calculator provides the well-known ratio of body weight to height.

Kawamoto [21] indicated that the success rate for decision support use is substantially higher for CDS that provision a recommendation versus an assessment. We found that just 16%

Table I - Percentage breakdown of output types of calculators. Note that a calculator may present multiple output types.

	Service 1 (n=138)		Service 2 (n=498)		Service 3 (n=130)	
Output Types	**Count**	**Percent of total**	**Count**	**Percent of total**	**Count**	**Percent of total**
diagnosis	2	1.45	27	5.42	6	4.62
advice/recommendation	42	30.43	37	7.43	7	5.38
probability	23	16.67	41	8.23	4	3.08
classification	75	54.35	195	39.16	69	53.08
discrete data	33	23.91	249	50.00	56	43.08

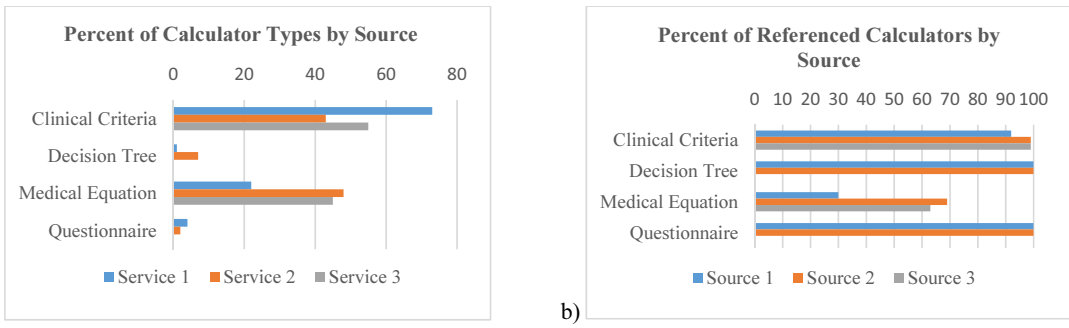


Figure 1 – Calculator Types and References by source. a) shows that Clinical Criteria and Medical Equations are the most popular types of medical calculators. In b) we find that Medical Equations are far less referenced than the other types.

(121/766) of calculators fall in the advisory category. With the majority of analyzed calculators not providing recommendations, there is a lower potential for significant adoption of medical calculators.

Calculator Categorization

Calculators in this study were analyzed and categorized into four major types:

1. Clinical Criteria – These are clinician facing calculators typically implemented as a scoring system. Answers to specific questions accrue points, with the total then looked up in a table to define the calculator output. These can require any combination of the four Input Types. For example, Total Cholesterol required by the ACC/AHA 2013 Cardiovascular Risk Assessment [28] accepts discrete data input. The Wells Score System for Deep Vein Thrombosis [29] asks for non-discrete data elements through questions such as "An alternative diagnosis is more likely than deep-vein thrombosis." The Multiple Myeloma Diagnostic Criteria [30] has input with logical computation on discrete data elements (M Protein: IgG > 3.5 g/L). The Head CT Rule for Minor Head Injury [31] requests obscure data elements such as "Inability to bear weight right after the injury as well as in the emergency department". Combinations of any of the input types may also be requested, as in the Metabolic Syndrome Criteria [32]: "Blood pressure >=130/>=85 or on blood pressure prescription"
2. Medical Equation – All inputs are Discrete Data Elements. The result of the calculator is found by computing a formula with the appropriate values. For example, the Cockcroft-Gault equation for estimating creatinine clearance is CreatClear = Sex * ((140 - Age) / (SerumCreat)) * (Weight / 72), where the value for Sex is 1 for male and 0.85 for female [33].
3. Questionnaire – Inputs can be any of the four Input Types and are designed to be answered either by a patient or in collaboration with a patient. A scoring system is usually employed, similar to Clinical Criteria. An example is the CAGE Questionnaire [34], which contains input prompts such as "Have you ever felt you needed to cut down on your drinking?"
4. Decision Tree – Inputs presented to users are dependent on answers to prior questions. A scoring system is used similar to Clinical Criteria. The PECARN Pediatric Head Injury/Trauma Algorithm [35] is an example of a decision tree.

Figure 1(a) shows the distribution of calculators by type across the three analyzed calculator services. Clinical Criteria calculators make up the majority of catalogued calculators. Because they can require Input Types other than Discrete Data Elements, additional steps may be required by the provider to search the EHR or other sources for relevant data and could reduce the likelihood of utilization. Medical Equations make up the next largest category. These are the only type that rely solely on Discrete Data Elements. Given the availability of EHR data, they can be automatically computed without interaction from a clinician. Questionnaires and Decision Trees make up a collective minority of the catalogued calculators. Both types are designed to be highly interactive and thus do not lend themselves well to automated computation.

Calculator References

The rate at which references were made available, for which types, and the accessibility of those references, were collected during calculator analysis. The availability and access to references fulfills a portion of the CDSS Taxonomy framework's "Explanation Availability of the Information Delivery axis" [22]. Clinicians can gain an understanding of the reasons behind a recommendation from the primary literature and is complimentary to the advisory content of medical calculators. The distribution of references by calculator type is presented in Figure 1(b). While the numbers of decision tree and questionnaire calculators were very small, we did note that Service 1 and Service 2 referenced 100 percent of these types. Clinical criteria calculators were referenced more than 90% of the time, with two services approaching full coverage. Medical equations were the least referenced type of calculator across the three services we analyzed.

We found that each of the three services presented references in distinct ways. One service listed references in citation style, while the other two attempted to provide URL links and categorization of the references. A primary concern uncovered in our analysis was the accessibility of the references. We conducted a detailed analysis of the largest calculator service that provided URL links (Table 2). A deeper analysis of the NCBI links showed that they all led to PubMed, a site which makes freely available basic information on articles, such as

Table 2 – Reference links provided by Service 2

Domain	Count
Internal Site Reference	47
Other URL	55
No URL Provided	65
www.ncbi.nlm.nih.gov	589

publication year and abstract, but not the full text. Lack of access to full text references could be an important factor in the adoption of newly implemented medical calculators.

An analysis of the publication year of the NCBI references supported a trend towards older publications (Figure 2), with the median publication year being 2002. Growth of implemented calculators follows an exponential curve until 2006. In the same year, there is a change in the rate of medical calculator implementations. Because this analysis represents a single point in time snapshot of medical calculator implementations as of March 2015, and implementation dates of online medical calculators are not available, we can only hypothesize that the reason for the change in calculator implementation rate is a lag from publication to implementation of approximately 9 years. Studies of medical research publication to widespread practice implementation show a similar lag of 17 years [36] to 24 years [37].

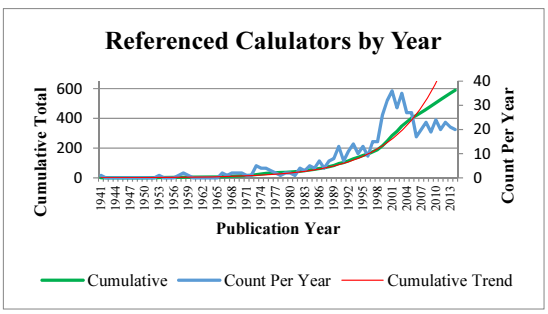


Figure 2 – Number of referenced calculators by year.

Discussion

Our analysis shows that less than half of available medical calculators lend themselves to fully automatic calculation of results, with past research indicating that the adoption of CDS increases with automatic provisioning [21]. The ability of a calculator to have its results displayed automatically is rooted in the decisions made during the research that produced the calculator publication. While any calculator may be included in provider workflow at the "right time", only a minority of calculators could automatically provide the resulting answer without requiring a clinician to manually input data that may already be available in the EHR. Medical equations are the single type of calculator capable of providing the result without the interaction of the user- this is due to the inputs requiring only discrete, structured data. Clinical criteria may be automated but may be challenging to develop due to the varying types of inputs that could be required. The other types of calculators (e.g. decision trees and questionnaires) are less suitable for automatic calculation due to their interactive nature. Thus, as new predictive models are developed, careful consideration should be given to the type of calculator that could be implemented. Medical equations and clinical criteria could be the preferred implementation if adoption and dissemination are desired for the model.

The advisory nature of current medical calculator outputs is also not consistent with prior studies that suggest recommendations lead to better adoption [21]. Only a small percentage of the calculators we studied (16%) provisioned results in an advisory fashion. Two of the most active forms of delivering medical calculator results included suggesting a diagnosis, and dispensing advice or recommendations for treatment. While we surmised that many factors play into the ability to provide advisory results, e.g. validation studies, liability, and confidence, it nevertheless is a factor related to adoption and should be considered in the development and publication of new predictive models.

Finally, 83% of implemented medical calculators in this study provided reference materials. The high rate of reference availability could prove a useful method of introducing new evidence-based medicine directly in the clinical workflow as embedded medical calculators; however, the inaccessibility of full text references may be problematic. It requires further study to determine whether or not reference availability would have an impact on perceptions of calculator credibility. The noted median year of publication of medical calculators was 2002, which highlights a potentially missed opportunity to leverage EHR deployed CDS as a means to introduce new evidence based medical literature.

Conclusion

This paper presents a taxonomy of medical calculators that can be used to inform future research in medical calculators and predictive algorithms. Researchers ultimately may be best positioned to impact the future of CDS adoption by becoming more cognizant of the types of data used to build these models, and the advisory nature of the results, and by being conversant in the fundamental structure of a medical calculator. These decisions may influence the speed at which new predictive models are implemented and delivered as automatic decision support within EHRs. EHR vendors and implementers should take note of the five rights of CDS, relevant usability and automation concerns, and disparities between different levels of clinical experience to design calculator workflows that are deployed automatically to end users. As CDS becomes more accepted as part of the delivery of medicine, evidenced by recent opinion [38] and the creation of "npj Digital Medicine" [39], insights into the issues surrounding integration of CDS into clinical workflow will help drive adoption of new technologies. We believe that future medical calculators will go beyond regression analysis and include more complex data, longitudinal data, and data from outside the EHR. Techniques such as deep learning, explainable AI, and big data technologies will make available more decision support that is based on discrete data in an EHR and can be automatically provisioned as medical calculators. Such disruptive and cutting-edge research will radically change medical practice in the coming decades, and contributions in this area must continue to push the comfort zones of the medical community. Building a solid understating in this area, as the collective research on medical calculators does, is necessary to prepare for such a future of digital medicine.

Acknowledgments

This study was supported by the Shumaker Endowment for Biomedical Informatics. The authors wish to thank Dr. Thomas Selva, MD and Dr. Stevan Whitt, MD for their support and guidance. The authors have no conflicts of interest to declare.

References

[1] P. Kierkegaard, Electronic health record: Wiring Europe's healthcare, *Comput. Law Secur. Rev.* **27** (2011) 503–515.
[2] K.D. Mandl, J.C. Mandel, et al., The SMART Platform: early experience enabling substitutable applications for electronic health records., *J. Am. Med. Inform. Assoc.* **19**

(2012) 597–603.
[3] D. Bender, and K. Sartipi, HL7 FHIR: An Agile and RESTful approach to healthcare information exchange, in: Proc. 26th IEEE Int. Symp. Comput. Med. Syst., IEEE, 2013: pp. 326–331.
[4] CDS Hooks, (2018), http://cds-hooks.org/ (accessed November 24, 2018).
[5] R.J. Koopman, K.M. Kochendorfer, et al., A diabetes dashboard and physician efficiency and accuracy in accessing data needed for high-quality diabetes care., *Ann. Fam. Med.* **9** (2011) 398–405.
[6] A.X. Garg, N.K.J. Adhikari, et al., Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review., *JAMA*. **293** (2005) 1223–38.
[7] S.Y. Park, S.Y. Lee, et al., The effects of EMR deployment on doctors' work practices: A qualitative study in the emergency department of a teaching hospital, *Int. J. Med. Inform.* **81** (2012) 204–217.
[8] D.A. Hanauer, G.L. Branford, et al., Two-year longitudinal assessment of physicians' perceptions after replacement of a longstanding homegrown electronic health record: does a J-curve of satisfaction really exist?, *J. Am. Med. Informatics Assoc.* **0** (2016) 1–8.
[9] I. Sim, P. Gorman, et al., Clinical decision support systems for the practice of evidence-based medicine., *J. Am. Med. Inform. Assoc.* **8** (2001) 527–34.
[10] J.A. Osheroff, J.M. Teich, et al., A roadmap for national action on clinical decision support, *J Am Med Inf. Assoc*. **14** (2007) 141–145.
[11] D.F. Sittig, A. Wright, et al., Grand challenges in clinical decision support, *J. Biomed. Inform.* **41** (2008) 387–392.
[12] J.S. Ash, D.F. Sittig, et al., Some unintended consequences of clinical decision support systems., *AMIA Annu. Symp. Proc.* (2007) 26–30.
[13] H. van der Sijs, J. Aarts, et al., Overriding of Drug Safety Alerts in Computerized Physician Order Entry, *J. Am. Med. Informatics Assoc.* **13** (2006) 138–147.
[14] C.A. Aakre, M.A. Dziadzko, et al., Towards automated calculation of evidence-based clinical scores., *World J. Methodol.* **7** (2017) 16–24.
[15] C. Aakre, M. Dziadzko, et al., Automating Clinical Score Calculation within the Electronic Health Record, *Appl. Clin. Inform.* **8** (2017) 369–380.
[16] J.A. Osheroff, E.A. Pifer, et al., Improving outcomes with clinical decision support: an implementer's guide, HIMMS, 2012.
[17] A.H. Morris, Decision support and safety of clinical environments, *Qual. Saf. Heal. Care*. **11** (2002) 69–75.
[18] B.D. Lau, A.H. Haider, et al., Eliminating Health Care Disparities With Mandatory Clinical Decision Support: The Venous Thromboembolism (VTE) Example., *Med. Care*. **53** (2015) 18–24.
[19] K.G. Shojania, A. Jennings, et al., The effects of on-screen, point of care computer reminders on processes and outcomes of care, *Cochrane Database Syst. Rev.* (2009).
[20] K. Kawamoto, and D.F. Lobach, Clinical decision support provided within physician order entry systems: a systematic review of features effective for changing clinician behavior., in: AMIA 2003 Annu. Symp. Proc., 2003: pp. 361–5.
[21] K. Kawamoto, C.A. Houlihan, et al., Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success., *BMJ*. **330** (2005) 765.
[22] A. Berlin, M. Sorani, et al., A taxonomic description of computer-based clinical decision support systems, *J. Biomed. Inform.* **39** (2006) 656–667.
[23] M.A. Dziadzko, O. Gajic, et al., Clinical calculators in hospital medicine: Availability, classification, and needs, *Comput. Methods Programs Biomed.* **133** (2016) 1–6.
[24] D.T. Durack, A.S. Lukes, et al., New criteria for diagnosis of infective endocarditis: utilization of specific echocardiographic findings, *Am. J. Med.* **96** (1994) 200–209.
[25] B.F. Gage, Y. Yan, et al., Clinical classification schemes for predicting hemorrhage: results from the National Registry of Atrial Fibrillation (NRAF)., *Am. Heart J.* **151** (2006) 713–9.
[26] W.A. Knaus, E.A. Draper, et al., APACHE II: A severity of disease classification system, *Crit. Care Med.* **13** (1985) 818–829.
[27] V. Apgar, A proposal for a new method of evaluation of the newborn infant, in: Anesth. Analg., 2015: pp. 1056–1059.
[28] D.C. Goff, D.M. Lloyd-Jones, et al., 2013 ACC/AHA guideline on the assessment of cardiovascular risk, *Circulation*. (2013) 01--cir.
[29] P.S. Wells, D.R. Anderson, et al., Value of assessment of pretest probability of deep-vein thrombosis in clinical management., *Lancet*. **350** (1997) 1795–8.
[30] B.G. Durie, and S.E. Salmon, A clinical staging system for multiple myeloma. Correlation of measured myeloma cell mass with presenting clinical features, response to treatment, and survival., *Cancer*. **36** (1975) 842–854.
[31] I.G. Stiell, G.H. Greenberg, et al., A study to develop clinical decision rules for the use of radiography in acute ankle injuries, *Ann. Emerg. Med.* **21** (1992) 384–390.
[32] S.M. Grundy, J.I. Cleeman, et al., American Heart Association; National Heart, Lung, and Blood Institute. Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement, *Circulation*. **112** (2005) 2735–2752.
[33] D.W. Cockcroft, and H. Gault, Prediction of Creatinine Clearance from Serum Creatinine, *Nephron*. **16** (1976) 31–41.
[34] J.A. Ewing, The CAGE questionnaire: Validation of a new alcoholism screening instrument, *JAMA*. **252** (1984) 1905–1907.
[35] N. Kuppermann, J.F. Holmes, et al., Identification of children at very low risk of clinically-important brain injuries after head trauma: a prospective cohort study., *Lancet (London, England)*. **374** (2009) 1160–70.
[36] E.A. Balas, and S.A. Boren, Managing clinical knowledge for health care improvement, *Yearb. Med. Informatics 2000 Patient-Centered Syst.* (2000) 65–70.
[37] D.G. Contopoulos-Ioannidis, G.A. Alexiou, et al., Life cycle of translational research for medical interventions., *Science*. **321** (2008) 1298–9.
[38] A.L. Beam, and I.S. Kohane, Big Data and Machine Learning in Health Care, *JAMA*. **319** (2018) 1317.
[39] S.R. Steinhubl, and E.J. Topol, Digital medicine, on its way to being just plain medicine, *Npj Digit. Med.* **1** (2017) 5.

Address for correspondence

Chi-Ren Shyu, PhD. shyuc@missouri.edu
Director and Shumaker Endowed Professor of Informatics Institute
University of Missouri, Columbia, MO, 65211, USA

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190294

Towards a Digital Lean Hospital: Concept for a Digital Patient Board and Its Integration with a Hospital Information System

Massah Hamidi[a], Piratheepan Mahendran[b], Kerstin Denecke[b]

[a] Institute for Medical Informatics, Bern University of Applied Sciences, Biel, Switzerland,
[b] Spitalzentrum Biel, Biel, Switzerland

Abstract

Lean management applied to healthcare aims at creating maximum value for patients by reducing waste and waits. It puts the patient's needs center stage, emphasizes employee involvement and continuous improvement. To realize this, visual tools such as the huddle board or the patient board are implemented in hospital wards. The boards are currently realized by whiteboards or flipcharts, which leads to duplicated data entries and loss of information. The objective of this work is to introduce a concept for digitalizing the patient board and integrate with a hospital information system (HIS) for improved data availability. Data on appointments, personnel planning and master patient data can be directly accessed from the HIS database. A digital patient board has several benefits: data can be collected from information systems, making it obsolete to record information several times. Even more functionalities, in particular, those supporting the communication between a patient and a healthcare team can be included by means of a digital board which improves patient involvement.

Keywords:

Health Communication, Hospital Information Systems, Patient Care Management

Introduction

To reduce costs, healthcare providers try to improve their outcomes while simultaneously achieving greater efficiency. One prominent method is lean management; it promises improved quality, capacity, and safety while limiting costs [1; 2]. Lean is a management philosophy that helps to create maximum value for patients by reducing waste and waits. It puts the patient's needs center stage, focuses on employee involvement and continuous improvement [3]. Originally developed for Toyota Production Services, lean has been applied to various industries and is now conceptually described as "an integrated socio-technical system whose main objective is to eliminate waste by concurrently reducing or minimizing supplier, customer, and internal variability" [4].

Visual means such as the huddle board or the patient board are used as tools to implement lean principles into practice [5; 6]. The patient is placed in the middle of all processes. By means of a patient board, be it a flipchart, whiteboard or a customized markerboard, installed in a patient's room, each patient is continuously informed about the course of the stay. On that particular board, a patient's needs and requests regarding the treatment process can be recorded manually during consultation with a doctor (for example, the date of discharge). In addition, the patient can also see what time the nursing staff will check the next time, who is responsible for him or her on a particular day or which examinations will take place on the current day. This information board creates transparency for the patient. Daily objectives defined by physiotherapists, physicians or nurses motivate to do suggested exercises continuously.

There are commercial tools available to support patient communication such as the communication boards from ID Signsystem [10] which are not digital solutions, but customized markerboards. MyCareBoard is a digital whiteboard distributed commercially by the company Lincor [11]. This board is developed for inpatients. It shows the latest laboratory data, provides access to social media tools and shows contact persons for the day. MyCareBoard provides many functionalities that go beyond the lean principle which renders the interaction much more complex. Since the inpatients in our collaborating hospital are often above 60 years, we believe that – in contrast to MyCareBoard - our digital patient board should focus on the provision of relevant information and on a well-defined set of functionalities.

Sehgal et al. found out that whiteboards placed in patient rooms have the potential to improve teamwork, communication and patient care [12]. They can improve patient experience and coordination of care between patients, their family members and hospital staff. Dry-erase whiteboards were used at the bedside but led to frustration for nursing staff, nursing leadership, hospital administrators and patients. Studies showed that one major reason is that these whiteboards fail because pens are unavailable when needed [12].

The use of a whiteboard version of a patient board was evaluated in a project at the Biel hospital center [4-7]. It was found that patients appreciate the visual representation of their daily routine and the planned appointments shown on the patient board. Some weak points were that data must be recorded several times on the different boards, in the HIS and on the patient board. Furthermore, data protection was not given since the current board is placed visibly to all visitors in the patient room.

To overcome these limitations, we introduce a concept for digitalizing the patient board and integrating it into the hospital workflow. Furthermore, we introduce our prototype for a digital patient board. The objective of a digitalized patient board is to increase data availability for patients, improve patient involvement, patient experience and provide a means for asynchronous information exchange and communication with a patient.

Methods

Development Process

Requirements were assessed in discussions with the leading nurses, the head of the nursing department and an IT manager of a Swiss hospital. In an iterative process accompanied by stakeholder discussions, a concept was developed to replace the current patient board realized as a whiteboard by means of a digital solution meeting the above-mentioned requirements. The interfaces to existing information systems were analysed. Finally, a prototype of a digital patient board was implemented considering the requirements from the orthopedic ward. Feedback on the system was retrieved in discussion with nurses and patients (see the section on usability test results).

Lean Management on the Partner Hospital

The hospital *Spitalzentrum Biel* involved in this work adapted the Seattle model of a lean hospital. Among other methods, they apply lean in their orthopedic and gynecology ward. Each ward is split into four zones with a zone comprising 4 patient rooms. In the daily personnel planning, nursing staff is assigned to the single zones. Over the day, nurses only address tasks within the assigned zone. Every hour, they go through the bedrooms in their zone and address the patient issues.

The lean principle is applied to support planning and communication among others by implementing three different boards (flowboard, huddle board, patient board). They are available as whiteboards on which hospital staff can write relevant information. The flowboard is located in each zone of a ward and is used by caregivers. They meet to plan their hourly rounds to the patients and allocate tasks. This information is written on the board. All professional groups of all zones meet twice a day at the huddle board to discuss the daily routine on the ward. The huddle board supports the communication and planning of the tasks in the ward.

The patient board is placed in the patient rooms. By means of a flipchart/whiteboard, each patient is informed about the course of his or her stay. A mobile care trolley was made available to the nursing staff for the hourly rounds. This is a mobile workstation for care to avoid unnecessary interruptions in the working process with patients as well as long distances. It contains a laptop and the most important materials as well as medications [7].

Results

In this section, we introduce the concept for digitalizing the patient board and integrating it with the system and process landscape of a hospital. Further, we describe the functionalities and components of our prototype of the digital patient board.

Requirements

A main requirement for digitalization of the patient board is to provide the same content as the one available on the existing whiteboard patient board, but in a digital format. Information should be accessible through a web application running on two devices: on a touch screen placed in the patient room and a tablet application accessible through a tablet. Nurses, physicians and patients interact with the touch screen. They should be able to enter and modify data. Wherever possible, relevant data items have to be retrieved from existing information systems and shown on the patient board. The information should be displayed in a suitable way for patients and health professionals.

The tablet version is only designed for use by patients. It should enable the patient to enter notes on the daily physiotherapeutic goal and to ask questions. Furthermore, the patient should have access to responsible care providers, appointments, daily physiotherapeutic objectives and the specified requests and needs. Data security has to be ensured in both devices.

Concept

Our concept for digitalizing the patient board and integrating it with the information system environment consists of several devices (Fig. 1).

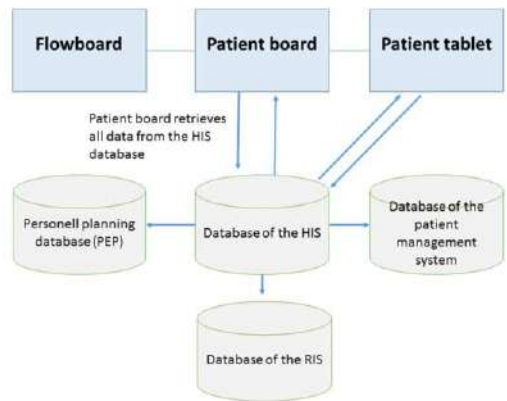


Figure 1: System overview: The patient board is the central data access screen. It is connected to the database of the hospital information system. A subset of the data is available through the patient application using a patient tablet.

The **patient tablet** with patient application is designed for use by patients during their stay at the hospital and allows, for example, accessing the individual daily, physiotherapeutic goal and a progress line.

The **patient board** is a web application accessible through a touch screen located in the patient room. It provides data access for doctors, nurses and patients.

The **patient portal** is an application for patients to be used inside or outside the hospital, mainly providing information on the stay.

Table 1 summarizes the type of information that is available through the different devices. Figure 1 shows the interactions between the different devices (flowboard, patient board and patient tablet) with the HIS and connected information systems. In this work, we are focusing on the digital patient board. However, the flowboard is included in the concept since it is currently under development.

All relevant data shown on the patient tablet or patient board is retrieved and stored in the HIS and is accessed by the different devices listed above. The information shown on the patient board is taken from the HIS database which in turn has interfaces to other databases. For example, from the personnel planning system PEP, the responsible physician and nurse and their period of service, are taken. This information can be shown on the digital patient board in order to select the responsible physician according to the shift planning. In future, additional boards that allow showing the result from shift planning directly on the patient board can be integrated. From

the radiology information system (RIS), scheduled radiology examinations of a patient are collected. The patient master data and case data are taken from the patient administration system. In future work, we will also digitize the flowboard that will also be connected to the patient board and HIS.

Table 1: Information available on different devices

Device	Information provided
Flowboard	Patient ID, surname, first name, patient room number, upcoming examination (where, when) shown in a flow chart
Patient board	Date, surname, first name, date of birth, language, responsible physician and nurses, upcoming examinations, goals, medication, admission date, discharge date
Patient tablet	Responsible physician and nurses, upcoming examinations, physiotherapeutic goals, request and needs from the patient

Patient Board and Patient Application

The electronic patient board was implemented as a web application, specifically implemented for the orthopedic ward of the collaborating hospital. We developed the patient board in two languages (French and German). Languages can be added easily in future, addressing the fact, that inpatients might neither speak German nor French.

When the application is started, the functionalities differ slightly depending on the platform used and access rights. Accessed through the touch board (referred to as patient board in Figure 1) by health professionals, data can be manipulated. More specifically, patient master data can be accessed and individual data records such as the therapeutic goal of the patient board can be modified. Functionalities of the touch screen application for the nurses and physicians include:

- Entering dietary restrictions, responsible care providers, patient needs,
- Specifying physiotherapeutic goal,
- Changing mobilisation schema (Figure 3),
- Entering discharge date,
- Retrieving examination and treatment appointments from relevant information systems.

The system can only be accessed by authentication with a password or patient barcode.

When data is accessed by a patient through a patient tablet, the individual progress is shown on the patient board; questions and needs can be formulated and stored for indirect communication with the healthcare team. In this way, the patient is provided with a platform for recording requests that cannot be addressed immediately by the care team. Contact persons of the healthcare team are listed on the screen. Additionally, the patient application displays a progress curve based on the mobilization scheme and daily goals of the physiotherapy. The display can be personalized so that a patient only sees data that is of importance for him. This ensures that patients are not overloaded by information. In order to make entries in the patient tablet app, patients must authenticate themselves with an integrated barcode scanner. The scanned bar/QR code is compared to the tablet UUID stored in the database and assigned to the patient. Figure 2 shows a screen of the patient application.

Figure 2. The digital patient board shows a progress line with respect to the mobilization progress, daily goal (Tagesziel), needs (Bedürfnisse), responsible nurses and clinicians, or appointments. Elisabeth Brönnimann is a fictitious character.

Figure 3: Mobilisation schema

Usability Test and Evaluation Results

To assess the usability of our digital version of the patient board in its appearances and for the different user groups (tablet vs. touch screen and patient view vs. health professional view), we performed a usability test. Nurses and potential patients have been interviewed. With "potential patients" we refer to persons that were not inpatients at the time of evaluation but are likely to become one day. They were recruited from the authors' social and work environment.

At the beginning of the test, patients were introduced to the topic of lean management and the general idea of the patient board. Afterwards, they were confronted with the app running on a tablet and a set of six tasks which had to be solved. The task included: creating, sharing and deleting a note, scanning a barcode, asking a question, opening the daily goal view and the progress bar.

Nurses assessed the patient board application that we opened in a tablet. Their tasks comprised: specifying the type of diet, adding and removing employees, specifying a mobilization goal and treatment phase, specifying a physiotherapeutic objective, adapting the date of discharge, and answering patient requests.

All participants were asked to think aloud while performing the tasks. A questionnaire with 11 statements and a 5-item Likert scale had to be completed afterwards (see Figures 3 and 4). The statements addressed usability issues (ease of use, understandability etc.) and layout.

18 potential patients and six nurses were involved in the usability test and completed the questionnaire. All participants were able to solve the tasks. The age distribution of the patients is as follows: age group 18-24 (four users), 25-44 (three users), 45-64 (seven users), 65-79 (four users).

The layout and design were well received by the potential patients. They felt comfortable while using the app and

confirmed that they could deal with the app intuitively. 72% claimed that they would be willing to use the application. The overall feedback was rather positive. The complete results from the questionnaire answered by patients are shown in Figure 4.

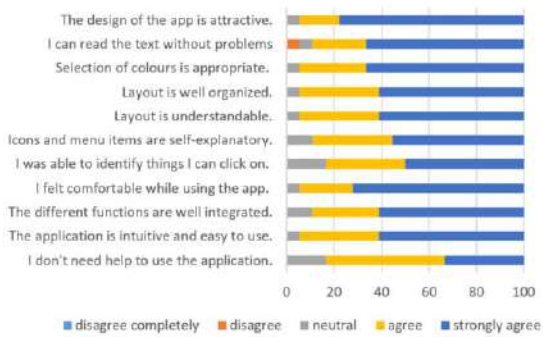


Figure 4: Evaluation results for patients (values in %)

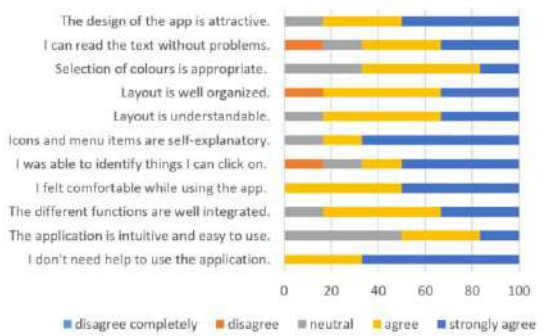


Figure 5: Evaluation results for nurses (values in %)

Compared to the patient-user group, nurses were more critical and had problems with some tasks (see Figure 5). For example, 50% of the nurses did neither agree nor disagree with the statement that the app can be used intuitively. One person had problems in dealing with the app, in particular with identifying things and reading text. However, most of the nurses confirmed they had not needed help to use the application.

Discussion

Benefits of the Digital Patient Board

Improving patient experience and coordination of care between the patients, their family members and staff is important for good outcomes in clinical care. In this paper, we introduced a concept and first prototype for a digital patient board integrated into the information system environment of a hospital. The usability test showed that patients and nurses are aware of the potentials of a digital patient board. Studies showed that patients appreciate having key information organized and available to them at a glance, leading to an increase in patient satisfaction scores [9]. Our digital patient board can facilitate in the future, consistent, standardized, and transparent communication among patients, caregivers in hospitals, medical centers and long-term care facilities.

Our implementation of a digital patient board provides features that are currently not provided by the whiteboard version, but that improve the communication between patient and healthcare team. In particular, the patient is better involved in the care process by having the opportunity to ask questions through the board. Based on the information shared by the patient, employees can respond directly to their respective needs. The patient always has an overview of his treatment and goals in mind.

One open challenge is to decide for an appropriate degree of transparency: since the interest and intellectual capacity of patients differ, the amount and kind of data displayed for the patient has to be considered carefully so as to avoid misunderstandings. Initial interviews with hospital staff show that the need for information also depends on the hospital and wards.

Another important issue is data integrity and data separation. Data integrity is ensured by the user logins and role management i.e. nurses will see more information and have more interaction possibilities when accessing the patient board with their credentials than patients.

The data flow via the corresponding interfaces between information systems in hospitals and the patient board will in future allow employees to dedicate themselves directly to communicating with the patient without having to re-enter existing information such as responsibilities or appointments on the patient board. Waiting times for the patient are reduced, employees are relieved and data integrity can be guaranteed. The economic effect still has to be proven in corresponding evaluations.

The functionalities of the patient board as tested in our prototype have been integrated into a mobile version of KISIM. KISIM is the HIS of the cooperating hospital developed and distributed by a Swiss software company CISTEC AG (http://www.cistec.ch/). KISIM mobile will provide access to the above-listed functionalities of the patient application. The data will be retrieved and stored in the KISIM database.

Problems of the Usability Test

There are some issues that influenced the usability test results. Nurses were not only focusing on their role as nurses but partially provided feedback from a patient perspective, i.e. they made comments on things that they assume to be not intuitive or understandable for patients instead of focusing on their role as nurses.

Another issue is that the “potential users” were not hospitalized and were recruited from the authors’ environment. Regarding usability, we consider these persons as representative since we selected persons randomly not requesting specific technical skills. However, people might conceive things differently when they are really inpatients. We tried to balance the age groups of persons involved in our test and could not determine differences in the feedback depending on the age of the user. As mentioned earlier, the inpatients of the collaborating hospital are older than 60 years. A field study has to be performed to ensure that inpatients can deal with the systems without problems.

A comparison between the whiteboard version and the digital version was not performed. Since the digital version of the patient board contains the same information as the current markerboard version of the patient board, we believe that the information is at least understandable to a similar extent. However, we did not verify this within our usability test.

Future Extensions and Integration with other Digital Boards

As mentioned in the concept section, we envision to digitize additional lean boards. There is already one commercial

product available that digitizes the huddle board (http://www.cistec.ch/abteilungsboard/). It supports an interactive shift scheduling in a ward. Additional functionalities of that board are:

- Providing an overview of the rooms and beds as well as the patients.
- Information from the patient record can be displayed directly on the board (configurable).
- Shift planning: By integrating the planning information from the PEP into the board, the employees of a shift are automatically displayed. These can then be assigned to the patients.
- Planning entries/exits.

The board has been implemented in one Swiss hospital [8] and has been in use since July 2018. We are currently digitalizing the flowboard. Once this is completed, all three boards can be integrated ensuring the best possible data integration and reuse. Questions concerning storage and remarks from patients in the HIS could help in understanding communication problems among patients and healthcare staff.

Conclusions

Lean management is going to be implemented in hospitals to better deal with available resources. In its current implementations, however, optimization potentials exist that could be addressed by means of digital solutions. We introduced our concept for a digital patient board as one crucial component of the lean hospital. The usability test confirmed good usability. Data can be automatically retrieved from information systems.

In future, in particular, the patient tablet as one out of several devices of our concept can be equipped with additional features, such as menu selection, or features for patient education. However, it has to be ensured that patients only see content that is relevant for them and which they can handle. We believe that the standard version of the digital patient board should focus on the information provided by the current patient board. Advanced users should be enabled to use additional functionalities.

Acknowledgements

We would like to thank all the persons involved in the usability test. Further, we thank the collaborating hospital and staff for the productive meetings and helpful comments.

References

[1] G.S. Kaplan, Virginia Mason Health System. The Lean Approach to Health Care: Safety, Quality, and Cost, *NAM Perspectives* **2** (2012) doi:10.31478/201211a

[2] D. Zrymiak, Lean Hospitals: Improving Quality, Patient Safety, and Employee Engagement, Second Edition, *Quality Management Journal* **23** (2016), 70–71.

[3] T. Rotter, C. Plishka, A. Lawal, L. Harrison, N. Sari, D. Goodridge, et al., What Is Lean Management in Health Care? Development of an Operational Definition for a Cochrane Systematic Review, *Eval Health Prof* (2018).

[4] R. Shah and P.T. Ward, Defining and developing measures of lean production, *J Oper Manage* **25** (2007), 785–805.

[5] D. Walker, Lean Hospital: Das Krankenhaus der Zukunft. MWV, (2015).

[6] A.K. Lawal, T. Rotter, L. Kinsman, N. Sari, L. Harrison, C. Jeffery, et al., Lean management in health care: definition, concepts, methodology and effects reported (systematic review protocol), *Syst Rev* **3** (2014).

[7] A. Angerer and T. Brand, Lean Hospital - Optimierung einer Bettenstation am Beispiel des Spitalzentrums Biel, *das Krankenhaus* **3** (2016), 196–199.

[8] Spitäler Frutigen Meiringen Interlaken AG. Bericht Pilot elektronisches Huddleboard, https://www.spitalfmi.ch/fileadmin/images/spital_interla-ken/fachbereiche/pflege/fmi_int_pdl_Bericht_Pilot_elektronisches_Huddleboard.pdf [accessed October 24, 2018]

[9] S. Singh et al., It's the writing on the wall: Whiteboards improve inpatient satisfaction with provider communication, *American Journal of Medical Quality* **26** (2011), 127-131

[10] ID Signsystem: Communication Board, http://www.id-communicationsystems.com/ [accessed November 2, 2018]

[11] Lucor: My Careboard, https://www.lincor.com/news/my-careboard/ accessed November 2, 2018]

[12] N.L. Sehgal, A. Green, A.R. Vidyarthi, M.A. Blegen and R.M. Wachter, Patient whiteboards as a communication tool in the hospital setting: a survey of practices and recommendations, *J Hosp Med* **5** (2010), 234-239. doi: 10.1002/jhm.638

Address for Correspondence

Prof. Dr. Kerstin Denecke
Bern University of Applied Sciences
Institute for Medical Informatics
Quellgasse 21, Biel, Switzerland
Contact via e-mail: kerstin.denecke@bfh.ch

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190295

SIMENS-LIS4SC, a Laboratory Information System for Biological Tests of Sickle Cell Screening and Healthcare

Al Hassim Diallo[a] , Gaoussou Camara[b] , Jean Baptiste Lamy[d], Moussa Lo[a], Ibrahima Diagne[a,c], Demba Makalou[c], Mamadou Diop[c], Dominique Doupa[a,c]

[a] Université Gaston Berger de Saint-louis, B.P. 234 Saint-Louis, Sénégal
[b] LIMA, Université Alioune DIOP de Bambey, B.P. 30 Bambey, Sénégal
[c] Centre de Recherche et de Prise en Charge Ambulatoire de la Drépanocytose, Sénégal
[d] LIMICS, Université Paris 13, 93017 Bobigny, INSERM UMRS 1142, Sorbonne Universités, Paris, France

Abstract

Neonatal screening and ongoing follow-up of children with sickle cell disease are essential to reduce the mortality caused by this disease. To ensure care continuity, it is essential to include in the patient's record the history and details of biological tests. Thus, it is necessary to provide a Laboratory Information System for electronic management of biological test prescription and results, and the laboratory system must integrate well with Health Information Systems. In this paper, we propose a Laboratory Information System for the management of biological tests for the neonatal screening and healthcare of sickle cell disease in Senegal.

Keywords:

Clinical Laboratory Information Systems; Anemia, Sickle Cell; Neonatal Screening

Introduction

Large amounts of data pass regularly through Senegal's health system between medical services and biological laboratories. This data includes on one hand the tests issued by the physicians and on the other hand, the test results produced by the laboratories. In addition, there is data specific to laboratory activities such as sampling data, the test results, and the test reports, etc. This data is usually recorded in paper documents and is accessed manually. This situation also leads to recurrent back and forth by patients or their carers between medical services and laboratories. Thus, the labs face enormous problems related to (i) collection, storage and processing of test prescriptions and results, (ii) data security, and (iii) possible impaired data integrity because of the frequent use of archives. It then becomes necessary to implement a Laboratory Information System (LIS) integrated into existing Health Information Systems (HIS) to overcome all these difficulties.

The purpose of the LIS is to manage the laboratory's workflow, including processing, storing, and managing biological analysis data in order to provide accurate results for clinical decisions [1,2]. In Senegal, the integration of a LIS with an HIS represents a new scientific and organizational dimension in the local medical practice. Indeed, despite several initiatives including the National Medical Information System for Senegal (SIMENS) project [3], a modular HIS, initially designed for medical services in health facilities of level three in the sanitary pyramid of Senegal [4,5], a LIS has never been proposed.

In this article, we propose a laboratory information system for the management of screening tests for sickle cell disease and all other tests (hematology, biochemistry, parasitology, and bacteriology) necessary for proper monitoring and good management of positively screened patients. The LIS is designed as a module of SIMENS and is called SIMENS-LIS4SC (SIMENS LIS for Sickle Cell) and is integrated with the HIS of the Center for Research and Ambulatory Care of Sickle Cell Disease (CERPAD) in the Saint-Louis region of Senegal. CERPAD aims at proposing a model for neonatal screening and early healthcare of sickle cell adapted to Senegal's public health system. Indeed, sickle cell disease is a major public health problem occurring in approximately 300,000 births annually worldwide [6]. According to a systematic review on sickle cell disease for children under five years od [7], both the highest prevalence and highest mortality of sickle cell is in Africa, and there is a need for national comprehensive newborn screening to identify patients, and to develop holistic care programs to provide therapeutics and education for families and children with the disease. In Senegal, there are no published studies on sickle cell prevalence. Few local and specific studies, such as one in Senegal, [8] reveal that sickle cell disease mainly concerns children and adolescents.

In the results, we cover the first phase of the project, which began in April 2017. The health services involved were the maternity wards of the Saint-Louis' Regional Hospital Center (CHRSL) and the reference health center of the city of Saint-Louis. We present the different interfaces designed for managing the biological test data of the CERPAD laboratory. We illustrate the system contribution in terms of quick and easy access to statistical data for decision-makers through a reporting and dashboard module. A preliminary evaluation was also conducted to show the LIS assessment by the different CERPAD actors involved in the laboratory tests processes.

Methods

The neonatal screening program included every newborn with the consent of their parents. Specially trained midwives and gynecologists informed parents about the process. Materials collected included for each newborn, a blood drop sample and an information sheet regarding the baby's medical data, the parents' marital status, contacts, and socio-professional status. In addition, for newborns suspected of having sickle cell disease or homozygous C after the initial screening, a second venous blood sample was collected for detailed analysis.

The initial blood sample was used to perform hemoglobin typing by the isoelectrofocusing method. The second sample was used for performing capillary electrophoresis of hemoglobin, to identify and confirm the sickle cell status. The center followed monthly newborns with a major sickle cell syndrome profile (SS, SC, SE, SDPunjab, SOArab, SLepore, Sβ°thalassemia, Sβ+thalassemia) or a CC status during systematic visits. They also received emergency care during acute attacks or other complications related to the disease.

Recommended during the follow-up for diseased patients, mandatory biological and radiological examinations evaluated systematically and periodically the patients' health state. These examinations' objective was to detect complications in early stages in order to propose preventive treatment before any organ deterioration or the appearance of functional repercussions. A doctor from the HIS could also prescribe these examinations during the patient consult and could send patients directly to the nurse for sampling. Samples were then passed to the lab technician to trigger the biological test process of sample sorting. The biological examinations were proteinuria (after 24h), blood count and reticulocyte level (every month from birth), serum iron, ferritin, lactate dehydrogenase, and irregular agglutinins (every year since birth), and micro-albuminuria and creatinine (every year from 5 years old). The radiological examinations were transcranial ultrasound (every year from 2 years old), abdominal ultrasound (every year from 5 years old), retinal angiography, and cardiac ultrasound (every year from 10 years old). Radiological examinations were performed in an external department to CERPAD and were not yet taken into account in the SIMENS-LIS4SC.

An application dedicated for data management of laboratory tests for sickle cell disease screening and healthcare was implemented. It became a part of a module integrated in SIMENS, acting as a laboratory information system and connected to the HIS module of SIMENS dedicated to sickle cell neonatal screening and healthcare. The SIMENS-LIS4SC had two main inputs (neonatal screening test request from maternity wards and biological test request for the diseased patient follow-up and healthcare from the physician) and one output, which were the test results integrated in the patient record. The application was implemented using an Agile development process [9]. We used PHP and the ZEND 2 framework to develop the LIS, as these technologies were already a part of SIMENS. For data management, we used the MySQL relational database management system. We worked with a medical and medico-technical team of doctors, nurses, laboratory technicians, biologists, and administrative secretaries who were very involved in the software design. Regular meetings made it possible to clearly identify needs, discuss the various user profile requirements, present intermediate results to gather opinions and suggestions, and define the following work, etc. To guide the design of new features and interfaces, workflow schemes and other essential information were regularly shared.

As a preliminary evaluation, we gathered the opinion of 9 users using a qualitative form including 18 questions. These users included a secretary, nurse, technician, biologist, and physician using the LIS for almost 2 years. For each question, users had to answer using a 4-value scale (fully agree, partially agree, partially disagree, fully disagree) and were allowed to give no opinion.

Results

Workflow

Figure 1 shows the proposed workflow for laboratory tests management of sickle cell neonatal screening, follow-up, and healthcare. The maternity wards intervened in the sampling phase for the newborn screening, the secretary while creating the patient's record, and the nurse for filling the sample data from the maternity wards. The nurse could have taken samples at the CERPAD center if necessary, for instance when the physician prescribed additional biological tests for the diseased patients follow-up and healthcare. When the samples were transmitted to the laboratory, the technician performed some checks before completing the tests and filling the results. Then, the biologist interpreted, validated and sent the results to the secretary for printing.

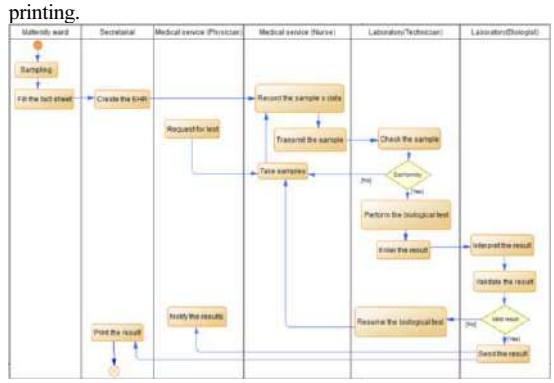


Figure 1– Workflow of Laboratory Tests Realization

The SIMENS-LIS4SC

In this section, we focus on the LIS interfaces that highlight the management of the sickle cell biological test data within the CERPAD for the neonatal screening and healthcare. Any of the identity data in screenshots does not refer to any real patient.

First, the screening process consisted for the secretary, to create a record in the Electronic Health Record (HER) for the patient based on his information sheet data from the maternity ward. The secretary was also responsible for recording biological test requests, such as hemoglobin screening as shown in Figure 2.

Figure 2– Screening Test Request from the Secretary Interface

Then, the nurse filled the data related to the samples (Figure 3) taken on the patient before sending them to the laboratory technician who performed the verification.

Figure 3– Sample Data Entry by the Nurse

Verification step shown in Figure 4 consisted of checking the state of the samples, the conformity of the material used, etc. In the event of a problem, the technician could request through SIMENS the resumption of the samples concerned.

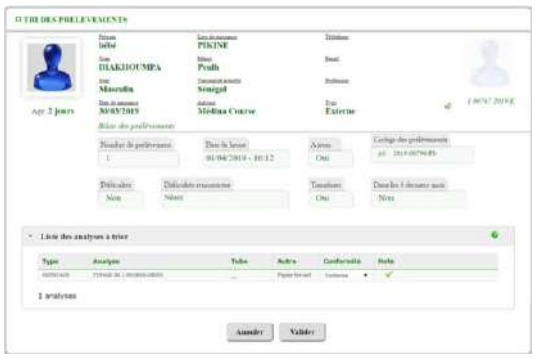


Figure 4– Samples Checking by the Technician

The technician was also responsible for recording the results of the biological tests as shown in Figure 5.

Figure 5– Test results entry by the technician

Finally, the biologist interpreted and validated the results for printing. The results were integrated into the patient's EHR. The secretary could access the results and print them as shown in Figure 6 for transmission to physician and newborn's parents.

Figure 6– Example sickle cell screening result print

In addition to accessing the laboratory test results directly in the patient record, the physician could prescribe complementary tests directly during the consultation. The Figure 7 shows how the SIH alerted the physician about the mandatory exams to prescribe for the follow-up of the diseased patient.

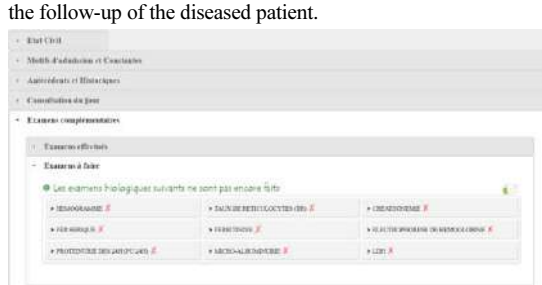


Figure 7– Complementary Tests Prescribed by the Physician

Example of Statistical Data on the Screening Tests

The neonatal screening program started in April 2017 with a two-month experimental phase. This phase also made it possible to test and maintain the SIMENS HIS and LIS modules for sickle cell disease screening.

Table 1– Summary of the screening from 06-2017 to 08-2018

Period	AA	AC	AD	AS	CC	SC	SS	Total
2017-06	226	10	0	23	0	0	0	259
2017-07	236	5	1	22	0	0	2	266
2017-08	48	1	0	6	0	0	0	55
2017-09	146	4	0	13	0	0	0	163
2017-10	286	5	0	29	0	0	1	321
2017-11	245	2	0	29	0	0	0	276
2017-12	261	7	0	21	1	0	0	290
2018-01	308	6	0	26	0	0	0	340
2018-02	251	7	0	30	0	0	0	288
2018-03	216	6	0	31	0	0	1	254
2018-04	208	2	0	19	0	1	0	230
2018-05	174	6	0	16	0	0	0	196
2018-06	219	3	0	19	1	0	0	242
2018-07	241	5	0	23	0	1	2	272
2018-08	43	0	0	1	0	0	0	44
Total	3108 (88.90%)	69 (1.97%)	1 (0.03%)	308 (8.81%)	2 (0.06%)	2 (0.06%)	6 (0.17%)	3496

Table 1 shows the evolution of the data over time and their distribution according to the various screened sickle cell profiles, covering a period of 15 months from June 2017 to August 2018. 3496 newborns were screened during this period. 3180 (green columns) of them were healthy and non-carriers. 316 (orange and red columns) were carriers of the gene Hemoglobin S. Among carriers, we had healthy carriers (AS) and diseased patients (SS, SC, SE, SDPunjab, SOArabe, SLepore, Sβ°thalassemia, Sβ+thalassemia). In this first phase, we found 8 (red columns) diseased patients of SS profile.

The figure 8 shows a dashboard of the distribution of (1) the overall screened newborns between internal and external status, (2) the diseased patients between gender and (3) the diseased patients according to SS, SC and CC profiles.

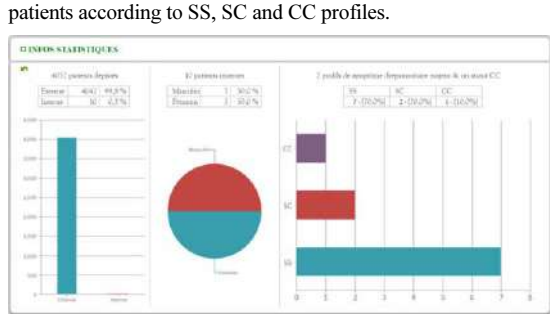


Figure 8– Patients Distribution According Internal and External Status, Gender and Sickle Cell Diseased Profiles

Table 2– Evaluation results and analysis

The system facilitates my work										
The system improves the quality of my care									**Key:**	
The system improves our ability to coordinate continuity of care									Fully agree	
The system improves medical information sharing									Partially agree	
The system improves decision-making									Partially disagree	
The system is easy to use									Fully disagree	
When using the system, the number of clicks is acceptable									No opinion	
System's response time is acceptable										
System's session opening is acceptable										
The system preserves patient's security and confidentiality										
The system is reliable										
The global quality of the system is excellent										
The information in the system is exhaustive										
The information in the system is quickly available										
The information in the system is relevant										
The information in the system is quickly available when needed										
The form and layout of the information in the system is acceptable										
The information in the system allows me to make decisions										

The figure 9 shows a dashboard of the distribution of only validated results by the biologist of the overall screened newborns between (1) gender, (2) sickle cell all profiles and (3) ethnic groups.

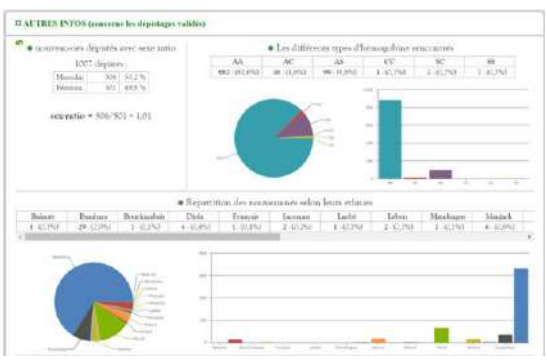


Figure 9– Patients' Distribution of Validated Results, According to Gender, Sickle Cell All Profiles, and Ethnic Groups

Evaluation of SIH SIMENS-LAB

The qualitative evaluation yielded very good results (see table 2). All users found that the system was easy to use, reliable, secure, and that it facilitated their work. Moreover, most of them considered that it was able to improve the quality and the continuity of care, and, to a lesser extent, the decision-making process. Decision-making seemed to be the weakest point of the system, though, with two users finding that the system did not improve decision making, and 3 expressing no opinion.

Discussion

In this section, we discuss the related works on Laboratory Information Systems which have various features :

- The automatic validation of analysis results, which consists in checking the plausibility of the results by confronting them all with a certain number of previously defined parameters in the system. In some laboratories in Europe, it is necessary to validate more than ten thousand results per day. The system ensures the same quality of validation throughout the day. For example, Valab software is a reference application in biological validation [10].
- The management of electronic medical records, sampling, data capture, validation, printing of results, and rapid communication of test results to connected health centers are executable tasks with e-Chasqui. This system is designed to support the national network of anti-tuberculosis laboratories in Peru. It improves the quality of care and ensures effective follow-up of TB patients [11,12].
- The management of HIV testing, treatment of HIV-positive patients, and patient data security is achievable through the NETLAB system. It is a software developed to improve the treatment of HIV in Peru. It manages 100 different diseases for which the National Institute of Public Health and the network of health laboratories in Peru provide test prescriptions and results. It ensures, among other things, the communication of laboratory results, in complete confidentiality to laboratory staff, health providers, and patients living with HIV [13].

In addition, there are other large-scale solutions for performing more complex tasks than in the examples mentioned above. For example, Delphic is a SIL for end-to-end information management for hospital and community labs. It covers all phases of the laboratory work process. It is adaptable to various laboratory environments and fully customizable, making it ideal for multi-laboratory organizations [14]. The BK-LIS is a LIS for patient data management of laboratories in health facilities in Vietnam. It is designed for the management of Hematology, Biochemistry, Immunology, Microbiology, Urine, Endocrine, and Cell Analysis.[15].

As far as we are concerned, information systems are being implemented especially for sickle cell disease, the most common genetic disease in the world. It is generally found in all continents but, the highest prevalence rates are observed in Africa. Thus, the Laboratory LARTIC (Support Laboratory Research and ICT) of the Faculty of Medicine of Antsakaviro of Madagascar in collaboration with the NGO Fight against Sickle Cell Disease in Madagascar, France and the Faculty of Medicine of Antananarivo have set up the openMRS-Sickle-Cell, an application based on OpenMRS for the network fighting against sickle cell disease in Madagascar. In Madagascar, where the prevalence of hemoglobin S trait carriers is estimated at

approximately 2,000,000 distributed in highly endemic areas that are often remote and almost inaccessible, the system allows the centralization of clinical data of patients screened and followed up [16].

In the context of CERPAD where we developed our LIS, we could design a system from scratch for systematic neonatal screening and ongoing follow-up of patients. We integrated social specificities of the West African region, such as the genetic roles played by ethnic groups, in a long-term multidimensional database for research purposes. This research should allow a better characterization of the disease taking into account the socio-demographic and environmental parameters in Saint-Louis of Senegal.

Conclusion

SIMENS-LIS4SC fulfills several functions allowing the realization of the different laboratory tasks for biological tests related to the management of sickle cell disease. It allows the laboratory to take advantage of the automation of tasks. The SIMENS-LIS4SC also makes data entry, reporting and archiving much easier within the CERPAD. Physicians can quickly access results and analysis reports interpreted and validated by the biologists, which greatly speeds up diagnosis and improves treatment. This system also provides efficiencies and decreases patient expenses by reducing the number of redundant tests. Researchers can use the data collected to carry out clinical, epidemiological, and social studies.

The SIMENS-LIS4SC is reusable in other sickle cell screening programs. It is also adaptable for any other disease screening and could extend to hospital level, by adding relevant tests.

Acknowledgements

This work is part of the SIMENS project which is supported by the African Center of Excellence in Mathematics Computer Science and ICT (CEA-MITIC), the Pierre Fabre Foundation, and the Senegalese High Study and Research Ministry through the Scientific and Technic Research Impetus Funding (FIRST).

References

[1] M.F. Collen, R.E. Miller, Clinical Laboratory (LAB) Information Systems, in: M.F. Collen, and M.J. Ball (Eds.), Hist. Med. Inform. U. S., Springer London, (2015) 525–591, doi:10.1007/978-1-4471-6732-7_12.

[2] I.C. Cucoranu, Laboratory Information Systems Management and Operations, *Clin. Lab. Med.* **36** (2016) 51–56, doi:10.1016/j.cll.2015.09.006.

[3] G. Camara, A.H. Diallo, M. Lo, J.-N. Tendeng, S. Lo, A National Medical Information System for Senegal: Architecture and Services, *Stud. Health Technol. Inform.* **228** (2016) 43–47.

[4] M. de la santé et de la prévention du Sénégal, Plan National de Développement Sanitaire (PNDS) 2009-2018, (2009).

[5] P.Y.B. Mané, Efficience et équité dans le système de santé du Sénégal, phdthesis, Université Claude Bernard - Lyon I, (2013).

[6] F.B. Piel, S.I. Hay, S. Gupta, D.J. Weatherall, T.N. Williams, Global burden of sickle cell anaemia in children under five, 2010-2050: modelling based on demographics, excess mortality, and interventions, *PLoS Med.* **10** (2013), doi:10.1371/journal.pmed.1001484.

[7] E. Wastnedge, D. Waters, S. Patel, K. Morrison, M.Y. Goh, D. Adeloye, I. Rudan, The global burden of sickle cell disease in children under five years of age: a systematic review and meta-analysis, *J. Glob. Health*. **8** (2018), doi:10.7189/jogh.08.021103.

[8] L. Thiam, A. Dramé, I.Z. Coly, F.N. Diouf, N. Seck, D. Boiro, A.A. Ndongo, I. Basse, B. Niang, I. Deme/Ly, A. Sylla, I. Diagne, O. Ndiaye, Profils épidemiologiques, cliniques et hématologiques de la drépanocytose homozygote SS en phase inter critique chez l'enfant à Ziguinchor, Sénégal, *Pan Afr. Med. J.* **28** (2017), doi:10.11604/pamj.2017.28.208.14006.

[9] K. Schwaber, M. Beedle, Agile Software Development with Scrum, 1st ed., Prentice Hall PTR, Upper Saddle River, NJ, USA, (2001).

[10] L. Prost, E. Rogari, How autoverification through the expert system VALAB can make your laboratory more efficient, *Accreditation Qual. Assur.* **7** (2002) 480–487, doi:10.1007/s00769-002-0544-1.

[11] J.A. Blaya, S.S. Shin, M. Yagui, C. Contreras, P. Cegielski, G. Yale, C. Suarez, L. Asencios, J. Bayona, J. Kim, H.S.F. Fraser, Reducing Communication Delays and Improving Quality of Care with a Tuberculosis Laboratory Information System in Resource Poor Environments: A Cluster Randomized Controlled Trial, *PLoS ONE*. **9** (2014), doi:10.1371/journal.pone.0090110.

[12] J.A. Blaya, S.S. Shin, M.J. Yagui, G. Yale, C.Z. Suarez, L.L. Asencios, J.P. Cegielski, H.S. Fraser, A web-based laboratory information system to improve quality of care of tuberculosis patients in Peru: Functional requirements, implementation and usage statistics, *BMC Med. Inform. Decis. Mak.* **7** (2007) 33, doi:10.1186/1472-6947-7-33.

[13] P.J. García, J.H. Vargas, P. Caballero N, J. Calle V, A.M. Bayer, An e-health driven laboratory information system to support HIV treatment in Peru: E-quity for laboratory personnel, health providers and people living with HIV, *BMC Med. Inform. Decis. Mak.* **9** (2009) 50, doi:10.1186/1472-6947-9-50.

[14] M. Gill, Delphic-a new comprehensive laboratory computer system, *Pathology (Phila.)*. **13** (1981) 634, doi:10.1016/S0031-3025(16)38353-2.

[15] D.H. Vu, D.T. Nguyen, Design of laboratory information system for healthcare in Vietnam BK-LIS, in: Proceedings of the Int. Conf. Commun. Electron. 2010, (2010) 110–114, doi:10.1109/ICCE.2010.5670692.

[16] H.F. Andriambololoniaina, R.L. Randriamboavonjy, A.N. Ratovohery, M. Razafinimanana, P. Jeannot, D. Zanamiarana, W.R. Razafindrakoto, R. Olivat, R. Fahafahantsoa, Implémentation de l'application OpenMRS ® pour le réseau de lutte contre la drépanocytose à Madagascar : Un modèle de contribution scientifique pour des enquêtes nationales dans les pays francophones en voie de développement, in: Pré-Symp. Francoph. À MEDINFO 2013, Copenhague, Denmark, (2013).

Address for correspondence

Mr. Al Hassim Diallo, Gaston Berger University – B.P. 30 – Saint-Louis – Sénégal, alhassimdiallobe@gmail.com, +221773139352.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190296

Portable Health Clinic: An Advanced Tele-Healthcare System for Unreached Communities

Rafiqul Islam[a], Yasunobu Nohara[a], Md Jiaur Rahman[b], Nazneen Sultana[b], Ashir Ahmed[c], Naoki Nakashima[a]

[a] Medical Information Center, Kyushu University Hospital, Fukuoka, Japan
[b] Grameen Communications, Dhaka, Bangladesh
[c] Department of Advanced Information Technology, Kyushu University, Fukuoka, Japan

Abstract

The Portable Health Clinic (PHC) system endeavors to take healthcare facilities along with remote doctors' consultancy to the doorsteps of the unreached people using an advanced telemedicine system. Thus, the necessity of having physical healthcare peripheries specially in the developing countries can be mitigated. The PHC system promotes preventive healthcare by encouraging regular health checkups so that diseases can be prevented as well as their severity can be mitigated, leading to a reduction on healthcare expenses. Thus, the number of patients along with excessive workload on existing healthcare human resources can be minimized. The current project in rural Bangladesh alone has served more than 41,000 people so far by the PHC system and a simple analysis of this data shows some significant findings on regional health status. A simple expansion of this program, covering a wider service area, can produce a big data to reflect the whole country`s health profile.

Keywords:
Preventive Healthcare, Telemedicine, Triage

Introduction

Healthcare facility is a basic right for all human beings. Unfortunately, the shortage of qualified doctors and health workers, insufficient medical facilities and lack of healthcare awareness remain as some major obstacles for ensuring a standard level of healthcare service in the developing countries [1, 2]. Under this circumstance, telemedicine with preventive healthcare could be considered as a key to overcome this situation. Keeping this in mind, the Portable Health Clinic (PHC) system has been developed as an advanced telemedicine system for the rural communities in Bangladesh [3, 4]. This system, also called "Doctor in Box", enables bringing healthcare services to the doorsteps of the rural communities at a affordable price (Figure 1).

Nowadays, the prevalence of non-communicable diseases like Diabetes Mellitus and Hypertension has increased to a cautious extent. Being a developing country, Bangladesh is not an exception. From our studies, we have learned that not only the urban people, but the people in suburban and rural areas are equally affected by these diseases. These diseases and the consecutive complications can be effectively prevented by taking cautions beforehand. Prevention is more important in countries like Bangladesh because of the limited ability of people to spend on health bills and the absence of provision from government. Preventing such diseases to occur or diagnosing it at early stage can help people to save substantially on medical bills. For this, regular screening of health status is important which can be facilitated by the PHC. Thus, the PHC system has been developed in a preventive healthcare approach with a special focus on non-communicable diseases [5, 6]. One of the main obstacles for ensuring basic healthcare service in developing countries is the poor doctor-patient ratio. According to World Health Organization (WHO), where it requires at least 12 doctors for standard healthcare service to every 10,000 populations, Bangladesh have just 4 doctors. However, it is not that easy to increase the number of doctors in a short period of time. When it is not easy to improve this doctor-patient ratio by increasing the number of doctors alone, decreasing the number of patients is very important. The PHC system also aims to contribute in improving this doctor-patient ratio by introducing preventive healthcare to reduce patients by advance intervention to the problem.

Figure 1. Portable Health Clinic ("Doctor in Box")

The PHC system made it possible to provide primary healthcare services to the doorstep of the rural communities through a telemedicine system. However, to ensure better consultancy by the remote doctor, accurate and a wide range of diagnosis reports of the patient are required to be available to the remote doctor. Unfortunately, most of the developing countries do not have enough quality diagnosis laboratories in the rural areas with qualified pathologists for producing reliable reports. Therefore, this work introduces a new module to the PHC system called the "Tele-Pathology" system that enables rural laboratory technologists to gain assistance from a remote professional pathologist using an online tele-healthcare system and receive the verified report from the pathologist [7]. This involves a very simple technology using available devices like a camera and microscope, and it can be easily replicated.

While providing this primary healthcare service in rural Bangladesh, a huge need of eye care services was identified.

Considering this demand in rural Bangladesh, this work also added another new module with the PHC system called the "Tele-EyeCare" system for ensuring eye care services [7].

The main objective of this work is to improve the PHC system and expand the service scope so that it can respond to the wide range of demands of rural patients with quality consultancy aided by accurate diagnosis. Addition of Tele-Pathology and Tele-EyeCare modules will contribute significantly in this direction.

Methods

PHC System Structure & Operations

The PHC system consists of 4 components: 1) PHC box with various medical sensors, internet enabled tablet pc and printer, 2) health worker 3) online dataserver for sharing and preservation of health data and 4) remote doctor call center (Figure 2). The health worker brings this PHC box to the patient to measure the vital information and upload this data, along with the medical history of the patient to the online server using the system application (app). The remote doctor gains access to this data and makes a video call to the patient for further verification. Finally, the doctor produces an online prescription and preserves it in the server under the patient`s personal file. Then, the health worker prints the prescription from the server and passes to the patient with detail explanation instantly (Figure 3). The whole process to serve one patient takes about 15 minutes excluding doctor`s consultancy time.

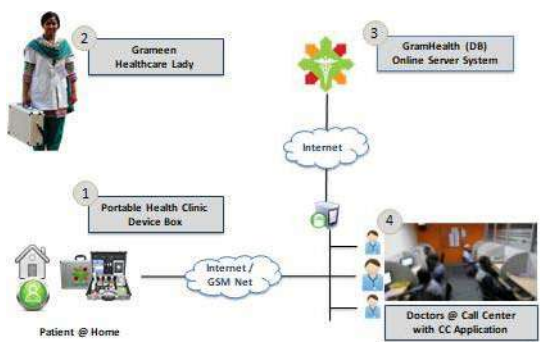


Figure 2. PHC System Structure

The PHC system introduces a triage process based on the concept of "B Logic" for the people of Bangladesh. It classifies the subjects under investigation in four categories, namely, (i) green or healthy (ii) yellow or caution (iii) orange or affected and (iv) red or emergent, based on the gradual higher risk status of health [8-11]. The subjects under orange and red are primarily diagnosed as in the risky zone who need doctor`s consultancy. However, the major part of the subjects who are diagnosed in the alarming zone (yellow) can be served by the trained health workers without medication and they can be prevented from shifting to the risky zone (orange and red) being under the guided lifestyle. This reduces the pressure on the doctors, enabling them to focus on the risky patients who deserve better attention.

Tele-Pathology System

The Tele-Pathology module of the PHC system enables the rural diagnostic centers operated by laboratory technologists (diploma) for producing quality pathological report with the support of the qualified remote pathologist (Figure 3). At present, this system is capable of a blood hematological (CBC) test, routine examination of urine, routine examination of stool and a skin scarping test. In this system, the rural laboratory technologist (1) collects the sample, prepares the physical report, produces the test slide, (2) takes a number of microscopic images of the slide with varying positions and then (3) uploads the images along with the physical report to the online server. The remote pathologist then (4) diagnoses the sample based on the microscopic slide images with the reference of physical report, finalizes the pathology report and preserves to the online server.

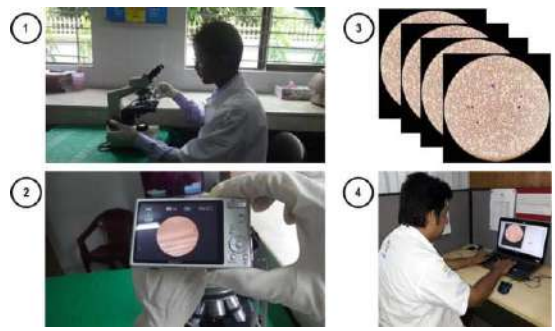


Figure 3. Tele-Pathology System

In case a pathological report is required by the call center doctor (physician), the doctor can gain access to the report from the online server for preparing a prescription. If needed, the laboratory technologist can also download this report to deliver to the patient.

Tele-EyeCare System

The Tele-EyeCare module of the PHC system ensures primary eye care services for the rural communities. This module has added a Digital Fundus Camera to the PHC box for retinal imaging (both Mydriatic & Non-Mydriatic) of the patients. This enables the health workers to have a better view of the retina and the peripheral for identifying the problem with certainty. Thus, the trained local health workers check and classify the patients as per severity of the problem. Low risk patients with simple complications are served by the health workers. However, in case of doubtful cases, the patients` vital information, initial primary checkup reports and ophthalmic images are shared with a remote ophthalmologist using the online server (Figure 4).

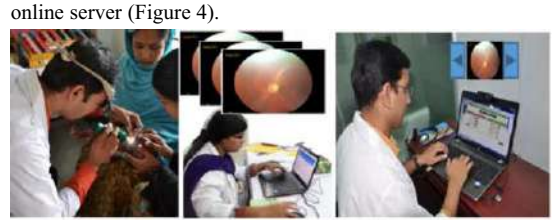


Figure 4. Tele-EyeCare Service System

The ophthalmologist then checks the initial primary checkup reports produced and shared by the local health worker, investigates the ophthalmic images, directly talks to the patients over video conference system, reconfirms their status and finally, provide online prescriptions. If needed, the ophthalmologist can ask for further investigation of the patient by the health worker and can also provide glasses prescription with the support of the health workers. Thus, this system enables to provide basic eye care to the ordinary patients by the health workers and special care for the critical patients by

professional ophthalmologist who really deserve special attention.

Village Service Delivery Model

In village service delivery model, preferably one female health worker works in a village as she can gain better access to the female patients due to privacy reasons. Usually, she uses a local village medicine shop (pharmacy) as her service point for daily service and general village patients come there for PHC service (Figure 5). However, she also visits door to door in case of elderly or disabled patients, pregnant women, emergency patients or special on-call service.

Figure 5. Village Service Delivery Model

Urban Service Delivery Model

Although, it was developed for the low income rural communities, it is found equality useful for the urban, rich, and aged community for home delivery service. The main clients are aged people who are suffering from non-communicable diseases like hypertension and diabetic that need regular checkups. Although, they are financially rich but as it is difficult for them to visit hospital regularly due to physical stress, trouble to arrive at doctor`s appointment at expected time, traffic congestion, etc. So the PHC home delivery service is offered to the pre-registered urban patients and the health workers visit the patients as per schedule.

Personal Health Record

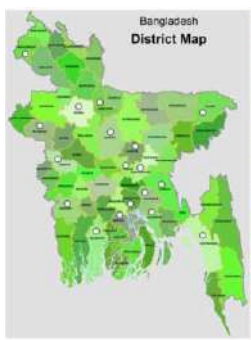
The PHC system preserves all medical data of the patients in the online server so that the doctor can refer previous data in need. All checkup data, pathological reports, medical histories and prescriptions are preserved sequentially. Also, all the patients are provided with their respective user id and password for their access to their personal, health record which is maintained with high security and privacy (Figure 6). They can change their profile information but only monitor the health records with graphical representations.

Figure 6. Personal Health Record & PHC Prescription

The prescriptions produced in the PHC system are unique in shape and format considering the targeted patients of the rural communities (Figure 7). It contains both the measured health data in the left side and doctor medication with advice on the right side. As most of the patients do not understand the significance of digits, it shows all health data with corresponding color sign. The same four colors of the triage system (green, yellow, orange and red) are used against each data point so that they can easily understand the severity of any particular item.

Results

The PHC healthcare services have been offered in 18 districts of Bangladesh at more than 70 service points all over the country (Figure 7). Until April 22, 2018, the number of patients served by the PHC system in Bangladesh alone was 41,949. Besides, the PHC activities are also continued in Cambodia, India, Thailand and Pakistan under the supervision and management of local partners.



As on April 22, 2018					
Year	Total Checkup	Health Status			
2010	377	103	15	248	10
2012	6,412	1,056	4,035	1,084	237
2013	13,277	1,595	6,781	4,301	546
2014	5,288	1,004	2,278	1,659	326
2015	5,547	1,330	1,429	2,016	772
2016	7,007	1,875	1,614	2,255	1,069
2017	2,853	366	800	868	361
2018	1,188	95	410	325	121
Total	41,949	7,424	17,362	12,756	3,442
		24,786 (61%)		16,198 (39%)	

Figure 7. PHC Service Statistic in Bangladesh

The diagnosed major diseases found in PHC healthcare services are hypertension (1,944), diabetes (1,365), anemia (647), and ophthalmic problem (138).

The Tele-Pathology service started in 4 rural diagnostic centers under 4 districts (Barisal, Bogura, Manikganj and Thakurgaon) of Bangladesh. So far, this system has served patients with a total of 1,610 Hematological (CBC) reports and 918 Routine Examination of Urine using remote pathologists.

The Tele-EyeCare mobile service has recently started in 1 rural center called “Vision Center” in the Nator district of Bangladesh on a test basis. So far, it has already served 2,410 checkups for 2,046 patients using remote ophthalmologists. Soon this service will be extended to other parts of the country.

Discussion

Out of a total 41,949 PHC healthcare patients, 61% (24,786) was green and yellow patients who were served by the health workers alone. The rest of the patients 39% (16,198) were served by doctors who needed medication. So, the PHC system can reduce the work load of a professional doctor by 61% that can be managed by local health workers. Thus, the best use of the valued resource of a doctor can be ensured and only the people who really need expert`s consultancy can avail it.

Due to its easy operation, a huge number of health data can be collected by the PHC system and this big data can be used for countrywide disease pattern analysis. For example, this research has found some significant differences of health parameters in different areas. In one area, we have found a

significantly small number of anemic cases among the adult women compared to the rest of the country. In another area, we have identified very high urine protein compared to the rest. Thus, the detail analysis of this data may show some significant findings including environmental issues.

Similar to the shortage of physicians, there is a huge shortage of ophthalmologists and it is just 0.063 ophthalmologist for every 10,000 populations in Bangladesh. The Tele-EyeCare module of the PHC system can highly contribute to ensure the best use of this valued resources for dealing with complicated cases and manage ordinary cases by health workers. For the further advancement of the Tele-EyeCare system, a new development has been started using Artificial Intelligence (AI). It will use ophthalmic image recognition technology with the aid of neural network and deep learning for automatic diseases identification. This will facilitate both the village health workers and ophthalmologists for better and prompt services.

Conclusions

At present, the PHC tele-healthcare system offers a unique opportunity for ensuring better healthcare service covering primary healthcare, eye care and pathological services to the unreached rural communities. However, this modular system will gradually be expanded in other healthcare service areas to cover common healthcare issues based on the local demand. Now, we are working on an Obs & Gyne module and Dental Care module to be added soon to the PHC system.

The concept of the PHC system came from the local demand of Bangladesh. However, since most of the developing countries are facing the same problem and having similar situations, an easily replicable PHC system can be in good use there. So far, this system has been replicated in India, Pakistan, Cambodia and Liberia with some localizations. However, there is still plenty of opportunities for further improvement of the system and expansion of the service to the other parts of the world.

Since the aging communities are increasing in the developed countries, they are also facing a similar crisis of doctor shortage in their rural areas. Therefore, there will arise a huge demand of PHC services in the developed countries as well [12]. To address this demand, the PHC system can be further improved with the aid of technologies. One of these attempts is to include Bluetooth enabled medical sensors so that the measured vital data from the sensors will be automatically transferred to the online server to avoid typing error from manual data entry [13]. Also, there is a requirement to develop the PHC box as an integrated unit so that it will be low cost, handy and light weight for easy operation by a rural health worker.

Acknowledgements

This research has been supported by Grameen Communications of Bangladesh and Kyushu University of Japan. The authors sincerely acknowledge the contribution from both the sides.

References

[1] S.M. Ahmed, M.A. Hossain, A.M. Rajachowdhury, A.U. Bhuiya, The health workforce crisis in Bangladesh: shortage, inappropriate skill-mix and inequitable distribution, *Human Resources for Health*, **9(1)**, 3, 2011.

[2] D.T. Jamison, J.G. Breman, A.R. Measham, G. Alleyne, M. Claeson, D.B. Evans, P. Jha, A. Mills, P. Musgrove, *Disease Control Priorities in Developing Countries, Second Edition*, World Bank and Oxford University Press, Washington, DC, 2006.

[3] A. Ahmed, A. R. Hargrave, Y. Nohara, R. Islam, P. Ghosh, N. Nakashima and H. Yasuura, "Portable Health Clinic: A Telemedicine System for UnReached Communities" in *Smart Sensors and Systems, Springer International Publishing*, 447-467, 2015.

[4] P. Ghose, R. Islam, N. Nakashima, Y. Nohara, A. Ahmed, S. Shimizu, Effectiveness of Preventive Healthcare through Telemedicine in Bangladesh, *Proceedings of 18th ISfTeH International Conference*, 2013.

[5] F. Yokota, A. Ahmed, K. Kikuchi, M. Nishikitani, R. Islam, N. Nakashima, Diabetes, obesity, and hypertension in Bheramara Kushtia District, Bangladesh - Results from Portable Health Clinic Data, 2013-2016, *Proceeding of Social Business Academia Conference*, 2016.

[6] Y. Nohara, E. Kai, P. Ghosh, R. Islam, A. Ahmed, (another 4 authors), S. Shimizu, K. Kobayashi, Y. Baba, H. Kashima, K. Tsuda, M. Sugiyama, M. Blondel, N. Ueda, M. Kitsuregawa, N. Nakashima, Health Checkup and Telemedical Intervention Program for Preventive Medicine in Developing Countries: Verification Study, *Journal of Medical Internet Research*, Vol.17, No.1, 2015.

[7] R. Islam, A. Ahmed, N. Nakashima, S. Shimizu, GramHealth: An Integrated Tele-Healthcare System for Unreached Communities, *Proc. of 11st Asian Telemedicine Symposium*, 2017.

[8] Metabolic syndrome criteria of International Diabetic Federation http://whqlibdoc.who.int/publications/2011/9789241501491_eng.pdf (accessed on April 1, 2019)

[9] Criteria of Obesity using BMI by the National Institutes of Health, USA http://www.nlm.nih.gov/medlineplus/ency/article/007196.htm (accessed on April 1, 2019)

[10] Guideline of American Heart Association http://www.heart.org/HEARTORG/Conditions/HighBloodPressure/AboutHighBloodPressure/Understanding-Blood-Pressure-Readings_UCM_301764_Article.jsp (accessed on April 1, 2019)

[11] Global Guideline for Type 2 Diabetes of the International Diabetes Federation https://www.idf.org/e-library/guidelines/79-global-guideline-for-type-2-diabetes (accessed on April 1, 2019)

[12] T. Hasegawa, R. Suzuki, T. Sakamaki, et.al. Summary of a multicentre prospective clinical study of home telemedicine. *Japanese Journal of Telemedicine and Telecare*, **13(2)**, 84-87, 2017-08.

[13] M. Hasan, F. Yokota, R. Islam, A. Fukuda, A. Ahmed, Errors in Remote Healthcare System: Where, How and by Whom? *Proc. 2017 IEEE TENCON*, 170-175, 2017.

Address for Correspondence

Rafiqul Islam (Maruf), Ph.D.
Medical Information Center, Kyushu University Hospital
3-1-1, Maidashi, Higashi-Ku, Fukuoka, 812-8582 Japan
Tel: +81 92-642-5881; Fax: +81 92-642-5889
Email: rimaruf@med.kyushu-u.ac.jp

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190297

Wireless Sensor Network for Fall Prevention on Geriatric Wards: A Report

Nico Jähne-Raden[a], Henrike Gütschleg[a], Marie Cathrine Wolf[a], Ulf Kulau[b] and Klaus-Hendrik Wolf[a]

[a] Peter L. Reichertz Institute for Medical Informatics, TU Braunschweig and Hannover Medical School, Germany,
[b] Institute of Computer Engineering, TU Braunschweig, Germany

Abstract

With regard to the growing number of older adults, it needs smart solutions for fall prevention. Especially at geriatric institutions, the risk of falling is very high and frequently leads to injuries, resulting in serious consequences. We present the Inexpensive Node for Bed Exit Detection (INBED), a comprehensive signaling system for fall prevention. The INBED system is based on a wireless sensor network infrastructure via IEEE802.15.4 and highly-specialized open hardware in-house developed wearable. The device, which will be attached to the patients, can detect several types of movement. Occurring events are forwarded to the nursing staff immediately by using the self-organizing and scalable network including wide area network integration. The system can help to relieve the staff while the personal freedom of movement and privacy of patients is increased. With this development, the energy-efficient, simple and intuitive mechanisms of proximity communication can be combined with broadband benefits.

Keywords:
Accidental falls, Accident prevention, Geriatrics.

Introduction

As a result of demographic change, the number of people to be cared for, and thus also the challenges and demands on health care, are increasing. Falls in older adults are frequent and fatal events with mostly serious consequences for the individual as well as considerable financial expenditures for the society.

For geriatric wards the average fall ratio per 1000 bed days is more than twice as high as on other clinical wards, which has been validated by the German quality report benchmark in geriatric health care from 2007 [1]. This can be significantly higher if you look on specialized clinical departments such as gerontopsychology wards. A large part of the falls occurs as a result of getting up in the immediate vicinity of the bed [2]. Usually high fall risk patients are encouraged to call the nurses for help if they wish to get up, this directive is not always followed (e.g. due to cognitive impairments). The situation in outpatient areas are similar and again most of the falls occur during rising events for transitions.

The increased falling risk of older adults is mainly caused by chronic diseases and predisposing factors like gait and balance disorders [3]. Furthermore, dementia and motor deficits patients fall twice as often as patients with normal cognitive abilities [4]. One of the most common consequences for elderly fallers is the femoral neck fracture [5]. The incidence among the over-65s in Germany for such an injury is 600 to 900 fractures per 100,000 people per year [6]. Beside the mental and physical strains, fall events are associated with immense consequential costs. The estimated overall annual costs of treatment of fall-related hip fractures in German hospitals and rehabilitation centers are up to 2.77 billion Euro [7; 8].

We designed in several iterations a highly specialized Wireless Sensor Network (WSN) system in different architectures to recognize an attempt to leave the bed for fall prevention. On the one hand, the system should not restrict the patient's movement. On the other hand, it has to recognize potential fall risks in a timely manner, so that an active intervention by nursing staff is possible to maintain the patient's care and safety. Furthermore, the diversity of clinical structures is quite challenging. Other more practical issues are the availability and amount of power sockets, the type of communication infrastructure (Wi-Fi, Ethernet, etc.) and individual restrictions of the specific country (e.g. hygiene requirements, care regulations or patient protection laws, limitations of the use of protocols e.g. Bluetooth or Zigbee). Hence, the following objectives for the development of the INBED fall prevention systems are:

- Development of a reliable and cost-effective solution for a sensor-based bed-exit/rising detection system (primary function) based on initial research results with different sensor prototypes.
- Recognition of restlessness conditions of patients prior to a rising attempt (optional primary function), so that the nursing staff can intervene immediately.
- Detection of falls (secondary function) and provision of virtual risk areas, such as stairs (secondary function).
- State-of-the-art technological concept, focusing on a minimal price for the end customer without reducing the quality of the overall system or individual functional areas.
- Creating an adaptive, scalable and modular communication infrastructure to include the wearables at a ward.

Related Work

In the last years, several studies showed a positive impact of bed exit systems on the fall rate of older adults [4; 9]. Currently, there are two main types of systems available. Distinguished in body worn and ambient variants, they can occur in different forms, e.g. mattress pad systems, ground pressure mats, infrared systems or garment clipping sensor systems [10]. Those systems have advantages and disadvantages depending on the symptoms of the patient and the surrounding. Hygiene requirements that demand special cleaning can exclude the use of system components if they cannot be cleaned to rule. The false positive alarm caused by objects like mobile dining tables, suit cases or by the patient's restlessness leads to dissatisfaction of the nursing staff. Furthermore, one of the main reasons for an

increased risk of falling is dementia [4] which often involves underweight and restlessness [11] whereby the reliable functionality of the system can be affected as well.

For example if a patient weighs under 50 kg, proper function of pressure sensors is not guaranteed so that mattresses or ground pressure mats (cf. [12]) are useless. In addition, mobile dining tables or suitcases can lead to false positive alarms by standing on such floor sensors, which can affect a discontent of the nursing staff. Moreover, special cleaning procedures of nursing and/or clinical facilities can lead to the exclusion of such system components. Having this in mind, the implementation of a proper fall prevention system should be suitable to diagnoses and symptoms of the particular patient condition. This is also in line with the National Institute for Clinical Excellence findings; *"To be effective, they [Bed-Exit Alarms] need to be implemented with care and with a clear understanding of their limitations"* [10].

In general, video monitoring would be a proper solution but besides legal restriction relating to the usage of camera recordings in public, the basic acceptance is an important field. The acceptance of cameras surveillance impartial is quite low in a major group (in general under 50 percent) (see [13; 14]).

In recent years the Peter L. Reichertz Institute for Medical Informatics have worked on the field of fall prevention by developing several prototype systems. This led to first clinical study with the aim of measurement of the reduction falls on a geriatric ward, by close monitoring with a portable proprietary sensor system (www.shimmersensing.com). For this purpose, a bed exit alarm with this wearable was developed as a core component, which reliably detects rising attempts [9].

With the results and experience of this first clinical trial and close collaboration of various clinicians and nursing staff, we were able to further improve the system and adapt it to the conditions of clinical use. Last but not least, this includes the development of various hardware-software-components for the implementation of an adaptive, scalable and modular overall system for fall prevention, which will be presented in this paper. In addition, compared to existing solutions the INBED system is highly cost efficient and comprises the specific requirements of clinical use cases.

INBED system

The first version was tested within a larger 15-month study. For this iteration of the system, we designed the base station using an Arduino with an Atmel ATmega168 micro controller unit (MCU) and a Bluegiga WT11 Bluetooth module. If regular messages from the sensor are missing or the sensor detects a rising attempt, it triggers an alarm and the base station activates the nurse call. For each patient, one pair, that can easily be connected to the nurse call system at bedside, is needed [9].

The wearable part on the fall prevention system was the Shimmer sensor system. Via Bluetooth connection, a variety of receivers can be used to integrate the new modality into the clinic's system. Porting the algorithms to the hardware was straightforward. The Shimmer device provides enough resources, making it possible to detect additional states of the patient [15].

With the gathered information and data from the 2012 study, we could determine new challenges of the clinical everyday life and master them through further development of the system.

In order to cover all requirements, we decided to develop our own scalable system with a wearable core component, the INBED, in close cooperation with the clinical partners. The wearable itself is a small, affordable wireless sensor board based on an ATmega2564rfr2 System on Chip where the MCU as well as the transceiver unit are fully integrated on a single chip combined with a Bosch BMX055 Microelectromechanical systems. In Figure 1 the board can be seen, including an image of a coin scale and in comparison, with the relay node.

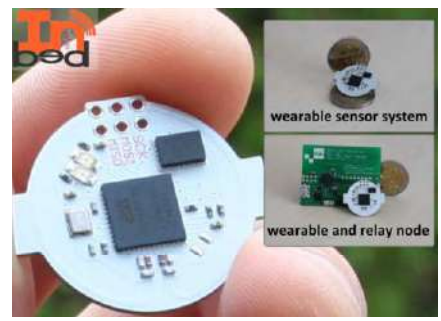


Figure 1– Current INBED wearable board with relay

The ATmega2564rfr2 includes both a low-power 8-bit MCU based on the AVR enhanced RISC architecture as well as a fully IEEE 802.15.4 compliant radio transceiver for the 2.4 GHz ISM band. To reduce the costs, a simple PCB dipole antenna is used. For the sensing part we used a small footprint Bosch BMX055 9-axis sensor module (3-axis accelerometer, 3-axis gyroscope, 3-axis magnetometer). The BMX055 is connected to the ATmega2564rfr2 via Inter-Integrated Circuit (I2C)-bus. By using the accelerometer, the BMX055 implements the option of a free fall detection which can trigger an interrupt. However, many other interrupts can be configured to trigger on several events, e.g. thresholds. Hence, by extensively using the pre-configurable interrupts, the energy efficiency of the entire wearable can be increased as reactive programming models can be applied.

The wearable is powered by a standard CR2032 battery. Thus, the cost overhead and efficiency limitations when using voltage regulators is excluded. To guarantee that the voltage level never undershoots the required level for reliable operation, the ATmega2564rfr2 integrates mechanisms like brown-out detection and a battery monitor. The voltage can be sensed continuously to allow an early notification when a battery runs out of energy.

All components are mounted on a 2-layer PCB with the dimension of 20mm diameter plus two 1mm wings and 6mm in height. The wearable detects several patient related movement events, like rising attempts, and sends a signal to the nursing staff immediately, using a relay node based on IEEE802.15.4. Finally, a signal processing base station (Raspberry Pi 3) to create a modular communication network. The base station provides an alarm which is displayed (optical, acoustical or haptic) on the user interfaces (hardware-enhanced mobile phone) and the staff terminals [15].

Overview and Scenario

The main goal of the INBED system is to inform the nursing staff about pre-fall patient movement events in appropriate time, that they can help the patient to get up and assist the walk (e.g. for the toilet) and prevent falls.

For the application of the system the INBED wearable will be attached to the patient's upper leg on the upper half of the thigh. This position is optimal to detect rising event by angle changes of the legs without affecting patient's comfort to much. Furthermore for a functional system, relay nodes have to be positioned

on the ward at strategic points, like patient rooms, at a maximum distance of 20m from one another to form a suitable network for data transmission and node locating. Finally, the base station has to be set up. It should be equipped with some patient information and at least with system information, which relay nodes is where at the ward and which wearable node is in circulation.

For a better description of the system and its functions, we will use a fictional but typical clinical scenario described in Figure 2.

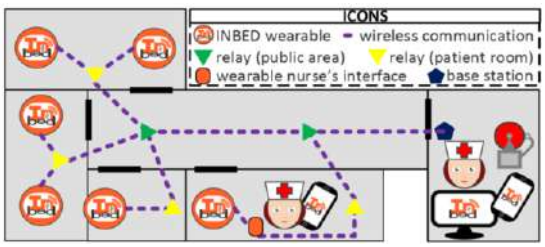


Figure 2– Floor plan of an exemplary ward with installed INBED system

The setting: The initial situation is to serve a regular geriatric ward. Here, the focus should be on patients with the following characteristics: high fall risk, motor functional deficits, sensory restricted, basically self-sufficient. In case of problems (pain, urgency) these patients are advised to contact the nursing staff to reduce the risk of falls. But in the majority of cases this instruction is ignored. Hence, in this scenario the patients are equipped with INBED wearables, which has been attached to the thigh by means of a regular band aid, at admission on the ward. A relay node has been installed in the rooms where the patients are located, as well as at certain distance on the ward floor, and at strategic points in the ward, like exits, elevators or public common area and the drug delivery. The base station is located in the nursing room and connected to the ward Wi-Fi. Ward internal mobile devices are available as an interface for the nursing staff. The model from Figure 2 can be assumed as a regular station. The overall system deploys an IEEE802.15.4 communication network and is therefore not dependent on external infrastructure. The network is structured as a tree with the base station as root node, relay nodes as intermediate nodes and the INBED wearables as leaf nodes while the wearables for the nursing staff can join the tree at any point. In every patient's room lays at least one patient, wearing an INBED device. The nurse's room is equipped with alarm interface devices, like a ward computer and the internal ward patient alarm system. Moreover, the simplified communication paths of the system can be seen, including broadcast communication of the wearables to fix patients room relay nodes as well as the unicast communication between the relay nodes and the base station. As can be seen one nurse is close to a patient, equipped with a smart phone including a system interface, which is utilized as a mobile relay node.

The procedure: A typical scenario might be, that a patient will wake up in the night e.g. due to a strong urinary urge. The patient begins to move in bed, due to a certain degree of physical restlessness is associated with the urge to urinate. This unusual high variance of the movement of the patient triggers a restlessness event. The event message is sent via a broadcast and received by the patient room relay node (nearest node). For each generated event alarm by the wearable will send various encrypted and pseudonymized information. These are information about the origin (ID of the sender), the nature of the incident (code of the event) and logistic data for the processing or function provision (battery voltage, etc.). While the INBED wearables send their messages via omnidirectional broadcast, the relay nodes forward their messages via directed unicast towards the base station.

A message contains message prefix, which shows the state of the current message within the system structure (new or already known), event identifier, device ID (wearable device), battery voltage (wearable devices), first relay node that received the massage (ID) and received signal strength indicator (RSSI) of the first receive. If the message is received by the base station, the contained data will be stored in a database which provides the information on several nursing staff devices (e.g. smart phones).

This alarm is displayed prioritized with the database's join information, affected patient, room, event type. The event is displayed via the ward computer as well as on the mobile devices of the nursing staff. Moreover, both acoustic or haptic alarm were optionally used. While the nursing staff is often quite busy and ignore restlessness warning and decides to continue with the current work, the restlessness warning can be deactivated via all interfaces. Further restless warning will be triggered by the system and notified by the staff, if the urge to urinate persists. In doing so, the nursing staff obtains knowledge about the urgency of an intervention on the patient.

With increasing urgency, the patient decides to go to the toilet without contacting the nurses. This triggers a rising-up-event which is distributed according to the restlessness event. At the base station, the rising-up alarm will be triggered with a higher priority compared to the restlessness alarm. Due to multi-morbid and motor-functionally restrictions of geriatric patients, the rising from reclined position will usually take some time (cf. [12; 14]). With the information about the patient's attempt to get up, the nursing staff is called to go to the patient as soon as possible. The alarm cannot be cancelled via application buttons.

When the nursing staff arrives at the patient, they can take care of his/her needs. By measuring the signal strength of the wearable's signal, the proximity to the nursing staff, can be estimated. Then an existing alarm will be switched off when staff is near an alarming device. Figure 3 shows the overall functions of the whole INBED system.

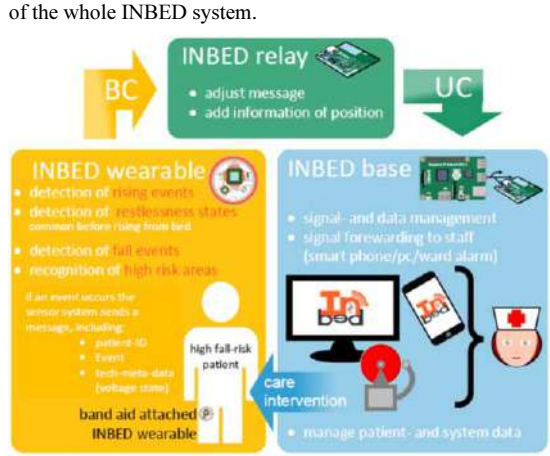


Figure 3– INBED system functionality diagram

Communication and Energy Efficiency: The past ten years research show a widespread of IEEE 802.15.4-based communication in small range Personal Area Network and for sensor communication [16; 17]. However, during the first bed-exit

study an intuitive shortrange one-on-one Bluetooth communication of the wearable part to the base station, that triggers an alarm via nurse call were used. The unidirectional communication via the already installed, analogue and functional nurse call system allows only alarm triggering and not the raw data transmission. The follow-up system consists of commercial-of-the-shelf components on an open hardware-design. This enables short-range communication via its own modular, intuitive and highly energy-efficient IEEE802.15.4-network and is therefore largely. Its own network provides additional side-effects, such as indoor tracking or risk area detection. Due to the bidirectional connection, an adjustment of thresholds while operation is possible [15].

Testing: The functionality test of the entire system was performed under real conditions, with all components. Three test subjects (m=1, w=2, age=28±3), for three trials per subject and event were equipped with an INBED wearable before different scenarios have been investigated. The test cases include both, intended triggering of events (e.g. by standing up, falling, restlessness, error triggering) and testing of the normal case without explicit supervision. For the tests the wearable was attached to the upper front of the thigh as described before.

To underline the suitability for daily use the test series were performed in a clinical environment. For the validation of rising detection, the test subjects started from a lying position and slowly began to rise. Finally, they stood up and get out of the clinical bed, as can be seen in Figure 4. The first three phase images show the approach to the edge of the bed and the beginning of the uplift of the upper body. The phase images four to six show the overcoming of the bed edge with the body and the final raising to the standing position (#7).

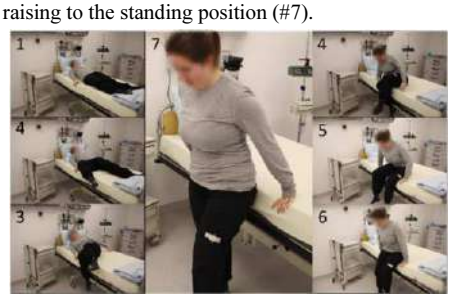


Figure 4: Procedure of a general rising out of bed, represented by seven phases.

We also validated the detection of fall events with six subjects (m=2, w=4, age=28±3), for five trials per subject, as well as restlessness events with two subjects (m=1, w=1, age=26±1), two nights with four wearables per subject, and the corresponding alarms. It should be mentioned, that these sequences of falls were exemplary but representative fall event in clinical environments like the rising before. Within the Figure 5 the raw acceleration data of the three axes are shown, as well as the restlessness counts at the bottom of the diagram and at the point when the interrupt occurred a dotted line can be seen. The counts of the diagram are weighted by the kind of the movement related to the intensity of the motion itself (calm motion = 1, stronger/quicker motion = 2).

The rise up is complete with the final standing of the subject. Thus, a measurement of a fixed period of time is not practical. In the shown case, the recording is about ten seconds long. The restlessness is transmitted after about one second and after about three seconds the rising was transmitted as an alarm.

For the fall test it can be said that the procedure is more or less the same as can be seen at rising with the slight difference that the trial isn't ending by standing but by a subject fall. While a fall the slight decreasing acceleration (near 0.4g) can be followed by high g peak by subject impact.

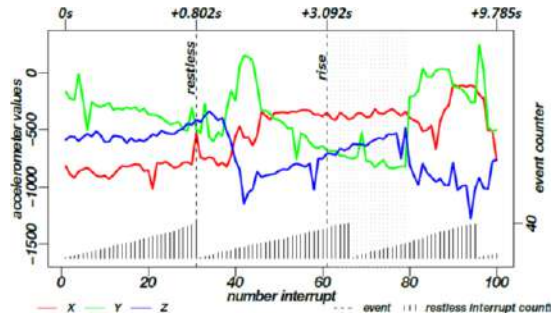


Figure 5– Diagram of the recorded INBED data during rising (raw triaxial acceleration data lines) – in lower area the restlessness counter lines - dashed line show the triggering of the alarms (restlessness and uprising) - grey dotted lines represent standing

For the test of restlessness events, the subjects went to a lying position and then turn from side to side. The recording of the uprising or falling test process differs from the restlessness in terms of duration, this is due to the functional determinism of rising. The recordings of the restlessness measurement are about 35 seconds long for the test trials. The triggering of the corresponding event is expected during the repeated rotation process. Within the tests the actual alarm triggered after about 20s under described circumstances. But it has to be mentioned that the trials aren't representative for every "real" geriatric patient hurry.

Prior to the test runs, we performed a learning phase based on collected data of rising events from different subjects within a cross-validation. According to the person's movement, interrupts are trigger the wearable to wakes up. Subsequently it reads out information about the three axes and the type of interrupts, e.g. if restlessness occurred. Due to the constant restlessness caused by the subject's movements, the INBED node does not have the opportunity to let the restlessness count decay again, by fading. Thus, the count is reset only after reaching the restless threshold. Besides the more detailed described tests above, further trials have been performed.

The testing of the main feature, the rise detection, was evaluated by a comprehensive assessment with 224 different subjects (mostly young students with knowledge of the functionality). The measurement was totally anonymous by means no subject meta data was stored. All subjects were placed in a laying position on a regular hospital bed like height cot.

Parallel to the described testing, several overall system interviews and demonstrations, for health care professionals and clinical IT experts, were done to improve the system and the system settings.

To test the communication features, as well as the energy-efficiency we investigated the communication range of the wearables. Therefore, INBED wearables were equipped with batteries and iteratively moved remoter away from a relay node. On average a communication range of 25m is reached for line of sight while about 15m-25m are usual in a common building structure. The different results are related to the specific structure of the sending path's obstacles, like walls, water pipes or power lines.

Results

For the main functionality the rising detection a sensitivity of 100% in total only slight differences in detection speed were mentioned (few seconds) could be achieved. Within the learning process we identified an angle threshold of 63−117° in sagittal axis.

For the fall events, an optimum value for the free fall phase is ≤ 0.4g and for the impact variance of ≥ 1.3g. However, problems were encountered, while heavy stomping gait pattern and individual steps can also be detected as a false-positive fall.

Within the restlessness detection tests no false-positive events were recorded. Any larger movement (e.g., rotation) seen in the reference camera footage has also been recognized by the INBED system as restlessness event and any documented false-positive rising was be detected by the system, as well.

To evaluate the energy efficiency of the nodes, we assume a battery power of 230mAh, a clock frequency of 8MHz of the MCU and a standard room temperature of less than 20°C. In "worst-case", which means a continuous transmission of a fixed message every 0.9s, the INBED wearable lasts about 52 hours (approx. 200,000 transmissions). The "best-case", where only life-sign messages occur, results in more than 19 days lifetime for the wearable.

Conclusion

The main functionality, the rising detection showed a high sensitivity (100%) for the tests. The specificity, under controlled circumstances, is also high (no false-positive while tests), but real clinical conditions will show reliable results. Former study results are promising (see [9]).

For the restlessness test a high specificity could be measured as well, but the arranged tests are maybe not that expressive, cause no older adult subjects were included for now, which may lead to inaccurate movements. Within a future study we will assess the specificity under real conditions.

The fall test results show quite good sensitivity in heavy falls, but the final sensor system setting can lead to false-positives while heavy gait or subject bumps. A strong bump of a patient, e.g. at a table, is interpreted with high probability also as a fall. However, this can be clarified by the staff by asking. In addition, such a strong impact can also lead to injury, hematomas, which may need treatment. The sensitivity is limited for short height falls, e.g. out of low-floor bed.

By a hybrid approach, both the advantages of short-range, e.g. energy efficiency, as well as the benefits of larger bandwidth of Wi-Fi from the base station can be used [15].

The further integration of other sensors, e.g. gyroscope is planned. Another goal is to further optimize internal processes by adding the possibility to adapt individual configuration values remotely. As the INBED system is operational in our second clinical long-term study, this time as a multi-centric and randomized study within two geriatric wards, starting soon.

References

[1] C. Heinze, R.J.G. Halfens, and T. Dassen, Falls in German inpatients and residents over 65 years of age, *Journal of clinical nursing* **16** (2007), 495-501.

[2] H. Nickens, Intrinsic factors in falling among the elderly, *Archives of Internal Medicine* **145** (1985), 1089-1093.

[3] L.Z. Rubenstein, K.R. Josephson, and D. Osterweil, Falls and fall prevention in the nursing home, *Clinics in Geriatric Medicine* **12** (1996), 881-902.

[4] C. Van Doorn, A.L. Gruber-Baldini, S. Zimmerman, J. Richard Hebel, C.L. Port, M. Baumgarten, C.C. Quinn, G. Taler, C. May, and J. Magaziner, Dementia as a risk factor for falls and fall injuries among nursing home residents, *Journal of the American Geriatrics Society* **51** (2003), 1213-1218.

[5] Dachverband Osteologie, Prophylaxe, Diagnostik und Therapie der Osteoporose bei Männern ab dem 60, *Lebensjahr und bei postmenopausalen Frauen. S3-Leitlinie des Dachverbands der Deutschsprachigen Wissenschaftlichen Osteologischen Gesellschaften eV–Kurzfassung und Langfassung. Internet: http://www. dv-osteologie. org/dvo_leitlinien/osteoporose-leitlinie-2014* (2014).

[6] R. Osterkamp, Population developments in Germany until 2050, *Der Chirurg; Zeitschrift fur alle Gebiete der operativen Medizen* **76** (2005), 10-18.

[7] E.-J. Weyler and A. Gandjour, Socioeconomic burden of hip fractures in Germany, *Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany))* **69** (2007), 601-606.

[8] L. Pientka and C. Friedrich, The costs of hip-fracture in Germany: a prospective evaluation, *Zeitschrift fur Gerontologie und Geriatrie* **32** (1999), 326-332.

[9] K.H. Wolf, K. Hetzer, H.M. Zu Schwabedissen, B. Wiese, and M. Marschollek, Development and pilot study of a bed-exit alarm based on a body-worn accelerometer, *Zeitschrift fur Gerontologie und Geriatrie* **46** (2013), 727-733.

[10] Anonymous, Bed-exit alarms: a component (but only a component) of fall prevention, *Health Devices* **33** (2004), 157-168.

[11] D. Cronin-Stubbs, L.A. Beckett, P.A. Scherr, T.S. Field, M.J. Chown, D.M. Pilgrim, D.A. Bennett, and D.A. Evans, Weight loss in people with Alzheimer's disease: a prospective population based analysis, *Bmj* **314** (1997), 178.

[12] E. Capezuti, B.L. Brush, S. Lane, H.U. Rabinowitz, and M. Secic, Bed-exit alarm effectiveness, *Archives of gerontology and geriatrics* **49** (2009), 27-31.

[13] M. Ziefle, S. Himmel, and W. Wilkowska, When your living space knows what you do: Acceptance of medical home monitoring by different technologies, in: *Symposium of the Austrian HCI and Usability Engineering Group*, Springer, 2011, pp. 607-624.

[14] K. Arning and M. Ziefle, "Get that Camera Out of My House!" Conjoint Measurement of Preferences for Video-Based Healthcare Monitoring Systems in Private and Public Places, in: *International Conference on Smart Homes and Health Telematics*, Springer, 2015, pp. 152-164.

[15] N. Jähne-Raden, H. Gütschleg, M.C. Wolf, and K.-H. Wolf, Advantages of Modular Hybrid Network Communication on Clinical Wards, *Studies in Health Technology and Informatics* **251** (2018), 31-34.

[16] V. Custodio, F. Herrera, G. López, and J. Moreno, A review on architectures and communications technologies for wearable health-monitoring systems, *Sensors* **12** (2012), 13907-13946.

[17] H. Tung, K. Tsang, K. Lam, B. Li, L. Yeung, K. Ko, W. Lau, and V. Rakocevic, A mobility enabled inpatient monitoring system using a ZigBee medical sensor network, *Sensors* **14** (2014), 2397-2416.

Address for correspondence

Nico Jähne-Raden, Hannover Medical School, Carl-Neuberg-Str. 1
D-30625 Hanover, Germany, Nico.Jaehne-Raden@plri.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190298

Transforming Nursing Documentation

Melinda L. Jenkins[a], Avaretta Davis[a,b]

[a] *Division of Advanced Practice Nursing, School of Nursing, Rutgers University, Newark NJ, USA,*
[b] *Veteran's Health Administration, Washington DC, USA*

Abstract

Graduate nursing education is positioned to transform nursing documentation so that it more fully describes nursing assessments, diagnoses, interventions, and outcomes to measure improvements in care. Learning to document with structured nursing terminology is an integral part of "Information Technology for Evidence-Based Practice", a required online course taken by all students in the Rutgers Doctor of Nursing Practice program. Beginning with SOAP and terminology required for billing, students create a clinical note adding elements of the Nursing Minimum Data Set, using Clinical Care Classification terms. Next, students are asked to select a nursing-related clinical practice guideline, electronic clinical quality improvement measure, and a screening tool that applies to their encounter note. Then, they identify Patient Reported Outcome Measures as well as improvement activities in the CMS Quality Payment Program. The course is well-received; many graduate students now face changes in documentation and electronic tools and can predict future evolution.

Keywords:

Terminology, Reference Standards, Nursing Informatics, Graduate Nursing Education

Introduction

Clinical documentation by Advanced Practice Nurses is useful to measure and improve the quality of care. Graduate nursing education is positioned to transform nursing documentation so that it more fully describes nursing assessments, diagnoses, interventions, and outcomes. As electronic health records evolve, optimizing clinical documentation will demand linking clinical notes to structured terminologies to power essential utilities such as clinical decision support, reports for value-based payment, and efficient information exchange with consumers and other providers. This paper presents a method of teaching documentation using structured nursing terminology, blending it with medical documentation required for billing, and connecting it with clinical guidelines, quality measures, patient-reported outcomes, and population health. Examples of students' exercises are included.

Learning to document with structured nursing terminology is an integral part of "Information Technology (IT) for Evidence-Based Practice", a required online course taken by all students in the Rutgers Doctor of Nursing Practice (DNP) program. The course is designed to meet one of the American Association of Colleges of Nursing's DNP Essentials: IT competencies for the improvement and transformation of health care [1] and the American Organization of Nurse Executives' recommended information management and technology competencies for nurse leaders [2]. It illustrates the use of 3 of the 4 nursing care elements in the Nursing Minimum Data Set: Nursing Diagnosis, Nursing Intervention, and Nursing Outcome [3].

Furthermore, it promotes the goals of the Nursing Informatics Working Group of the International Medical Informatics Association [4] to:

- Educate/inform nurses regarding electronic health record standards
- Promulgate standards that enable representation of nursing-related measurements
- Educate/inform nurses regarding nursing-related measurement standards

Many advanced practice nursing curricula continue to teach the Problem-Oriented Medical Record and Subjective, Objective, Assessment/Diagnosis, and Plan (SOAP) format developed by Weed [5,6]. However, in early publications, Weed wrote, "…all narrative data presently in the medical record can be structured, and in the future all narrative data may be entered through series of displays, guaranteeing a thoroughness, retrievability, efficiency, and economy important to the scientific analysis of a type of datum that has hitherto been handled in a very unrigorous manner" [5, p.599].

For many years, SOAP notes documenting history, examination, and medical decision making have provided the foundation for billing using Evaluation and Management (E & M) codes developed by the Center for Medicare and Medicaid Services (CMS), appropriate to the level of care provided in a patient encounter [7]. While CMS has recently proposed consolidating encounter payment levels, the SOAP format will very likely continue to be standard practice [8]. Most SOAP components can be captured with SNOMED-CT (Systemized Nomenclature of Medicine – Clinical Terms), a reference terminology recognized by the American Nurses Association [9]; many nursing terminologies are also mapped to SNOMED-CT [10].

Structured Terminology: CCC

Beginning with SOAP and terminology required for billing (history, examination, International Classification of Diseases, Tenth Revision (ICD-10) coded diagnoses [11], Current Procedural Terminology (CPT) coded medical interventions [12], and E & M codes [7]), DNP students in the Rutgers IT course are asked to create a clinical note that adds elements of the Nursing Minimum Data Set, as captured by Clinical Care Classification (CCC) terms [13]. CCC coded nursing terminology has a four-level framework that consists of 21 care components within 4 healthcare patterns (Health Behavioral, Psychological, Functional, Physiological), 176 nursing diagnoses, 804 nursing interventions with 4 action types, and 3 potential outcomes (Improved, Stabilized, Deteriorated) for each nursing diagnosis.

Each student's note should include expected outcomes as well as a plan with five categories of interventions, as indicated for each diagnosis: Medications, Procedure/Lab/Radiology orders, Care coordination/Referrals, Patient Education and Patient self-management assignment, and Follow up/next visit. CCC nursing terminology is ideal for the documentation of expected outcomes and Care coordination/Referrals, Patient Education and Patient self-management assignment, and Follow up. Within the electronic health record, CCC captures nursing's contribution to care by mapping nursing interventions to patient outcomes.

A brief example of a student's note follows:

S. "Though I have not fallen for years, I feel unsteady at times." Meds: Carvedilol, 40 mg extended release PO qd. History of myocardial infarction 5 years ago. No surgeries. Last eye exam 2 months ago. Lives alone in a small house with stairs.

O. Vital signs: BP: 110/80, HR: 70, RR: 16, T: 98.6. Thin white woman, age 81. Alert and cooperative. Wearing flat shoes with good tread. Labwork WNL. Cardiac and Neuro exams WNL. Gait WNL. Fall risk score = 14

A. Hypertension (ICD 10: I11.0), Safety-Fall Risk (CCC: N33.6.1)

Goal: Improve Fall Risk

P. Refill medication (RxNorm: 860524) [14], Teach Environmental Safety (CCC: N42.1), Monitor/Follow up 2 weeks (CCC: N42.0)

Outcome: Fall Risk Improved

Clinical Guidelines, Quality Measures, Screening Tools

Next, students are asked to select a nursing-related clinical practice guideline that pertains to the documented scenario. Students are instructed to find a relevant guideline from their own practice or from a search of the literature or from ECRI Guidelines Trust [15]. An appropriate guideline for this patient encounter is from the American Geriatrics Society [16, 17].

After this step, students are asked to identify nursing-related electronic Clinical Quality Improvement (eCQI) measures from CMS that are related to the encounter. Advanced practice nurses are eligible clinicians for eCQI. They should revise their encounter note if necessary to include key data elements for the measure chosen. CMS Measure CMS139v7 is appropriate for this encounter [18]. It is the "percentage of patients 65 years of age and older who were screened for future fall risk during the measurement period." It matches National Quality Forum (NQF) measure number 0101.

Students are encouraged to search for a reliable screening tool related to the guideline, and appropriate for electronic documentation. In this example, one appropriate Fall Risk Screening algorithm is STEADI, that combines subjective and objective information [19]. Other appropriate screening tools include the Morse Fall Risk Scale [20] and the Get Up and Go Test [21,22]. The students have the opportunity to revise their encounter note to include data elements required for screening.

PROMIS/HCAHPS/ Health Status Indicators/MIPS

Students are asked to describe two or three Patient Reported Outcome Measures (PROMIS) that relate to the encounter and to explain why they picked these PROMIS measures [23]. In this example related to falls, a possible PROMIS measure might be a Mobility or Physical Function measure from PROMIS Item Bank v2.0. There are several brief checklists, available in both English and Spanish, that may be incorporated into electronic documentation. Students are asked to describe implications and opportunities that use of PROMIS brings to their practice. Ideally, the tool would be used during the patient assessment, with a follow up test to assess outcome after nursing intervention.

PROMIS individual patient measures are then compared with the institution-level Medicare Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) measures with which students working as RNs may be more familiar [24].

To address population health, and demonstrate the type of data freely available online, Students are asked to go to the County Health Rankings and Roadmaps website [25], select a state and a county, then compare two or three of its health status indicators with CDC-defined peer counties. Then each student describes the most likely improvement activity that nurses may perform to address the chosen indicators. Students are taught to explore the improvement activities in the CMS Quality Payment Program to find a related activity, if possible [26]. If not, students describe an alternative and feasible nursing intervention on the Merit-based Incentive Payment System (MIPS) list.

Conclusion

Overall, the student response to these assignments has been very positive, and many of their comments highlight the learning that has taken place. Some examples follow:

- The material is very different than the research & clinical courses we usually take, so it's refreshing.
- I did not care for health information technology prior to taking this course, and after taking it, I have discovered there are infinite resources and systems aimed to improve quality of healthcare; it was quite fascinating actually.

- This course has made me want to branch out and incorporate informatics into my career.
- I am so interested in IT changes happening in my workplace now, and while reading the textbook would jot down notes to discuss with my committee and hopefully we can do a hospital wide study about something related to nurse/provider perception of the new EMR, and patient engagement in the e-portal.
- Taking a course that gave insight as to how health data is entered, processed, managed, and accessed is essential to professional practice. It has helped me rethink EHRs and how something as simple as entering vital signs can later be accessed to analyze treatment protocols.

Evidence is emerging that standardized structured nursing documentation supports accurate and complete information in practice, data reuse and sharing, and improved efficiency, business analytics, and care quality [27]. It is imperative to provide a strong background in clinical informatics to graduate nursing students who will become leaders as expert clinicians and administrators. This course shows clearly how standardized nursing terminology is applied to determine the contribution of nurses in important measures related to patient outcomes.

Future plans for the course are to continually update it with clinically-relevant readings and assignments, and to incorporate evolving standards, such as the National Library of Medicine Value Set Authority [28] and assessment tools linked to Logical Observation Identifiers Names and Codes (LOINC) coding [29]. The goal is to transform documentation of nurses' work so that it will be recognized in measures related to quality of care and payment.

References

[1] American Association of Colleges of Nursing. Doctor of Nursing Practice Essentials; 2006. Available from: https://www.aacnnursing.org/DNP/DNP-Essentials

[2] American Organization of Nurse Executives. (2015). AONE Nurse Executive Competencies. Chicago, IL: Author. Available from www.aone.org

[3] Werley, HH et al. The Nursing Minimum Data Set: abstraction tool for standardized, comparable, essential data. Am J Pub Hlth. 1991; 81 (4): 421-426.

[4] International Medical Informatics Association. Nursing Informatics Working Group Standards; 2016. Available from: https://imia-medinfo.org/wp/ni-health-informatics-standards/

[5] Weed, LL. Medical records that guide and teach. New Eng J Med. 1968; 278 (11): 593–600.

[6] Weed, LL. (1968-03-21). Medical records that guide and teach. New Eng J Med. 1968; 278 (12): 652–657.

[7] Centers for Medicare & Medicaid Services. Evaluation and Management Services. 2017, Aug. Available from https://www.cms.gov/Outreach-and.../eval-mgmt-serv-guide-ICN006764.pdf

[8] Centers for Medicare & Medicaid Services. Proposed Policy, Payment, and Quality Provisions Changes to the Medicare Physician Fee Schedule for Calendar Year 2019. 2018, July. Available from https://www.cms.gov/newsroom/fact-sheets/proposed-policy-payment-and-quality-provisions-changes-medicare-physician-fee-schedule-calendar-year-3

[9] American Nurses Association. Position Statement: Inclusion of Recognized Terminologies Supporting Nursing Practice within Electronic Health Records and other Health Information Technology Solutions. Available from: https://www.aorn.org/-/media/aorn/.../posstat-endorsed-ehr-hit-terminologies.pdf

[10] SNOMED-CT. 5 Step Briefing. 2018. Available from: https://www.snomed.org/snomed-ct/five-step-briefing

[11] Centers for Disease Control and Prevention (CDC). National Center for Health Statistics: International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM). 2018. Available from https://www.cdc.gov/nchs/icd/icd10cm.htm

[12] American Medical Association. Current Procedural Terminology. 2018. Available from: https://www.ama-assn.org/practice-management/cpt-current-procedural-terminology

[13] Clinical Care Classification System. 2018. Available from: https://www.sabacare.com/

[14] National Library of Medicine. Unified Medical Language System (UMLS): RxNorm. Available from: https://www.nlm.nih.gov/research/umls/rxnorm/

[15] ECRI Institute. ECRI Guidelines Trust. 2018. Available from: https://www.ecri.org/components/GuidelinesTrust/Pages/default.aspx

[16] American Geriatrics Society and British Geriatrics Society. Summary of the Updated American Geriatrics Society/British Geriatrics Society clinical practice guideline for prevention of falls in older persons. J Am Geriatr Soc. 2011;59: 148–57.

[17] Huntzinger A. AGS releases guideline for prevention of falls in older persons. Am Fam Phys. 2010; 82(1):81-82. Available from https://www.aafp.org/afp/2010/0701/p81.html

[18] Electronic Clinical Quality Improvement (eCQI) Resource Center. Falls: Screening for future fall. 2018. Available from: https://ecqi.healthit.gov/ecqm/measures/cms139v7

[19] Stevens JA, Phelan EA. Development of STEADI: A fall prevention resource for health care providers. Health Promot Pract 2013;14:706–714.

[20] Morse, J. Enhancing the safety of hospitalization by reducing patient falls. American Journal of Infection Control. 2002; 30: 376-80

[21] Phelan EA, Mahoney JE, Voit JC, Stevens JA. Assessment and management of fall risk in primary care settings. Med Clin North Am. 2015;99(2):281-93.

[22] Podsiadlo D, Richardson S. The timed "up & go": A test of basic functional mobility for frail elderly persons. Journal of the American Geriatrics Society. 1991; 39: 142

[23] Health Measures. Patient Reported Outcome Measurement Information System (PROMIS). 2018. Available from: http://www.healthmeasures.net/explore-measurement-systems/promis

[24] Medicare.gov. Hospital Compare: Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS). Available from: https://www.medicare.gov/hospitalcompare/Data/Overview.html

[25] County Health Rankings and Roadmaps. 2018. Available from: http://www.countyhealthrankings.org/explore-health-rankings#county-select-38

[26] Quality Payment Program. 2018. Available from https://qpp.cms.gov/mips/explore-measures/improvement-activities?py=2018#measures

[27] Westra BL, Latimer GE, Matney SA, Park JI, Sensmeier J, Simpson RL, Delaney CW. A national action plan for sharable and comparable nursing data to support practice and translational research for transforming health care. JAMIA. 2015; *22*(3): 600-607

[28] National Library of Medicine. Value Set Authority Center. 2018. Available from: https://vsac.nlm.nih.gov/

[29] LOINC. Available from: https://loinc.org/

Address for correspondence

Melinda L. Jenkins, PhD, FNP.
Email: Melinda.jenkins@rutgers.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190299

Developing a Safety Case for Electronic Prescribing

Yan Jia[a], Tom Lawton[b], Sean White[c], Ibrahim Habli[a]

[a] Department of Computer Science, University of York, York UK,
[b] Bradford Royal Infirmary and Bradford Institute for Health Research, Bradford UK
[c] NHS Digital, Leeds UK

Abstract

It is now recognised that Health IT systems can bring benefits to healthcare, but they can also introduce new causes of risks that contribute to patient harm. This paper focuses on approaches to modelling and analysing potential causes of medication errors, particularly those arising from the use of Electronic Prescribing. It sets out a systematic way of analysing hazards, their causes and consequences, drawing on the expertise of a multidisciplinary team. The analysis results are used to support the development of a safety case for a large-scale Health IT system in use in three teaching hospitals. The paper shows how elements of the safety case can be updated dynamically. We show that it is valuable to use the dynamically updated elements to inform clinicians about changes in risk, and thus prompt changes in practice to mitigate the risks.

Keywords:

Electronic Prescribing, Medication Errors, Patient Safety

Introduction

Many countries have promoted Health Information Technology (HIT) as a primary means to improve the safety and efficiency of healthcare delivery. For example, the US government and European Commission have initiated policies to promote the adoption and use of HIT [1]. In the UK, several funding programmes have been launched to drive technology use within the National Health Service (NHS) [2]. One of the main types of HIT being targeted is Electronic Prescribing Systems (EPS) that involve "*the use of computing devices to enter, modify, review, and output or communicate, drug prescriptions*" [3].

There are many potential benefits associated with EPS, e.g. reduction in prescription errors as a result of fewer illegible orders, easier repeat prescriptions and better ability to track prescriptions [4]. However, the introduction of EPS also introduces new causes of risks, for example alert fatigue [5]. In order to realise the benefits of EPS, thorough risk assessment must be conducted. This should enable hospitals to evaluate whether the EPS will achieve safer care by reducing current clinical risk and also controlling the new risks associated with the introduction of the new technology.

In many engineering domains the use of a Safety Case (SC) is an established practice [6]. A SC is a structured argument, supported by evidence, that a system is acceptably safe in its context of use [7]. The SC is a risk management tool, providing rationale for accepting a system into service and enabling the relevant stakeholders to make informed decisions. The SC is particularly useful when it is produced at the same time as the system is designed and deployed, as it can help to inform design decisions and potential changes to the clinical workflows [8].

Previously, we reviewed the notion of hazard for HIT [9] and implemented a tool-supported methdology called the Safety Modelling, Assurance and Reporting Toolset (SMART) [10]. In this context, hazards are conditions or behaviour that can be observed at the level of the clinical system, and which have a clear link to patient harm, e.g. wrong medication. In this paper, we use SMART to support hazard analysis and develop a SC for an EPS deployed in three teaching hospitals. This paper focuses on three areas:

- Safety analysis: modelling causes and consequences of hazards, and hazard controls;
- Safety case: the use of the Goal Structuring Notation (GSN) [11] to present a safety argument, reflecting the hazard controls and risk acceptance;
- Through-life safety: an initial analysis of those aspects of the SC that are static, and those which can benefit from being updated dynamically.

Our approach involves proactive safety analysis, prior to deployment of EPS. In addition, we present a rationale for dynamically updating the SC, to support through-life safety. We refer to this as a Dynamic Safety Case (DSC) [12].

Methods

Setting: The study was undertaken in three teaching hospitals based on a large-scale HIT system, with a focus on the prescribing process as part of medication management. Figure 1 is an abstract model of the medication management process, of which prescribing is the first step, and the scope of our study.

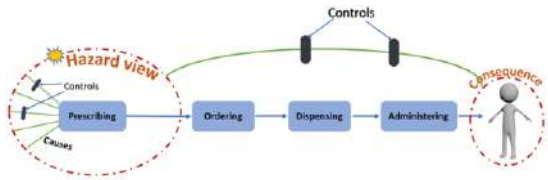


Figure 1 – Abstract Model of Medication Management

Data collection: a qualititve reseach method is used in this study. Data was collected from multi-disciplinary workshops (eight in total), which were organised to perform the hazard identification and risk analysis for the EPS. There were 3 clinical consultants, 2 nurses, 2 pharmacists, 2 safety engineers, 2 researchers and 2 systems engineers (representing the EPS supplier) involved in the workshops.

Data analysis: A combination of Software Hazard Analysis and Resolution in Design (SHARD) [13] and Systems Engineering Initiative for Patient Safety (SEIPS) [14] were used to stimulate the indentification of hazards, their causes and consequences, and the controls associated with the hazards. SHARD is suitable for identifying hazards and casuses of hazards from a software perspective, e.g. missing data, but it does not address complex interactions between software systems and humans. SEIPS provides a framework for a comprehensive consideration of work system design, which includes five elements: *person*, *tasks*, *technology and tools*, *environment* and *organisation*, and its impact on care processes and outcomes. This helps to address the interactions of software system and humans in a complex socio-technical context. Thus the two methods are complementary. The results of using SHARD and SEIPS are recorded using bow-tie diagrams. Finally, GSN was used to represent the safety arguments for EPS based on the results and findings from the workshops.

Results

Six main findings were identified, which are reported in this section.

The importance of a clear process model

To ensure the safe implementation of HIT systems, it is important to understand how HIT systems are used to support the clinical activities. As such, the first step was to define a clear process model to reflect the relationships between EPS, as a HIT system, and prescribing, as a clinical activity. During this task, one challenge arose, which is to what level of granularity the prescribing process should be modelled, e.g. a more detailed IT centered view, including "right click for more medication options", or a more abstract clinical activity level, such as "choose right medication". As a result of considering both the validitiy of the processes and the emphasis on clinical context for hazard identification, the multidisciplinary team constructed the clinical process model to describe the flow of clinical activities and decisions, linked to the specific functions in the EPS, as shown in Figure 2. This also reflects well established health informatics approaches to evaluating HIT systems, that HIT interventions should be clinically- and problem-driven rather than technology-driven [15]. As is shown in Figure 2, the "sign" clinical activity in this model is associated with the "sign medication" function in the EPS. This model provides an understanding of the interaction between a HIT system and its clinical activities which is necessary before being able to identify hazards.

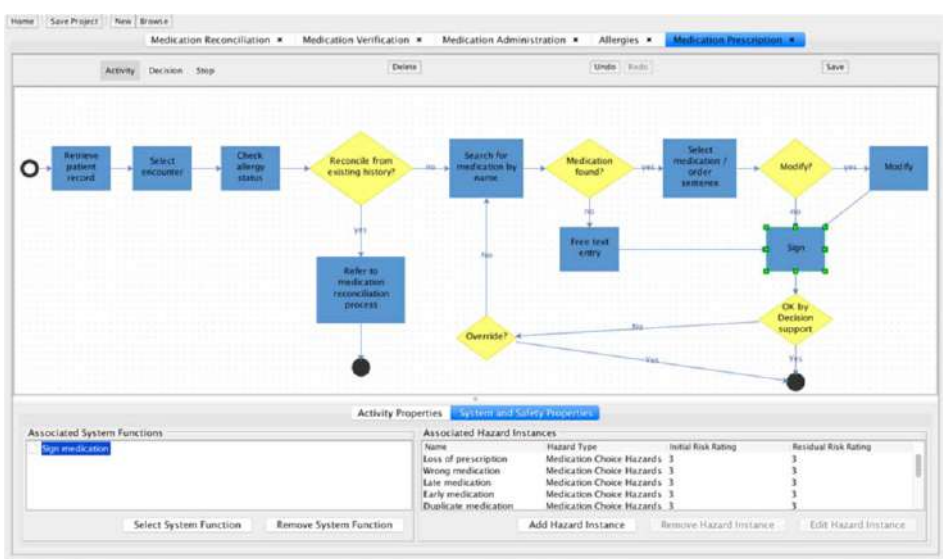
Identifying hazards in complex clincal settings

The most critical step in achieving and demonstrating the safety of HIT systems is to conduct a systematic process to identify potential hazards during the product development and then engineer them out or reduce their likelihood [16]. In order to carry out a thorough and proactive hazard identification, we first agreed on an overall hazard categorization based on the five rights in medication safety [17], producing five general categories:

1. Wrong patient selected in prescribing phase
2. Wrong medication dose prescribed
3. Wrong medication route prescribed
4. Wrong medication time prescribed
5. Wrong medication choice

Each hazard category can be refined further. We do this using the failure classes defined in SHARD, which provides a structured approach to the identification of potentially hazardous behaviour in software systems. By applying these failure classes (*omission*, *commission*, *early*, *late* and *incorrect*) to the hazard category 5 - *wrong medication choice*, defined above, we identified seven specific hazard cases, *loss of prescription*, *unintended medication*, *wrong medication*, *late medication*, *early medication*, *duplicate medication* and *adverse interaction*.

For brevity, the rest of the paper focuses on the wrong medication choice category, but the method described above would apply equally well to the other hazard categories.



Fgure 2- Prescribing Process in SMART

Modelling the causal chains (technical and non technical)

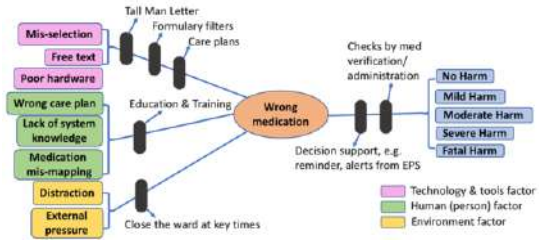


Figure 3 – Bow-Tie for Wrong Medication

In order to mitigate the hazards, it is important to identify their causes so we can identify effective controls. In a socio-technical system, it is important to recognise both technical and non - technical factors. Therefore, we used the SEIPS model to guide us to think systematically. The use of the SEIPS model both helps to identify causes of the specific hazard, and the controls. For each hazard, there are a range of controls that can counteract causes of the hazards, and others that can influence the consequence of the hazards. Figure 3 illustrates the causes and consequences and the controls related to the *Wrong Medication* hazard, arising out of the analysis workshops.

In Figure 3, we can see that there are 8 identified causes that contribute to *Wrong Medication* during prescribing (the left-hand side of the bow-tie). Among them, three causes belong to the technology factor (*technology and tools* in SEIPS), which is related to the EPS deployed in the hospital. A well designed EPS should have the ability to mitigate the causal factors, for example by integrating Tall Man Letter to make it noticeable or striking out inapplicable options when the prescribers are selecting medications. Another three causes are categorised as human factors (*person* in SEIPS). These causes can be controlled by providing education and training to clinical practitioners. The last two causes are work environment factors (*environment* in SEIPS). They are related to the local organisation and policy. These factors should be controlled by providing guidance or procedures by the local organisation, e.g. to close the ward at key times.

The classification of causal factors is useful, as recognising that different causal factors belong to particular categories helps to find the right control. It also helps to make the causes of hazards explicit and reveal the weak points of the system. For example, considering *Wrong Medication*, it seems that the organisation should also seek to reduce the pressure on clinical practitioners, e.g. by reducing the un-necessary and non-clinical related workload (*environmental* factor).

Difficulty of determining severity

Turning to the right-hand side of the bow-tie, we consider the consequences of the hazards. In the workshops, we found it very challenging to assess the potential harm concerning a particular hazard. For example, consider the *Wrong Medication* hazard; the medication type, the profile of the patients, the complexity of the clinical conditions and the state of the clinical setting would lead to different consequences.

In addition, we found it difficult to map the consequence of medication errors to severity of harm. From our literature review, we discovered that it is very hard to find information to make such connections. Studies such as [18] and [19] either just give a severity classification without a detailed description of how they mapped their patient results to the severity, or they focus on error types and their causes, but do not identify the severity of the patient outcome. Further work is needed on how to categorise the severity of harm and give concrete examples how to map the consequences (patient outcomes) to different severities. In order to illustrate this, we present examples of patient harm in Table 1 using the World Health Organization (WHO) severity classification [20]. This table is intended to be illustrative but refining and expanding it, e.g. by considering different aspects of human function such as vision and respiration, might aid in future hazard and risk assessment.

Clear arguments, the essense of reasoning

Based on insights gained in the workshops, we employed GSN to represent the safety argument for the EPS in three teaching hospitals. Figure 4 shows the top level of the safety argument and reflects the use of a Hazard Log to record information about all the hazards. A Hazard Log is a standard safety management tool for recording and tracking information about risks, used in other sectors, e.g. aerospace, but also applicable in healthcare and required by the NHS HIT standards [7; 21].

Table 1– Examples of Severity Classes

Severity	None	Mild	Moderate	Severe	Fatal
Summary	No symptoms (detected)	Symptoms short-term, requiring minimal intervention	Harm or loss of function may be long-term, requiring intervention	Life-saving intervention needed; long-term harm or permanent loss of function	Death caused or brought forward by the incident
Examples	Paracetamol given instead of priadel (loss of therapeutic effect)	Nausea, vomiting or diarrhoea from overdose of epirubicin	Digestive problems including ulcers and internal bleeding	Blindness due to prescribing a diuretic to patients with low blood pressure	Weekly dose of methotrexate given daily
	Use of antibiotics to treat viral infections (NB reduces utility of antibiotics)	Forgetting to specify maximum daily dosage for an "as required" drug	Hypotension due to overdose of lisinopril	Lung damage and possible sepsis giving oral treatment to patient with dysphagia	Ten times overdose of insulin
		Accidental sedation due to prescribing diazepam not diltiazem	Dyspepsia and ulcers from overuse of non-steroidal anti-inflammatory drugs for arthritis		Haemorrhage from incorrect use of warfarin

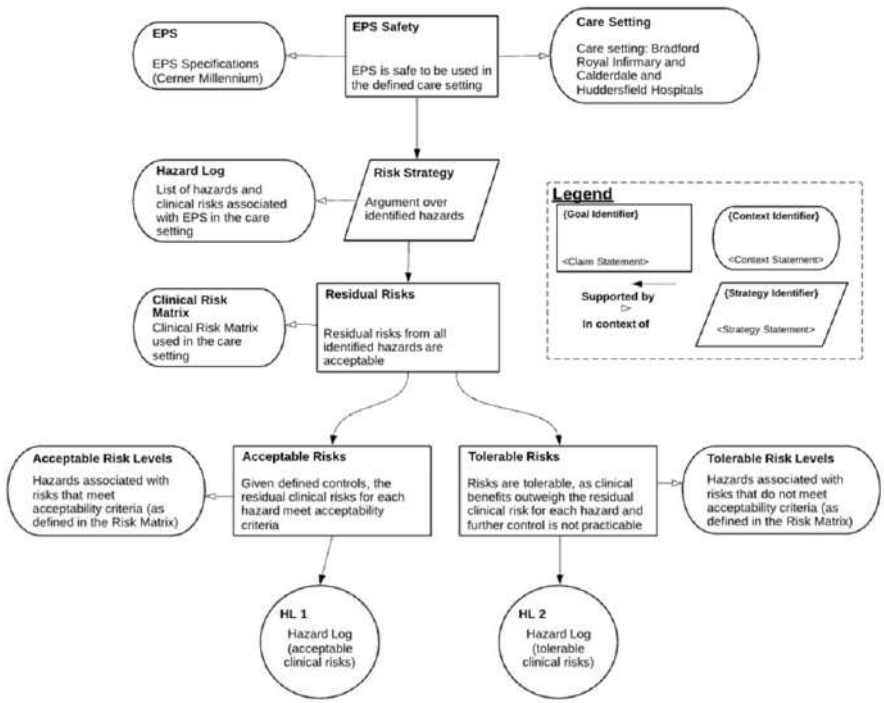


Figure 4 – Top level of the GSN for the EPS for the three Teaching Hospitals

It is often infeasible to eliminate risks, so the claim **Residual Risks** shows that the risks left after implementing the controls are acceptable and/or managed in the context of the Risk Matrix used at the hospitals; this breaks down into two sub-cases:

- Residual risks meet the acceptability criteria
- Residual risks do not meet the criteria, but can be accepted due to the clinical benefit

Evidence to support these claims can come directly from the Hazard Log.

For example, the risk of the "*Wrong Medication*" hazard has been reduced to an acceptable level, given the seven controls identified in Figure 3. The controls and risks are set out in the Hazard Log. Alternatively, we could have expanded the GSN argument further, replacing solution HL1 with more detailed sub-arguments for each hazard. For the "*Wrong Medication*" hazard, the solutions at the bottom of the argument would be evidence about the effectiveness of each of the controls.

The evidence about the controls on the left of the bow-tie is generated from analysis of the error records, which are derived from chart review, automated review of electronic records, review of incident reports, review of self-reporting, patient and staff interviews, and from direct observation [22]. For example, in order to assess whether the controls – Tall Man letters and Formulary filters – are effective, we should interrogate the error rates related to mis-selection of medication and free text.

Dynamic Safety Cases

It is well established that the safety of critical services is a dynamic property [23]. In healthcare, this dynamism is often attributed to variation in the health and care services, and their underlying systems, as well as in the environments within which they are deployed. Although variation can be seen as a negative attribute, e.g. a sign of noncompliance, increasingly more emphasis is placed on the necessity of variation to enable, and sometimes empower, people and technologies to adjust and adapt to ensure continuous safe care. The ability to adapt and adjust is a key enabler for resilience in healthcare [24]. Unlike traditional SC, which often remain static and are only updated in a reactive manner, both the justification and evidence base of the SC for a complex process such as prescribing should evolve based on real-time data that is collected, proactively, from diverse sources, particularly covering and combining clinical, organisational and technological factors.

The SC described in this paper will be extended and integrated with a new dynamic risk model and uncertainty assessment algorithms, based on Bayesian Networks, for proactively computing the confidence in, and updating the reasoning about, the safety of the medication services based on real-time data. This will be combined with a set of update rules triggering the provision of actionable suggestions to clinicians in response to changes in the services, clinical settings, the safety argument or the confidence in that argument. Thus clinicians will be able to take risk reduction action, i.e. adding new controls, based on leading indicators/precursors of problems before they develop into potential errors and patient harm.

Discussion

The paper summarises our work to date on developing a full SC for the entire HIT system in the three hospitals. Our results so far show that having a clear model of the medication process aids analysis, both in identifying hazards and their controls. In particular the clincal process models enable the workflows to be analysed at a level which is understandable by the users of the EPS, and which is also clinically meaningful.

The use of the bow-tie to model the causes and consequences of a hazard in the precribing process gives a direct visuliation of how a hazard is controlled, what can be the potential causes of this particular hazard, what consequences there can be for the patient through this hazard and what kind of controls we have to mitigate the hazard. This is a very useful basis for hazard and risk assessment. We see an opportunity to use this approach to improve analysis of medication safety, particularly assessment of risk, beyond just EPS, and this is an area for future research.

Further, we have shown how to use GSN to construct an explicit argument to justify the safety of EPS use in the context of the wider medication management system. The SC rests on evidence, some of which relates to the controls identified in the bow-tie diagram. SC are predictive, and the evidence is usually based on analysis prior to operation. A first step in making the SC representative of actual use of EPS would be to update the evidence to reflect the effectiveness of the controls, as they change over time. Ideally we would use this to show that the EPS plus controls is better (in terms of patient safety) than the previous manual system. However it is hard to obtain data that shows what happened before introduction of EPS, and hence to make such comparisons. In contrast, in future, it will be possible to see whether or not the controls are effective and the extent to which the risk is reducing over time, e.g. the frequency of over-riding alerts and error rates relating to mis-selection and free text are going down. Thus the SC can support management of risk through life, rather than just being a tool for deciding whether or not a system can be deployed. This is a key area of our future work, and should lead to development of DSC.

Conclusions

Management of risk associated with HIT systems is challenging as the technology is used in a complex socio-technical setting, and the staff using the systems are often under significant pressure, due to the volume of work, or the need to respond to patients' symptoms very quickly. We have presented our approach to assessing the safety of HIT systems, based on work on EPS in three UK hospitals, which draws on accepted practices in other domains, which we believe helps address the problems of managing safety in a healthcare setting.

The work enables causes and consequences of hazards to be analysed more directly than is possible with a purely statistical approach. Further, it enables the role and effectiveness of risk controls to be assessed. The work reported here is part of an ongoing research programme that will enable dynamic control over safety risk, by updating the evidence in the SC from analysis of operational data.

Acknowledgements

This work is partly funded by Bradford Royal Infirmary.

References

[1] European Commission, eHealth Action Plan 2012–2020. Innovative healthcare for the 21st century, in, 2012.
[2] Department of Heath, The NHS plan: a plan for investment, a plan for reform, in, 2000.
[3] J. Teich and J.J.W. Marchibroda, DC, Electronic prescribing: toward maximum value and rapid adoption. eHealth Initiative, (2004).
[4] E.L. Abramson, Causes and consequences of e-prescribing errors in community pharmacies, *Integrated pharmacy research & practice* **5** (2015), 31.
[5] T.H. Payne, EHR-related alert fatigue: minimal progress to date, but much more can be done, *BMJ Quality & Safety* (2018), bmjqs-2017-007737.
[6] M.A. Sujan, I. Habli, T.P. Kelly, S. Pozzi, and C.W. Johnson, Should healthcare providers do safety cases? Lessons from a cross-industry review of safety case practices, *Safety Science* **84** (2016), 181-189.
[7] NHS Digital, DCB0160: Clinical risk management: its Application in the Deployment and Use of health IT Systems, in, 2018.
[8] R. Bloomfield, N. Chozos, D. Embrey, J. Henderson, T. Kelly, and F. Koornneef, Using safety cases in industry and healthcare, *Health Foundation, London* (2012).
[9] I. Habli, S. White, M. Sujan, S. Harrison, and M. Ugarte, What is the safety case for health IT? A study of assurance practices in England, *Safety Science* **110** (2018), 324-335.
[10] I. Habli, Safety modelling, assurance and reporting toolset (SMART) for digital health, *Informatics for Health, Manchester, UK* (2017).
[11] The GSN Working Group Online, GSN Standard, in, 2018.
[12] E. Denney, G. Pai, and I. Habli, Dynamic safety cases for through-life safety assurance, in: *Proceedings of the 37th International Conference on Software Engineering-Volume 2*, IEEE Press, 2015, pp. 587-590.
[13] D.J. Pumfrey, *The principled design of computer system safety analyses*, University of York, 1999.
[14] P. Carayon, A.S. Hundt, B. Karsh, A.P. Gurses, C. Alvarado, M. Smith, and P.F. Brennan, Work system design for patient safety: the SEIPS model, *BMJ Quality & Safety* **15** (2006), i50-i58.
[15] H. Fraser, E. Coiera, and D. Wong, Safety of patient-facing digital symptom checkers, *The Lancet* (2018).
[16] E.H. Shortliffe and M.J. Sepúlveda, Clinical Decision Support in the Era of Artificial Intelligence, *JAMA* (2018).
[17] M. Grissinger, The five rights: a destination without a map, *Pharmacy Therapeutics* **35** (2010), 542.
[18] J. Kelly and D. Wright, Medicine administration errors and their severity in secondary care older persons' ward: a multi‐centre observational study, *Journal of clinical nursing* **21** (2012), 1806-1815.
[19] K.L. James, D. Barlow, R. McArtney, S. Hiom, D. Roberts, and C. Whittlesea, Incidence, type and causes of dispensing errors: a review of the literature, *International journal of pharmacy practice* **17** (2009), 9-30.
[20] W.H. Organization, International Classification for Patient Safety Key Concepts and Preferred Terms, in, WHO Geneva, 2009.
[21] NHS Digital, DCB0129: Clinical Risk Management: its Application in the Manufacture of Health IT Systems, in, 2018.
[22] M.P. Tully and B.D. Franklin, *Safety in medication use*, CRC Press, 2015.
[23] C. Vincent and R. Amalberti, Safety in healthcare is a moving target, in, BMJ Publishing Group Ltd, 2015.
[24] R.L. Wears, E. Hollnagel, and J. Braithwaite, *Resilient health care, volume 2: the resilience of everyday clinical work*, Ashgate Publishing, Ltd., 2015.

Address for correspondence

Yan Jia, Department of Computer Science, University of York, York UK. email: yj914@york.ac.uk; tel. 01904325492

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190300

Towards an Open-Source Oncology Electronic Medical Records System for Low-Resource Settings: Development of Chemotherapy Management in OpenMRS

Johnblack K. Kabukye[a,b], Alexis M. Casaceli[c], Ellen Ball[d], Mario De Armas[c], Ronald Cornet[b]

[a] *Uganda Cancer Institute, P.O. Box 3935 Kampala, Uganda*
[b] *Department of Medical Informatics, Amsterdam UMC - Location AMC, Amsterdam, Netherlands*
[c] *IBM, Armonk, NY, United States*
[d] *Partners In Health, Boston, MA, United States*

Abstract

Cancer is a major public health challenge in low and middle income countries (LMICs). In this paper, we describe work in progress to develop functionality within an open-source electronic medical records system to support safe and standardized cancer care in low-resource settings. We engaged cancer care providers from LMICs to elicit and prioritize requirement, and following a rapid application development approach we developed chemotherapy prescription and documentation functionality within OpenMRS.

Keywords:

Developing Country; Electronic Health Records; Neoplasms

Introduction

Cancer is a major cause of mortality and morbidity worldwide. As of 2018, there are an estimated 14 million cancer cases and 9.6 million cancer deaths worldwide, according to the WHO [1]. Cancer causes more deaths than HIV/AIDS, Malaria, and Tuberculosis combined.

The cancer burden is heaviest in low and middle income countries (LMICs) where it is fast replacing infectious diseases as a leading public health challenge. Sixty percent of cancer cases and over 70% of cancer deaths are occurring in LMICs. Yet cancer has not been receiving the needed attention and priority, e.g., in the form of funding. Cancer receives only 2% of the funding that is put towards other diseases in LMICs [2].

Leveraging health information technologies such as electronic medical record (EMR) systems in oncology in LMICs can contribute to the improvement of cancer care outcomes through improved care coordination, standardization of treatment, facilitation of guideline adherence and computerized clinical decision support (CDS) [3]. Cancer is a complex family of diseases and cancer care is also complex. There are hundreds of cancer types, each with different risk factors, prevention strategies, presentation, staging, investigations and treatment approaches, including chemotherapy, radiotherapy, and surgery [4]. In addition, health care systems in LMICs are varied, often with limited availability of cancer care specialists and services such as radiotherapy and pathology [2; 5]. Medical errors, e.g., in chemotherapy administration, are common in such complex care environments [6] and the lack of standardization of care affects outcomes. Lack of good quality data for research and planning is also a major challenge for cancer care in LMICs [3; 7] that EMRs could help address.

Unfortunately, adoption of EMRs in oncology in LMICs remains low. For example, the majority of EMRs in Africa are implemented within HIV/AIDS care programs, and only 27% in non-HIV related programs [8]. There have been few reports on EMR implementations in cancer, e.g., in Rwanda [7], but these implementation projects are still in their infancy.

The complexity of oncology is a factor in this low EMR adoption because oncology is thought to require complex and expensive software systems [9] which are not affordable in LMICs, especially considering the lack of prioritization and funding in LMICs.

Open source software systems offer a potential solution to the prohibitive cost, in addition to fostering collaboration and enhancing the use of interoperability standards [10; 11]. Open source systems are arguably more secure since the source code is reviewed by many independent members of the open source community. OpenMRS (https://openmrs.org/) is one such open source EMR platform that is widely used in LMICs particularly in the management of HIV/AIDS, tuberculosis, malaria, and maternal and child health [12]. It is a robust, modular and scalable web-based platform built in Java. It uses MySQL databases and has an extensive data dictionary called CIEL (Columbia International eHealth Laboratory) which maps to ontologies such as ICD-10, SNOMED CT, RxNORM, and LOINC. OpenMRS also uses interoperability standards such as HL7 and FHIR. It is freely available for download and has a large community of active developers and implementers in over 40 countries.

In this paper, we describe work in progress to leverage the OpenMRS platform to develop a fully functional oncology EMR that meets the needs and requirements of cancer care in LMICs [13]. The initial work focused on developing a chemotherapy management module to facilitating efficient, standardized, and safe chemotherapy ordering, documentation, and tracking by doctors and nurses.

Methods

Stakeholder Engagement and Requirements Elicitation

Stakeholders were purposively identified to include general doctors, oncologists, nurses, and others working in cancer care, especially those in LMICs or those familiar with this context. Organizations represented included Partners in Health (PIH), Uganda Cancer Institute, Dana-Farber Cancer Center and UNC Project Malawi. We also involved software engineers from the OpenMRS community (mostly from PIH), the OpenMRS leadership team, and IBM Health Corps. These regularly engaged in a variety of meetings and communication interactions: via email, video conferences, OpenMRS Talk (http://talk.openmrs.org), OpenMRS design forums, and face-to-face meetings in Boston, MA, to elicit and prioritize requirements, discuss design considerations and review sketches and prototypes. The technical members of the team also visited and engaged with those on the ground at the PIH's cancer treatment facility in Mirebalais, Haiti to review the current workflows, paper forms, and treatment protocols, and to interact with target end-users.

Development Process

We followed the rapid application development approach. The software engineers and OpenMRS implementers were co-located in Boston, MA over approximately one month. We reviewed the existing code (freely available in GitHub), the OpenMRS data model and concept dictionary to determine what needed to be modified and what could be reused. Then we developed mock-ups and prototypes for the user interfaces implementing the chemotherapy protocols. These were iteratively reviewed by target end users and changes made immediately. Several modules that are already available in the OpenMRS platform were used e.g., the metadata module for adding new concepts, and the OpenMRS API, a REST web service used to implement the technical workflows.

Results

Requirements

Several requirements for comprehensive oncology support were elicited, including oncology-specific documentation (e.g., tumor description), exporting data to cancer registries, computerized clinical decision support in terms of diagnosis, reminders, etc.

Chemotherapy management was prioritized of all the requirements because it is a high-risk part of the oncology workflow – complex drug combinations, tightly controlled doses, severe toxicity and need for several safety checks [14]. Currently, chemotherapy ordering at the represented cancer centers is paper-based and relies on verbal instructions between care team members.

For the chemotherapy management module, requirements, include the following: (i) provide an oncology-specific patient overview with patient identification, cancer diagnosis, cancer journey summary (e.g., On CHOP cycle 3 of 6); (ii) show list of appropriate chemotherapy regimens and pre and post medication for the cancer diagnosis, (iii) automatically calculate body surface area (BSA) from height and weight, and (iv) calculate doses and generate a prescription when the doctor chooses a regimen (v) allow the prescriber to modify the dose by a given percentage but provide a reason for the modification (vi) track cumulative doses of certain drugs (e.g., Anthracyclines) and notify the prescriber when ceiling doses are reached.

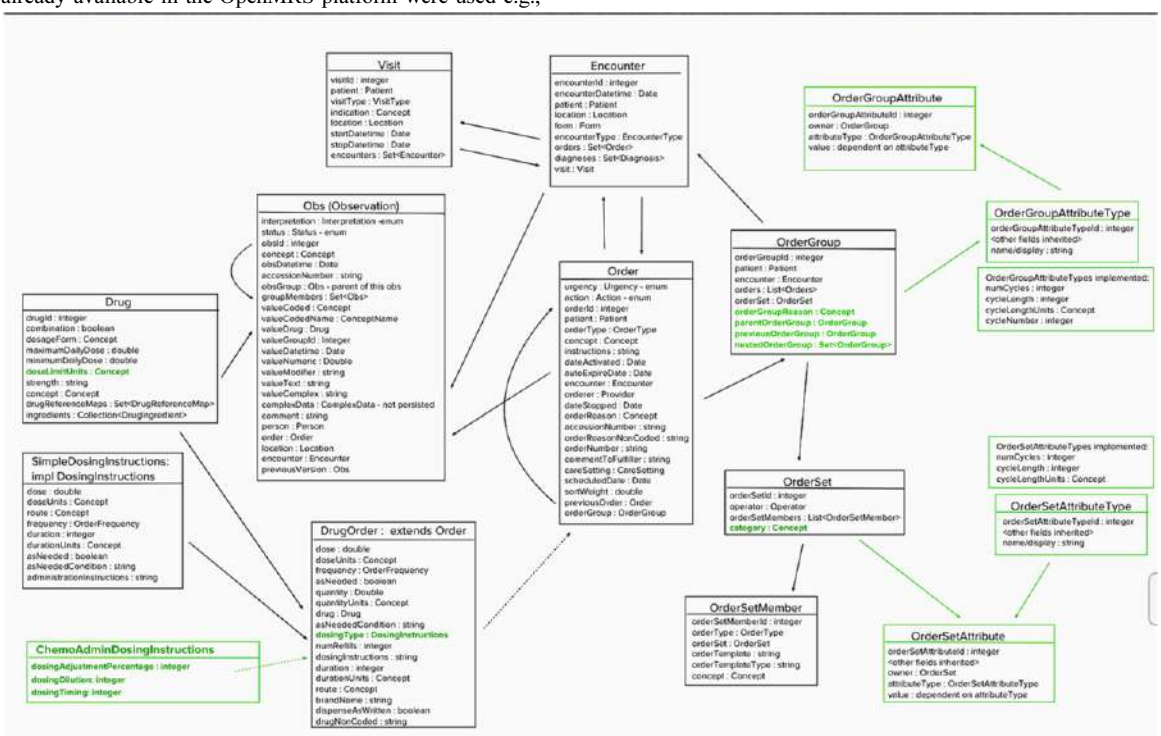


Figure 1 – Data model with chemotherapy related concepts added (in green).

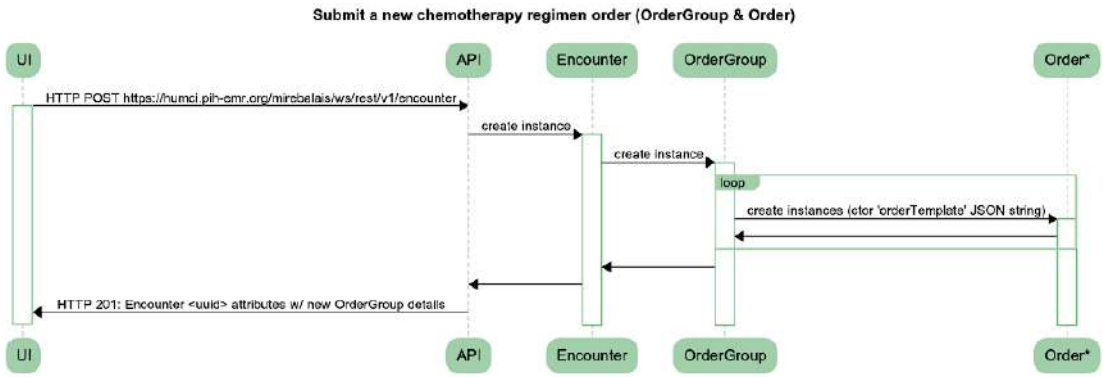


Figure 2 – OpenMRS API Invocation to create a chemotherapy order

Changes in OpenMRS

The oncology module required introduction of new concepts that are currently not available in the OpenMRS data model and concept dictionary. **Figure 1** shows the data model with chemotherapy-related concepts that were added during this project colored in green. These concepts are related to drug combinations and the cyclic nature of chemotherapy, as well as the requirement to track number of cycles in regimens, cycle numbers (e.g., 3 of 6), length of cycles, administration group (e.g., chemotherapy, pre medication, post medication), and maximum lifetime dose.

The *OrderGroup* (attribute of *Order*) and *Drug* objects that already exist in OpenMRS were modified to enable capture of information about chemotherapy regimen, such as *cycleNumber*, *cycleLength*, *cycleLengthUnits*, and *DoseLimitUnits*. *ChemoAdminDosingInstructions* was also added to The *DrugOrder* object (extension of *Order* object) to capture special instructions that are not captured by *SimpleDosingInstructions*. Classification into pre-medication, chemotherapy, and post-medication leverages the *OrderType* concept field in *Order*.

We also used the CIEL dictionary currently in PIH's version of OpenMRS, which contains oncology concepts, constrained to the workflow and context, e.g., mappings to ICD-O, the oncology subset of ICD. While each facility can add concepts to CIEL, a team of terminology experts at Open Concept Lab (https://openconceptlab.org/) coordinates and distributes CIEL, and encourages all local additions to be added centrally to ensure consistence.

After addition of the concepts, we hardcoded templates for twelve chemotherapy protocols currently used at Mirebalais for management of lymphomas, breast cancer and other solid tumors. The templates were made as metadata files in the YAML syntax, describing each chemotherapy regimen's medications and cycle details. YAML is data serialization language similar to XML but with a minimal syntax that focuses on data interchange rather than document mark-up. We chose to make the regimen templates in this format to facilitate ease of future addition of new chemotherapy treatments to the initial set, and to ease update of the delivered chemotherapy regimens by trained doctors, clinicians, and non-software programmers.

To allow changes to the chemotherapy regimen templates, or add new ones easily by different cancer centers, we developed the Yet Another Automated Regimen (YAAR) management tool, which can process the YAML files and manage the OpenMRS REST API back-end messaging. The tool was implemented in Python and is meant to be extended as requirements evolve. As a future goal, the tool should be extended to support additional treatment types (new schemas) beyond the chemotherapy regimens targeted in this project.

User Interface Design

We used the React UI framework (Material UI) to develop the front end. This first displays the oncology-specific patient summary as per the requirements above, with an option select "Prescribe chemotherapy". The prescriber is then taken to a screen that shows chemotherapy regimens and pre and post medication, their schedule, doses and information such as dose modifications or cumulative dose. With a few clicks the prescriber selects the appropriate regimen, schedule and doses, adds notes if necessary and submits the order.

Figure 2 shows a sample OpenMRS API invocation to create a new *OrderGroup* and *Orders* during chemotherapy ordering, and **Figure 3** shows some of the screens for the chemotherapy prescription interface. We also made a nurses' interface that allows the nurse to document the treatment after it has been administered as well as details such as patient reaction to medication or toxicity, track cycle numbers, etc.

All the code, documentation, workflow analysis and mock-ups, as well as links to other project resources are available in GitHub (https://github.com/openmrs/openmrs-module-oncology).

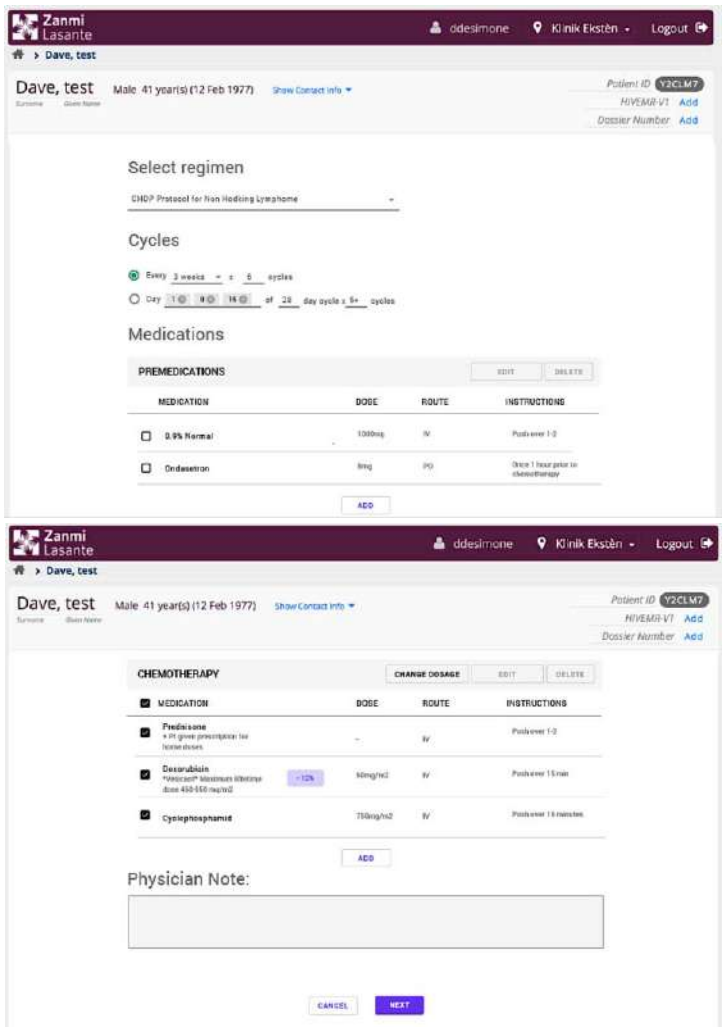


Figure 3 – Examples of user interface screens for chemotherapy prescription workflow

Discussion

In this paper, we describe collaborative efforts towards developing a fully functional oncology EMR for low resource settings based on the open source EMR OpenMRS. We began by focusing on chemotherapy management, and currently we have a functional prototype based on the workflow and clinical services at the PIH Mirebalais cancer center. We implemented twelve chemotherapy regimens and documentation requirements, but the product is scalable using the chemotherapy template authoring (YAAR) tool which allows new regimens to be added so that the system can be used in different contexts or at other cancer care centers. The prototype has been evaluated as part of the development process by oncologists and other cancer care providers from Uganda, Rwanda, Malawi, Haiti, and the US but the module is yet to be implemented at point of care. The twelve chemotherapy protocol templates that were configured were based on the care process at Mirebalais cancer center and that is where initial implementation is planned, but with minimal changes the module can be used elsewhere. Evaluations and further development for implementation in Uganda, Rwanda and Malawi are ongoing, and feedback from these clinical applications will inform future iterations of development.

This project also provides a basis for development of clinical decision support such as automatic body surface area calculation, automatic dose calculation, safety checks and alerts for dose (including lifetime cumulative doses) and schedules basing on protocols integrated in the system, as well as better clinical data capture and (re)use for forecasting and planning.

Future features will include the addition of pharmacist workflows to complete the chemotherapy administration workflow, as well as other forms for complete oncology documentation e.g., diagnosis, staging and treatment, follow up and interfacing with cancer registries.

Some challenges faced during the project included difficulty in gaining consensus on data model changes to support cancer specific concepts, and a complex code base that was hard for new developers to learn. Current challenges include dedicated software development teams and project management to advance the work quickly.

Conclusions

This project demonstrates the potential for collaborative development of affordable open source systems for addressing the cancer epidemic in LMICs. The system will facilitate a safe and standardized prescription of chemotherapy as well as proper documentation and monitoring. This is important because chemotherapy is very toxic, and cancer care is complex and prone to medical errors, which can be detrimental for the patient.

In this project, we also developed tools which can be used by other cancer care providers to implement their own protocols. Using the rapid application development approach allowed us to accomplish this project in a relatively short time. Constant interaction with the end users at all stages of the project was crucial for a proper understanding of the requirements and the oncology domain in general. Co-location of the technical team was advantageous since some members of the team were new to OpenMRS.

Acknowledgements

This work was financed by IBM Foundation in form of man-hours of user experience/interface (UX/UI) designers, business analysts and software developers as well as funding for design workshops and meetings. We appreciate the contributions of the following:

IBM team: Alexis Casaceli, Rebecca Dahlman, Mario De Armas, Ian Lewis, Elma Maksumic, Pravesh Patel, Oleksandr Sabov, Joel C. Mangan.

Partners In Health contributors: Ellen Ball, Ruth DeMuse, David DeSimone, Dayo Fadelu, Mark Goodrich, Lauren Greenberg, Michael Seaton.

OpenMRS Community: Darius Jazayeri, Andrew Kanter, Burke Mamlin.

References

[1] World Health Organization, WHO | Cancer, *WHO*, (2018). Available at: http://www.who.int/mediacentre/factsheets/fs297/en/. (Accessed: 13th April 2018)

[2] A. Moten, D. Schafer, and M. Ferrari, Redefining global health priorities: Improving cancer care in developing settings, *Journal of global health* **4** (2014).

[3] S.B. Clauser, E.H. Wagner, E.J.A. Bowles, L. Tuzzio, and S.M. Greene, Improving modern cancer care through information technology, *American journal of preventive medicine* **40** (2011), S198-S207.

[4] American Society of Clinical Oncology (ASCO) & National Cancer Institute (NCI), *Clinical Ocology Requirements for the EHR (CORE) White Paper* (2009).

[5] T.D. Dye, S. Bogale, C. Hobden, Y. Tilahun, V. Hechter, T. Deressa, M. Bizé, and A. Reeler, Complex care systems in developing countries: breast cancer patient navigation in Ethiopia, *Cancer: Interdisciplinary International Journal of the American Cancer Society* **116** (2010), 577-585.

[6] K.A. Elsaid, S. Garguilo, and C.M. Collins, Chemotherapy e-prescribing: opportunities and challenges, *Integrated pharmacy research & practice* **4** (2015), 39.

[7] N.M. Tapela, T. Mpunga, N. Karema, I. Nzayisenga, T. Fadelu, F.R. Uwizeye, L.R. Hirschhorn, M.A. Muhimpundu, J.P. Balinda, and C. Amoroso, Implementation science for global oncology: the imperative to evaluate the safety and efficacy of cancer care delivery, *Journal of Clinical Oncology* **34** (2015), 43-52.

[8] M.O. Akanbi, A.N. Ocheke, P.A. Agaba, C.A. Daniyam, E.I. Agaba, E.N. Okeke, and C.O. Ukoli, Use of electronic health records in sub-Saharan Africa: progress and challenges, *Journal of Medicine in the Tropics* **14** (2012), 1.

[9] J. Warner and E. Hochberg, Where is the EHR in Oncology?, *Journal of the National Comprehensive Cancer Network* **10** (2012), 584-588.

[10] P.C. Webster, The rise of open-source electronic health records, *The lancet* **377** (2011), 1641-1642.

[11] S. Sfakianakis, C. Chronaki, F. Chiarugi, F. Conforti, and D. Katehakis, Reflections on the role of open source in health information system interoperability, *Yearbook of medical informatics* **16** (2007), 51-61.

[12] N.A. Mohammed-Rajput, D.C. Smith, B. Mamlin, P. Biondich, B.N. Doebbeling, and O.M.C. Investigators, OpenMRS, a global medical records system collaborative: factors influencing successful implementation, in: *AMIA annual symposium proceedings*, American Medical Informatics Association, 2011, p. 960.

[13] J.K. Kabukye, S. Koch, R. Cornet, J. Orem, and M. Hagglund, User Requirements for an Electronic Medical Records System for Oncology in Developing Countries: A Case Study of Uganda, in: *AMIA Annual Symposium Proceedings*, American Medical Informatics Association, 2017, p. 1004.

[14] L.N. Shulman, R.S. Miller, E.P. Ambinder, P.P. Yu, and J.V. Cox, Principles of safe practice using an oncology EHR system for chemotherapy ordering, preparation, and administration, part 1 of 2, *Journal of oncology practice* **4** (2008), 203.

Correspondence

Johnblack K. Kabukye, *MBChB, MSc*
email: j.k.kabukye@amc.uva.nl

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190301

Scoring Patient Fall Reports Using Quality Rubric and Machine Learning

Melanie Klock[a], Hong Kang[b], Yang Gong[b]

[a] *College of Natural Sciences, the University of Texas at Austin, Austin, Texas, USA,*
[b] *School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA*

Abstract

Patient falls, a subcategory of patient safety events, cause further harm and anxiety to patients in healthcare systems. Patient fall reports are a valuable resource to identify safety issues that demand further attention. Still, the main challenge for patient fall reports is the lack of quality and detail in writing. A method of evaluating patient fall reports would help us better understand the root causes of falls and prevent their recurrence to improve patient safety. Employing the Agency for Healthcare and Quality rubric for assessing the quality of fall reports, we compared three different machine-learning models and identified the most effective method for scoring fall reports using AHRQ's rubric. The results of this study are intended to be applicable in healthcare facilities to score reports during reporting for reporters to improve report quality. The ultimate goal is to increase learning from fall reports for better prevention of patient falls.

Keywords:
Patient Safety; Falls; Machine Learning

Introduction

Measuring and analyzing incidents that could result in unintended harm towards patients, or patient safety events (PSE), are essential in forming strategies to improve patient safety. A PSE is gathered through a process called event reporting, which happens on a voluntary and limited basis [1]. However, the reports often vary in level of detail, bias, and ambiguity. Event reporting has proven to be successful in increasing safety in other hazard-prone industries, such as aviation [2]. Even with a plethora of technology and data, healthcare lags behind in collecting and applying PSE reports. In the 1999 report, *To Err Is Human*, the Institute of Medicine places incident reporting as one of the most urgent areas of healthcare to improve [3]. Clearly, managing an efficient PSE reporting system is crucial for clinicians and hospitals to learn from past events and form solutions to improve patient safety.

Patient falls, the most common type of PSE, are defined by the Agency for Healthcare Research and Quality (AHRQ) as "an unplanned descent to the floor with or without injury to the patient." There are between 700,000 and 1,000,000 patient falls each year in the United States [4]. A patient fall can have various consequences such as additional pain or injury. It may also lead to additional hospital stay, on average about 6.3 more days, costing the patient about $14,000 [5]. Even if there is no injury, the fall can cause anxiety or embarrassment to the patient. Insofar, the healthcare community has heavily emphasized report collection, but poor processing of safety reports remains a key challenge in PSE reporting [6]. Current reporting systems are suffering from low-quality reports, which has greatly damaged the learning value of previous events. Poor quality reports are not only unhelpful but also potentially dangerous, as missing information could cause hospitals to make incorrect decisions about the patient. High quality reports, on the other hand, provide hospitals with accurate and complete data to prevent future patient falls and increase safety. Quality improvement remains a problem because reporters have no way of receiving immediate feedback on their reports. Therefore, a tool of rapidly assessing reporting quality is necessary. If we could alert reporters of their report quality during the writing process, they could improve their description of the event and submit higher quality reports.

Although AHRQ has developed a rubric to identify if patient fall reports contain enough information for effective analysis [7], no real-time scoring system currently exists. In our previous study, we proposed a PSE similarity searching model that provides the reporter with suggestions on solutions and prevention, increasing the motivation of the reporter to include as many details as possible [8]. We also designed a user-centered, self-learning reporting system that assembles a large PSE knowledge database [9]. Therefore, a tool that scores reports would further upgrade our system by providing numeric scores to reporters and further increasing their motivation to improve the quality of their report. Scoring would also build a higher-quality database with more effective self-learning.

Instead of traditional human review, machine learning is a plausible alternative to score patient fall reports in terms of efficiency. Natural language processing (NLP) allows computers to understand human language and has been used to extract medication information from electronic health records (EHR), a similar data type to PSE [10]. Two classic machine learning methods, support vector machine (SVM) and random forest, have been successful in classifying hospital stay records for disease prediction [11]. Recurrent neural networks (RNN), a type of deep learning model, have recently been a popular approach for more complex text classification tasks [12]. In the healthcare field, RNN have displayed higher accuracy than other machine learning methods when learning diagnoses from EHR [13]. We hypothesize that machine learning, especially deep learning, will be highly effective in scoring patient fall reports. Different machine learning algorithms can have varying levels of success depending on the initial dataset and the classifier's ability to gain an effective generalization without overfitting [14]. Comparison of multiple algorithms is therefore necessary for finding the most accurate scoring method specific to the fall reports dataset.

This study compared the performance of three different machine learning models and identified the most promising for the classification and scoring of patient fall reports. We first processed the fall reports from a Patient Safety Organization (PSO) institute using NLP techniques and labeled them using

the AHRQ rubric. Then classifiers based on SVM, random forest, and RNN models were evaluated by measuring the model performance on AHRQ rubric labeling. We finally proposed a report quality scoring system and presented both high- and low-quality report examples. The results of this study could be used to give reporters immediate feedback and encourage improvement of reports before submission. Ultimately, better quality patient fall reports would help healthcare providers identify what causes falls and how to prevent them.

Methods

Construct a Database of Labeled Patient Fall Reports

Patient fall reports are commonly written as first-person narratives describing the event and any details that the reporter deems pertinent to the fall. The reporter is usually a nurse but can be a staff, volunteer, or any type of person witnessing the event. Therefore, fall reports can vary between reporters of different professions or backgrounds. The AHRQ toolkit for preventing falls in hospitals [7] includes a specialized rubric for assessing patient fall report quality, which has the potential to measure the fall reports from 13 perspectives (Table 1).

Table 1 – 13 Perspectives Measuring Patient Fall Reports

Labels
1. Witnessed/Not witnessed
2. Outcome of investigations recorded
3. Type of injury
4. Buzzer/bell available within reach before fall
5. Fall from Bed
6. Whether bedrails were in use
7. Floor wet/dry/talcum powder
8. Footwear
9. Walking aid in use/in reach
10. Mental state
11. First fall this admission or repeat fall
12. Days since admission
13. Medication affecting risk of falls

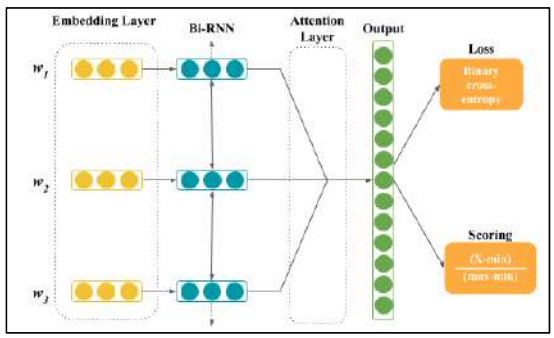
We used 2,046 de-identified fall reports from a PSO institute to extract and classify the unstructured portions of the reports using the categories shown in Table 1. We scored each report based on if each category was included ("1") or not included ("0") in the report. Two reviewers labeled the reports. First, the reviewers labeled the same 100 reports and resolved any disagreements in scoring through group discussion. Since the divergence rate was minimal at 0.15%, the reviewers then labeled the remaining reports separately.

Construct Machine Learning Models

The three models were constructed to perform multi-label, multi-class classification, where multiple labels could be assigned to one report. The labels were then used to calculate a final score reflecting the report quality. SVM and random forest classifiers are both suited for multi-label, multi-class classification and were constructed using the labeled patient fall reports. Each report was transformed into vectors and weighted using term frequency-inverse document frequency (TF-IDF). The SVM utilized the one-versus-rest strategy to achieve multi-class capabilities.

The tokenization of RNN model was conducted by using the Keras preprocessing module to transform report into a sequence of numbers corresponding to the frequency of each word. To simplify the input for the model, the sequences were padded to the length of the longest report (N=436). The model contained an embedding layer, hidden layers, an attention layer, and an output layer. Figure 1 shows the structure of the RNN.

Figure 1– Flowchart Representation of RNN



The first layer of the model was the word embedding layer, where each tokenized word was expanded into a 100-dimensional vector from the GloVe pre-trained library [15]. The GloVe library is trained on six billion tokens from public domains, helping the model become more generalized even if training data is limited. The word embeddings were set as trainable to fit the patient safety scenarios. After the embedding layer, the model contained a bi-directional long short-term memory (Bi-LSTM) layer with a hidden size of 128. An attention layer was placed on the top of Bi-LSTM layer to assign and merge weights of word. To prevent overfitting, we applied dropout with a dropout rate of 0.5 within the Bi-LSTM layer. The output layer used the sigmoid activation function and produced a list of 13 labels. The RNN was compiled on the Keras API with TensorFlow backend. Hyperparameters were tuned on the patient fall data to optimize performance.

Train and Test the Model

Due to the limited size and imbalance of our dataset, the structures of the models were first tested on single-label Yelp reviews from the Yelp Dataset Challenge in 2018 [16]. The Yelp Dataset Challenge is a public dataset of business reviews released by Yelp for academic and educational purposes and is especially useful to test RNN models [17]. A subset of 3,500 positive and 3,500 negative labeled reviews was chosen to be larger and more balanced than the patient fall dataset. After testing on the Yelp dataset, the models were optimized on the patient fall dataset. Both datasets were shuffled and split into training (70%), validation (10%) and testing (20%) sets. All three models were evaluated using the performance metrics accuracy and F_1 score.

Scoring the Reports

For each report, the probability estimates for the 13 categories were taken from the highest performing model among the three and summed. The quality score of report i was calculated using the scaling function:

$$Score_i = \frac{(E_i - E_{min})}{(E_{max} - E_{min})}$$

where E_i was the sum of probability estimates of report i, and E_{min} and E_{max} were the minimum and maximum sums among all the reports.

Results

Distribution of Labels

We labeled 2,046 reports based on the 13 categories from the AHRQ rubric on patient fall reports. There was high variation between frequencies of positive labels in the reports. The most frequent label was "Type of injury," seen in 1,290 reports. The least frequent label was "Days since admission," seen in only five reports. The average number of positive labels each report contained was 2.189 labels out of 13, with a standard deviation of 1.386.

Model Performance

The three models were first tested on the Yelp dataset, then the patient fall dataset. Table 2 shows the average accuracy and F_1 score for each model on the two datasets. The RNN model showed the highest performance on the patient fall dataset, but the SVM showed the highest performance on the Yelp dataset.

Table 2 – Yelp Performance

Method	Yelp Dataset		Patient Fall Dataset	
	Accuracy	F_1	*Accuracy*	F_1
SVM	87.1%	0.870	89.9%	0.648
Random forest	73.4%	0.714	87.6%	0.279
RNN	90.0%	0.889	86.3%	0.397

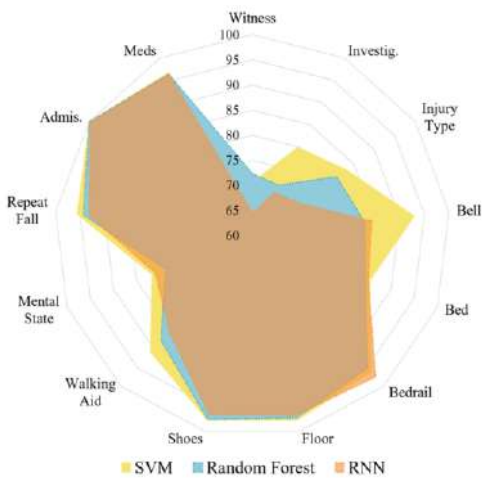


Figure 2 – Accuracy for Patient Fall Reports

Figure 2 shows the accuracy scores for each rubric category. For all rubric categories, SVM showed equal or better accuracy scores compared to random forest or RNN. Figure 3 shows the F_1 scores for each rubric category. For Figure 3, some rubric categories have no F_1 score if the model never assigned positive labels for that category. The SVM model showed much larger F_1 scores for all rubric categories for the patient fall dataset.

In contrast to the Yelp dataset, the SVM model was the highest performer in both accuracy and F_1 score for the patient fall dataset. SVM was also most successful in giving positive labels for minority classes during testing. The *r*andom forest and RNN models gave no positive labels for multiple rubric categories (i.e., Admission, Footwear, Bedrails, Medication, Floor, Repeat Fall). The SVM model gave no positive labels for only one rubric category (Admission).

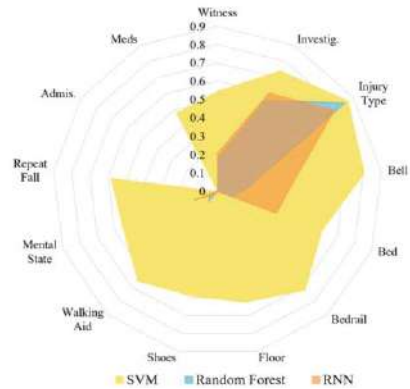


Figure 3 – F_1 Score for Patient Fall Reports

High and Low Quality Report Examples

To display the effectiveness of the SVM model, a high quality and low quality report from the database are presented in Table 3. The SVM model's scoring reflects the difference in the level of detail between the two reports.

Table 3 – High- and Low-Quality Report Examples

Report 1	**Label:** 1, 2, 3, 10, 13	**Score:** 0.843
"Patient received meds at 2100, including Ambien 5 mg for rest/sleep. Patient found on floor at 2120, confused with slurred speech. Staff assisted patient to stand, was able to bear full weight without difficulty or pain. No contusions or lacerations seen upon visual exam. Able to walk without pain or difficulty. Vs 99.8 100 22 B/P was 130/50. No change in LOC. Asleep in bed at 2200."		
Report 2	**Label**: 4	**Score**: 0.100
"Patient was up in chair; upon entering room found patient sitting upright on floor directly in front of chair with legs extended reading the TV guide. When patient questioned if she fell states no I slid out of chair. When questioned patient why she was trying to get up states I was going to throw some trash away. Call light was beside patient at time of incident."		

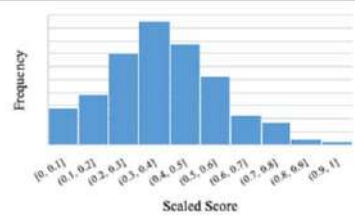


Figure 4 – Distribution of Scaled Scores

Scoring Distribution

Figure 4 shows the distribution of scaled scores made by the SVM model.

Discussion

Implement a Scoring System for Fall Reports

Patient fall reports are an important resource to understand the causes and prevention of patient fall events. Unfortunately, event reporting systems suffer from a lack of high-quality reports. Without high-quality reports, learning from past fall events becomes speculative and inefficient. This study identifies the most promising machine learning method to score patient fall reports. Providing immediate scores to reporters would increase the quality of patient fall reports and ultimately advance the learning and prevention of patient fall events.

Importance of Scoring Patient Fall Reports

Regardless of specific model, this study shows that machine learning is effective towards scoring patient fall reports. This discovery has implications for the healthcare environment because it addresses the need for higher-quality patient fall reports and provides a first step towards improvement. The majority of reports scored between 0.3 and 0.4, implicating that most patient fall reports are lacking in detail and quality. However, it is crucial that reports contain such information that could aid in post-fall patient care or discussions on prevention strategies. The two reports shown in the results section (Table 3) display how level of detail affects how patient falls can be studied. The high score report gives information on medication, where certain medications can lessen the patient's spatial awareness, induce dizziness, and therefore cause fall events. Healthcare providers could learn from this report and assess if that medication is dangerous for patients. The lower score report contains some details, like the patient's motivation for getting up, but not the specific ones included in the rubric that would be helpful for analyzing the cause of the fall. The two reports were of the same approximate length but of very different quality, signifying that a high-quality report does not have to take longer than a low-quality report to write. In the case of the low-quality report, real-time scoring would have helped the reporter become aware of missing information and improve the report quality before submission. Any future analysis of the report would become more meaningful and actually aid in patient fall prevention, whereas the original low-quality report would have been less usable.

Challenges of Scoring Patient Fall Reports

While machine learning has been utilized for similar data types like EHR, its application for patient fall reports presents several challenges. The effectiveness of machine learning depends heavily on the quality of the training dataset. If the training dataset lacks key data, the model's performance will be severely limited. For patient fall reports, both the size and quality of the current dataset need improvement for machine learning to reach its full capabilities. Specifically, deep learning requires a larger dataset of information to learn, since its neural structure is more complex than classic machine learning models. The current patient fall dataset has an imbalanced representation of the 13 rubric categories, with minority classes marked positive in under 100 reports. Machine learning models have difficulty classifying minority classes because the model rarely sees positive labels during training. The model will be applied to classify reports negatively for minority classes not based on learning, but on probability.

As a result, we used both accuracy and F_1 scores as performance metrics. Since the dataset contained more negative labels than positive labels, a model could increase its accuracy by selecting the negative label without learning anything. Therefore, accuracy cannot be the defining metric, but rather, the F_1 score better represents performance based on false positives and negatives. We also applied scaling to the model's scoring function to account for minority classes. With the average report having only two positive labels out of 13, many scores would be around or under 0.2 without scaling. Scaling allows reporters to understand the relative differences between high and low quality reports, whereas unscaled scores would have been too small to differentiate. The scaled scores will change dynamically as the database size increases, especially when higher quality reports are added. This makes the scoring system not just a tool for measuring report quality, but also for helping reporters improve quality relative to other reports.

Evaluate Models

We tested all three models on Yelp reviews before patient fall reports, since the Yelp dataset was larger and balanced. The Yelp dataset represents an expected improvement of the patient fall dataset in size and quality, which is especially needed for deep learning models. By using both datasets, we were able to see which model performs best currently and which model has the greatest potential for the future of patient fall reporting. The SVM model was identified as the best method for the current patient fall database. The SVM model was also more successful in learning positive labels for minority classes. Due to a lack of positive labels, Random forest and RNN labeled the entire testing dataset negative for multiple categories. Those categories correlated with a lower frequency of positive labels. SVM had this problem with only one category, "Days since admission", which had only five positive labels out of 2,046 reports.

However, for the Yelp dataset, the RNN model displayed the highest performance among the three models. This suggests that if the collection and quality of patient fall reports increases in the future, the RNN could become a better method than the SVM. Two specific features of the RNN support this statement. Firstly, the RNN might be better equipped for multi-label, multi-class classification. Deep learning has been shown to be better at capturing label dependencies for multi-label, multi-class classification on large-scale data [18]. Secondly, the RNN model used pre-trained word embeddings, while the SVM used TF-IDF vectors. TF-IDF weights use only the words in the training data, while pre-trained word embeddings represents all words from a larger vocabulary. If new words appear in the testing data, pre-trained word embeddings would be better at handling those words than TF-IDF vectors.

Limitations

The voluntary reports are usually submitted by frontline staff rather than patient safety experts. Therefore, it is inevitable that reports contain incomplete or biased information about patient fall events due to differences in background and training of reporters. Another limitation is the quantity and balance of the dataset used to train the models. All three models will improve with a larger and more balanced training dataset, especially the RNN. An improvement in the balance of the dataset starts with circulating knowledge on the AHRQ rubric and confirming that reporters understand the rubric when writing. It is important to focus on this aspect of event reporting in addition to the collection of reports. Our scoring system could aid in this process, since reporters would need to understand the AHRQ rubric so they could include the details necessary to improve their score.

Future Work

Any of the three models will be embedded to our reporting systems [9] and could be consolidated into existing computerized reporting systems in hospitals and other healthcare facilities. This would allow staff in hospitals to fill out reports per regular protocol and receive a score based on the quality of their reports before submission. Because of the model's scoring function, healthcare providers would see an increase in report detail and could better address what causes falls in their facility. In the future, an interface could be added to inform the reporter why the report received a certain score and specific areas of the rubric to improve upon. This would also teach the reporter the standards for patient fall reports and cause reporters to submit better reports in the future. More generally, while we focused on patient fall events in this study, there is still the challenge of improving the quality of other PSE reports. The models proposed in this study could act as a starting point for applying machine learning to other PSEs and improving the PSE reporting system in general.

Conclusions

We proposed a machine learning approach for classifying and scoring patient fall reports based on the AHRQ rubric. We identified the most promising machine learning method for both the current dataset and for the future of patient fall reporting. This study addresses the need for better-quality patient fall reports by creating a way that reporters can get immediate feedback on the quality of their reports. The improvement of patient fall reports would help us better understand causes the falls and develop strategies of fall prevention.

Acknowledgements

This project is supported by UTHealth Innovation for Cancer Prevention Research Training Program Post-Doctoral Fellowship and Summer Fellowship (Cancer Prevention and Research Institute of Texas grant #RP160015), and by the Agency for Healthcare Research & Quality (1R01HS022895). The Titan Xp GPU used for this research was donated by the NVIDIA Corporation.

References

[1] Reporting Patient Safety Events, in, Agency for Healthcare Research and Quality, July 2018, https://psnet.ahrq.gov/ primers/primer/13/reporting-patient-safety-events.

[2] C. Macrae, The problem with incident reporting, BMJ *Qual Saf* **25** (2016), 71-75.

[3] L.T. Kohn, J.M. Corrigan, and M.S. Donaldson, To Err is Human: Building a Safer Health System, *Institute of Medicine Committee on Quality of Healthcare in America* (2000).

[4] Preventing Falls in Hospitals, in, Agency for Healthcare Research and Quality, Rockville, MD, January 2013, https://www.ahrq.gov/professionals/systems/hospital/fallpxtoolkit/index.html

[5] C.A. Wong, A.J. Recktenwald, M.L. Jones, B.M. Waterman, M.L. Bollini, and W. Dunagan, The cost of serious fall-related injuries at three Midwestern hospitals, *The Joint Commission Journal on Quality and Patient Safety* **37** (2011), 81-87.

[6] I. Mitchell, A. Schuster, K. Smith, P. Pronovost, and A. Wu, Patient safety incident reporting: a qualitative study of thoughts and perceptions of experts 15 years after 'To Err is Human, *BMJ Qual Saf* **25** (2016), 71-75.

[7] Tool 5A: Information to Include in Incident Reports, Agency for Healthcare Research and Quality, Rockville, MD, January 2013, https://www.ahrq.gov/professionals/systems /hospital/fallpxtoolkit/fallpxtk-tool5a.html.

[8] H. Kang and Y. Gong, A Novel Schema to Enhance Data Quality of Patient Safety Event Reports, *AMIA Annual Symposium Proceedings* (2017), 1840-1849.

[9] H. Kang and Y. Gong, Design of a User-Centered Voluntary Reporting System for Patient Safety Events, *Studies in Health Technology and Informatics* **245** (2017), 733-737.

[10] H. Xu, S. Stenner, S. Doan, K. Johnson, L. Waitman, and J. Denny, MedEx: a medication information extraction system for clinical narratives, *Journal of the American Medical Informatics Association* **17** (2010), 19-24.

[11] M. Khalilia, S. Chakraborty, and M. Popescu, Predicting disease risks from highly imbalanced data using random forest, *BMC Medical Informatics and Decision Making* **11** (2011).

[12] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature* **521** (2015), 436-444.

[13] F. Ma, R. Chitta, J. Zhou, et al, Dipole: Diagnosis Prediction in Healthcare via Attention-based Bidirectional Recurrent Neural Networks, *Proceedings of 23rd ACM SIGKDD* (2017), 1903-1911.

[14] A. Khan, B. Baharudin, L. H. Lee, and K. Khan, A Review of Machine Learning Algorithms for Text-Documents Classification, *Journal of Advances in Information Technology* **1** (2010), 4-20.

[15] J. Pennington, R. Socher, and C. D. Manning, GloVe: Global Vectors for Word Representation (2014).

[16] Yelp Dataset Challenge, 2018, from https://www.yelp.com/ dataset/challenge.

[17] D. Tang, B. Qin, and T. Liu, Document Modeling with Gated Recurrent Neural Network for Sentiment Classification, *Proceedings of the 52nd Annual Meeting of ACL* **1** (2015), 1555-1565.

[18] J. Nam, J. Kim, E. L. Mencia, I. Gurevych, and J. Fürnkranz, Large-scale multi-label text classification-revisiting neural networks, *ECML PKDD* (2014), 437-452.

Address for Correspondence

Yang Gong, MD, PhD

7000 Fannin St Suite 600, Houston, TX 77030, USA

Yang.Gong@uth.tmc.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190302

A Collaborative Decision Support Tool for Managing Chronic Conditions

Nadin Kökciyan[a], Martin Chapman[b], Panagiotis Balatsoukas[b], Isabel Sassoon[a], Kai Essers[a], Mark Ashworth[b], Vasa Curcin[a,b], Sanjay Modgil[a], Simon Parsons[a], Elizabeth I Sklar[a]

[a] Department of Informatics, King's College London, London, United Kingdom,
[b] Department of Population Health Sciences, King's College London, London, United Kingdom

Abstract

This paper describes work to assess the feasibility of using a decision support tool to help patients with chronic conditions, specifically stroke, manage their condition in collaboration with their carers and the health care professionals who are looking after them. The system contains several novel elements: the integration of data from commercial wellness sensors, electronic health records and clinical guidelines; the use of computational argumentation to track the source of data and to resolve conflicts and make recommendations; and argumentation-based dialogue to support interaction with patients. The proposed approach is implemented as an application that can run on smart devices (e.g. tablets). The users have personalised dashboards where they can visualise their health data and interact with a conversational chatbot that provides further explanations about their overall well-being.

Keywords:

Decision Support Systems, Clinical; Artificial Intelligence; User-Computer Interface

Introduction

The aim of the CONSULT (Collaborative mObile decisioN Support for managing mULtiple morbidiTies) project is to explore the feasibility of employing a collaborative decision-support tool to help patients suffering from chronic diseases to self-manage their treatment plans. By 'collaborative,' we mean that the patient, carers, and medical professionals work as a team to decide on the best treatment plan for the patient. To establish feasibility, we are developing a system, called CONSULT, which connects a patient to wireless sensors that are gathering data about them, provides real-time updates of data from their electronic health record (EHR), and provides recommendations and explanations based on clinical guidelines. Separately, CONSULT provides a connection for a patient's general practitioner (GP) to have access to information being gathered about the patient. Feasibility is being assessed both at a technical level, in terms of whether it is possible to construct a working system that connects these disparate elements together, and at a usability level, in terms of whether all the parties find the system to be helpful. We are not, at this stage, assessing clinical benefit.

The CONSULT system exhibits the following novel features: (1) integration of data from commercial wellness sensors, patient's EHR, inputs from health care professionals (HCPs), and treatment guidelines to produce an adaptive care plan customised to the patient's current circumstances; (2) application of *computational argumentation* to structure and track the data from these disparate sources and identify reinforcing and conflicting information; and (3) interaction with patients via *argumentation-based dialogue* to ensure understanding of the information gathered in (1) and to address, and potentially resolve any conflicts found in (2). The users have personalised dashboards where they can visualise their health data and interact with the system.

Methods

Motivation

The CONSULT project was motivated by evidence that engaging patients in the self-management of chronic conditions can be beneficial to their well-being [1-3]. Clinical colleagues suggested that a suitable target population for a study in self-management would be stroke survivors, with the aim of the study being secondary stroke prevention. This suggestion was supported by an initial focus group with patients/carers and HCPs. In this focus group, stroke survivors reported a desire to receive additional support, beyond what can be provided by HCPs. In addition, HCPs at the focus group were keen to leverage new technology to help monitor patients.

The CONSULT System Overview

An overview of the CONSULT system architecture is shown in Figure 1. There are seven primary building blocks that make up the system: (a) patient input sources, including biometric data gathered by commercial wellness sensors and a patient's EHR; (b) user interfaces, including an interface for patients, as well as an interface for HCPs and a third interface for system administrators (orange blocks); (c) web-facing servers for gathering input data and supporting user interfaces (red blocks); (d) internal databases for storing raw data (blue blocks); (e) data mining processes, aggregating raw data and extracting natural language from arguments (yellow blocks); (f) aggregated data, including the output of the data mining and argumentation processes (pink blocks), and (g) a computational argumentation engine and associated sub-components, including inputs of computational guidelines, and drug interactions (green blocks). In the following sections, we describe the multiple information sources shown in Figure 1 in more detail, before detailing how these information sources are combined for the purpose of decision support. We then describe how a user can engage with the system.

Step (1.b) information provided by NICE guideline CG127:
"*Offer step 1 antihypertensive treatment with a CCB to people aged over 55 years and to black people of African or Caribbean family origin of any age.*
If a CCB is not suitable, for example because of oedema or intolerance, or if there is evidence of heart failure or a high risk of heart failure, offer a thiazide-like diuretic."

Formal Representation in the Argumentation Engine:
$(age \geq 55) \vee black\text{-}african \vee black\text{-}caribbean \rightarrow offer(C, S_1, d)$
$\neg tolerated(C) \rightarrow \neg offer(C, S_1, d) \wedge offer(D, S_1, d)$
$oedema \vee heart\text{-}failure \vee highrisk\text{-}heart\text{-}failure \rightarrow offer(D, S_1, d)$

Figure 2 - Example Representation of Step 1 of NICE Guideline CG127

Information Sources

Clinical Guidelines

Clinical guidelines are documents that help HCPs and patients to decide on appropriate treatments. However, guidelines are mostly expressed in natural language. Clinical guidelines should be represented in a structured way in order to automate the reasoning process in decision support systems. We represent domain-specific knowledge (e.g. the hypertension domain) using a logical language. We also use existing semantic representations of guideline information (e.g. drug interactions [4]) in the reasoning process.

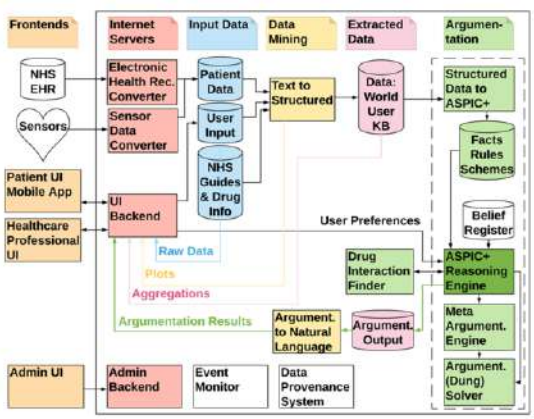


Figure 1 - Abstract CONSULT System Architecture

Domain-Specific Representation: In order for CONSULT to reason about treatment plans, we represent knowledge in the hypertension domain using first order logic [5]. For example, in Figure 2, we represent part of the hypertension treatment guideline CG127 published by NICE [6]. The information provided for this particular step is represented in terms of logical rules. Patient characteristics, such as ethnicity or experienced side effects, could change the treatment plan. Hence, we take a similar approach to represent the relations between possible treatment plans and side-effects formally.

Use of External Ontologies: Patients typically deal with multiple comorbidities, which makes the reasoning process difficult, as conflicts among recommendations may arise. This requires the representation of: (1) recommendations that can be made for each condition (as described previously) and (2) potential interactions among such recommendations. Zamborlini et al [4] introduce a semantic approach to detect interactions among recommendations by combining multiple guidelines. For CONSULT, we packaged Zamborlini's work as a web service, allowing us to create additional computational forms of guidelines in the semantic format required for the identification of interactions. This information is then used as an additional data source—*Drug Interaction Finder* in Figure 1—for the argumentation engine. Specifically, new guidelines are authored as *quad triples*, added to a triplestore, and then processed by a logic-based reasoner in order to identify interactions of interest to the argumentation engine. The web service allows the argumentation engine to interrogate various stages of the interaction identification reasoning process, such as which recommended drugs have been identified as being in conflict.

Electronic Health Records

CONSULT's next information source is a patient's EHR specifically their demographic information, blood pressure history, medication history and details of long term conditions. To integrate with a given patient's EHR, we rely on an endpoint provided by the vendor responsible for storing that patient's record. Typically, this endpoint is a local application programming interface (API) provided by each individual installation of the vendor's EHR software within a GP clinic. Therefore, CONSULT leverages the *data node connector* approach proposed in the TRANSFoRm project [7], by installing software within each GP practice that is designed to access this local API and transmit extracted EHR data to the rest of the CONSULT system for reasoning. Other models of EHR data access leveraged include direct collection of data on multiple patients from an external server provided by the vendor and simulated local API access using the N3/HSCN network. In any route to access the data, issues of governance are handled directly with the GP practice and the patient.

As each EHR vendor uses different formats to structure and code their data, once EHR data has been collected for a given patient, it needs to be standardized in order to enable the rest of the CONSULT system to be agnostic to the vendors from which the EHR data is derived. We choose Fast Healthcare Interoperability Resources (FHIR) standard as this format [8], and structurally transform each EHR to FHIR using a semi-automated matching and mapping process, while relying on services such as METMAPS [9] for code transformations, specifically to SNOMED which is used as part of the FHIR standard. Once transformed, data is inserted into a FHIR server, enabling the CONSULT system to operate as an *application* under the SMART-ON-FHIR paradigm [10].

Wireless Sensors

Our final data source is a patient's biometric health measures, extracted from a range of wearable devices. We primarily aim to acquire data on a patient's current blood pressure, pulse rate, activity and heart rate, since these are the most important measures for stroke patients. However, we do not ignore additional data that is also sent by the devices alongside this primary data (e.g. sleep quality). We employ a range of devices for this purpose, ranging from devices that are readily accessible to consumers (e.g. wrist worn devices) to more specialist medical devices (e.g. dedicated heart rate

monitoring devices), where the former is advantageous as it increases the accessibility of the system, and the latter potentially offers greater accuracy and frequency of readings. A separate study involving some of the authors evaluates the quality of sensor data produced by consumer medical devices.

In general, we aim to make the remainder of the CONSULT system as agnostic as possible to the hardware from which the readings originate. To do this, we build integration components for each wearable vendor's API—typically a remote REST endpoint, or a simpler data store—and then, upon the receipt of new sensor readings, convert this data from its vendor specific format to FHIR, which is also designed to represent and store patient health measures. This information can then be accessed by the rest of the system in the same manner as the EHR data via our FHIR server.

Integrating Data for Decision Support

The different patient data sources available to CONSULT (Figure 1) are exploited and combined to present an up-to-date view of the patient's situation and to support any reasoning in support of recommendations made. The data is merged and transformed to monitor how the patient's latest readings compare to the patient's baseline. In cases where there is deviation, relevant alerts notify GP and patient accordingly.

The argumentation engine in CONSULT is the component where recommendations are made. This engine generates possible arguments and conflicts between them (e.g. conflicts in treatment guidelines that arise in the management of multiple morbidities). It also computes treatment options to follow by providing further explanations for each option. We use argument schemes [11] and critical questions to automatically construct arguments and identify conflicts between them. Argument schemes are semi-formal representations of the structures of common types of arguments. They explain the construction a particular argument. The argumentation reasoning engine, based on ASPIC+ [12], uses the received data to instantiate argument schemes and attack schemes in a metalevel argumentation framework [13; 14], and it constructs arguments and attacks to support any self management or treatment query related to the patient [5]. Such queries are submitted through the personalised dashboards of CONSULT, where argumentation results are shared in a human-understandable way, and stakeholders can interact with CONSULT to understand the decisions made by the argumentation engine.

User Interface and Interaction Scenarios

The interface for the CONSULT system has two main components: (1) a dashboard component that visualises longitudinal personal health data, presents tailored health recommendations to patients for disease self-management, and communicates the effect of different treatment and preventive interventions on their health risk (e.g. the risk of experiencing another stroke); (2) a conversational agent (chatbot) component the role of which is to provide patients with alerts and explanations about their health state (e.g. an increase in systolic blood pressure beyond the ideal reference range), to present treatment recommendations for self-managing their condition (e.g. which over-the-counter painkiller is the most indicated for reducing their backache given their current blood pressure levels, treatment plan and clinical guidelines), or to allow users to perform, in an interactive environment, simple health information-seeking tasks (e.g. in the form of acquiring links to authoritative health literature and websites about a specific medication, measurement, or condition).

In the remainder of the paper, we refer to the following example of Martin, a 60-year-old male who has suffered a stroke, and who is using the CONSULT system and a variety of wellness sensors to monitor his own health. Martin and his GP interact through the CONSULT system.

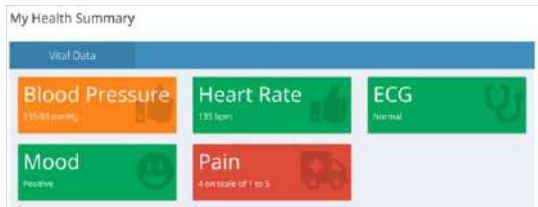


Figure 3 - The Dashboard (Overview) for an Android Tablet

Data Summary

The dashboard component of CONSULT contains an overview and a preview interface personalised for patients/carers and HCPs [15]. The overview interface, depicted in Figure 3, displays a summary of the most recent measurements for all types of personal health data collected from the patient (e.g. blood pressure, heart rate, sleep activity, pain, stress, mood). For the representation of this information in the dashboard, we use a tile-based design where each tile provides information about each health data type. Moreover, we use colour-coding to make clear immediately to the user when a specific measurement is outside the normal range [16]. For example, for blood pressure, the colour green was used to indicate that the latest measurement was within the specified normal range, the colour orange indicated pre-hypertension levels —as depicted in Figure 3— while red required attention. By selecting a tile from the overview interface, the user can access longitudinal health data about the specific measurement in the preview interface. A typical preview interface provides users the opportunity to view their data at specific time intervals (e.g. hourly, daily, weekly, monthly or yearly), as averages or all raw measurements (using line graphs for averages, and scatter plots for raw measurements). Also, for each time interval, the user is provided with additional descriptive information, such as the average, minimum and maximum value. In addition to personal health data, the dashboard provides users with the opportunity to use a risk calculator and to visualise (using *cates* plots) the effect of specific treatment and life-style interventions on their current risk of experiencing another stroke [17].

To improve the legibility and readability of content, these features were used: clean typeface (Arial), large default font size (12<), high contrast between characters and background (plain background and use of balanced colour saturation and luminance for text and graphs), writing that corresponds to a US sixth-grade reading level (equal to year 7 for England). In terms of accessibility, to improve access for colourblind users, both colour and symbols/labels were used to show that a value is within or outside normal range, or the selection of few well-contrasting colours instead of multiple colours.

Treatment Recommender

If Martin's blood pressure is not under control, then as part of the consultation with his GP, there needs to be a decision as to how to modify his treatment. The CONSULT system can support this by presenting the GP with relevant, summarised and up-to-date patient data, along with recommendations for possible treatments that consider these data, the patient's EHR, their preferences, and clinical guidelines. The treatment recommendations are generated through the argumentation

reasoning engine. A more in-depth description of the approach CONSULT takes when reasoning with the different possible and at time conflicting treatment options is described in [5].

Interacting with the CONSULT ChatBot

The conversational component of the CONSULT system serves two main purposes. The first purpose is to provide a patient the opportunity to seek evidence-based advice about a health problem. For example, Martin may be suffering from back pain and CONSULT, using a chatbot, can advise him on what he can do, as depicted in Figure 4. The chatbot is aware of the patient's latest wellness sensor readings, the data in their her – so will not recommend a treatment that caused side effects, for example – and clinical recommendations. These interactions are supported by argumentation-based dialogue [18; 19]. Additionally, the patient may have questions regarding their current treatment plan (e.g. why a particular medication has been prescribed). All the explanations are generated by the argumentation engine and displayed on the personalised dashboard. The second purpose of the conversational component is to alert the patient to an irregularity in one or more of their recent measurements and initiate a conversation, the purpose of which is to find a possible solution—suggesting the patient to review her blood pressure readings—or to advise the patient to contact a HCP.

Figure 4 - The interaction between Martin and the Chatbot

Results

We are currently in the process of evaluating the CONSULT system. Our intention was to design an application broad enough to accommodate the needs of people suffering from different chronic conditions, we have been focused on the context of stroke patients [1]. Based on our previous experience with this group of patients and the strong links to the South London Stroke Register (SLSR), we identified patients with different characteristics in terms of risk factors, comorbidity or demographic groups. The focus groups also involved co-design activities, following a design thinking approach [20], that resulted in user-generated versions of how information should be displayed. We will conduct additional user studies to evaluate the usability of the proposed system to answer the following questions: (1) do the stakeholders of the CONSULT system (patients and HCPs) use the system? (2) do they think that it is useful to assist them in making decisions? and (3) do they like interacting with the system through the chatbot? Our initial focus groups have already allowed us to explore the answers to some of these questions [1].

Additionally, in [5], we have shown that argumentation is promising in explaining decisions to help HCPs and patients choose a treatment plan together. The use of argument and attack schemes specialised for the medical domain will be a next step to consider to generate better explanations [14]. The CONSULT system collects data from multiple information sources; as such, it is important to represent the interactions between these sources. One way of doing this is the use of commitments, which help the system to automatically decide what information source to trust and reason accordingly [21].

Discussion

Several works combine patient and clinical data collected from a variety of sources for the purposes of decision support, however many do not consider the number and variety of sources that are integrated by the CONSULT system. Systems that use a subset of the sources found in CONSULT include those that rely predominantly on sensor data, such as the system proposed by Groat et al. [22], which integrates data collected from glucose and exercise monitors to determine if patients are adhering to self-reported self-management behaviours. Others rely predominantly on a patient's medical history, such as the system proposed by Evans et al. [23], which aims to identify COPD in patients through a range of offline sources, including EHR data and echocardiograms, and the system proposed by Mosa et al. [24], which aims to identify patients at risk of CINV by mining EHR data.

With respect to reasoning with data sources, various works focus on developing argumentation-based systems for clinical decision support. Atkinson et al propose the DRAMA agent to reason about patient treatment [25]. This is similar to our setting as it deals with treatment recommendations and makes use of argument schemes to construct arguments; however, each argument is associated with a value and the argumentation results change according to the prioritisation of such values. In *arguEIRA* [26], the authors make use of argumentation to detect and label anomalies in patient's reactions to treatments in the intensive care unit. In *Carrel+* [27], the goal is to develop an argumentation based tool where agents conduct a deliberation dialogue to decide on the organ transplant viability. In contrast to these works, we consider data coming from multiple information sources rather than a centralised database. CONSULT also goes further than previous work in the degree to which it allows stakeholders to interact with the system to understand the argumentation reasoning better. CONSULT also provides dashboards to help patients self-manage their conditions and so provides a health monitoring facility that goes beyond the previous work cited.

Conclusions

CONSULT is one of the few systems to take a collaborative approach to the management of chronic disease. It is also the first decision support system to make recommendations by combining multiple information sources, data science techniques, agreement technologies and an interactive chatbot. We implement our proposed approach as a mobile application for Android tablets to help stroke patients and HCPs make decisions during the treatment process.

Future work focuses on the full evaluation of the system as a feasibility study for the deployment of this kind of technology. The main questions that need to be resolved are the technical feasibility of successfully operating a system that connects patient, sensors, EHR and GP together in real-time, and the

feasibility of having patients, carers and medical professionals use the system without finding it burdensome. We believe that the principles behind CONSULT can be adapted to help with a number of chronic diseases, and we hope to explore this hypothesis in future work.

Acknowledgements

CONSULT is funded by the UK Engineering & Physical Sciences Research Council (EPSRC) under grant #EP/P010105/1.

References

[1] T. Porat, N. Kökciyan, I. Sassoon, A.P. Young, M. Chapman, M. Ashworth, S. Modgil, S. Parsons, E. Sklar, and V. Curcin, Stakeholders' views on a collaborative decision support system to promote multimorbidity self-management: barriers, facilitators and design implications, in: *American Medical Informatics Association (AMIA) Annual Symposium*, 2018.

[2] C.M. Boyd and M. Fortin, Future of multimorbidity research: how should understanding of multimorbidity inform health system design?, *Public Health Reviews* **32** (2010), 451.

[3] P. Fraccaro, M.A. Casteleiro, J. Ainsworth, and I. Buchan, Adoption of clinical decision support in multimorbidity: a systematic review, *JMIR medical informatics* **3** (2015).

[4] V. Zamborlini, M.d. Silveira, C. Pruski, A.t. Teije, E. Geleijn, M.v.d. Leeden, M. Stuiver, and F.v. Harmelen, Analyzing interactions on combining multiple clinical guidelines, *Artificial Intelligence in Medicine* **81** (2017), 78-93.

[5] N. Kökciyan, I. Sassoon, A.P. Young, M. Chapman, T. Porat, M. Ashworth, V. Curcin, S. Modgil, S. Parsons, and E. Sklar, Towards an Argumentation System for Supporting Patients in Self-Managing their Chronic Conditions, in: *Proceedings of the AAAI Joint Workshop on Health Intelligence (W3PHIAI 2018)*, 2018.

[6] National Institute for Health and Care Excellence (NICE), Hypertension in adults: diagnosis and management, in: *https://www.nice.org.uk/guidance/cg127*, [accessed on March 28, 2019], 2011.

[7] B.C. Delaney, V. Curcin, A. Andreasson, T.N. Arvanitis, H. Bastiaens, D. Corrigan, J.F. Ethier, O. Kostopoulou, W. Kuchinke, M. McGilchrist, P. van Royen, and P. Wagner, Translational Medicine and Patient Safety in Europe: TRANSFoRm-Architecture for the Learning Health System in Europe, *BioMed Research International* **2015** (2015), 1-8.

[8] D. Bender and K. Sartipi, HL7 FHIR: An Agile and RESTful approach to healthcare information exchange, in: *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, 2013, pp. 326-331.

[9] S.F. Liang, J.F. Ethier, T. Porat, A. Tapuria, B.C. Delaney, and V. Curcin, MeTMapS-Medical terminology mapping system, *CEUR Workshop Proceedings* **1692** (2016).

[10] J.C. Mandel, D.A. Kreda, K.D. Mandl, I.S. Kohane, and R.B. Ramoni, SMART on FHIR: a standards-based, interoperable apps platform for electronic health records, *Journal of the American Medical Informatics Association* **23** (2016), 899-908.

[11] D. Walton, C. Reed, and F. Macagno, *Argumentation Schemes*, Cambridge University Press, 2008.

[12] S. Modgil and H. Prakken, A general account of argumentation with preferences, *Artificial Intelligence* **195** (2013), 361-397.

[13] S. Modgil and T. Bench-Capon, Metalevel Argumentation, *Journal of Logic and Computation* **21** (2011), 959-1003.

[14] A.P. Young, N. Kökciyan, I. Sassoon, S. Modgil, and S. Parsons, Instantiating Metalevel Argumentation Frameworks, in: *Proceedings of the 7th International Conference on Computational Models of Argument (COMMA)*, 2018, pp. 97-108.

[15] B. Brown, P. Balatsoukas, R. Williams, M. Sperrin, and I. Buchan, Interface design recommendations for computerised clinical audit and feedback: hybrid usability evidence from a research-led system, *International Journal of Medical Informatics* **94** (2016), 191-206.

[16] B.J. Zikmund-Fisher, A.M. Scherer, H.O. Witteman, J.B. Solomon, N.L. Exe, B.A. Tarini, and A. Fagerlin, Graphics help patients distinguish between urgent and non-urgent deviations in laboratory test results, *Journal of the American Medical Informatics Association* **24** (2016), 520-528.

[17] A. Fagerlin, B.J. Zikmund-Fisher, and P.A. Ubel, Helping patients decide: ten steps to better risk communication, *Journal of the National Cancer Institute* **103** (2011), 1436-1443.

[18] E. Sklar and M.Q. Azhar, Argumentation-based Dialogue Games for Shared Control in Human-Robot Systems, *Journal of Human-Robot Interaction* **4** (2015), 120-148.

[19] P. McBurney and S. Parsons, Dialogue Games for Agent Argumentation, in: *Argumentation in Artificial Intelligence,* G. Simari and I. Rahwan, eds., Springer US, Boston, MA, 2009, pp. 261-280.

[20] K. Dorst, The core of 'design thinking' and its application, *Design studies* **32** (2011), 521-532.

[21] I. Sassoon, N. Kökciyan, S. Parsons, and E.I. Sklar, Towards the Use of Commitments in Multi-agent Decision Support Systems, in: *Workshop on Dialogue, Explanation and Argumentation at Human Agent Interaction (HAI)*, 2018.

[22] D. Groat, H. Soni, M.A. Grando, and D.R. Kaufman, Design and Testing of a Smartphone Application for Real-Time Self-Tracking Diabetes Self-Management Behaviors, *Applied Clinical Informatics* **9** (2018), 440-449.

[23] R.S. Evans, J.F. Lloyd, V. Flint, B.D. Horne, S. Rea, D.S. Collingridge, S.E. Abplanalp, A.A. Fazili, and D.P. Blagev, Decision Support to Help Identify Patients with Chronic Obstructive Pulmonay Disease Exacerbation, in: *American Medical Informatics Association (AMIA) Annual Symposium*, 2018, pp. 1687-1689.

[24] A.S.M. Mosa, A.M. Hossain, and I. Yoo, A Dynamic Decision Support System for Preventing Chemotherapy-Induced Nausea and Vomiting, in: *American Medical Informatics Association (AMIA) Annual Symposium*, 2018.

[25] K. Atkinson, T. Bench-Capon, and S. Modgil, Argumentation for Decision Support, in: *Database and Expert Systems Applications*, 2006, pp. 822-831.

[26] M.A. Grando, L. Moss, D. Sleeman, and J. Kinsella, Argumentation-logic for creating and explaining medical hypotheses, *Artificial Intelligence in Medicine* **58** (2013), 1-13.

[27] P. Tolchinsky, U. Cortes, S. Modgil, F. Caballero, and A. Lopez-Navidad, Increasing human-organ transplant availability: Argumentation-based agent deliberation, *IEEE Intelligent Systems* **21** (2006), 30-37.

Address for correspondence

Simon Parsons, Department of Informatics, King's College London, Bush House, Strand Campus, 30, Aldwych, London WC2B 4BG, simon.parsons@kcl.ac.uk.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190303

Usability Across Health Information Technology Systems: Searching for Commonalities and Consistency

Ross Koppel[a,b] and Craig Kuziemsky[c]

[a]*University of Pennsylvania, Philadelphia, PA, USA*
[b]*University at Buffalo (SUNY), Buffalo, NY, USA*
[c]*Telfer School of Management, University of Ottawa, Ottawa, ON, Canada*

Abstract

Usability of health information technology (HIT) remains a predominant concern — one often exacerbated by clinicians' need to access information created by many different professionals in different settings, often using very dissimilar EHRs or even different configurations of the same EHR. Because of these variations, we argue that we must no longer think of usability as anchored in one setting, one EHR, one data standard, or one type of clinician. Rather, usability must be understood as a collective and constantly evolving process. This paper seeks to address that reality by 1) substantially expanding our previously-developed conceptual matrix of the wide range of settings and interfaces comprising modern HIT and 2) presenting actual examples of EHR usability issues with similar data but very different displays or processes.

Keywords:

Ergonomics, patient safety, evaluation studies

Introduction

Healthcare information technology (HIT) usability issues remain a predominant source of medical errors and other undesirable outcomes [1; 2]. While research has identified usability issues in a single system or setting, the challenge of usability across a range of systems remains problematic and far less rigorously examined. Patient care increasingly occurs across multiple providers, settings, and HIT systems. Thus, usability must be considered not just for one system, but across several systems and users. Functions or features in HIT (e.g. data retrieval or display) are seldom consistent across systems, which often leads to errors [3]. Clinicians may have to use several different systems, meaning they will encounter variability in how one logs in and out or enters and searches for data. They will also have to reconcile a wide assortment of fonts and displays. Ash, Berg and Coiera describe how unintended consequences from HIT usage – especially entering and retrieving information – can lead to 'silent errors' that remain prominent today [4; 5].

We cannot uncouple HIT usability considerations in one system from the broader context of how healthcare delivery occurs. The reality of modern healthcare delivery is that providers may have to use multiple systems and will encounter numerous ways of displaying data and of searching for and retreving information [6; 7]. While it may be infeasible to implement a standardized HIT across the many different settings where healthcare delivery occurs, there is a need to understand usability variations to proactively manage unintended consequences. While there is a wide body of research that has looked at ways to improve the usability of individual HIT systems (e.g. [8]), there is a paucity of studies that have looked at cross-system usability to better understand the issues clinicians encounter when they have to use multiple systems.

In previous work, we developed a matrix of eleven usability dimensions and contextual differences to stimulate discussion about usability issues across providers and settings [9]. This paper extends that work by expanding the usability dimensions to fifteen and also by identifying specific usability issues for each dimension. We assembled these examples from our decades of observations, from the human-computer interaction literature, and from lists of problems reported to the IT departments at several hospital systems (e.g. [3; 10-12]). We then provide a set of case examples to illustrate how our matrix helps to better contextualize multi-system usability issues.

Results

Expanded Matrix of Usability Dimensions and Contextual Differences

Table 1 shows our expanded matrix of usability dimensions and contextual differences. The matrix identifies three usability categories or dimensions: 1) Displays, Navigation, and Screen Rules; 2) Implementation, Staffing, and Cost; and 3) Authentication, Staff Access Rules, and Logins/Logouts. We believe these three overarching groups to be significant for cross-system usage. Each group encompasses a set of usability challenges within the dimension and specific factors that account for these challenges. For example, for the first usability dimension (Displays, Navigation, Screen Rules), the first usability issue is differences across electronic health record (EHR) systems and versions. Usability testing needs to account for contextual differences that will exist due to different EHR vendors and versions.

Following the expanded matrix we provide a set of case examples that help to visualize and understand how the specific usability issues described in the matrix vary across different clinical systems. We then illustrate our expanded matrix's utility with examples of multi-system usability issues. Each example refers to specific usability dimensions and issues from Table 1.

Table 1 – Expanded Matrix of Usability Dimensions and Contextual Differences

Usability Category	Usability Challenge	Specific Usability Factors
Displays, Navigation, Screen Rules	**1.** Different EHR systems & versions present data in very different ways. (Often, it is the cost or fear of chaos that inhibits shifting and upgrading.)	EHR vendor & version
	2. Inconsistent and confusing data displays. Fonts, colors, metrics, interfaces and more vary dramatically across systems. Providers become comfortable viewing data in a specific context and may be confused when the display changes. (Figure 1)	Usability displays (font, size, metrics, color schemes, color intensity)
	3. Finding patients and data – there is a high degree of inconsistency in navigation and search functions for patient location data and clinical data (e.g. lab or medication data). Problems locating patients or medical data cause inefficiency and, at worst, can be lethal. **a.** Patient-finding method: by last name, first name, MRN, MDs, unit, room #, team name, etc. **b.** Finding labs, meds, problem lists, etc. (Many different lab and test names for same item; listings can be sorted chronologically, in reverse chronological order, by requester name, by lab facility or lab tech name, alphabetically, by organ system, etc. [Figure 2])	
	4. Rules on the number of screens and patient charts that can be open at one time vary. Each new chart or screen increases the probability of entering an order into the wrong patient chart or reading data from the wrong patient chart. (Figure 3) **a.** ID safety, patient name and photo on each screen: name on every screen, photo on page, photo size, display clarity, position on screen **b.** Number of screens and patient charts open at one time	
Implementation, Staffing, Cost	**5.** The literature reflects myriad conflicts between medical staff and corporate leadership or consultants selected by management. **a.** EHR implementation authority: in-house, system-wide (enterprise) **b.** Role of implementation consultant(s): consultants largely determine configuration, consultants mostly advisory, consultants absent or not in authority	
	6. There may be a false belief that implementation of new EHRs will reduce the need for, size of, and proximity of IT teams.	IT team location (on site, not on site) & size (expanded, stay same size)
	7. Practice size, type and combinations thereof are major factors determining EHR cost, design and configuration.	Clinician type (e.g. MD, RN, NP), practice size (if outpatient), practice type (inpatient vs. outpatient)
Authentication, Staff Access Rules, Logins/Logouts	**8.** Need for repeated logins and complex authentication requirements cause frustration and errors via workarounds and interruptions. Circumvention of access rules creates opportunities for wrong patient errors and unauthorized access. **a.** Authentication (login rules): username, password, card, biometric, two-factor authentication, combination of multiple methods **b.** Number of logins: by type (e.g. for each patient, to give drugs, to order tests), total number per day or hour	
	9. Automatic logout times cause interruptions and prompt workarounds, such as the use of Styrofoam cups to defeat proximity sensors	Automatic logout times (too quick leads to thought & work interruptions, too long leads to security risks)
	10. Access rules that are inconsistent with clinical need and workflow are a major frustration and lead clinicians to share passwords and ID cards. More directly, they may create wrong patient errors by causing confusion about patients and associated data.	Access rules (by role [e.g. MD, RN], status [e.g. admin, clinical], patient, unit)
	11. Often governments or agencies (the FDA, for example) set rules for data formats and drug use. There are also often data formatting requirements from participating labs, drug naming or drug categorizing companies, and other linked facilities. (Figure 4)	Data interoperability requirements (formats can be set by health information exchange policies, cooperating labs or suppliers, governments or other agencies)

Displays, Navigation and Screen Rules

Example of Issue 2 – Inconsistent Displays and Confusing Screens

Figure 1 shows an EHR screen for ordering vitamin K to counteract excess anticoagulation medications causing unwanted bleeding in a patient. The interface presents several problems that make the simple task of selecting a delivery method for vitamin K difficult. The clinician is presented with a screen that offers little guidance or organization, yet offers many choices and options for this critical substance. An alternative presentation of the options, perhaps with decision trees and visually comprehensible categories of conditions, would make the process safer and more efficient.

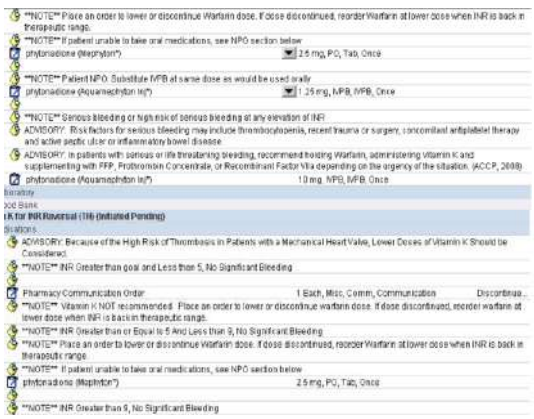


Figure 1 – This EHR screen is for selecting vitamin K to stop bleeding. Chaotic and confusing, it hinders a critical decision.

Example of Issue 3b – Finding Medication Data

Medication ordering errors are the most common form of medical error. There are roughly 4,500 medications in the average hospital formulary, with many more available through pharmacies. Variations in presentation of brand name vs. generic name, doses, routes, schedules, and the like add complexity and danger. Figure 2 provides examples of two different systems. The system in Figure 2A lists the drug name, dose and route in a single field, while the system in Figure 2B shows the generic name, drug name, route and dose in four separate fields. The different displays also have different methods by which drugs can be searched. In the system in Figure 2A, drugs can only be searched and selected alphabetically because there is only one field with all the details. The system in Figure 2B allows drugs to be searched or ordered by any of the fields (e.g. by drug or generic name).

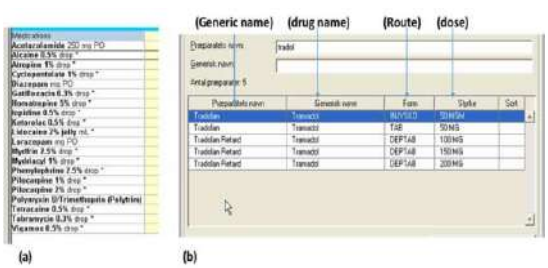


Figure 2 – Medication listings from two different systems

Other EHR systems have yet different configurations of the fields, and some have additional fields such as manufacturer or Anatomical Therapeutic Chemical (ATC) codes [13]. The different order, number of fields, and ordering and display capabilities puts the burden on individual providers to develop heuristics or other strategies to retrieve and interpret necessary information. This can present usability challenges that lead to medical errors and other unintended consequences.

Example of Issue 4b – Number of Screens Open at One Time

Figure 3 highlights the difference between viewing data on a single screen or via multiple screens. By separating the combined chart (Figure 3A) into two screens (Figures 3B and 3C) on the EHR, the pediatrician may be prevented from observing the relationship between the two variables. In this case, the dip ("dent") in weight that is not present in height may indicate gastrointestinal problems, abuse, malnutrition, or other maladies that would be missed due to the isolated charts. (The "dent" in the weight graph is exaggerated to illustrate the data in the small-format illustration; in actual practice, the chart takes up the full screen or page.)

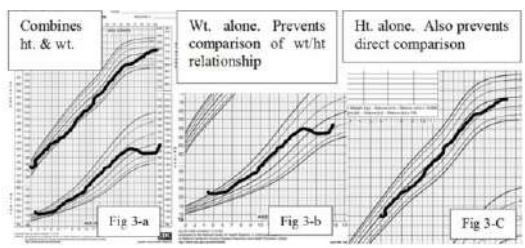


Figure 3 – Standard weight/height chart used by pediatricians for over 100 years (A); display of separate weight chart (B) and height chart (C) in some EHRs

Implementation, Staffing, Cost

Example of Issue 7 – Practice Size, Type and Combinations

Practice configurations can vary significantly depending on context. Usability challenges may be less pronounced in an independent primary care clinic where providers use a single system. Hospitals, however, can present specific usability challenges, as providers often have to access different systems depending on the context of care delivery. Specialty areas such as the operating room often have their own HIT systems.

A recent study of an implementation of a perioperative system provides an example of this issue [14]. Because the system in the study was a real-time system for perioperative data only, other data, such as lab or historical patient data, were not part of the perioperative system; therefore, clinicians needed to access this data in other systems. For example, a patient's pre-op data (e.g. diagnosis, surgical and anesthesia history) and their lab and radiology data were only viewable in the hospital EHR system. This led to two issues. First of all, the EHR system was not routinely used by surgical staff prior to implementing the perioperative system, meaning users had to learn how to use a whole new system. Secondly, and more significantly, the dual systems created a chasm between data from the EHR and data from the perioperative system. Users often needed information from both systems simultaneously. For example, on the day of surgery, a patient's historical data is in the EHR, while the active charting on the patient's case is occuring in the perioperative system. As a result, clinicians

have to look at records in two different systems, which creates visualization problems as they need to toggle between two different interfaces. Clinicians commented that they preferred to view the data side by side; thus, they were required to devise innovative ways to acess data across the two systems, such as using an iPad to access the perioperative system and the hospital terminal to access the EHR system.

Authentication, Staff Access Rules, Logins/Logouts

Issue 8a – Authentication and Login Rules

A common challenge to providers moving across settings is the need to log in and out of multiple HIT systems. This issue highights common authentication problems; namely, that many types can be used, including passwords, fingerprints, ID cards, or retinal scans. If providers must use multiple systems with different login systems, usability issues emerge.

Fingerpint readers, for instance, may become bacterial reservoirs, and they cannot be used by clinicians wearing gloves. Many clinicians are reluctant to place their eyes on iris readers, or their eyes may appear to change in response to diabetic retinopathy or cataracts. Clinicians who forget ID cards can be prohibited from accessing their EHR for an entire day.

Issue 9 – Automatic Logout Times Cause Interruptions

Automatic logout is designed as a security feature, but it can also lead to usability problems. In the aforementioned perioperative case study, the system had an automatic logout as a security feature. However, the usability impact of the logout feature was dependant on which perioperative area the clinician worked in. In high-traffic areas, such as surgical day care centers, nurses quickly moved from patient to patient, with only a few fields of data entry for each patient. In this case, the automatic logout was not a problem. However, in the operating room, the anesthesiologist would have the perioperative system open for the duration of the patient's surgery. Some anesthesiologists commented that there would be blocks of time during the surgery when they would not be entering data for a long enough period that the system would log them out. They would then have to log back in and find the patient record yet again. Worse, sometimes an anesthesiologist would pre-configure some of the upcoming data entry and would then lose that configuration when the automatic logout occurred.

Issue 10 – Access Rules

Patients with contagious infections are placed in isolation rooms. To enter those rooms, clinicians wear special gowns and gloves. Those clinicians also cannot roll the mobile computer workstations into those rooms with them. Therefore, they often rely on a colleague in another room or the hallway to access that patient's chart for them. However, the only way they can enter orders and data is by giving their password to the non-isolated clinician colleague or by leaving their computer ID card with the cooperating colleague before entering the room. This violates computer access rules and makes tracking the authors of orders impossible.

Example of Issue 11 – Governments or Other Agencies Often Set Rules for Data Formats and Drug Use

Figure 4 describes an issue at a major hospital with a home-built EHR system. There was insufficient funding to update the vaccine list for pediatricians, so vaccines that were no longer approved and no longer available remained on the EHR menus. If a pediatrician were aware of the change, they could simply prescribe a different vaccine. If the pediatrician were not aware of the change, however, they would prescribe it but the pharmacy would be unable to administer it. Worse, the pediatrician would assume the child was protected when they were not.

Vaccine	Vaccine Schedule and Status
~~ABCD~~	Found ineffective. Use XYZ drug instead
EFGH	Administer at 6 mos. and at 18 mos.
~~IJKL~~	Don't administer. Replaced by UVWX.
MNOP	Administer at 3 mos.
~~QRST~~	Prohibited. Do not use

Figure 4 – Discontinued vaccines remain on drug menus

Discussion

It has been said that consistency is one of the most important design principles for achieving usability [15]. However, the range of existing settings, care delivery models and HIT systems renders it improbable that we will have consistent HIT design guidelines any time soon. Rather, we must develop approaches to best manage the diversity of HIT systems that exist. This paper expands our previous multi-setting, multi-system, and multi-user matrix of usability dimensions [9]. In our ongoing work, we seek to encourage a more panoptic design of HIT software by incorporating the need to focus on usability across several facilities and many software vendors' products, addressing many clinicians' multifaceted needs when they confront substantial interface and functional differences. Frameworks for multi-dimensional usability are needed in settings where there are multiple systems for the same task, such as e-prescribing [16]. An expansion of usability considerations is also needed as we move toward greater collaborative care delivery and must train users in the use of HIT features across multiple settings and systems.

In this paper, we used examples to illuminate how this new and expanded matrix may be operationalized, illustrating how information retrieval methods for clinical data in one system may be inefficient or even impossible in another, and different usability issues users will encounter across different systems. We also highlight usability differences that can occur in the same settings because of different system uses and users across units or departments. Our work seeks to demonstrate how information retrieval differences can lead to serious usability and medical errors. It also emphasizes the need for ongoing user training and education in HIT, as expertise on one HIT system does not assure competence on other systems. Further, the factors differentiating how HIT systems are actually used will encompass not only the vendor's designs, but also the work of local IT teams, local implementation teams, implementation consultants, local regulations, corporate rules, and more. In addition, more modifications will be required due to changing patient and clinician populations, new medications, new mobile applications, patient-provided data, and new procedures.

The usability dimensions and issues we identified focus on our prior studies, which is a limitation of this paper. Other usability dimensions and issues will no doubt exist in other contexts and settings. Our next steps will be to conduct field studies to test our usability matrix across different providers, settings, HIT systems, and care delivery models.

Conclusions

The distributed nature of HIT systems and functions ensure they will always be emergent, interactive and multifaceted. Our approaches to studying usability and HIT must be equally emergent and multifaceted. The old model of a clinician learning and using one system is already superannuated by modern medical practice. As systems struggle to achieve interoperability and effective sharing of data, the need for cross-system commonalities across HIT systems will only increase. When one gets into a car, one assumes the gas pedal is on the right and the brake is on the left; such basic assumptions do not apply to EHRs. In some EHR systems, information access may require a patient's hospital room number; in another system the clinician may need to know the patient's medical record number. In some EHR systems, lab reports are found via the name of the test; in others, via the name of the laboratory or the requesting physician.

Given these variations, we assert that usability must encompass analyses and evaluations of many EHR interfaces as used by many different clincians in a range of settings with diverse implementations. Usability must be conceptualized on both an individual system level and as a collective reality.

References

[1] A. Kushniruk, C. Nohr, S. Jensen, and E.M. Borycki, From Usability Testing to Clinical Simulations: Bringing Context into the Design and Evaluation of Usable and Safe Health Information Technologies. Contribution of the IMIA Human Factors Engineering for Healthcare Informatics Working Group, *Yearb Med Inform* **8** (2013), 78-85.

[2] R. Marcilly, L.W. Peute, M.C. Beuscart-Zephir, and M.W. Jaspers, Towards Evidence Based Usability in Health Informatics?, *Stud Health Technol Inform* **218** (2015), 55-60.

[3] S.W. Smith and R. Koppel, Healthcare information technology's relativity problems: a typology of how patients' physical reality, clinicians' mental models, and healthcare information technology differ, *J Am Med Inform Assoc* **21** (2014), 117-131.

[4] J.S. Ash, M. Berg, and E. Coiera, Some unintended consequences of information technology in health care: the nature of patient care information system-related errors, *J Am Med Inform Assoc* **11** (2004), 104-112.

[5] P.L. Elkin, M.C. Beuscart-Zephir, S. Pelayo, V. Patel, and C. Nohr, The usability-error ontology, *Stud Health Technol Inform* **194** (2013), 91-96.

[6] T. Payne, J. Fellner, C. Dugowson, D. Liebovitz, and G. Fletcher, Use of more than one electronic medical record system within a single health care organization, *Appl Clin Inform* **3** (2012), 462-474.

[7] J.R. Vest and K. Simon, Hospitals' adoption of intra-system information exchange is negatively associated with inter-system information exchange, *J Am Med Inform Assoc* **25** (2018), 1189-1196.

[8] P.Y. Yen and S. Bakken, Review of health information technology usability study methodologies, *J Am Med Inform Assoc* **19** (2012), 413-422.

[9] R. Koppel and C. Kuziemsky, Interface Usability Across and Within EHR Vendors and Medical Settings: The Often Unexamined Need for Interface Similarities, *Stud Health Technol Inform* **234** (2017), 183-187.

[10] K. Stowers and M. Mouloua, *Human Computer Interaction Trends in Healthcare: An Update*, 2018.

[11] P. Carayon and K.E. Wood, Patient safety - the role of human factors and systems engineering, *Stud Health Technol Inform* **153** (2010), 23-46.

[12] C.E. Kuziemsky and A. Kushniruk, Context mediated usability testing, *Stud Health Technol Inform* **205** (2014), 905-909.

[13] A. Kushniruk, J. Kaipio, M. Nieminen, H. Hypponen, T. Laaveri, C. Nohr, A.M. Kanstrup, M. Berg Christiansen, M.H. Kuo, and E. Borycki, Human Factors in the Large: Experiences from Denmark, Finland and Canada in Moving Towards Regional and National Evaluations of Health Information System Usability. Contribution of the IMIA Human Factors Working Group, *Yearb Med Inform* **9** (2014), 67-81.

[14] C.E. Kuziemsky, A Model of Tradeoffs for Understanding Health Information Technology Implementation, *Stud Health Technol Inform* **215** (2015), 116-128.

[15] J. Nielsen, Enhancing the explanatory power of usability heuristics, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Boston, Massachusetts, USA, 1994, pp. 152-158.

[16] Z. Ahmed, Y. Jani, and B.D. Franklin, Qualitative study exploring the phenomenon of multiple electronic prescribing systems within single hospital organisations, *BMC Health Serv Res* **18** (2018), 969.

Address for correspondence

Ross Koppel, Ph.D.

Email: rkoppel@sas.upenn.edu

Phone: 215 898 7679

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190304

Preliminary Assessment of the Interoperability Maturity of Healthcare Digital Services vs Public Services of Other Sectors

Angelina Kouroubali[a], Anastasia Papastilianou[b], Dimitrios G. Katehakis[a]

[a] Institute of Computer Science, Foundation for Research and Technology - Hellas, Heraklion, Greece,
[b] National Centre for Public Administration and Local Government, Athens, Greece

Abstract

The development of electronic services for healthcare presents challenges related to the effective cooperation of systems and stakeholders in a highly regulated environment. Assessing the interoperability maturity of the provided services helps to identify interoperability issues in public administration. This paper presents a typical healthcare digital service: the inpatient admission in a public hospital in Greece. The Interoperability Maturity Model (IMM) is applied to assess its maturity, identify improvement priorities, and compare it with digital services of the healthcare sector. An analysis is also performed to compare a group of fourteen healthcare digital public services with sixty-seven public services of other sectors in the country. The IMM is a useful tool to facilitate awareness raising and priority setting concerning interoperability in public administration. What is discovered, through this preliminary assessment, is that healthcare digital services seem to have higher overall interoperability maturity than those of other sectors in Greece.

Keywords:

Health Information Interoperability, Health Information Systems, Public Health Administration

Introduction

Electronic Health or eHealth is a term incorporating different concepts, including health, technology, and business. Several definitions published include these concepts with varying degrees of emphasis. Health, as used in these definitions, usually refers explicitly to healthcare as a process, rather than to health as an outcome [1]. In eHealth, technology is portrayed as a means to expand, to assist, or to enhance human activities, rather than as a substitute for them. As such, eHealth can not only benefit citizens, patients and healthcare professionals, but also health organizations, businesses and public authorities [2]. Despite the opportunities and benefits, major barriers hamper the wider uptake of eHealth. One of the major ones is the lack of interoperability between eHealth solutions, which is far more than just data exchange. Interoperability in eHealth is about the delivery of contextually relevant understandings efficiently and securely to facilitate care coordination, irrespective of application, vendor or device [3,4]. It is about improving healthcare.

The healthcare sector has many digital services and tools that are interrelated with the Electronic Health Record (EHR) of the citizen, and several public services such as ones related to identification and authentication, health insurance validation, coordinated care and others. Interoperability in eHealth is challenging for various reasons, including the fact that different products and solutions in the market do not follow well-known standards and interoperability guidelines [4-6]. It is important to create the necessary conditions and frameworks to guide the market towards interoperability solutions that follow specifications that facilitate interaction with existing healthcare services and necessary public services, integrated care pathways and shared workflows [7]. Healthcare digital services must interact not only with many other digital services offered by administrative bodies within the sector, but also across different sectors. In addition, the need arises for cross-border sharing and utilization of services, data, and business processes [8].

Although shared and reusable data are gradually being introduced to the healthcare domain through increasingly interoperable systems, measurement of the interoperability maturity of those services has not been common. Interoperability assessment methods involve the use of maturity models as a framework to describe the way a service is carried out within the same or across domains [9].

The Interoperability Solutions for Public Administrations, Businesses and Citizens programme (ISA²) (https://ec.europa.eu/isa2/isa2_en) in the European Union (EU) supports the development of digital solutions that enable public administrations, businesses and citizens to benefit from interoperable cross-border and cross-sector public services. In recognition of the importance of creating and promoting interoperable public digital services, IMM [10] was developed, as part of the ISA² programme in order to assess the interoperability readiness of the digital public services and raise awareness of the need for interoperable solutions. The IMM helps public service owners to evaluate, improve and consider all key interoperability aspects of the public service. As an evaluation tool, the IMM can be useful for national and cross-border services. The interoperability maturity of public services has been assessed for 17 Trans European systems and the Swedish and Cyprus public administrations to evaluate a number of public services provided at national and local levels [11]. Interoperability assessment models have been used also in the United States and Australia [12,13].

This paper introduces the IMM and applies it to a specific, typical healthcare scenario found in a Greek hospital, to assess its interoperability readiness as a digital service for shared activities. This scenario is subsequently compared to interoperability readiness of other digital healthcare services. In addition, further analysis identifies the maturity levels of healthcare digital services compared to other digital public services in Greece. Healthcare in Greece is provided by the National Health System (NHS). As a public service, it does not

exist in isolation but as part of the wider national public administration. The paper concludes with discussion of the key findings and ideas for future research.

Methods

Methods used for this work were based on the IMM. The IMM was used to evaluate the inpatient admission service delivered through applications developed by the Foundation of Research and Technology – Hellas (FORTH) at the NHS of Greece. In addition, an analysis was performed to aggregate the interoperability maturity of 14 healthcare services in total. The inpatient admission service results were compared to the results of the other (13) healthcare services. Finally, all healthcare services results were compared to the maturity evaluation of 67 digital public services of other sectors in Greece. The methodology is described in detail in the next paragraphs.

Interoperability Maturity Model

IMM has been designed to help public service owners at different government levels (i.e. local, regional, national, and cross-border) to evaluate the current interoperability maturity level of a public service and gain insight into the improvement priorities that are needed to reach the next level of interoperability maturity. A public service is a service that addresses the public interest and is delivered by a public administration to citizens (A2C), to business organizations (A2B) and/or to other public administrations (A2A). A *process trigger* initiates the public service that consists of several process steps. IMM measures the level of interaction of a public service with services of other organizations towards the realization of mutually beneficial and agreed common goals through the exchange of information and reuse of services [10]. Interoperability is distinguished in three domains: *service delivery* which refers to the way the public service delivers its outcome to the end-user, *service consumption* which looks at the services or data that are being reused from other public administrations and businesses where the assessed public service has the role of the consumer, and *service management* which refers to the coordination of all interactions with the internal and external environment. Each domain can receive a score from 1 to 5. Explanations of the scores are provided in table 1.

Table 1– Domain Scores for Interoperability Maturity

Score	Interoperability Maturity
1	*ad hoc: poor interoperability* - referred to a service that cannot be considered interoperable
2	*opportunistic: fair interoperability* - the digital public service implements some elements of interoperability best practices
3	*essential: essential interoperability* - the digital public service implements the essential best practices for interoperability
4	*sustainable: good interoperability* - all relevant interoperability best practices are implemented by the digital public service
5	*seamless: seamless interoperability* - the digital public service is a leading interoperability practice example for others

The healthcare digital service that is analyzed and presented in this paper refers to the admission of a patient in the clinic of a public hospital in Greece. The scenario is presented and its maturity model is assessed. The IMM questionnaire was filled out to assess the interoperability readiness of the service for the three domains of interoperability: *service delivery*, *service consumption* and *service management*. The IMM was used vs the more lite version called Interoperability Maturity Assessment of a Public Service (IMAPS) which is a compact self-assessment online survey [11]. IMM was selected because it provides a comprehensive toolset for a detailed and in-depth analysis of the service landscaping. Based on the assessment a tailor-made set of recommendations is provided to the service owner. IMAPS has been introduced by ISA^2 at a later stage to support a faster self-assessment with easy to understand recommendations.

Assessment and Analysis

The IMM assessment took place as part of the work performed for the course *interoperability evaluation of digital public services* that was delivered by the National Centre for Public Administration and Local Government in Greece. The course introduced the IMM to public employees who were asked to apply it to a digital service and assess its maturity. The course was delivered 15 times through the years 2016-2018 and had overall 425 participants working in small teams. As part of the coursework, 81 different digital public services have been assessed. The quantitative analysis was performed on 81 public services, which were separated into two groups. One group had the digital services that were related to healthcare (14), and the other group had the rest of the public services (67). For the healthcare services, a comparison was made between the inpatient admission services and the rest (13). A descriptive analysis compared the services for the three domains of interoperability: *service delivery*, *service consumption* and *service management*.

Limitations of the Study

The study is considered as a preliminary assessment of healthcare services as the sample size is quite small. The interpretations of the analysis cannot be widely generalized but provide an indication of the interoperability maturity of the digital services of the public sector in the country. In addition, the IMM has an inherent limitation in that the assessment that occurs relies widely on self-reporting and interpretation of the questions. It is important to take into account that the results are based on qualitative critical evaluation. This limitation has been partly minimized with the intervention of experienced trainers during the assessment exercise of small groups of experts.

Results

Digital Healthcare Service in a Greek Public Hospital

The Patient Administration family of applications (ICS–A), is part of the FORTH Integrated Care Solutions (ICS) suite (https://www.ics.forth.gr/ceha/index_main.php?l=e&c=664). The ICS family of applications supports, among others, all patient management processes for both inpatients and outpatients in a healthcare unit, including all medical, nursing and administrative processes in the hospital. ICS applications support patient admission, transfer and discharge, hospitalization data logging, scheduling of appointments and surgeries, waiting lists, ordering and recording of medical acts, billing, payment collection, and electronic reimbursement [14]. ICS applications interact with digital services within and across the healthcare enterprise using interoperability services. They contribute to the improvement of organizational performance and cost savings by applying the *data entered-once* and *used-anywhere* key concepts.

The Inpatient Admission Service

The inpatient admission office application supports all the business processes related to inpatient admissions. It follows the movement of the patient during hospitalization, from admission to discharge. The *service process trigger* occurs when a patient or caregiver arrives at the patient admissions office for admission. The administrative staff searches for a patient record in the local hospital database and then in the regional central patient registry. If the patient record is not found, the administrative staff registers a new patient record. The patient demographic data, the patient insurance data and the insurance coverage validity are acquired online from the National Electronic Confirmation Service (ATLAS) [15], using the National Social Security Number (AMKA) as an identifier. The administrative staff fills out additional patient information and creates the patient record. If the patient record already exists in the Hospital Information System (HIS), the administrative staff checks for the validity of patient insurance coverage by retrieving data from ATLAS. Then the patient admission to the hospital is registered and the HIS sends a notification to the service of the National Organization for Health Care Services (EOPYY) [16]. Then, the ward management system at the clinic where the patient is admitted receives the corresponding patient admission notification and relevant data. The inpatient admissions office application also communicates directly or indirectly with third party systems including laboratory and radiology, pharmacy, enterprise resource planning, and business intelligence systems.

The family of applications described are based upon an open, scalable and evolvable architecture that integrates distributed information and knowledge in a flexible manner, focusing on the timely and effective delivery of the appropriate information to all authorized users. Being the outcome of applied research, it encompasses both state of the art trends and real-world requirements for effective use [14,17].

Interoperability Maturity Assessment

The assessment of the inpatient admission service scored 2.80 for the dimension s*ervice delivery*, 4.30 for s*ervice consumption*, and 4.45 for s*ervice management*. The overall interoperability maturity score was 3.98. The results are presented in table 2.

Table 2– IMM Assessment of Inpatient Admission Service

Maturity Dimension	Score	Interoperability
Service Delivery	2.80	*essential*
Service Consumption	4.30	*sustainable*
Service Management	4.45	*sustainable*
Overall Maturity	3.98	*sustainable*

The inpatient admission service consumes the digital electronic confirmation services of ATLAS for demographics and insurance coverage, and the inpatient admission announcement service provided by EOPYY. The ATLAS service has been implemented as a base registry for AMKA and insurance coverage. The inpatient admission announcement has been implemented as a service for cost control within the NHS. Both services are consumed, when needed, by digital services of the public sector. The inpatient admission service maturity level for *service consumption* is considered *sustainable* as it consumes all available digital services that exist.

Interoperability Maturity of Healthcare Public Services

The evaluated services in the healthcare sector are provided by various health administration authorities and other organizations such as hospitals and pharmacies. The list in table 3 contains services that are related to the healthcare and social welfare of citizens as evaluated during the course work.

Table 3– List of Assessed Healthcare Services

#	Service Description
1	Hematological examinations
2	Hospital patient admission ticket issue for outpatients
3	Hospital admission ticket submission to the Business Intelligence System of the Ministry of Health
4	Health insurance record
5	Registration of dependent family members for insurance eligibility assignment
6	Electronic submission of hospitalizations to EOPYY
7	Inpatient admission service
8	Issue of a European Health Insurance Card (EHIC)
9	Disability Certification
10	Registration of new members for the acquisition of AMKA
11	Electronic decision to grant Medicare benefit
12	Digital signatures for citizens
13	Updating of contact information for an insured citizen
14	Registration and information retrieval on insurance eligibility

Focusing on the assessment of evaluated healthcare services, the value of the overall interoperability maturity of healthcare services is 3.32 as shown in table 4, and is characterized in terms of IMM as interoperability level 3: *essential*.

Table 4– IMM Assessment of Healthcare Services

Maturity Dimension	Score	Interoperability
Service Delivery	2.69	*essential*
Service Consumption	3.55	*sustainable*
Service Management	3.49	*sustainable*
Overall Maturity	3.32	*essential*

The interoperability maturity level for the *service delivery* dimension was 2.69, for *service consumption* dimension was 3.55, and *service management* dimension was 3.49. Looking more closely at the internal dimensions of health services maturity level, it is noted that in healthcare, the highest level (3.55) is in *service consumption* versus the lowest in *service delivery* (2.69). The low score for *service delivery* can be attributed to the fact that healthcare digital public services are mostly delivered from administrative personnel on behalf of citizens. These services are usually applications within healthcare organizations. There are no multichannel options for the delivery of the services. In addition, the services are not part of a service catalogue. As a result, the score for *service delivery* in healthcare remains low. On the other hand, *service consumption* is high as there are specific digital services such as base registries and services for financial management that have to be consumed by the majority of healthcare services.

Interoperability Maturity of the Inpatient Admission Service vs Other Healthcare Services

The overall maturity level of the inpatient admission service is 3.98, which is *sustainable*. Analysis of the rest of the 13

healthcare services showed an overall interoperability maturity of 3.27 which is characterized as *essential* (figure 1).

It is seen from the interoperability assessment that the described scenario scores higher than other healthcare services examined.

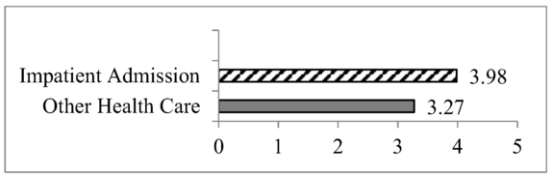


Figure 1 – Interoperability Maturity of Inpatient Admission vs Other Healthcare Services

Interoperability in Healthcare Services vs Public Services of Other Sectors

The average value of the interoperability maturity level of healthcare services is 3.32, which is not significantly higher than the interoperability maturity score of all other public services whose maturity is 3.29, as shown in Figure 2.

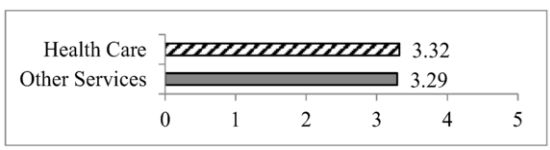


Figure 2 – Interoperability Maturity of Healthcare vs Public Services of Other Sectors

The healthcare services and public services of other sectors have, in terms of IMM, the same interoperability maturity at *essential* level. Looking more closely at the internal dimensions of healthcare services maturity versus those of all other services, as shown in Figure 3, it is noted that the *service delivery* dimension scores slight lower in healthcare than in other public services. This could be attributed to the fact that public services of other sectors could be used directly by citizens and are part of service catalogues.

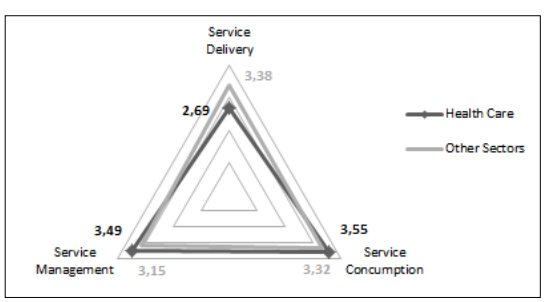


Figure 3 – Interoperability Maturity Dimensions of Healthcare Services vs Public Services of Other Sectors

Discussion

Based on the analysis of the real data generated during the conducted coursework, it has been shown that healthcare services score slightly higher in overall interoperability maturity to that of public services of other sectors. In the healthcare sector, several digital services have been implemented in the past years for the entire public sector. These implementations were reinforced through the economic crisis and the need for better control of costs within the NHS. Base registries for the healthcare sector, such as the electronic confirmation service ATLAS, are supporting instruments for public administrations.

The EU, through the ISA^2 programme, encourages cross-border public administrations to develop integrated digital services for national administrations in support of the vision for the European single digital market. To work towards this direction a number of instruments need to be in place to support this work. Development, implementation and use of these instruments in the European public sectors can help administrations reach the next level of maturity in the digital services they provide. Some of these instruments are presented in the following paragraphs.

Health Interoperability Framework: The framework gives specific guidance on how to set up interoperable digital public services. A framework can integrate the fundamentals and set the standards of healthcare services. The new European Interoperability Framework (new EIF) in combination with the healthcare specific interoperability framework provide an appropriate guide for establishing a national interoperability framework for healthcare [7].

European eHealth Governance: The eHealth Network is the main decision body on eHealth at a European level. It gathers representatives of the Member States at a high level, on a voluntary basis, to define a common vision and strategy for eHealth across Europe [18]. The eHealth network was established by article 14 of Directive 2011/24/EU on patients' rights in cross-border healthcare. The eHealth Network identifies areas for cooperation and meets two times a year to agree on common priorities.

Health Quality Assessment: Certification of healthcare processes will allow continuous monitoring in healthcare organizations. As far as interoperability and conformance testing are concerned, it is very important that a compliance strategy is in place as well as a roadmap for the development and maintenance of national specifications and interoperability principles, standards-related rather than self-defined. The creation of a mechanism for compliance control and certification of relevant software is considered critical.

Service Catalogues: The general service catalogue for public administration services can also help the healthcare sector by considering all possible extensions.

Healthcare Process Modeling: It is important to incorporate process modeling in healthcare service delivery using international standards, such as the graphical representation for business process model and notation (BPMN), and templates of descriptions of high level use cases and realization scenarios [7]. Standardized process modeling will facilitate the unified description of digital services for easier consumption and sharing.

Healthcare Service Vocabulary: The process of providing cross-border healthcare services across EU Member States is complex, due to the heterogeneity of the actors, information and services of the different Member States. The complexity of exchanging data may lead to semantic interoperability conflicts. The core public service vocabularies can be extended and used, in healthcare as in all other areas, to reduce these semantic conflicts (https://joinup.ec.europa.eu/solution/core-public-service-vocabulary).

Healthcare Learning Programs: Education and training about interoperability, interoperability assessment, and sector specialized interoperability challenges for end users, policy makers and public employees, are essential. These learning

courses providing the basic and advance understanding of interoperability challenges, needs and issues in the public sector.

Conclusions

Assessment of interoperability maturity and specific measures for interoperability enhancement can contribute to cost reduction and greater integration through reuse of available services. Orchestration of services is an effective manner to maximize service outcomes and benefits for citizens and public administrations. IMM is a way to investigate deeply how a service performs and relates to other digital public services. It helps service owners to gain a better understanding of the interoperability maturity of a digital public service. An evaluation strategy needs to be closely linked with specific policies to support the continuation of the assessment, governance and implementation of digital public services nationally.

The maturity of healthcare digital public services show a higher maturity in service consumption and service management, as well as to the overall interoperability maturity, compared to other public services. This indicates that the healthcare sector has made improvements towards harmonizing with relevant EU directives for integrated public services. Future work will involve the application of IMM in a wider range of public services as well as the tracking of interoperability maturity of public services across time.

There is significant space for improvements towards the development and implementation of seamless interoperable services for citizens, administrations and businesses in the public sector across Europe. Further work is necessary to validate the results amongst a larger number of services. In addition, further research can be conducted to evaluate similar public services across countries in Europe.

Acknowledgements

This work has been funded, in part, from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688095 (SlideWiki) (https://slidewiki.eu/), and by the Center for eHealth Applications and Services (https://www.ics.forth.gr/ceha) of the Institute of Computer Science, of FORTH. The authors would like to acknowledge Fokion Logothetidis, Ioannis Petrakis, and Nikolaos Stathiakis for their contribution in the definition and assessment of the hospital scenario presented.

References

[1] Pagliari, C., Sloan, D., Gregor, P., Sullivan, F., Detmer, D., Kahan, J. P., Oortwijn, W., and MacGillivray, S. (2005). What is eHealth (4): a scoping exercise to map the field. *J Med Internet Res*, **7**(1).

[2] Cunningham, S. G., Wake, D. J., Waller, A., & Morris, A. D. (2014). Definitions of eHealth in *eHealth, Care and Quality of Life*, Springer, Milano, 15-30.

[3] Stroetmann, K. A. (2014). Health System Efficiency and eHealth Interoperability–How Much Interoperability Do We Need? *New Perspectives in Information Systems and Technologies*, **2**, 395-406.

[4] Kouroubali, A., Starren, J., Barrows, R. C., & Clayton, P. D. (1997). Practical lessons in remote connectivity. Proceedings : a conference of the American Medical Informatics Association. *AMIA Fall Symposium*, 335–339.

[5] Katehakis, D. G., Kondylakis, H., Koumakis, L., Kouroubali, A., & Marias, K. (2017). Integrated Care Solutions for the Citizen: Personal Health Record Functional Models to Support Interoperability. *EJBI*, **13**(1), 41-56.

[6] Hammond, W. E. (2017). A New World for Better Health. *EJBI*, **13**(1), 3-8.

[7] Katehakis, D. G., Kouroubali, A., & Fundulaki, I. (2018) Towards the Development of a National eHealth Interoperability Framework to Address Public Health Challenges in Greece. Semantic Web Technologies for Health Data Management, http://ceur-ws.org/Vol-2164/paper6.pdf [accessed on April 01, 2019].

[8] Katehakis, D. G., Masi, M., Wisniewski, F., & Bittins, S. (2016). Towards a Cross-domain Infrastructure to Support Electronic Identification and Capability Lookup for Cross-border ePrescription/Patient Summary Services. *pHealth proceedings.* 152-157.

[9] Leal, G. D. S. S., Guédria, W., & Panetto, H. (2019). Interoperability assessment: A systematic literature review. *Computers in Industry*, **106**, 111-132.

[10] Interoperability Maturity Model, https://joinup.ec.europa.eu/document/interoperability-maturity-model [accessed on April 01, 2019].

[11] ISA2 Interoperability Maturity Assessment of a Public Service https://ec.europa.eu/isa2/solutions/imaps_en [accessed on April 01, 2019].

[12] Carvalho, J. V., Rocha, Á., & Abreu, A. (2019). Maturity Assessment Methodology for HISMM-Hospital Information System Maturity Model. *Journal of medical systems*, **43**(2), 35.

[13] Interoperability Framework - Interoperability Maturity Model v1.0 in Australia https://developer.digitalhealth.gov.au/specifications/ehealth-foundations/ep-1143-2006/nehta-0062-2007 [accessed on April 01, 2019].

[14] Katehakis, D. G., Halkiotis, S., & Kouroubali, A. (2011). Materialization of Regional Health Information Networks in Greece: Electronic Health Record Barriers & Enablers. *J Healthc Eng*, **2**(3), 389-403.

[15] Hellenic National Insurance Registry, ATLAS, https://www.atlas.gov.gr/ATLAS/Pages/Home.aspx [accessed on April 01, 2019].

[16] National Organization for Health Care Services Homepage, http://www.eopyy.gov.gr [accessed on April 01, 2019].

[17] Katehakis, D. G. (2018). Electronic Medical Record Implementation Challenges for the National Health System in Greece. *IJRQEH*, **7**(1), 16-30.

[18] eHealth Network: https://ec.europa.eu/health/ehealth/policy/network_en [accessed on April 01, 2019].

Address for correspondence

Angelina Kouroubali, FORTH-ICS, N. Plastira 100, Vassilika Vouton, GR 700 13 Heraklion, Crete, Greece, kouroub@ics.forth.gr, +302810391680.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190305

Enterprise Architecture in Hospitals: Resolving Incongruence Issues

Anne Kristin Ajera, Eli Hustada, Polyxeni Vassilakopouloua

[a] *Department of Information Systems, University of Agder, Kristiansand, Norway*

Abstract

Enterprise Architecture allows addressing technologies and processes in a holistic way and mirrors choices related to process standardization and data integration. It has the potential to address long-standing problems in health information systems related to fragmented IT portfolios, immature IT infrastructures, and silo-structured organizing. Nevertheless, uptake of Enterprise Architecture in hospitals has been slow. To understand the issues related to this slow uptake we have undertaken an interview study with architects and managers. The issues identified reveal a level of incongruence between healthcare as a domain and the practice of EA. Specifically, by analyzing the experiences of architects and managers we identified four different areas of such incongruence that create the need to reconcile a) Bottom-up vs. Top-down Planning b) Clinical vs. Systems´ Knowledge, c) Local vs. Global Arrangements and d) Patient Safety vs. Patient Privacy. Building on prior related research we propose ways for resolving the incongruence issues identified.

Keywords:

Architecture; Health Information Systems; Qualitative Research

Introduction

An organization's Enterprise Architecture (EA) describes in a hierarchical way its processes, the data and applications that support these processes, and all related information and communication technology (ICT) arrangements [3]. Practitioners and researchers have advocated EA as a systematic approach for designing, planning, and implementing process and technology changes [5; 21]. The EA approach addresses technologies and processes in a holistic way and mirrors choices related to standardization and integration [17]. During the last two decades, hospitals started EA initiatives aiming to address long-standing problems in health information systems related to fragmented IT portfolios, immature IT infrastructures, and silo-structured organizing.

Norway has formally adopted EA as a strategic tool for hospital information systems and processes [14]. Nearly all hospitals in the country are public and organized as health trusts that can include several local hospitals. The trusts are allocated to four independent regional health authorities under the jurisdiction of the Ministry of Health and Care Services. The strategic coordination, prioritization, and consolidation of key ICT issues across the regions are performed by a separate organizational entity dedicated to healthcare ICT called the National ICT (NICT). Four regional health authorities is jointly owned by NICT. The Norwegian government has issued several white papers that describe how ICT shall be used to achieve health policy objectives and deliver more effective and efficient services, with emphasis on quality and patient security [9]. Two of the most important white papers that guide the development of eHealth services are "The Healthcare Coordination Reform" (2009) and "One Citizen – One Record" (2012). The first one addresses issues related to collaboration, while the second one, sets the targets for the evolution of healthcare services [2; 8].

Norwegian hospitals are supported by complex information infrastructures that evolved over many years. The earliest use of electronic documentation of patient information in health services dates back to the 1970s while the first implementations of applications for entire hospital coverage started in the 1980s [7; 16]. With the adoption of EA, structured, comprehensive and aligned blueprints for current and future states of hospital systems and processes can be developed. Furthermore, EA can provide guidance for implementing processes and technology changes to operationalize strategies. Nevertheless, despite the potential benefits and the state mandate for introducing EA, there have been significant delays and challenges.

To gain insights about EA introduction in Norwegian hospitals, we performed interviews with key actors at the local, regional and national level across the hospital sector. Prior research has shown that the introduction of EA is far from straightforward and pointed to the importance of a favourable organizational culture [1; 15; 23]. Intrigued by this previous research finding, we specifically investigated the following research question: *are there inherent issues related to incongruence between healthcare as a domain and the practice of EA?*

The rest of the paper is structured as follows. First, we describe the empirical setting and explain our research method. Second, we provide the results. Third, we discuss the results and drawing from prior related research we propose ways for resolving the incongruence issues identified. We conclude the paper by pointing to the contributions and limitations of our research.

Methods

Data collection

Semi-structured interviews were conducted with Enterprise Architects and Managers involved in the introduction of EA in Norwegian hospitals at the local, regional, and national level. In total 17 interviews were performed between November 2016

and August 2017 (Table 1). All the informants interviewed had at least two years of experience in their current position. At the national level, most of the interviewees had been working since NICT started (2.5 years ago) while all were experienced before joining NICT. At the regional level, the enterprise architects had been employed for an average of about 5 years, and the managers for about 7 years. At the local level, the enterprise architects had been employed for an average of 10 years and the managers for 8 years on average. Interview questions explored the experiences of participants. The interviews included topics on how EA was used, the role of enterprise architects, and issues about national coordination and collaboration in eHealth. All interviews were audio recorded and transcribed.

Table 1– Interviews Performed

Level	Informants	# interviews
National	5 Enterprise Architects	5
Regional	4 Enterprise Architects and 2 Managers	7
Local	2 Enterprise Architects and 3 Managers	5

In Figure 1 we present the overall structure of the hospital sector in Norway (as described in the introduction) marking the specific units where interviews have been performed.

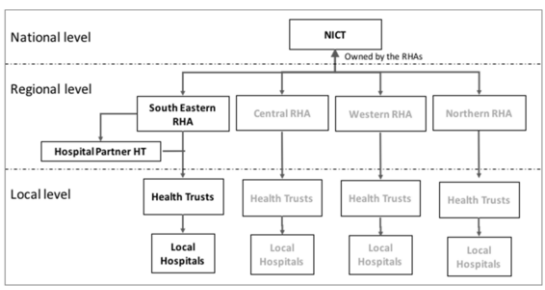


Figure 1–Overview of the Norwegian Hospital Sector (units covered by interviews are marked in bold)

Analysis

An iterative approach was followed for the data analysis. The coding followed the principles of first- and second-cycle coding [13]. The first cycle was done in NVivo where all anonymized transcripts were entered. The transcripts were further analysed and grouped into Excel forms. In the second cycle, the data were discussed, organized, and compared in an iterative process to identify emerging patterns. This initial analysis yielded a wide range of challenges, some related to the technological landscape, the financing mechanisms in place, the different logics of actors involved and the governance arrangements. Furthermore, we also encountered issues related to inherent healthcare characteristics that complexify the introduction of EA. Drawing from these initial findings, and building on previous research that points to the significance of idiosyncratic characteristics (such as culture) for the introduction of EA, we focused our analysis on issues related to incongruence between healthcare as a domain and the EA practice.

Results

We identified four key themes related to healthcare characteristics that create friction in the introduction of EA. In the paragraphs that follow we present the findings from the interviews organized according to these key themes.

Clinical vs Systems Knowledge

Several informants expressed concerns related to the instances where core hospital processes need to be redefined. In such instances, systems´ knowledge encounters medical knowledge. One of the architects wondered about the role of systems experts since they "*do not treat patients*". Another architect said, "*administrative tasks can be standardized, but for clinical ones we cannot suggest the best way*". One of the managers pointed to the singularity of healthcare domain related to the key role of the medical staff for setting the norms for patient care irrespectively of hierarchical positioning: "*the tax director does not understand that the health director cannot decide*". Another architect pointed to issues of distributed control in the domain "*you have so many strong doctors … you have strong departmental directors…there are many little kings*". Overall, hospitals unlike other large-scale organizations such as banks or public administrative services where EA has been successfully introduced, are characterized by strong professional practices that have a decisive role in process and systems optimization.

Local vs Global Arrangements

One of the architects lucidly expressed a key challenge with EA work: "*they want their local systems, not regional ICT services*". The health trusts want autonomy to meet local needs and there is a power struggle with the regional authorities. The regional authorities do own the health trusts, but they do not treat patients directly, so they do not experience day-to-day operational challenges. A manager explained that health trusts are similar to individual companies; they do not really have a "*mother-daughter*" relationship within the overall healthcare structure. One of the enterprise architects said, "*a model that ensures both standardization and taking care of local wishes may be beneficial*". Another architect pointed to concrete differences between two major hospitals: "*hospital A is much more IT mature than B, doctors and many nurses are involved in the IT department, they are much closer to the users*". A manager explained that although the needs may be similar in different regions there are different cultures across hospitals and these cultural differences matter.

Bottom-up vs Top-down Planning

EA is a plan-driven approach, however, holistic planning of ICT within healthcare can be especially challenging as explained by one of the architects interviewed "*the biggest challenge is to maintain local understanding…it should be taken from patients to clinicians and upwards not top-down*". Another architect explained that it is important to follow a bottom-up approach because clinicians often have good solutions to their own problems. Nevertheless, one of the managers pointed to the importance of top-down approaches that allow better coordination, the manager pointed to the need for a stronger central role "*everybody wants new systems,*

nobody wants to change". Balancing bottom-up with top-down planning has proven to be challenging for the EA initiatives.

Patient Safety vs Patient Privacy

An architect pointed to the need to consider data flows not only in terms of security and privacy but also in terms of patient safety. The standard regulations and procedures for systems analysis and risk containment are addressing issues related to data security and patient privacy but not issues of patient safety. Architects are not experts in issues related to patient safety. Another architect explained how important it is to make available patient data when patients move from region to region, but this is not straightforward to implement. A manager pointed also to the need for data exchanges between primary care and hospitals. Overall, the visions for data integration are bringing up unresolved issues related to balancing patient safety and privacy.

Table 2 provides an overview of the results, mapping the four key themes identified to key EA features that are challenged by the singularities of the healthcare domain. Specifically, the practice of EA is plan-driven and oriented towards process standardisation, and data integration and these characteristics relate to friction experienced when introducing EA in healthcare.

Table 2– Incongruence Themes Identified

EA characteristics	Healthcare characteristics	Incongruence themes identified
Plan-driven	National strategies but also pivotal bottom-up initiatives	Bottom-up vs Top-down Planning
Process standardisation	Processes inscribe clinical knowledge Local variation	Clinical vs Systems´ Knowledge Local vs Global Arrangements
Data integration	Sensitive but also mission-critical data	Patient Safety vs Patient Privacy

Discussion

The issues identified reveal a level of incongruence between healthcare as a domain and the practice of EA. The friction themes identified, can be related to prior research findings in Health Informatics. Healthcare is "work regarded as unusually complex, uncertain, and of great social importance." To ensure the best possible outcomes under these difficult circumstances, "the strategy pursued is to couple capability with discretion in one responsible actor and place him or her as close as possible to the problem situation … legitimate control over the nature and quality of professional practice is vested in the professional staff, not in the administration" [18]. Prior research [20] has identified that in healthcare universality is always "local universality" in the sense that it "always rests on real-time work and emerges from localized processes of negotiations and pre-existing institutional, infrastructural, and material relations". Furthermore, along the same line of thought, researchers [4] have conceptualized the distinction between conjoint and context-dependent design negotiations showing how in successful standardization processes stakeholders define and agree on boundary factors (elements that are meaningful across borders) while creating possibilities for local reconfigurations.

Prior research has also pointed to the fact that frequently, advancements in healthcare come out of practice-driven initiatives without a predetermined strategy and without the initial support of management; a novelty in healthcare usually entails extensive work over lengthy periods of time by different participants [10; 11; 19]. An approach that balances between bottom-up and top-down approaches for the evolution of Health IT Systems has been proposed in the literature [6]. The approach is labelled "middle-out" and is described as a situation where governments provide incentives and support that encourage clinical providers to acquire systems that are technically or functionally compliant, and to pursue innovations that keep their systems compliant over time. Such an approach entails specifying commonly agreed compliance requirements.

Ensuring data security and privacy while catering for patient safety is one of the big challenges when it comes to streamlining data flow and pursuing data integration in healthcare. A possible way to address this challenge is by enabling a more active role for the patients. A patient-centric logic not only spans the whole spectrum of patients' needs from preventive healthcare, to treatments and long-term care but also presumes a more active role for the patients themselves. Patients can contribute through information sharing, self-service, and assisting healthcare staff acting as resource integrators [12]. To do this, they need to be able to access, manipulate and contribute data. Moreover, they need to stay informed about who is accessing their data and be able to manage access. Enabling patients to control data flows entails ensuring the clarity, user-friendliness, and transparency of patient-oriented data handling solutions.

Overall, healthcare is a safety-critical domain that requires everyday coping with uncertainty. Patients can have unique combinations of conditions and this explains the acknowledged need for medical discretion. Reliability in healthcare is not only the outcome of protocols and formal procedures but also, of an acquired capacity to perform even though working conditions fluctuate and are not always known in advance. This capacity frequently found in high-reliability settings is a mix of risk anticipation and containment encapsulated in the term "mindfulness" [22]. Mindfulness is analysed to: a) preoccupation with failure; healthcare practitioners are concerned with success as much as with failure – Hippocrates' Oath "do no harm", b) reluctance to simplify; simplifications are avoided, c) sensitivity to operations; attention is given to process dependability under diverse circumstances, d) commitment to resilience; healthcare professionals need to be able to recover from mishaps and cope with surprises pursuing alternative means to goals, and paying attention both to error prevention and containment, and e) deference to expertise; the need for a wide array of specialisms is acknowledged and respected. EA initiatives need to retain and possibly further enhance mindfulness in the domain.

Implications for Practice

The findings suggest that in order to advance with EA in hospitals, it is important to resolve key issues related to the characteristics of EA and the singularities of the domain. There is limited prior research on the domain-specific characteristics

that hinder the introduction of EA in healthcare. This is work that we have begun to undertake. The table that follows (Table 3) is building upon prior related research to propose a set of resolutions for the incongruence themes identified. These proposed resolutions can guide practitioners involved in the introduction of EA in hospital settings.

Table 3– Resolutions for Incongruence Themes Identified Directions for Mindful EA Initiatives

Incongruence Themes	Proposed Resolutions	Brief Description
Bottom-up vs Top-down Planning	Middle-out	Incentivize clinicians to acquire compliant systems (based on common requirements).
Clinical vs Systems´ Knowledge Local vs Global Arrangements	Configurable solutions catering for both standardisation and local needs	Technology accommodating clinical discretion and local variation.
Patient Safety vs Patient Privacy	Stronger patient role – patient data access management	Patients can contribute acting as resource integrators.

Limitations and Strengths

The study has been conducted within Scandinavian healthcare so we cannot judge to what extent the findings reflect friction and tension with the introduction of EA in hospitals in different countries that have different institutional characteristics. However, the issues identified relate to the characteristics of healthcare as a practice in general and are also found in prior research conducted in different countries. A strength of our study is that it links to prior related research to propose concrete resolutions. Furthermore, it provides detailed qualitative data on the experiences of architects and managers involved in the implementation of EA at different levels (local, regional, national) and can provide insights for those that seek to contribute to such initiatives.

Conclusions

This study suggests that there are inherent issues in the introduction of EA in hospitals that relate to some level of incongruence between EA and the key characteristics of the healthcare domain. Specifically, by analyzing the experiences of architects and managers we identified four different areas of such incongruence that create the need to reconcile a) Bottom-up vs. Top-down Planning b) Clinical vs. Systems´ Knowledge, c) Local vs. Global Arrangements and d) Patient Safety vs. Patient Privacy. Our study can be used as a basis for further research towards the development of a measurement instrument to assess trade-offs in the four incongruence areas identified. As these areas relate to the nature of healthcare delivery, prior research has brought insights related to potential resolutions. Although the introduction of EA in hospitals is far from straightforward, working towards resolutions for the specific incongruence themes identified can be a basis for mindful EA initiatives. Mindful EA can be a healthcare-specific EA approach for systematically designing, planning, and implementing process and technology changes.

Acknowledgements

We would like to thank the informants that shared their experiences and views. We would also like to thank the anonymous reviewers for their constructive comments that helped us improve our manuscript.

References

[1] S. Aier, The role of organizational culture for grounding, management, guidance and effectiveness of enterprise architecture principles, *Information Systems and E-Business Management* **12** (2014), 43-70.

[2] C. Bergland, På vei mot én innbygger - én journal, (2018). https://vimeo.com/258664516 (accessed March 14, 2018).

[3] S.A. Bernard, *An introduction to enterprise architecture*, *3rd Edition*, AuthorHouse, Bloomington IN, United States, 2012.

[4] P. Bjørn, S. Burgoyne, V. Crompton, T. MacDonald, B. Pickering, and S. Munro, Boundary factors and contextual contingencies: configuring electronic templates for healthcare professionals, *European Journal of Information Systems* **18** (2009), 428-441.

[5] R.V. Bradley, R.M.E. Pratt, T.A. Byrd, C.N. Outlay, and D.E. Wynn Jr, Enterprise architecture, IT effectiveness and the mediating role of IT alignment in US hospitals, *Information Systems Journal* **22** (2012), 97-127.

[6] E. Coiera, Building a national health IT system from the middle out, *Journal of the American Medical Informatics Association* **16** (2009), 271-273.

[7] P. Doupi, E. Renko, and S. Giest, S., Ehealth Strategies - Country Brief: Norway, (2010). http://ehealth-strategies.eu/database/documents/Norway_CountryBrief_eHStrategies.pdf (accessed March 25, 2018).

[8] Directorate of eHealth, IKT-organisering i helse- og omsorgssektoren, (2017). https://ehelse.no/Lists/Publikasjoner/Attachments/12/Rapport - IKT-organisering i helse- og omsorgssektoren.pdf (accessed November 30, 2018).

[9] Directorate of eHealth, Styringsdokument (2017). https://ehelse.no/styringsdokument (accessed March 10, 2018).

[10] A. Essén, and S. Lindblad, Innovation as emergence in healthcare: unpacking change from within, *Social Science & Medicine* **93** (2013), 203-211.

[11] M. Janssen, A. Stoopendaal, and K. Putters, Situated novelty: Introducing a process perspective on the study of innovation, *Research policy* **44** (2015), 1974-1984.

[12] J.R. McColl-Kennedy, S.L. Vargo, T.S. Dagger, J.C. Sweeney, and Y. van Kasteren, Health care customer value cocreation practice styles, *Journal of Service Research* **45** (2016), 370-389.

[13] M.B. Miles, A.M. Huberman, and J. Saldaña, *Qualitative Data Analysis: A Methods Sourcebook*, *3rd Edition*, Sage Publications, Thousand Oaks, California, USA, 2014.

[14] NICT, Nasjonal IKTs strategiplan for 2013-2016, (2012). https://nasjonalikt.no/documents/saksdokumenter for

styremøte og pf/strategi nikt 2013 - 2016.pdf (accessed May 28, 2015).
[15] E.I. Niemi, and S. Pekkola, Enterprise Architecture Benefit Realization: Review of the Models and a Case Study of a Public Organization, *ACM SIGMIS Database: the DATABASE for Advances in Information Systems* **47** (2016), 55-80.
[16] Norwegian Center for Electronic Patient Records, EPR Monitor: Overview of prevalence and use of ICT in healthcare services, (2009). http://hiwiki.idi.ntnu.no/images/7/79/EPJ-monitor-2008-hovedrapport.pdf (accessed May 28, 2015).
[17] J.W. Ross, P. Weill, and D. Robertson, *Enterprise architecture as strategy: Creating a foundation for business execution*, Harvard Business Review Press, Boston, MA, United States, 2006.
[18] W.R. Scott, Managing Professional Work: Three Models of Control for Health Organizations, *Health Services Research* **17** (1982), 213-240.
[19] T. Thune, and A. Mina, Hospitals as innovators in the health-care system: A literature review and research agenda, *Research Policy* **45** (2016), 1545-1557.
[20] S. Timmermans, and M. Berg, Standardization in action: Achieving local universality through medical protocols, *Social Studies of Science* **27** (1997), 273-305.
[21] V. Venkatesh, H. Bala, S. Venkatraman, and J. Bates, Enterprise architecture maturity: the story of the veterans health administration, *MIS Quarterly Executive* **6** (2007), 79-90.
[22] K.E. Weick, K.M. Sutcliffe, and D. Obstfeld, Organizing for high reliability: Processes of collective mindfulness, *Research in Organizational Behavior* **21** (1999), 81-123.
[23] S. Weiss, S. Aier, and R. Winter, Institutionalization and the effectiveness of enterprise architecture management, in *Thirty Fourth International Conference on Information Systems*, Association for Information Systems, 2013, 1 - 19.

Address for correspondence

Associate Professor Polyxeni Vassilakopoulou
Information Systems Department, University of Agder
Universitetsveien 25, 4630, Kristiansand, Norway
Email: polyxenv@uia.no

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190306

Standardizing Key Issues from Hospital Through an Electronic Multi-Professional Discharge Checklist to Ensure Continuity of Care

Anne Kuusisto[a], Anne Joensuu[b], Minna Nevalainen[c], Terhi Pakkanen[b], Paula Ranne[b], Juha Puustinen[b,d,e]

[a] *Administration Centre, Satakunta Hospital District, Pori, Finland,*
[b] *Medical Care Area, Satakunta Hospital District, Pori, Finland*
[c] *Regional Council of Satakunta, Pori, Finland*
[d] *Pori Social Security Center, Pori, Finland*
[e] *Faculty of Medicine, University of Helsinki, Helsinki, Finland*

Abstract

This article describes development of the multi-professional discharge checklist and its implementation into the nursing documentation system (NDS) as part of the patient's overall care plan. The aim was to harmonize patient's admission and care period documentation and to improve the quality of electronic nursing discharge summaries. The ultimate goal was to ensure continuity of care. The multidisciplinary discharge checklist was developed in two phases to support the discharge of elderly patients (over 65 years). First, the information content of the checklist was defined, and second, it was integrated into the NDS. Focus groups of social and healthcare professionals (n = 82) in specialist health care, primary health care and social services defined the information content and participated in the feedback and checking rounds. The development work should continue. Particular attention should be given to the technical performance of discharge checklists in the NDS.

Keywords:
Continuity of Patient Care, Checklist, Patient Discharge

Introduction

Continuity of patient care is not always realized when the patient moves from hospital to home or primary health care; as a result, up-to-date information is not available to the professionals [1]. Poor communication and lack of documented guidelines or checklists have increased the incidence of handoff errors [2] and adverse events, with adverse events associated with information and medication being the most common [3]. The lack of standardization of information availability forces professionals to search for information in multiple places and pages of documents, which is not an efficient use of professionals' time and adds to their cognitive load [4]. Standardization of these procedures is thus a Joint Commission requirement for accredited hospitals [5].

In Finland, nursing documents are produced, stored and presented using the Nursing Documentation System (NDS) which is part of the Electronic Health Record (EHR). The Finnish nursing model applied is based on the nursing process, a nationally defined nursing core data set and the Finnish Care Classification (FinCC). The FinCC consists of the Finnish classification of nursing diagnoses (FiCND), the Finnish classification of nursing interventions (FiCNI) and the Finnish classification of nursing outcomes (FiCNO) [6,7]. An important objective of the implementation of the nursing model has been to enable multi-professional collaboration and exchange of information between health professionals [8].

An electronic nursing discharge summary (ENDS), which is part of the Finnish nursing model, is recorded at the end of the care period or as a mid-term review in long-term care [6,7]. It is stored in the Patient Data Repository and available for professionals and patients [9]. The aim of the ENDS is to ensure the continuity of patient care when the care responsibility passes from one organization to another, and to gather the key information of the service event into a readable and comprehensible form [6]. According to a recent study, nurses working in primary health care valued the ENDS's efforts to improve information flow and cooperation, but the data content was estimated to be poor [1].

Safe discharge from the hospital is an important way of supporting home living and safe rehabilitation after hospitalization. A recently published systematic review stated that inadequate or delayed information transfer between hospital and primary health care professionals was common [10]. To facilitate the interface between specialized health care and primary health care, so-called discharge checklists are recommended [11-16]. The checklist is a tool for systematic work. It seeks to prevent memory lapses and to avoid human error [17,18]. They include important issues related to the patient's discharge [17,19], such as things that describe the patient's functional performance ability [20,21].

Discharge checklists have made it possible to ensure that all relevant issues related to the discharge have been taken into account [12]. Acute medicine discharge checklist has helped to add structure to the complex interprofessional communication which is essential to safe discharge transitions [22]. In a systematic review, they have been found to improve the quality of medical case summaries [16]. Patient's own checklist that engages the patient can make discharge safer by targeting gaps in patient education and correcting potential adverse events [23].

Checklists remain paper-based [23] and their functionality is limited [15,17]. Recently, the integration of the electronic discharge checking lists, especially in medical processes, has been shown to be beneficial [15,18]. Technologies must be highly usable to ensure high-quality and safe delivery of care without unnecessary increases in workload [24]. Nurses have had to integrate data and information for summaries from EHRs and devices by hand, typically by remembering data [25]. Technical implementation of checklists in the EHR system has

proved to be challenging. Extra difficulties have been caused by the fact that they have had to be added separately "by hand" to each patient's patient record [15]. Besides causing extra work, transferring and duplicating patient data in the patient information system manually is prone to changes in information [26] and thus a threat to patient safety [3]. Structured documentation has improved the quality of documentation, i.e., recording of nursing activities and results, supported the exchange of information, and contributed to the continuity and coordination of care and the reuse of information [27].

According to a systematic review, mere transfer of the discharge checklist into electronic format may not ensure its successful implementation. In order for it to be successful, it is essential to integrate it into existing work processes [18]. Hospitals should develop documented guidelines, promote cooperation and clarify work processes [2] so that the information content of the discharge checklists can be immediately updated and shared among the various occupational groups involved in the care of the patient [15].

According to our knowledge, a multidisciplinary discharge checklist has not previously been implemented in the day-to-day Nursing Documentation System. The purpose of this article is to describe the development of the information content of the multiprofessional discharge checklist into a NDS to harmonize patient's admission and care period documenting and to improve the quality of electronic nursing discharge summaries. The ultimate goal is to ensure continuity of care.

Material and Methods

Setting

In Finland, municipalities are responsible for organizing social welfare and primary health care. Hospital districts (n = 21) organize specialized health care [28]. This development project was carried out in the Satakunta Hospital District (SHD) in 2014–2016 as part of a project funded by the Ministry of Social Affairs and Health for the discharge and rehabilitation processes of older people. In 2018, SHD provides specialized health care services to approximately 223,000 residents in co-operation with primary health care and social services [29]. The NDS has been in use in SHD since 2015.

The multidisciplinary discharge checklist was developed within the project in two phases to support safe discharge of elderly patients (over 65 years). In the first phase, the information content of the discharge checklist was defined, and in the second, it was integrated into the NDS.

Phase I. Definition of the Data Content of the Multi-Professional Discharge Checklist

The development of the information content of the discharge checklist was preceded by a comprehensive literature review on the continuity of care [1]. In addition, we acquired some examples of paper checklists used elsewhere in Finland. The data content of the discharge checklist was developed using focus group interviews in workshops related to a development project in elderly care. This method can be used at the beginning of development projects, whereby focus groups can highlight, for example, the wishes or needs concerning the information content of the system or the user interface solution[30].

Focus groups of social and healthcare professionals (n = 82) in specialist health care, primary health care and social services defined the information content of the discharge checklist and participated in the feedback and checking rounds. Participants in the working groups were found and selected on the basis of volunteering and interest expressed. The sizes and configurations of the focus groups at the meetings varied (Table 1).

*Table 1 – Composition of the Multidisciplinary Team (n = 82) by Profession**

Professional title	n
Nurse	35
Doctor	11
Physiotherapist	6
Rehabilitation counselor	6
Social worker	6
Practical nurse	5
Home care supervisor	2
Senior manager of the elderly	2
Project worker	2
In addition, one representative of each: occupational therapist, fitness instructor, senior care supervisor, service counselor, psychologist, nurse and elderly care professional	7

* The number of professional groups is indicative

The aim of the workshops was to find out key information for the patient's discharge to ensure continuity of care from the perspective of different professionals and organizations. The workshops were designed in a multi-professional team. There were always two team leaders, one asking questions and the other taking notes. In the workshops, small group work based on patient cases and free-form ideas was used.

The material consisted of examples of paper checklists used elsewhere, written output from workgroups, and notes from team leaders. The material was analyzed and categorized thematically. The results made up the thematic areas of the discharge checklist.

The paper discharge checklist was piloted between July 1, 2015 and May 16, 2016 in several units in specialist health care and primary health care at SHD. The checklist was updated three times. The functionality of the paper checklist was evaluated using the so-called principles of continuous evaluation (oral, written, Webropol®).

Phase II. Integrating the Multidisciplinary Discharge Checklist into Nursing Documentation System (NDS)

The fourth version of the checklist was implemented into special health care NDS as a template for the electronic form. The use of a template is recommended to improve the usability of NDSs, multi-professional co-operation and the utilization of information [8,31,32]. The template content was constructed from the FinCC classification [6,7]. The template contains the default parts. It can also be edited for individual patients [33].

An electronic discharge checklist was tested between May 16, 2016 and September 30, 2016 in specialist health care in elderly psychiatry, rehabilitation, pulmonary diseases, neurology and internal disease ward patients over 65 years of age because it was a project of older people. The use of the checklist was evaluated face-to-face in discussions and feedback sessions (n = 4). Additionally, five random samples were taken from each of the units participating in the pilot (n = 5) to assess whether the checklist was filled or used and whether the information was transferred to the electronic nursing discharge summary.

Results

Information Content of the Multi-Professional Discharge Checklist

After continuous dialogue with feedback and three check rounds, the content of the paper-based multi-professional discharge checklist consisted of the following content areas: (1) housing, home care and follow-up care, (2) functional ability at discharge, (3) medication, (4) social benefits and (5) patient involvement.

Multi-Professional Discharge Checklist as Part of NDS

The contents of the paper-based discharge checklist were transferred in structured form, as a so-called template, into the NDS (Figure 1). The categories corresponding as closely as possible to the data content were selected from the FinCC classification. The template serves as a plan for co-ordination of care and follow-up care as well as a daily documenting base.

At the beginning of the discharge checklist, a space is reserved for the planned discharge date. It is recommended that the discharge date is recorded immediately at the beginning of the care period to ensure a safe and systematic discharge.

The planned care activities in the structured discharge checklist are shown in bold in Figure 1. They are in accordance with the Finnish classification of nursing interventions (FiCNI version 3.0): (1) involvement in planning and implementation of care, (2) supporting patient coping, (3) supporting patient self-help, (4) providing aids for day-to-day activities, (5) orientation tracking, (6) mental state monitoring, (7) pharmacotherapy and (8) planning for continuity of care. After the end of the project, nutrition-related guidance was added to the discharge checklist.

The planned care activities, such as orientation tracking, are specified by keywords (oriented/forgetful). The contents of the discharge checklist can be modified in accordance with each patient's personal care needs and a particular care activity can be included in the NDS as an exclamation mark (for example: blood glucose drops easily <3.0 mmol/L). (Figure 1.)

CARE PLANS

COORDINATION OF CARE AND CONTINUITY OF CARE, NEED FOR FURTHER CARE

30.10.2018-

Objective of care: planned discharge date:

Planned care activities

Involvement in planning and implementation of care: Patient/relative/sender

Supporting patient coping: Type of accomodation/home life/services/social situation/family carer

Supporting patient self-help: Movement/toilet functions/personal hygiene/dressing/eating

Providing aids for day-to-day activities: Aids/care accessories

Orientation tracking: Oriented/forgetful

Mental state monitoring: Normal/depressed/anguished

!***Pharmacotherapy**: Patient/relative/continued care/**WARFARIN**/medicines as needed

Planning for continuity of care: Information on the discharge/further examinations/instructions for further care

Figure 1 – The Information Content of the Electronic Multi-Professional Discharge Checklist (FiCNI, version 3.0)

Discussion

In this paper, we described how the information content of a multi-professional discharge checklist was developed into NDS to harmonize documenting patient's admission status and hospital period and to improve the quality of electronic nursing summaries. The ultimate goal was to ensure continuity of care. The starting point for this development project was that *"cannot cope at home"* is not a sufficient description of admission to hospital. Instead, the preparation of the patient's discharge should be started as early as possible so that the patient and his/her family or friend can stay at home as well as possible after acute care. For example, the patient's ability to function or any need for aids or care accessories should be clarified as early as possible.

Improving effectiveness of communication among caregivers and the safety of medication use have been included as items for hospital accreditation [5]. The use of structured documentation has improved the quality of the documentation [27], and the use of discharge checklists can reduce the potential for preventable adverse events associated with transfer of data [2]. Standardized checklists have previously shown benefits in patient care, especially in medicine. They have helped to standardize and harmonize good practices in the organization of continuity of care [13,16,23].

Previously, professionals had to search numerous pages of documentation for information that might not be there [4]. Matters that are relevant for the discharge are often only in the hands of professionals. This introduces patient safety concerns of increased risk for errors [3]. According to our knowledge, an electronic multi-professional discharge checklist is not available elsewhere in NDS. In this project, it was desired that the content of the discharge checklist in NDS was constructed from the patient's point of view and that it would be generic to the social and healthcare professionals involved in the patient care regardless of profession. The aim was to avoid situations where no single document contains all the information needed for the patient's discharge.

The discharge checklist was implemented as a template on the NDS, based on the recommendation of an IT vendor. The categories describing the data content of professionals were found quite well from the FinCC classification [6,7]. On the other hand, there were problems with the compilation of the data content into electronic format. We knew that lack of written guidelines has led to errors in patient transfer situations [2]. That is why we provided end users with technical and substantive written guidelines in support of the discharge checklist. The guidelines for using and filling in the discharge

list were not detailed because different units care for patients with various disorders and the patients have different personal care needs.

In our project, the technical implementation of the discharge checklist into the NDS was challenging. Past experience has shown that usability problems with these records can have unintended consequences that harm patients and cause additional workload for nurses and other clinicians [24]. In our project, professionals made many suggestions for improvement which were passed on to the IT system supplier. Based on initial user experience, they suggested, for example, that the discharge checklist should open automatically for all patients over the age of 65. Furthermore, it was recommended that the position of the discharge checklist that has been added to the NDS should be interchangeable. The information contained in the checklist should be automatically transferred to the electronic nursing discharge summary. End users also hoped that the content of the discharge checklist could be "hidden" and opened as a drop-down menu as display space is limited [compare 15].

A systematic review by Kattel et al. [10] has suggested dynamic documentation development and implementation of electronic discharge software that can automatically populate sections of discharge summariers. These recommendations are noteworthy. Patient data transfer within the information system is a risk for patient safety [26]. Unless technology supports the work processes [18], commitment to the discharge checklist can be challenging and the desired benefits will be missed.

Strengths and Limitations

The representativeness of the material was good. The workshops had a wide representation of different professional groups in specialist health care, primary health care and social work. It is noteworthy that in this project, customers and patients (families) did not participate in defining the data content of the discharge checklist although the original aim was to build a content-enhancing patient perspective.

Recommendations for Future Work

Patient discharge cannot be designed solely for professionals. In the future, patients and families should be engaged and included in the development work. It would also be desirable to link patients' self-generated entries to the discharge checklist. The contents of the checklist and its introduction should also be looked at in relation to other retrieval situations than discharge from the special health care ward (e.g. from nursing institution to hospital).

Conclusions

In this article, we described how the key issues from the hospital were standardized through an electronic multidisciplinary discharge checklist in NDS to ensure continuity of care. We found that based on preliminary user experience, the electronic multidisciplinary discharge checklist was perceived as a useful tool. More research is needed, however. We suggest that development work should continue. Technological developers should be aware of the different functional needs that must be taken into account when data contents are configured in the EHR system. In the future, particular attention should be focused on the technical performance of discharge checklists in the NDS.

Acknowledgements

We gratefully acknowledge the funding of this research by Satakunta Hospital District (EVO 83013) in Finland.

References

[1] A. Kuusisto, Securing the continuity of patient care by means of an electronic nursing discharge summary, Dissertation, University of Eastern Finland, 2018.

[2] M. Yu, H.Y. Lee, G. Sherwood, and K. Eunman, Nurses' handoff and patient safety culture in perinatal care units: Nurses' handoff evaluation and perception of patient safety culture at delivery room and neonata unit in South Korea, J Clin Nurs **27** (2018), e1442-e1450.

[3] C. Byrne, H. Sierra, and R. Tolhurst, Does a checklist reduce the number of errors made in nurse-assembled discharge prescriptions? Br J Nurs **26** (2017), 464-467.

[4] P. Sockolow, R. Hellesø, and M. Ekstedt, *Digitalization of patient information process from hospital to community (home) care nurses: international perspectives*, in: A.K. Rotegård et al., eds, International Medical Informatics Association (IMIA) and IOS Press, 2018.

[5] The Joint Commission on the Accreditation of Healthcare Organizations, National Patient Safety Goals Effective January 2018: Hospital Accreditation Program, 2018.

[6] P. Liljamo, U.M. Kinnunen, and A. Ensio, FinCC classification system, user's guide. FiCND 3.0, FiCNI 3.0, FiCNO 1.0, National Institute for Health and welfare (THL), Classifications, terminologies and statistic guidelines 2/2012, Helsinki, 2012.

[7] U.M. Kinnunen, K. Junttila, P. Liljamo, A.L Sonninen, M. Härkönen, and A. Ensio, FinCC and the National Documentation Model in EHR - User Feedback and Development Suggestions, Stud Health Technol Inform **201** (2014), 196-202.

[8] A. Kuusisto, P. Nykänen, and J. Kaipio, Usefulness of nursing documentations in multi-professional collaboration and information exchange in Finland, J Nurs Care **3** (2014).

[9] Kanta services, available at: https://www.kanta.fi/en/what-are-kanta-services, accessed 21 October 2018.

[10] S. Kattel, D.M. Manning, P.J Erwin, H. Wood, D.T. Kashiwagi and M.H. Murad, Information transfer at patient discharge: A systematic review, J Patient Saf, available at: https://www.ncbi.nlm.nih.gov/pubmed/26741789, accessed 21 October 2018.

[11] T.A. Cortes, S. Wexler, and J.J. Fitzpatrick, The transition of elderly patients between hospitals and nursing homes – improving nurse-to-nurse communication. J Gerontol Nurs **30** (2004), 10-15.

[12] H.D. Hadjistavropoulos, S. Garrat, J.A. Janzen, M.D. Bourgault-Fagnou, and K. Spice, Development and evaluation of continuity of care checklist for improving orthopaedic patient discharge from hospital, Orthop Nurs **13** (2009), 183-193.

[13] C. Soong, S. Daub, J. Lee, C. Majewski, E. Musing, P. Nord, R. Wyman, G.R. Baker, N. Zacharopoulos, and C.M. Bell, Development of a checklist of safe discharge practices for hospital patients, J Hosp Med **8** (2013), 444-449.

[14] A. Cresswell, M. Hart, O. Suchanek, T. Young, L. Leaver, and S. Hibbs. Mind the gap: Improving discharge communication between secondary and primary care, BMJ Qual Improv Rep **4** (2016).

[15] T. Garg, J.Y. Lee, K.H. Evans, J. Chen, and L. Shieh, Development and evaluation of an electronic health record-based best-practice discharge checklist for hospital patients, Jt Comm J Qul Patient Saf **41** (2015), 126-131.

[16] M. Unnewehr, B. Schaaf, R. Marve, J. Fitch, and H. Friederichs, Optimizing the quality of hospital discharge summaries – a systematic review and practical tools, Postgrad Med **127** (2015), 630-639.

[17] N.A. Harun, A.Y. Finlay, M. Salek, and V. Piguet, The development and clinical evaluation of a ´traffic-light´design dermatology outpatient discharge information checklist, Br J Dermatol **175** (2016), 572-582.

[18] H.S. Kramer and F.A. Drews, Checking the lists: A systematic review of electronic checklist use in health care, J Biomed Inform **71** (2017), S6-S12.

[19] T.P. Diaz-Montes, L. Cobb, O.A. Ibeanu, P. Njoku, and M.A. Gerardi, Introduction of checklists at daily progress notes improves patient care among the gynecological oncology service, J Patient Saf **8** (2012), 189-193.

[20] S. Al-Damluji, K. Dzara, B. Hodshon, N. Punnanithinont, H.M. Krumholz, S.I. Chaudhry, and L.I. Horwitz, Hospital variation in quality of discharge summaries for patients hospitalized with heart failure exacerbation, Circ Cardiovasc Qual Outcomes **8** (2015), 77-86.

[21] P. Scully, B. O'Donnel, C. Petrs, M. O'Connor, and D. Lyons, Older patient hospital admissions following primary care referral: the truth is in the referring, Ir J Med Sci **185** (2016), 483-491.

[22] K. Drake, M. McBride, J. Bergin, H. Vandeweerd, and A. Higgins, Ensuring safe discharge with a standardized checklist and discharge pause, Nursing **47** (2017), 65-68.

[23] M. C. Gao, P.B. Martin, J. Motal, L.F. Gingras, C. Chai, M.E. Maikoff, A.M. Sarkisian, N. Rosenthal, and B.M. Eiss, A multidisciplinary discharge timeout checklist improves patient education and captures discharge process errors, Qual Manag Health Care **27** (2018), 63-68.

[24] K.D. Lopez, and L. Fahey, Advocating for greater usability in clinical technologies, Crit Care Nurs Clin North Am **30** (2018), 247-257.

[25] G. Alexander and N. Staggers, A systematic review of the designs of clinical technology. Findings and recommendations for future research, ANS Adv Nurs Sci **32** (2009), 252- 279.

[26] V. Jylhä, Information management in health care: a model for connecting information culture and patient safety, Dissertation, University of Eastern Finland, 2017.

[27] K. Saranto, U.M. Kinnunen, E. Kivekäs, A.M. Lappalainen, P. Liljamo, E. Rajalahti, and. H. Hyppönen, Impacts of structuring nursing records: a systematic review, Scand J Caring Sci **28** (2014), 629-647.

[28] Ministry of Social Affairs and Health, Social welfare and health care system in Finland, responsibilities, available at: https://stm.fi/en/social-and-health-services/responsible-agencies, accessed 21 October 2018.

[29] Satakunta Hospital District, Best service and trusted care for you (video), available at: http://www.satshp.fi/Sivut/default.aspx, accessed 21 October 2018.

[30] D. McGonigle, K. Mastrian, and J.A. Effken, The Human Technology Interface, in: D. McGonigle and K.G. Mastrian, eds., Jones & Bartlett Learning, 2018.

[31] P. Nykänen, J. Viitanen, and A. Kuusisto, Evaluation of the national nursing model and four nursing documentation systems in Finland – lessons learned and directions for the future, Int J Med Inform **81** (2012), 507-520.

[32] A. Kuusisto, J. Kaipio, and P. Nykänen, The National Nursing Model from the nursing practice perspective – Results from a Finnish evaluation study, available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3799135/, accessed 21 October 2018.

[33] J.R. Treadwell, S. Lucas, and A.Y. Tsou, Surgical checklists: a systematic review of impacts and implementation, BMJ Qual Saf **23** (2014), 299-318.

Address for Correspondence

Anne Kuusisto PhD, RN
Satakunta Hospital District
Sairaalantie 3
28500 PORI
FINLAND
Email: first name.surname@satshp.fi

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190307

Contents of Informational and Management Continuity of Care

Anne Kuusisto[ac], Paula Asikainen[ab], Kaija Saranto[c]

[a] *Administrative Center, Satakunta Hospital District, Pori, Finland*
[b] *University of Tampere, Tampere, Finland*
[c] *Department of Health and Social Management, University of Eastern Finland, Kuopio, Finland*

Abstract

Continuity of patient care (COC) is considered an essential feature of good quality care, but the ambiguity of the concept has given rise to methodological challenges in scientific studies. This study has a strong link to the functional definitions of electronic health records (EHR). In order to evaluate how COC is achieved, through a discharge summary, for example, the contents of COC should be defined. Conceptual consensus on COC as a multidimensional concept has increased. This study was conducted to provide an overview of the dimensions and descriptions of informational and management continuity of care. A scoping review was conducted. We found that informational continuity of care refers to data tool, data content, data structures or information quality related processes. Management continuity of care refers to information flow, co-operation, co-ordination, multiprofessionality or management processes. We identified the need to define next the contents of relational and cross-border continuities.

Keywords:

Continuity of Patient Care, Electronic Health Records, Patient Discharge Summaries (MeSH)

Introduction

Continuity of patient care (COC) is considered an essential feature of good quality care, but the ambiguity of the concept has given rise to methodological challenges in scientific studies. COC is more often default than defined. [1] This study has a strong link to the functional definitions of electronic health records (EHR). Unless COC is clearly defined, it can not be measured [1; 2], achieved [3], or continuity-promoting methods [4] such as shared and synchronized EHRs can not be developed [5].

Continuity of care is a global priority for reorienting health services to the needs of people [5]. Through COC, the reduction in the risk of re-hospitalization is significant both for individual patients and in terms of the effectiveness of the service system [6; 7]. COC is associated with information flow, collaboration [2], patient safety issues [8], improved care results [9] and reduced costs [10].

Surprisingly, COC is a broad concept that has been loosely defined and used without a stable or clear aim [2]. This, in turn, leads to challenges when it comes to identification of effective methods for improving COC [4]. Originally identified in Canada [1], the classification into relational (also called inter-personal), informational and management continuity of care has been quoted a great deal [e.g. 2; 5; 11; 12; 13; 14; 15]. Conceptual consensus on COC as a multidimensional concept has increased, but only single standardized measures exist [13]. Different measures are needed to measure different dimensions of COC [2]. The vagueness of the methods of measuring COC makes it difficult to compare studies [9].

Electronic nursing discharge summary (ENDS) is a data tool that is supposed to comprise shared use of information to help maintain patient continuity and safety, collaboration between professionals, and thus provide good care results [15]. In order to evaluate how COC is achieved when using ENDS, the concept of COC needs to be defined. It is obvious that there is a need to determine explicitly what informational and management continuity of care mean. The purpose of this paper is to provide an overview of the dimensions and descriptions of informational and management continuity of care based on previous studies. The research questions are: (1) What are the dimensions and the descriptions of informational continuity and (2) What are the dimensions and descriptions of management continuity. The aim is to present the contents of informational and management continuity of care.

Methods

A scoping review, which is a systematic approach for synthesizing research evidence, was conducted in 2016 (Figure 1). Searches were updated in 2017–2018. The scoping review was used to investigate the scope, nature and gaps in the COC research and literature and to identify and present relationships between concepts [16].

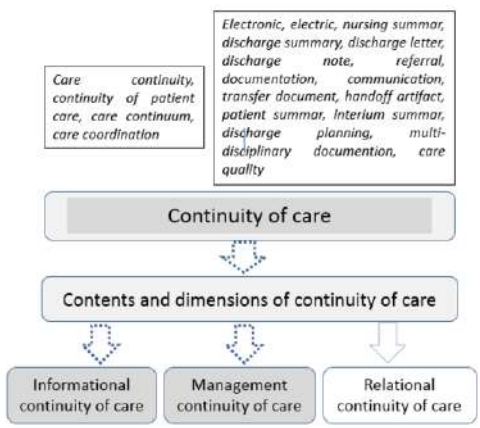


Figure 1. Search Areas and Search Terms for Databases and Search Services

Table 1. The Dimensions and Descriptions of Informational Continuity of Care

Data tool	Data content	Data structures	Information quality
Paper[18, 19]	Administrative data[32]	Standards-compliant interoperability[30, 31, 32]	Spelling and grammar[18]
Telephone[18]	Demographic data[32]	Structural change of information[15, 30, 31, 34, 36, 37]	Sufficiency of information[18]
Mail and e-mail[6, 20]	Clinical data[32]	Need for structuring[20, 36]	Correctness of information[30]
Fax[6, 20]	Medication data[13, 20, 24, 26]	The benefits of structuring[20, 37]	Patient-oriented approach[18, 33]
Dictation[20]	Basic patient knowledge[19, 20]	Challenging structuring[15]	
Video[21]	Identification of provider[19, 20]	Lack of structuring[36]	
Electronic summary[20, 22]	Incomplete recording of investigation results[6, 20]		
Automatic summary[23, 24]	Need for care[33]		
The patient carries a summary[20]	Goals for care[33]		
Telecommunications[25]	Nursing interventions[33]		
E-message system[26]	Care outcomes[33]		
Regional eHealth network[27, 28]	Multiprofessional information content[15]		
Regional information system[20, 29]	Patient summary[34, 35]		
National patient records archive[30, 31]			

Systematic searches were made in relevant databases (PubMed, Cochrane, Science Direct and CINAHL Complete) and search services (Google and Google Scholar) without search restrictions and with 2018 as start time limit. The data search was supplemented by a manual search to ensure comprehensive retrieval of information. The search terms were used alone and in combination (Figure 1). The abstracts of retrieved references were studied. The scoping review was not limited to particular countries, but only included literature published in Finnish or English.

When it was found that conceptual clarity of COC had been sought in multidimensional models, in particular the continuity of care trilogy [1; 5; 11; 13], searches were targeted at literature on the informational and management dimensions of COC. The aim was to refine the contents of COC and continue with the definition of the concepts. The literature was tabulated by author, publication year, country, purpose, material acquisition method, target group, time and key results. Detailed tables are available online at http://epublications.uef.fi/pub/urn_isbn_978-952-61-2707-1/ [15].

Inclusion criteria were articles about dimensions and descriptions of informational and management continuity of care during the patient transfer process. Peer-reviewed studies as well as expert articles and reports were included to ensure a meaningful and comprehensive review of the literature. The exclusion criteria were unpublished manuscripts (abstracts) and articles on the relational continuity of care, because this study focuses on the professional perspective.

The material was analyzed thematically [17]. After compilation and encoding, the material was compared and the dimensions of informational and management COC (e.g. data tool) and their descriptions (e.g., paper or emails) were searched and synthesized.

Results

What are dimensions and descriptions of informational continuity of care?

Table 1 shows the *dimensions and descriptions of informational continuity* of care based on the choices made by the researcher in the literature. The dimensions are data tool, data content, data structures and data quality.

There are many types of data tools. Information is forwarded on paper [e.g. 18; 19], by phone [e.g. 18], mail, e-mail or fax [6; 20]. Information is generated by dictation [20] and using video [21]. Electronic summaries are used for information sharing between organizations [20; 22] and the automatic composition of the data stored in EHR is sought [23; 24]. Summaries can be conveyed by the patient [20] or e.g. by means of telecommunication [25]. The common infrastructure between hospital and primary health care is sought from the e-message system [26], regional eHealth network [27; 28], regional information system [20; 29] and national patient records archive [30; 31].

Data content of the discharge summaries consisted of administrative, demographic and clinical data [32] and was found to be inadequate in medication [13; 20; 24; 26], basic patient knowledge, identification of the provider [19; 20] and investigation results [6; 20]. Information was insufficient regarding care needs, goals, nursing interventions and outcomes [33]. The discharge summaries written by doctors and nurses were professional-specific and partly overlapping [15]. One solution could be a patient summary [34; 35]

Structuring of information requires standards-compliant interoperability [32] and was studied from the perspective of its need [20; 36], benefits [20; 37], challenges [15] and deficiencies [33].

Description of the quality of the information refers to spelling and grammar, sufficiency [16], correctness of information [28] and patient-oriented approach [16; 30].

What are the dimensions and descriptions of management continuity of care?

Table 2 presents the *dimensions and descriptions of management continuity* of care based on the choices made by the researcher in literature. The dimensions are information flow, co-operation, co-ordination, multiprofessionality and management.

Table 2. The Dimensions and Descriptions of Management Continuity of Care

Information flow	Co-operation	Co-ordination	Multiprofessionality	Management
Health IT[14]	Partnership[13]	Coordinator role[13, 14]	Multiprofessional documenting[38]	Resourcing and organization[39]
Timeliness[8, 20, 22]	Networks[14]	Care plan[11, 13, 38]	Multiprofessional discharge planning[41]	Encouragement and support[37]
Information gap[13]	Inadequate cooperation[30]	Follow-up care plan[6, 20, 33]	Multiprofessional cooperation[42]	Politics and decision-making[43]
Reliability [10, 28, 30, 38]	Work practices[27]	Care, service and patient management plan[11]	Multiprofessional information exchange[42]	Knowledge management[44]
	Confidence[39]	Advance care plan[40]	Role clarity[13]	Financial aspects[14]
	Communication[12]	Discharge summary[14, 38]	Confidence in team[13]	
	Understanding other's work[38, 39]	Nursing discharge summary[14, 15, 29]		
		Multiprofessional summary[24, 36]		

The information flow was studied in terms of health IT [14] timeliness [8; 20; 22], information gap [13] and reliability. Reliability was analyzed with regard to access to patient information [38], the confidentiality and security of electronic data [10], sensitive data [28] and privacy risks [30].

Co-operation was looked at in terms of partnership [13], networks [14], inadequate cooperation [30], work practices [27], confidence [13; 39], communication [12] and understanding other's work [38; 39].

Co-ordination descriptions are coordinator role [13; 14] and different care plans, such as care plan [11; 13; 38], follow-up care plan [6; 20; 33], care, service and patient management plan [11] and advance care plan [40]. Different summaries include discharge summary [14; 38], nursing discharge summary [14; 15; 29] and multiprofessional summary [24; 36].

Multiprofessionality was described in the field of documenting [38], but also in terms of multiprofessional discharge planning [41], cooperation, information exchange [42], role clarity and confidence in team [13].

Management in relation to continuity of care was studied resourcing and organizing [39], encouragement and support [37], policy and decision-making [43], knowledge management [44] and financial aspects [14].

Discussion

The purpose of this study was to provide an overview of the dimensions and descriptions of informational and management continuity of care. This study was carried out because we think that the previously published and much cited continuity typology, i.e. information, management and relational continuity [1; 11; 13; 14], do not give a sufficient picture of the dimensions and descriptions of informational and management continuity. The pressure on data interoperability [30; 31; 32] and care integration has increased [3] since the publication of the continuity typology. This study thus supplements the continuity trilogy [1; 11; 13; 14] in terms of the contents of informational and management continuity of care.

Informational continuity (Table 1) has previously been defined as follows: "*The use of information on past events and personal circumstances to make current care appropriate for each individual"* [11]. In nursing literature, information transfer has been the most prominent of the content areas of COC [1]. Technology and different data tools play an important role in enabling informational continuity, and data tools provide access to information. Electronic information exchange between organizations is hampered by lack of interoperability and the fact that patient information systems do not "talk to each other" [42]. Traditional data tools are still in use. Exchange of information from hospital to home care takes place using paper [18] or fax [6; 20; 26], which are not secure data tools. In addition, many referrals from primary health care to specialized healthcare are still paper-based [19].

Informational continuity has been considered to comprise how well patients' health information follows them over time between different places of care and service providers [5; 11]. For years, a common infrastructure for facilitating information exchange between the hospital and the community has been sought from the regional information system [20; 29], eHealth network [27; 28], electronic message system [26], and the national patient records archive [30; 31]. If professionals have to use a number of separate data tools that are not part of the professional workflow they may not be used [27].

The data content of medical case summaries has seen a lot of development. Still, their contents are inadequate [6; 20]. However, there exist standards as to which data should be included in discharge summaries [32]. For the time being, discharge summaries are professional group-based and partly overlapping [35]. The solution could be a patient summary [31; 32] generated automatically from EHR [23; 24]. Getting an automatic summary of the data recorded in the IT system requires standards-compliant interoperability [30; 31; 32].

Descriptions of data structures are seen especially in articles from the 2010s. They have been studied e.g. in patient transfer and discharge situations [37] and nurses' handoffs [35].

Informational continuity is largely a combination of shared, synchronized care records and their accuracy [11]. From the point of view of the quality of the information, especially important are language, sufficiency [18] and correctness [30] of information, as well as the patient-oriented approach [18, 33].

Management continuity (Table 2) has previously been defined as follows: "*A consistent and coherent approach to the management of a health condition that is responsive to a patient's changing needs"* [11]. We found that management continuity of care is related to organization structures and care planning. It is emphasized when the patient moves across organizational boundaries and focuses, for example, on patient management plans [38; 40] and summaries [15; 29]. In Scandinavia and the USA, researchers studied information flow from hospital to home care and found that accurate and in-time information was difficult to achieve [8].

Partnership [13] and networks [14] are an important part of co-operation. From the technology point of view, work practices

involve the need to learn how to work with new data tools [27]. Unless technology supports work processes, commitment to the new data tools (e.g. discharge checklists) can be challenging and the desired benefits will be missed.

Co-ordination requires a coordinator [13; 14] and includes different care plans [6; 11; 13; 20; 33; 38; 40], summaries [14; 15; 24; 29; 36; 38] and workflows. For example, ENDS has promoted information flow, cooperation and getting a complete picture of the client [15]. In Finland, the Patient Data Repository is in use [31]. Today, many patients in need of social and health services can have a number of services, care, rehabilitation or other plans (Table 2).

Multiprofessionality in documenting has long been discussed [36]. Despite this, it still does not work very well [15; 33]. One solution could be role clarity and confidence in team [13].

We found a few references on management support [34] from the point of COC [14; 37; 39; 43; 44].

Strengths and limitations

Scoping review is best suited for examining complex and heterogeneous topics like continuity of care [16]. The reliability of this review was increased by the use of pre-defined inclusion and exclusion criteria and the utilization of methodical literature [16; 17].

An information technician checked the searches made [17]. Data retrieval was continued until references repeatedly recalled the same articles. The analysis proved to be the greatest challenge [see 17] and the reliability of the review may be slightly weakened by the fact that one researcher (AK) made the selection of articles.

Recommendations for the future work

Our findings are a starting point to expand on and provide further insight into the construct. We identified the need to define the contents of relational continuity, i.e., the relationship between the professional and patient and cross-border continuity, i.e., situations where the patient's care responsibilities change.

Conclusions

We found that informational continuity of care refers to data tool, data content, data structures or information quality related processes. Management continuity of care refers to information flow, co-operation, co-ordination, multiprofessionality or management processes. Continuity of care is a concept that underlies many applications in eHealth and is thus one of the core constructs of biomedical and health informatics. Further work is still needed but the foundations are laid with this study.

Acknowledgements

We gratefully acknowledge the funding of this research by Satakunta Hospital District (EVO 83013) in Finland.

References

[1] R. Reid, J. Haggerty, and R. McKendry, Defusing the confusion: concepts and measures of continuity of health care, *Canadian Health Services Research Foundation*, 2002.

[2] A.A. Uijen, H.J. Schers, F.G. Schellevis, and W.J. van den Bosch, How unique is continuity of care? A review of continuity and related concepts, *Fam Pract* **29** (2012), 264-271.

[3] A. Wierdsma, C. Mulder, S. de Vries, and S. Sytema, Reconstructing continuity of care in mental health services: a multilevel conceptual framework, *Journal Health Serv Res Policy*, **14** (2009), 52-57.

[4] M. Smeulers, C. Lucas, and H. Vermeulen, *Effectiveness of different nursing handover styles for ensuring continuity of information in hospitalised patients*, Cochrane database of systematic reviews, **24** (2014).

[5] World Health Organization, *Continuity and coordination of care: a practice brief to support implementation of the WHO Framework on integrated people-centred health services,* World Health Organization, 2018.

[6] S. Kripalani, F. LeFevre, C.O. Phillips, M.V. Williams, P. Basaviah, and D.W. Baker, Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care, *JAMA,* **297** (2007), 831-841.

[7] S. Shepperd, J. McClaran, C.O Phillips, N.A. Lannin, L.M. Clemson, A. McCluskey, I.D. Cameron, and S-L. Barras, *Discharge planning from hospital to home.* Cochrane database of systematic reviews, **20** (2010).

[8] P. Sockolow, R. Hellesø, and M. Ekstedt, *Digitalization of patient information process from hospital to community (home) care nurses: international perspectives*, in: A.K. Rotegård et al., eds, International Medical Informatics Association (IMIA) and IOS Press, 2018.

[9] *S.* Puntis, J. Rugkåsa, A. Forrest, A. Mitchell, and T. Burns, Associations between continuity of care and patient outcomes in mental health care: a systematic review. *Psychiatr Serv,* **66** (2015), 354-363.

[10] M. Alatzri, P. Heywood, R.D. Neal, and B. Leese, Continuity of Care: Literature review and implications. *Sultan Qaboos Univ Med J*, **7** (2007), 197-206.

[11] J.L. Haggerty, R. Reid, G.K. Freeman, B.H. Starfield, C.E. Adair, and R. McKendry, Continuity of care: a multidisciplinary review. *BMJ,* **327** (2003), 1219-1221.

[12] M. Gulliford, S. Naithani, and M. Morgan, What is 'continuity of care'? *Journal of Health Services & Research Policy*, **11** (2006), 248-250.

[13] J.L Haggerty, D. Roberge, G.K. Freeman, C. Beaulieu, and M. Bréton, Validation of a generic measure of continuity of care: when patients encounter several clinicians, *Ann Fam Med*, **10** (2012), 443-451.

[14] M. Przysucha, S. Vogel, J. Hüsers, S. Wache, B. Sellemann, U. Hübner, Requirements for Collaborative Decision Support Systems in Wound Care: No Information Continuity Without Management Continuity, *Stud Health Technol Inform,* **253** (2018), 133-137.

[15] A. Kuusisto, *Securing the continuity of patient care by means of an electronic nursing discharge summary*, Academic dissertation, University of Eastern Finland, 2018, available at: http://epublications.uef.fi/pub/urn_isbn_978-952-61-2707-1/urn_isbn_978-952-61-2707-1.pdf, [accessed on March 24, 2019].

[16] M.D.J. Peters, C.M. Godfrey, H. Khalil, P. McInerney, D. Parker, and C.B. Soares, *Guidance for conducting systematic scoping reviews. Methodology paper.* International Journal of Evidence-Based Healthcare. University of Adelaide, Joanna Briggs Institute, 2015.

[17] R. Whittermore and K. Knafl, The integrative review: updated methodology, *J Adv Nurs*, **52** (2005), 546-553.

[18] R.M. Olsen, O. Hellzén, L.H. Skotnes, and L. Enmarker, Breakdown in informational continuity of care during hospitalization of older home-living patients: a case study. *Int J Integr Care* **12**, 2014.

[19] P. Scully, B. O'Donnell, C. Peters, M. O'Connor, and D. Lyons, Older patient hospital admissions following primary care referral: the truth is in the referring. *Ir J Med Sci* **185** (2016), 483-491.
[20] S. Kattel, D.M. Manning, P.J Erwin, H. Wood, D.T. Kashiwagi and M.H. Murad, Information transfer at patient discharge: A systematic review, *J Patient Saf,* (2016), *available at: https://www.ncbi.nlm.nih.gov/pubmed/26741789,* [accessed on October 21, 2018].
[21] Y. Li, Hospital discharge: Written vs verbal information, *The Joanna Briggs Institute.* JBI Evidence Summary, 2017.
[22] S.F. Murphy, L. Lenihan, F. Orefuwa, G. Colohan, I. Hynes, and C.G. Collins, C.G, Electronic discharge summary and prescription: improving communication between hospital and primary care. *Ir J Med Sci* **186** (2016), 455-459.
[23] B. Polnaszek, J. Mirii, R. Roiland, A. Gilmore-Bykovskyi, M. Hovanes, and A. Kind, Omission of physical therapy recommendations for high-risk patients transitioning from the hospital to subacute care facilities. *Arch Phys Med Rehabil* **96** (2015), 1966-1972.
[24] F.H. Sakaguchi and L.A. Lenert, Improving continuity of care via the discharge summary. *AMIA Annual Symposium Proceedings Archive*, (2015), 1111-1120.
[25] V.A. Crooks and G. Agarwal, What are the roles involved in establishing and maintaining informational continuity of care within family practice? A systematic review. *BMC Fam Pract* **9** (2008).
[26] L. Melby, B. Brattheim, and R. Hellesø, Patients in transition - improving hospital-home care collaboration through electronic messaging: providers´ perspectives. *J Clin Nurs* **24** (2015), 3389-3399.
[27] J. Reponen, E. Marttila, H. Paajanen, and A. Turula, Extending a multimedia medical record to a regional service with electronic referral and discharge letters. *J Telemed Telecare* **10** (2004), 81-83.
[28] T. Schabetsberger, E. Ammenwerth, S. Andreatta, G. Gratl, R. Haux, G. Lechleitner, K. Schindelwig, C. Stark, R. Vogl, I. Wilhelmy, and F. Wozak, From a paper-based transmission of discharge summaries to electronic communication in health care regions, *Int J Med Inform* **75** (2006), 209-215.
[29] K.M. Unertl, K.B. Johnson, and N.M. Lorenzi, Health information exchange technology on the front lines of healthcare: workflow factors and patterns of use. *J Am Med Inform Assoc* **19** (2012), 392-400.
[30] T. Greenhalgh, K. Stramer, T. Bratan, E. Byrne, J. Russel, and H.W.W. Potts, Adoption and non-adoption of a shred electronic summary record in England: A mixed-method case study, *Br Med J* 2010.
[31] *Kanta services*, available at: https://www.kanta.fi/en/what-are-kanta-services, [accessed on October 21, 2018].
[32] American Society for Testing and Materials, *ASTM E2369-12 Standard Specification for Continuity of Care Record (CCR)*, West Conshohocken (PA): ASTM International; c2014, available at: http://www.astm.org/Standards/E2369.htm, [accessed on March 24, 2019].
[33] E. Carlsson, M. Ehnfors, and A. Ehrenberg, Multidisciplinary recording and continuity of care for stroke patients with eating difficulties. *J Interprof Care* **24** (2010), 298-310.
[34] *Smart Open Services for European Patients*, available at: https://old.ezdravotnictvo.sk/en/eHealth_Programme/Slovakia_in_International_eHealth_Projects/Pages/European-Patient-Smart-Open-Services-(epSOS).aspx, [accessed on October 23, 2018].
[35] J. O´Rourke, J. Abraham, L.A. Riesenberg, J. Matson, and K.D. Lopez, A Delphi study to identify the core components of nurse to nurse handoff, *J Adv Nurs* **74** (2018), 1659-1671.
[36] U. Hübner, D. Flemming, K.U. Heitmann, F. Oemig, S. Thun, A. Dickerson, and M. Veenstra, The need for standardised documents in continuity of care: Results of standardising the eNursing Summary. *Studies in Health Technology and Informatics* **160** (2010), 1169-73.
[37] K. Saranto, U.M. Kinnunen, E. Kivekäs, A.M. Lappalainen, P. Liljamo, E. Rajalahti, and. H. Hyppönen, Impacts of structuring nursing records: a systematic review, *Scand J Caring Sci* **28** (2014), 629-647.
[38] M. Kirsebom, B. Wadensten, and M. Hedström, Communication and coordination during transition of older persons between nursing homes and hospital still in need of improvement. *J Adv Nurs* **69** (2012), 88-895.
[39] R.M. Olsen, B.H. Østnor, I. Enmarker, and O. Hellzén, Barriers to information exchange during older patients' transfer: nurses' experiences. *J Clin Nurs* **22** (2013), 2964-2973.
[40] S. Johnson, P. Butow, I. Kerridge, and M. Tattersall, Advance care planning for cancer patients: a systematic review of perceptions and experiences of patients, families, and healthcare providers. *Psychooncology* **25** (2016) 362-386.
[41] K.M. Romagnoli, S.M. Handler, F.M. Ligons, and H. Hochheiser, Home-care nurses' perceptions of unmet information needs and communication difficulties of geriatric patients in the immediate post-hospital discharge period. *BMJ Qual Saf* **22** (2013), 324-332.
[42] J. Kaipio, T. Lääveri, H. Hyppönen, S. Vainiomäki, J. Reponen, A. Kushniruk, E. Borycki, and J. Vänskä, Usability problems do not heal by themselves National survey on physicians' experiences with EHRs in Finland. *Int J Med Inform* **97** (2017), 266–281.
[43] T. Oksholm, T. Rustoen, and M. Ekstedt, Transfer Between Hospitals is a Risk Situation for Patients After Lung Cancer Surgery, *Cancer Nurs* **20** (2017).
[44] V. Jylhä, *Information management in health care: a model for connecting information culture and patient safety,* Academic dissertation, University of Eastern Finland, 2017, available at: http://epublications.uef.fi/pub/urn_isbn_978-952-61-2384-4/urn_isbn_978-952-61-2384-4.pdf, [accessed on March 24, 2019].

Address for correspondence

Satakunta Hospital District
Sairaalantie 3
28500 PORI
FINLAND

Corresponding author:
Anne Kuusisto PhD, RN
first name.surname@satasairaala.fi.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190308

Differences Between What Is Said During the Consultation and What Is Recorded in the Electronic Health Record

Virginie Lacroix-Hugues [a,b], Sarah Azincot-Belhassen[a], Pascal Staccini[c], David Darmon[a]

[a]*Université Côte d'Azur, Département d'enseignement et de recherche en médecine générale, Nice, France.*
[b]*CHU de Nice, Département de Santé Publique, Nice, France.*
[c]*Université Côte d'Azur, Département d'Ingénierie du Risque et Informatique de Santé, Nice, France.*

Abstract

Electronic Health Records (EHRs) can be used for research but this raises the problem of data quality.

Objective: To evaluate the quality of the information recorded in an EHR by a general practitioner (GP) during a regular office consultation.

Method: 191 dialogs between the GP and patient were recorded and translated into the International Classification of Primary Care Second edition (ICPC-2) codes. Written information of the corresponding EHR was extracted and coded for comparison.

Results: The primary reason for the consultation was recorded in the EHR in 41.2% of the cases and the diagnosis in 44.1% of the cases. Diagnoses noted in the EHR were less often communicated to the patients than the primary reasons ($p<0.0001$).

Conclusion: There is a loss of information between the dialog during a consultation and what is reported in the EHR. Consequences in terms of continuity and safety of care can be expected.

Keywords:
Electronic Health Records; Information Management; Self Report.

Introduction

Although the first traces of the medical record appeared in the 9th century with the creation of interesting case registers, the medical records attached to each patient did not appear until the end of the 18th century [1]. However, their contents were significantly limited and the medical record, as we understand it today, dates from the 19th century with the creation of the modern hospital.

In France, the medical record has assumed an important place since the 1970s, but was declared mandatory only in 1995 (deontology code and Public Health Code). Their computerization began in the 1980s and, according to a survey conducted in 2017, 96% of the general practitioners (GPs) reported having software for the management of their patients [2].

This computerization of data provides a rich source of data that can be leveraged to support research and public health [3-5], and large databases in primary care, such as Clinical Practice Research Datalink (CPRD) [6], were created in Europe. This model of data collection is being developed in France with the project PRIMEGE PACA (health data warehouse prepared from the EHRs of GPs in southeastern France using International Classification of Primary Care Second edition coding). However, secondary use of EHR raises the problem of data quality and especially in terms of missing data [7, 8]. While, in French hospitals, quality and safety of care indicators are clearly defined by the High Authority of Health [9] and ensure a certain quality level of the patient record, this is not the case in ambulatory care. Indeed, if the deontology code makes the possession of a patient record mandatory, it specifies that it not be subject to any formalization. In addition, neither the Remuneration about Objectives of Public Health (ROSP) nor the "forfait structure" (funding to facilitate the organization and computerization of the medical office) for GPs mention any element concerning the quality of the EHRs [10].

The objective of this study was to evaluate the quality, mainly in terms of completeness, of the information recorded in an EHR by a general practitioner (GP) during a regular office consultation.

Methods

Protocol of the Study

Nine GPs from the south east of France using the same software were recruited on a voluntary basis. For each doctor, consultations were recorded with a Dictaphone. A total of 191 consultations (one per patient) were recorded between June 2016 and June 2017. EHRs corresponding to these consultations were retrieved for data comparison. Once all the data were collected, they were anonymized.

On the day of the data collection, an information notice was displayed in the waiting room and the patients who agreed to participate had to sign a consent form at the beginning of the consultation.

Data Processing

Firstly, the recordings were analyzed to identify the reasons for the consultation and the diagnosis. These elements were noted in an information grid and were compared with the content of the corresponding EHRs to check whether all data had been recorded. Reasons for consultation and diagnoses present in the EHRs without having been located in the recordings were also noted. The reasons and the diagnoses were classified into three categories: "oral only", "oral and written", "written only" and all the data was recorded in an Excel® file.

To be properly exploited, the collected data had to be standardized. To this end, the International Classification of Primary Care 2nd version (ICPC-2), recognized as the reference classification in primary care by the World Organization of Family Doctors (Wonca) and the World Health Organization (WHO) was used. It has a biaxial structure and consists of 17 chapters [11], each divided into 7 components (comp.) that address symptoms and complaints (comp. 1), process codes (comp. 2), infections (comp. 3), neoplams (comp. 4), injuries (comp. 5), congenital anomalies (comp. 6) and other diagnoses (comp. 7). All the reasons and the diagnoses where coded in ICPC-2 using Prometheus, an encoding help engine [12]. All these stages were performed by a resident in general medicine.

Statistical analysis

Statistical analysis was performed using R version 3.4.3 software (the R Foundation for Statistical Computing). Chi² and Fisher tests were used, and multivariate analysis was conducted using logistic regression. The Alpha risk value was set to 0.05.

Results

Description of GPs and Patients

Nine general practitioners, including five men and four women, were recruited. Their average age was 53.56 years and five were trainer GPs. One GP declared that all consultation elements are noted in his EHR. For the others, the reasons for consultation were primarily noted (75% of GPs), followed by diagnoses (62%) and biometrics (37%).

Of the 191 patients involved in the consultations, 58% were women and 41% were men, and for 1% of patients' gender was not specified in the EHR. They were, on average 51.70 years old, with extreme values of 2 and 90 years old.

Reasons for Consultation

A total of 683 reasons for consultation were identified, which is an average of 3.6 reasons per consultation. Most of them corresponded to ICPC-2 chapter A (general / unspecified: 41.43%), chapter L (musculoskeletal: 11.57%) and chapter R (respiratory: 9.81%) (Fig 1). In more than half the cases, reasons for consultation expressed during the consultation were not recorded in the EHR (375 reasons compared to 638 oral reasons, 58.78%). Those recorded in the EHR had been orally expressed in the majority of the cases (263 of 308 written reasons, 85.39 %).

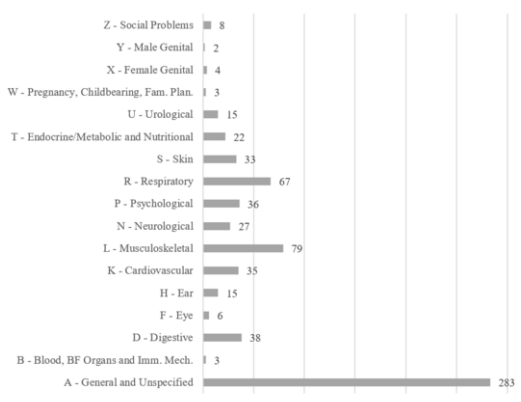


Figure 1: Reasons for Consultation by ICPC-2 Chapters

Diagnoses

A total of 159 diagnoses were identified, which is an average of 0.8 diagnoses per consultation. Most of them corresponded to ICPC-2 chapter A (general / unspecified: 22.01%), followed by chapter L (musculoskeletal: 18.24%) and chapter R (respiratory: 15.09%) (Fig 2). Most of the diagnoses noted in the EHR had not been orally expressed during the consultation (66 of 107 written diagnoses, 61.68%). Those specified during the consultation were recorded in the EHR in less than half of the cases (41 diagnoses compared to 93 oral diagnoses, 44.09%).

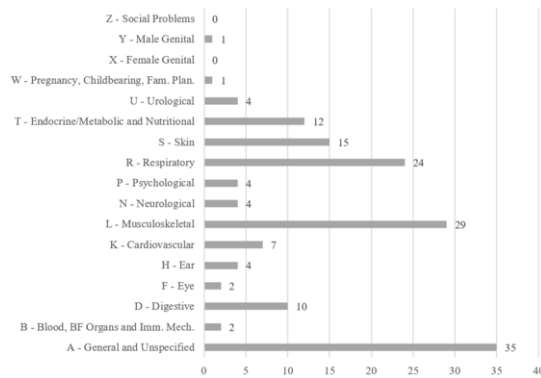


Figure 2: Diagnoses by ICPC-2 Chapters

Analysis

There was no significant difference between the transcription of the reasons for consultation and diagnoses ($p = 0.6527$), but diagnoses were present more often only in the EHR than the reasons ($p < 0.0001$).

Trainer GPs and older physicians had recorded the reasons in the EHR more often (OR = 1.864 [1.302 – 2.680] and OR = 1.052 [1.026 – 1.080]). The psychosocial reasons (chapters P and Z) were less often transcribed than the others (OR = 0.473 [0.217 - 0.965]). Regarding the ICPC-2 components, the reasons corresponding to symptoms and complaints (SPL) or process codes were transcribed more often than the others (OR = 1.834 [1.199 – 2.844]) (Table 1).

It was among new patients that we found the most reasons for consultation present only in the EHR (OR = 4.172 [1.567 – 10.614]). The ophthalmological (F), cardiovascular (K), metabolic / nutritional or endocrine (T), urological (U) or social (Z) reasons were found more often in the EHR without being addressed during the consultation than in other chapters of ICPC-2 (OR = 3.204 [1.387 - 7.122]). Concerning the ICPC-2 components, the reasons corresponding to the categories infections and other diagnoses were most often found only in the EHR (OR = 2.217 [0.981 – 4.781]) (Table 2).

With regard to diagnoses, those in the general / unspecified (A) category were less often recorded in the EHR than the others (OR = 0.263 [0.072-0.823]) (Table 3). We also observed that trainer GPs were less likely to note the diagnoses in the EHR without telling the patient during the consultation (OR = 0.379 [0.160-0.871]) (Table 4).

Table I. Transcription of the Reasons for Consultation

Oral reasons (n = 638)	Not transcribed		Transcribed		p
Patients					
Mean age	54.28 [51.69-56.86]		57.41 [54.62-60.19]		0.4853
GPs					
Gender					
- Female	157	54.70%	130	45.30%	0.4802
- Male	218	62.11%	133	37.89%	
Trainer GP					
- No	203	68.12%	95	31.88%	0.0007
- Yes	172	50.59%	168	49.41%	
Mean age	51.79 [50.68-52.90]		56.28 [55.37-57.19]		0.0001
Reasons for consultation *					
Psycho-social					
- Yes	31	73.81%	11	26.19%	0.0470
- No	338	57.39%	251	42.61%	
ICPC-2 component					
- SPL and proc.±	280	55.78%	222	44.22%	0.0058
- Others	89	68.99%	40	31.01%	

**7 not coded reasons (n = 631)*

± Symptoms/complaints and process codes

Table 2. Verbalization of Written Reasons for Consultation

Written reasons (n = 308)	Oral and written		Only written		p
Patients*					
Seniority					
- Old	247	87.28%	36	12.72%	0.0031
- New	15	62.50%	9	37.50%	
Reasons for consultation ±					
Chap F.K.T.U or Z					
- Yes	27	69.23%	12	30.77%	0.0049
- No	235	87.69%	33	12.31%	
ICPC-2 component					
- Infections / other diagnose	33	71.74%	13	28.26%	0.0474
- Others	229	87.74%	32	12.26%	

**1 reason is excluded because the data set was not available (n = 307)*

±1 not coded reason (n = 307)

Table 3. Transcription of the Diagnoses

Oral diagnoses (n = 93)	Only oral		Oral and written		p
GPs					
Paper files*					
- No	34	64.15%	19	35.85%	0.0808
- Yes	11	40.74%	16	59.26%	
Software considered adapted					
- No	4	100.00%	0	0.00%	0.9891
- Yes	48	53.93%	41	46.07%	
Diagnoses±					
Chapt. A					
- No	28	45.90%	33	54.10%	0.0287
- Yes	21	75.00%	7	25.00%	

**13 unspecified elements (n=80)*

± 4 diagnoses not coded (n = 89)

Table 4. Verbalization of Written Diagnoses

Written diagnoses (n = 107)	Oral and Written		Written only		p
Patients					
Mean age	47.76 [40.11-55.40)		57.17 [51.72-62.62]		0.1440
GPs					
MSU					
- No	15	27.78%	39	72.22%	0.0241
- Yes	26	49.06%	27	50.94%	
Diagnoses*					
ICPC-2 component					
-Infections/ neoplasms	16	55.17%	13	44.83%	0.3660
- Others	24	31.58%	52	68.42%	

** 2 diagnoses not coded (n=105)*

Discussion

Main findings

Reasons and diagnoses were recorded in the EHR in less than half the cases (41 % for reasons and 44% for diagnoses).

Incomplete medical records is a well-known problem. Long [13] compared the medical records of 17 patients with the care received when consulting for a sore throat, and he noticed that the items of the interview and the clinical examination were completed more often than they were noted in the medical record. While prescriptions are generally well recorded, that is not the case for lifestyle data [14,15], and several studies have also demonstrated that doctors have carried out more preventive action than was noted in the medical records [8,16].

Diagnoses were found more often in the EHR without being expressed to the patient than the reasons for consultation. This could be explained by still-hypothetical diagnoses for which the GP wishes to wait for confirmation before informing the patient, in order to not worry the patient unnecessarily.

The teaching activity of trainer GPs could explain their resolve to leave more information in their file and thus to transcribe more reasons for consultation. They also more often communicated their diagnosis to the patient. We can assume that they are more receptive of their obligation to inform the patient of the diagnosis and to the legal obligation to record it in their EHR.

Older physicians may be less concerned about profit and have longer consultations, which allows a more faithful transcription of the oral data. In addition, some may intend to prepare EHRs as complete as possible to facilitate the transition with their successor.

Differences between ICPC-2 chapters were found. Psychosocial reasons for consultation were less often noted in the EHR than in those related to other categories. This could be explained by the GPs favoring discussion and eye contact in this type of situation or by a desire not to record elements that may be stigmatizing for the patient. On the other hand, certain reasons may have been detected during the analysis of the recordings without having been perceived by the GP. Indeed, the difficulty to detect social problems has been described by Bentsen [17]. Conversely, reasons corresponding to the SPL or process codes components were more often transcribed, probably in a process of transparency and clarity for the record

to facilitate subsequent patient care. For diagnoses, those corresponding to ICPC-2 chapter A (general / unspecified) were less often transcribed. This could be related to the lack of specificity of this chapter, which may also correspond to a health concern, a request for information, a request for renewal, etc., that GPs do not consider noteworthy in their EHR.

The first consultation with a new patient is a moment of discovery during which many things are expressed. During this consultation, the GP probably employs a great deal of interpretation in his EHR to draw a clear patient profile, which could explain the presence of a greater number of reasons for consultation that have not been verbalized.

Limits

To prevent the patient from feeling intimidated and to keep the consultation as natural as possible, a Dictaphone was used instead of the researcher attending the consultation and transcribing it directly. However, we cannot rule out an observation effect (effect Hawthorne [18]). Indeed, if the physician is aware of being observed, it may influence the filling habits and behaviors during the consultation.

The recruitment of GPs and patients were made on a voluntary basis and we do not have information concerning the patients who refused registration. This is a measurement bias and we can presume that patients with "sensitive" reasons for consultation, such as risky behavior or psychological problem, for example, refused registration more often.

For the purpose of reproducibility and to limit the biases related to the use of multiple software, we selected GPs with the same software. However, the effect of the software's ergonomics in the filling of the EHR could not be studied.

This study compared only oral/written reasons for consultation and oral/written diagnoses, but we cannot assert that missing data was not recorded in an inappropriate field. For instance, a reason for consultation could have been recorded in the EHR as a diagnosis or a symptom.

Implications

A French Public Health Authority (ANAES) analyzed several studies and concluded that data contained in medical records are for internal use [19]. Doctors use them as a reminder and record only the information they consider necessary, without worrying about accurately recording their actions. However, in 1972, Weed said, "most doctors can remember three problems of a patient; the very good doctors retain five; but the average patient has eleven" [20]. This lack of completeness leads to several problems. Carsley et al. found that drug allergies are not recorded for a large proportion of patients [15]. If there is no information in the EHR, it is difficult to know if the patient does not have allergies, if the practitioner did not look for them or if forgot to record them. This situation raises safety issues. Continuity of care can also be affected. Indeed, Haggerty et al. argue that informational continuity of care represents one of three dimensions of care continuity and is defined as the use of information on past events and personal circumstances to make current care appropriate for each individual [21]. Furthermore, continuity of care has a positive impact on the care of chronically ill patients [22].

The prevalence of chronic diseases continues to grow, and GPs have to face a real "epidemiological transition" [23]. This increase of chronic diseases is associated with an increase in polypathology. A health survey conducted in France between 2002 and 2003 found that the number of reported diseases increased with age: four for 40-64 years old, five for 65-79 years old and six for those 80 years and older [24]. As a result of population ageing, GPs consultations become more intense and more complex. Given the consultation complexity, it seems unrealistic to expect GPs to record all the elements of the consultation. It will be challenging to find the right balance between a certain level of quality and EHR completeness to ensure safety and continuity of care without prolonged consultations.

However, the use of EHRs for evaluation and research will require a higher level of quality. According to Goldberg [25], limiting the risk of error and bias would be possible only under several conditions:

- Systematically record a minimum of information for each patient;
- Define a core of data to record;
- Record information using a recognized classification;
- Control real-time input and verify data (missing data, outliers);
- Check intra-observer and inter-observer variability.

Perspectives

Improving data quality in EHRs could serve two purposes:

- Obtain a minimum data quality to ensure safety and continuity of care;
- Collect data from general practitioners wishing to support research in primary care, and agree to follow a quality chart when recording data in their medical records.

To improve data quality, several options can be considered:

- Train GPs on ICPC-2 to encourage them to code their data and on the proper use of medical software;
- Involve patients so that they can complete their data [26];
- Develop algorithms to assign ICPC-2 codes to free text;
- Develop algorithms to restore some missing data (e.g., a diabetes diagnosis for patients receiving oral antidiabetic drugs).

Qualitative studies could be conducted to identify barriers and facilitators for recording consultation elements in the EHR. Individual feedback has already been identified as a low-cost tool to improve data quality [27,28]. Qualitative studies could also be used to define a quality chart in collaboration with GPs and to validate acceptable and achievable criteria.

Conclusions

There is a loss of information between the dialog during a consultation and what is reported in the EHR. Reasons and diagnoses were recorded in the EHR in less than half of the cases. This lack of completeness can be expected to affect safety and continuity of care.

Reachable and acceptable quality criteria must be defined with French GPs to improve the quality of their EHRs.

References

[1] A. Liever, G. Moutel, Le dossier médical : concepts et evolutions (droits des patients et impact sur la relation soignants-soignés) [last access 2018 oct 18]. Available from: http://www.ethique.sorbonne-paris-cite.fr/sites/default/files/dossier_11.pdf

[2] Les médecins à l'ère du numérique [last access 2018 oct 18]. Available from : https://www.ipsos.com/fr-fr/les-medecins-lere-du-numerique

[3] W.R. Hersh, Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance, *Am J Manag Care* **13** (2007), 277–8.

[4] C. Safran, M. Bloomrosen, W.E. Hammond, S. Labkoff., S. Markel-Fox, P.C. Tang, et al., Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper, *J Am Med Inform Assoc JAMIA* **14** (2007), 1–9.

[5] G.S Birkhead, M. Klompas, N.R. Shah, Uses of electronic health records for public health surveillance to advance public health, *Annu Rev Public Health* **36** (2015), 345–59.

[6] CPRD [last access 2018 oct 20]. Available from: https://www.cprd.com/home/

[7] B.A. Goldstein, A.M. Navar, M.J. Pencina, J.P.A. Ioannidis, Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review, *J Am Med Inform Assoc JAMIA* **24** (2017), 198–208.

[8] T.R. Dresselhaus, J.W. Peabody, M. Lee., M.M. Wang, J.Luck, Measuring Compliance with Preventive Care Guidelines, *J Gen Intern Med* **15** (2000), 782–8.

[9] rapport_2016_dpa_mco_vf.pdf [last access 2018 nov 13]. Available from: https://www.has-sante.fr/portail/upload/docs/application/pdf/2016-12/rapport_2016_dpa_mco_vf.pdf

[10] Rosp et nouveau « forfait structure » : la déclaration des indicateurs ouverte jusqu'au 5 février [last access 2018 nov 13]. Available from: https://www.ameli.fr/medecin/actualites/rosp-et-nouveau-forfait-structure-la-declaration-des-indicateurs-ouverte-jusquau-5-fevrier

[11] WHO | International Classification of Primary Care, Second edition (ICPC-2) [last access 2018 oct 30]. Available from : https://www.who.int/classifications/icd/adaptations/icpc2/en/

[12] Prometheus [last access 2018 nov 14]. Available from : http://www.promethe.org/

[13] T. Long, A limitation in the use of office records for health care evaluation, *Pediatrics* **56** (1975), 295-6.

[14] M. Pringle, P. Ward, C. Chilvers, Assessment of the completeness and accuracy of computer medical records in four practices committed to recording data on computer, *Br J Gen Pract J R Coll Gen Pract* **45** (1995), 537–41.

[15] S. Carsley, C.S. Birken, P. Parkin, E. Pullenayegum, K. Tu, Completeness and accuracy of anthropometric measurements in electronic medical records for children attending primary care, *J Innov Health Inform* **25** (2018), 963.

[16] A. Wilson, P. McDonald, Comparison of patient questionnaire, medical record, and audio tape in assessment of health promotion in general practice consultations, *Br Med J* (1994), 1483–5.

[17] B. Bentsen, The accuracy of recording patient problems in family practice, *J Med Educ* (1976), 311–6.

[18] Effet Hawthorne – Psychologue du Travail [last access 2018 oct 23]. Available from: http://www.psychologuedutravail.com/psychologie-du-travail/effet-hawthorne/

[19] Agence Nationale d'Accréditation et d'Evaluation en Santé, La tenue du dossier médical en médecine générale : état des lieux et recommandations, 1996 [last access 2018 oct 27]. Available from: http://www.has-sante.fr/portail/upload/docs/application/pdf/Dm.pdf

[20] B. Alderson, Help your patients, your colleagues and yourself. A review of medical records, *Aust Fam Physician* **20** (1991), 1004–5.

[21] J.L. Haggerty, R.J. Reid, G.K. Freeman, B.H. Starfield, C.E. Adair, R. McKendry, Continuity of care: a multidisciplinary review, *BMJ* **327** (2003), 1219–21.

[22] M.D. Cabana, S.H. Jee, Does continuity of care improve patient outcomes?, *J Fam Pract* **53** (2004), 974-80.

[23] J.P. Mackenbach, The epidemiologic transition theory, *J Epidemiol Community Health* **48** (1994), 329–31.

[24] C. Allonier, De quoi souffre-t-on ? Etat des lieux des maladies déclarées en France Enquête décennale santé INSEE 2002-2003, 2007.

[25] M. Goldberg, G. Breart, H. Ducrot, L. Dusserre, A. Letourmy, B. Morel, Le médecin et son ordinateur: l'in formatique au cabinet médical, Paris, 1994, 372P (Frison-Roche).

[26] L. Ward, M. Innes, Electronic medical summaries in general practice–considering the patient's contribution, *Br. J. Gen. Pract.* **53** (2003) 293–297

[27] S. de Lusignan, An educational intervention, involving feedback of routinely collected computer data, to improve cardiovascular disease management in UK primary care, *Methods Inf Med* **46** (2007), 57–62.

[28] S. De Lusignan, P.N. Stephens, N. Adal, A. Majeed, Does feedback improve the quality of computerized medical records in primary care?, *J Am Med Inform Assoc JAMIA* **9** (2002), 395–401.

Address for correspondence

Virginie Lacroix-Hugues

Département d'Enseignement et de Recherche en Médecine Générale, Université Nice Sophia Antipolis - Faculté de Médecine, 28 avenue Valombrose 06107 NICE Cedex 2, France

Email : lacroix-hugues.v@chu-nice.fr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190309

Improving the Performance of Clinical Decision Support for Early Detection of Sepsis: A Retrospective Observational Cohort Study

Ling Li[a], Kasun Rathnayake[a], Malcolm Green[b], Mary Fullick[b], Amith Shetty[c,d], Scott Walter[a], Jeffrey Braithwaite[a], Harvey Lander[b], Johanna I Westbrook[a]

[a] *Australian Institute of Health Innovation, Macquarie University, Sydney, New South Wales, Australia*
[b] *Clinical Excellence Commission, Sydney, New South Wales, Australia*
[c] *eHealth NSW, Sydney New South Wales, Australia*
[d] *Westmead Clinical School, University of Sydney, New South Wales, Australia*

Abstract

Sepsis remains a significant global health problem. It is a life-threatening, but poorly defined and recognized condition. Early recognition and intervention are essential to optimize patient outcomes. Automated clinical decision support systems (CDS) may be particularly beneficial for early detection of sepsis. The aim of this study was to use retrospective data to develop and evaluate seven revised versions of an electronic sepsis alert rule to assess their performance in detecting sepsis cases and patient deterioration (in-hospital mortality or ICU admission). Four revised options had higher sensitivity but lower specificity than the original rule. After discussion with clinical experts, two revised options with the highest sensitivity were selected. Further analysis on the number of alerts and time intervals between alerts and patient outcomes was conducted to decide the option to be implemented. This study has provided a data-driven approach to improve the CDS on early detection of sepsis.

Keywords:

Sepsis, Clinical Decision Support Systems, Early Diagnosis

Introduction

Sepsis is a life-threatening condition that arises when the body's response to an infection injures its own tissues and organs [1; 2]. Despite advances in care, existing epidemiological studies suggest that sepsis remains a huge burden. A recent systematic review extrapolated data from high-income countries to suggest global estimates of 31.5 million sepsis and 19.4 million severe sepsis cases, with potentially 5.3 million deaths annually [3].

Sepsis is one of the most pervasive, but poorly defined and recognized conditions. It has been called "one of the oldest and most elusive syndromes in medicine" [4]. Early recognition and intervention are essential to optimize patient outcomes. To improve early sepsis detection, several automated sepsis alert systems using electronic medical record (eMR) data have been developed for use in hospital intensive care units (ICU) and in non-critical care settings [5]. The St. John Sepsis Surveillance Agent is one such system. It was developed by Cerner Corporation, an American supplier of health information technology solutions, services, devices and hardware. The St. John Sepsis Surveillance Agent has been implemented in more than 550 hospitals in the United States [6; 7].

An updated version, the Modified St. John Rule, was developed by the Clinical Excellence Commission (CEC) in New South Wales (NSW), Australia. It includes additional clinical criteria for activating sepsis alerts in addition to SIRS (systemic inflammatory response syndrome) alerts. There are three panels in the Modified St. John Rule as shown in Figure 1: i) CEC – a sepsis alert if any of two clinical criteria are satisfied; ii) SIRS – a SIRS alert if three or more modified SIRS criteria are met; iii) Organ dysfunction – a sepsis alert if two or more SIRS criteria and at least one organ dysfunction criteria are satisfied. An alert would be triggered if any clinical criterion is satisfied when all relevant measurement(s) within the specified lookback time period (Figure 1) is(are) available.

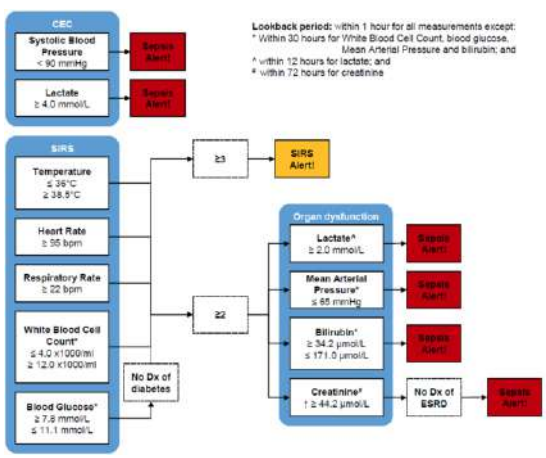


Figure 1 – Simplified Flow Diagram for the Modified St. John Rule. Note: ESRD=End-stage Renal Disease and Dx=Diagnoses.

The modified SIRS criteria used in the Modified St. John Rule were adapted from the original SIRS definition, which was introduced in the first two consensus sepsis definitions [8; 9]. The original SIRS criteria have been criticized for their poor specificity, with 90% of ICU patients and 50% of general ward patients meeting the criteria at some point during their hospitalization [2; 10]. These criticisms are consistent with our findings based on the evaluation of the Modified St John Rule [11], where more than half of the alerts were SIRS alerts. Of patients who experienced a SIRS alert, nearly half of them had two or more SIRS alerts during their hospital stay. The most recent sepsis definition (Sepsis-3) was developed in 2016 and has abandoned SIRS entirely. Sepsis-3 defines organ dysfunction as

an increase in total quick Sequential Organ Function Assessment (qSOFA) score of ≥ 2 [12; 13].

The Adult Sepsis Pathway is a paper-based tool developed for early detection of sepsis in adult patients and currently in practice in hospitals throughout NSW, the most populous state in Australia. The Adult Sepsis Pathway is part of the SEPSIS KILLS program [14; 15], aiming to reduce preventable harm to patients through improved recognition and management of severe infection and sepsis in NSW hospitals. This program has been associated with improved patient outcomes [16]. In this study, we proposed seven revised options of the Modified St John Rule by: i) removing the SIRS alerts and ii) incorporating clinical criteria and threshold values from the qSOFA criteria and the Adult Sepsis Pathway (details in the Methods section). For example, the clinical threshold for heart rate was "≥95 beats/minute" as per the Modified St. John Rule. In the revised options, we included the heart rate clinical thresholds as "≥ 95 beats/minute or ≤50 beats/minute", which was used in the Adult Sepsis Pathway. The aim of this study was to develop and evaluate a set of revised versions of the Modified St. John Rule in order to improve the early identification of patients with sepsis.

Methods

Study Design, Setting and Population

This retrospective observational cohort study included all adult patients (aged 18 years and over) admitted to an acute care teaching hospital in Sydney between December 2014 and June 2016. The hospital is a 570-bed tertiary urban hospital with 24,500 inpatient admissions annually. Sepsis cases were identified based on sepsis-related diagnosis codes based on the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification (ICD-10-AM) and assigned after discharge. The Classification of Hospital Acquired Diagnoses (CHADx) [17] was applied to patients' diagnoses to identify sepsis cases. Ethics approval was obtained from the Macquarie University Human Research Ethics Committee.

Revised Modified St. John Rules

Seven revised options were proposed to improve the performance of the Modified St. John Rule (Figure 2). The changes in these options were:

- Removing the SIRS alert (all seven options)
- Including at least one of five modified SIRS criteria satisfied plus at least one organ dysfunction as a trigger for a sepsis alert (options 2, 4, 5, and 7)
- Including base excess <-5.0 mEq/L as an immediate trigger for a sepsis alert (options 3, 4, 6 and 7)
- Including lactate ≥2 mmol/L as an immediate trigger for a sepsis alert (options 1, 5, 6 and 7)
- Adopting different clinical threshold values:
  - systolic blood pressure ≤100 mmHg (options 1, 5, 6, and 7)
  - heart rate≥ 95 beats/minute or ≤50 beats/minute (options 1, 5, 6 and 7)

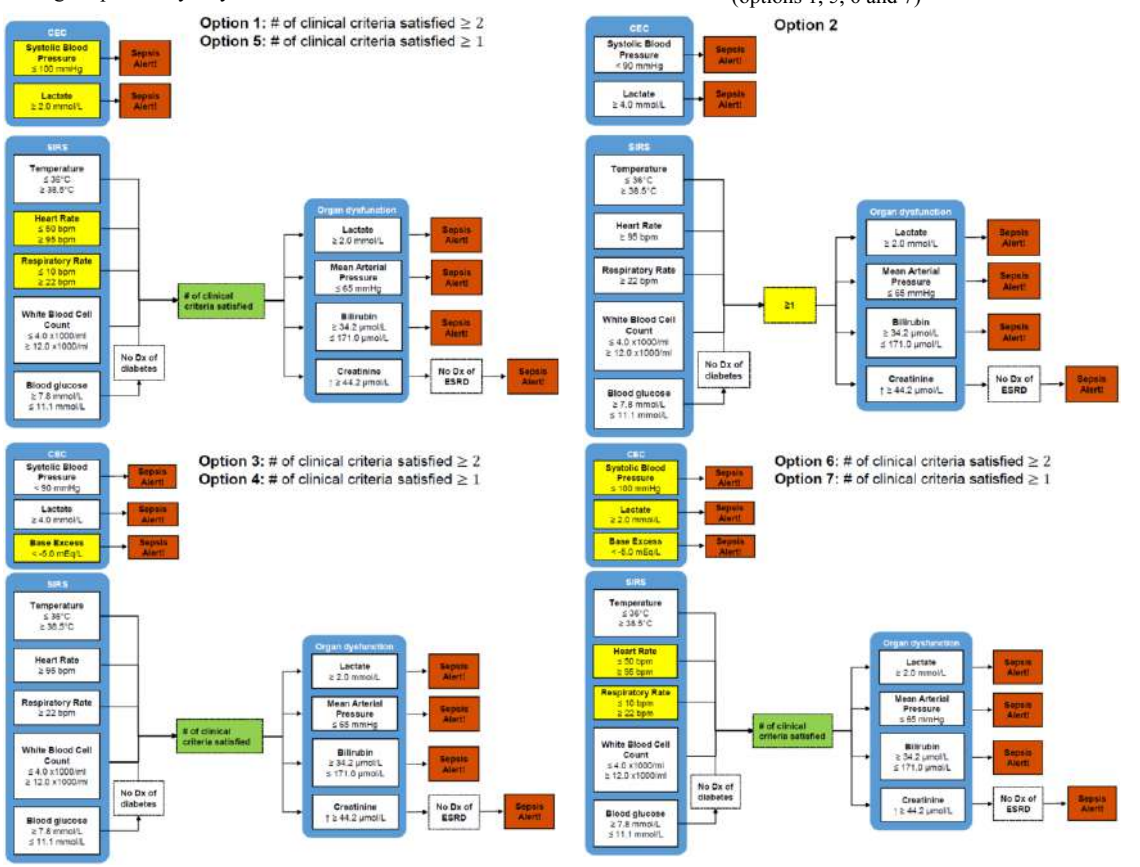


Figure 2 – Simplified Flow Diagrams of Seven Revised Versions of the Modified St. John Rule. Yellow and Green Highlights Indicate the Changes in Each Option. Note:#=Number, ESRD=End-stage Renal Disease and Dx=Diagnoses. See Figure 1 for the Lookback Periods.

- respiratory rate ≥ 22 breaths/minute or ≤10 breaths /minute (options 1, 5, 6 and 7)

Data Sources, Linkage and Management

Patient demographic data and admission related data, including vital signs, laboratory results and ICU admissions were extracted from different clinical information systems. Vital signs and laboratory tests were time-stamped. Data sets from different sources were linked using de-identified medical record numbers and related time stamps. ICD-10-AM diagnosis codes were used to identify patients with diabetes (E10 to E14) and end-stage renal disease (ESRD; N18.5).

Data Analysis

Two patient outcomes were used for assessing the performance of each revised option: i) ICD-10-AM coded sepsis and ii) In-hospital mortality or ICU admission. Performance metrics of seven options and the original Modified St John Rule were calculated, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and the area under the receiver-operating characteristic curve (AUROC). After discussion with clinical experts, two options with the highest sensitivity were chosen as best options for screening. Further analysis was conducted based on these two options to compare three different time intervals: i) time from admission to the first sepsis alert, ii) time from the first sepsis alert to the ICU admission, and iii) time from the first sepsis alert to death in hospital. Analyses were performed using R (version 3.5.0) and SAS (version 9.4).

Results

A total of 36,065 patient admissions for 28,957 unique patients were included in the study. Patients' median age was 55 years (inter-quartile range [IQR]: 38-71) and 41.3% (N=11,946) of these patients were male. Among these patient admissions, 3.9% (n=1,402) involved admissions to ICUs during their hospital stays and a total of 483 patients died in hospital during the study period. The median length of stay (LOS) was 1.9 days (IQR: 0.3-4.9). A total of 3.5% of admissions (N=1,279) involved an ICD-10-AM coded case of sepsis during their hospital stay.

Detection of ICD-10-AM Coded Sepsis Cases

ICD-10-AM coded sepsis cases were used to assess the performance of these revised options for detecting sepsis (Table 1). Options 2, 3 and 4 had an increased specificity and PPV compared to the original Modified St John Rule. Option 2 had the highest specificity (96.26%) among all options. Option 3 flagged 30 false positives (FP) for every ten true positives (TP, i.e., FP/TP=1,529/502), compared to 40 for the original rule (i.e., FP/TP=2,492/611).

Options 1, 5, 6 and 7 identified more sepsis cases, i.e., true positives than the original Modified St. John Rule. Option 7 had the highest sensitivity (64.97%) among all options, closely followed by Option 6 (64.43%). Options 6 and 7 would correctly identify at least six sepsis cases for every ten admissions with a sepsis alert, compared to less than five out of ten for the original rule (sensitivity 47.77%). Both options 6 and 7 had the highest NPVs (98.39% and 98.40%).

Although these options generated more false positives resulting in lower specificity (Option 6: 79.76% and Option 7: 79.31%) than the original rule (92.84%), fewer false negatives (FN) meant fewer sepsis cases would have been missed than the other options and the original rule. For every ten correctly identified sepsis cases, five cases might be missed (i.e., FN/TP=448/831) for Option 7. In contrast, Option 2 would miss 20 false negatives, i.e., sepsis cases, for every ten correctly identified sepsis case (i.e., FN/TP=865/414); the original rule would miss 10 sepsis cases (i.e., FN/TP=668/611).

Detection of Deteriorating Patients

Death in hospital or ICU admission was used to assess the performance of these seven revised options for detecting patients' deterioration (Table 1). Similar to results for detecting sepsis cases, Option 2 had the highest specificity (97.01%) and Option 3 had the highest PPV (43.57%). Options 1, 5, 6 and 7 had higher sensitivity and AUROC than the original Modified St. John Rule. Options 6 and 7 had much higher sensitivity (71.74% and 72.56%) than the original Modified St John Rule (50.99%). For patients who experienced a sepsis alert based on options 6 or 7 during their hospital stays, at least seven out of ten patients would have died or been admitted to an ICU compared to five out of ten patients if based on the original rule.

Table 1–Revised Modified St. John Rule Options for Detecting i) ICD-10-AM Coded Sepsis and ii) In-hospital Mortality or ICU Admission (N=36,065). Note: PPV=Positive Predictive Value; NPV=Negative Predictive Value; AUROC= Area Under the Receiver-Operating Characteristic Curve; CI=Confidence Interval; Original: the Assessment Based on the Modified St. John Rule (Figure 1)

Options	Sensitivity (%, 95% CI)	Specificity (%, 95% CI)	PPV (%, 95% CI)	NPV (%, 95% CI)	AUROC (%, 95% CI)
	i) ICD-10-AM coded sepsis				
Option 1	62.16 (59.44-64.82)	80.07 (79.65-80.49)	10.29 (9.86-10.74)	98.29 (98.17-98.41)	71.12 (70.74-71.49)
Option 2	32.37 (29.81-35.01)	96.26 (96.05-96.45)	24.13 (22.42-25.92)	97.48 (97.39-97.57)	64.32 (63.92-64.71)
Option 3	39.25 (36.56-41.99)	95.60 (95.38-95.82)	24.72 (23.19-26.31)	97.72 (97.62-97.81)	67.43 (67.04-67.81)
Option 4	42.92 (40.19-45.69)	94.18 (93.92-94.42)	21.32 (20.07-22.62)	97.82 (97.72-97.92)	68.55 (68.16-68.94)
Option 5	62.86 (60.15-65.52)	79.59 (79.16-80.01)	10.17 (9.75-10.61)	98.31 (98.19-98.43)	71.23 (70.85-71.60)
Option 6	64.43 (61.73-67.05)	79.76 (79.34-80.19)	10.48 (10.06-10.92)	98.39 (98.26-98.50)	72.10 (71.73-72.46)
Option 7	64.97 (62.29-67.59)	79.31 (78.88-79.73)	10.35 (9.94-10.78)	98.40 (98.28-98.52)	72.14 (71.77-72.51)
Original	47.77 (45.00-50.55)	92.84 (92.56-93.11)	19.69 (18.63-20.80)	97.97 (97.87-98.08)	70.31 (69.93-70.68)
	ii)In-hospital mortality or ICU admission				
Option 1	67.38 (65.11-69.60)	80.87 (80.45-81.29)	15.00 (14.50-15.51)	98.02 (97.88-98.15)	74.13 (73.77-74.48)
Option 2	40.06 (37.73-42.42)	97.01 (96.82-97.19)	40.15 (38.16-42.17)	97.00 (96.88-97.11)	68.54 (68.15-68.92)
Option 3	51.45 (49.06-53.84)	96.66 (96.47-96.85)	43.57 (41.79-45.38)	97.55 (97.43-97.66)	74.06 (73.70-74.41)
Option 4	59.30 (56.94-61.64)	95.47 (95.25-95.69)	39.61 (38.13-41.11)	97.91 (97.79-98.02)	77.39 (77.04-77.73)
Option 5	68.60 (66.35-70.79)	80.42 (80.00-80.84)	14.93 (14.45-15.42)	98.08 (97.95-98.21)	74.51 (74.15-74.87)
Option 6	71.74 (69.55-73.86)	80.70 (80.28-81.12)	15.69 (15.21-16.19)	98.28 (98.14-98.40)	76.22 (75.87-76.57)
Option 7	72.56 (70.38-74.66)	80.26 (79.83-80.68)	15.54 (15.08-16.02)	98.32 (98.18-98.44)	76.41 (76.06-76.76)
Original	50.99 (48.60-53.38)	93.52 (93.25-93.78)	28.26 (27.04-29.52)	97.44 (97.32-97.56)	72.26 (71.89-72.62)

Table 2–Comparison between Options 6 and 7 on the Number of Sepsis Alerts and the Timing of the First Sepsis Alert during the Admission. Note: IQR=Inter-quartile Range and CI= Confidence Interval

Characteristics	Option 6	Option 7
Total number of alerts	37,170	43,616
Total number of admissions with at least one alert	7,863	8,029
Number of alerts per 100 admissions, mean (95% CI)	103 (98-108)	121 (115-127)
Admission to the first alert (hours), Median (IQR)	23.9 (10.1-60.3)	24.1(10.2-60.3)
Time from the first alert to death (hours), Median (IQR)	149.4 (41.4-315.3)	149.2 (46.5-314.3)
Time from the first alert to the first ICU admission (hours), Median (IQR)	233.3 (111.0-428.6)	235.7 (113.6-425.8)
Time from the first alert to death or ICU admission (hours), Median (IQR)	221.8 (98.1-391.1)	223.0 (101.8-390.5)

Further Comparison of Two Options with the Highest Sensitivity

Across the two study outcomes, options 6 and 7 had the highest sensitivity and AUROC. The only difference between two options was the number of SIRS criteria required to trigger a sepsis alert: at least two for Options 6 and at least one for Option 7 (Figure 2). As a result, more alerts would be triggered for more patients using Option 7 than that for Option 6 (Table 2). Option 7 had 121 sepsis alerts per 100 admissions (95% CI: 115-127) compared to many fewer alerts from Option 6 (103 alerts/100 admission, 95% CI: 98-108). In addition, 8,029 admissions would have experienced a sepsis alert if Option 7 was implemented compared to 7,863 admissions if Option 6 was in place.

Both options had very similar alert timing patterns. The median hours from admission to the first alert was around 24 hours and the median hours from the first alert to death in hospital or an ICU admission was around 220 hours (~9.2 days).

Discussion

Automatic clinical decision support systems (CDS) have the potential to provide notifications to clinicians to facilitate real-time early detection of sepsis cases. Following our previous evaluation of one such system [11], i.e. the Modified St John Rule, we proposed seven revised options to improve early identification of patients with sepsis. Using retrospective data, two of the revised options were found to have much higher sensitivities and AUROC in identifying both sepsis cases and deteriorating patients than the original Modified St John Rule. These two options would correctly identify at least six sepsis cases for every ten admissions with a sepsis alert compared to less than five out of ten for the original rule. Similarly, for every ten admissions with a sepsis alert, at least seven patients would be correctly identified as either dying or having an ICU admission compared to five out of ten for the original rule.

The study results were presented to a panel of clinical experts to decide which options should be selected for further implementation in the clinical information system. Given there is a tradeoff between sensitivity and specificity, the high sensitivity options had relatively low specificity. However, a highly sensitive test also means that there are few false negative results, and thus fewer sepsis cases would be missed. Sepsis is a life-threatening complication and any delay in treatment with effective antibiotics increases the risk of organ failure and death. Choosing the options with high sensitivity would save lives. The recommended option, i.e. Option 6, has been adopted by the Clinical Excellence Commission in New South Wales for further implementation and evaluation.

One strength of this study is that we used the composite outcome of in-hospital mortality or ICU admission in addition to ICD coded sepsis cases given the known limitation of ICD coding [18]. A major United States study recently published in the Journal of the American Medical Association found that estimates of the sepsis incidence based on ICD coding ranged between half and twice the actual clinical rate [18].

In medical informatics, large data have been collected for different purposes, including clinical care, administration, and research. We developed algorithms to retrospectively evaluate the performance of seven revised options using data collected routinely from different clinical information systems. This data driven approach allowed us to test different scenarios without utilizing resources and time to develop and implement the CDS.

Previous studies have demonstrated that Big Data techniques, such as machine learning, can be incorporated into electronic health records to predict clinically relevant outcomes in patients with sepsis [19]. Although early identification of sepsis is still challenging, Big Data techniques make it possible to utilize these large volumes of heterogeneous data to provide deeper insights and better understanding of poorly defined and recognized conditions, such as sepsis.

Conclusions

Sepsis remains a significant global health problem. This study has provided a data-driven approach to improve the CDS for early detection of sepsis patients. CDS for early detection of sepsis can be improved and assessed using routinely collected data from different clinical information systems. This data-driven approach would save valuable resources in the health care system and potentially save lives.

It is essential to combine early recognition of sepsis patients with early intervention in order to optimize patient outcomes. The CDS may integrate the early warning systems with the availability of rapid response teams designed to achieve earlier intervention [20]. However, it is important to note that rigorous studies are lacking to evaluate the benefits of these CDS systems [21].

Acknowledgements

This study was commissioned by the Clinical Excellence Commission in NSW, Australia and was conducted as part of project to evaluate the risk identification tools for the early detection of sepsis. MG, MF and HL are employed at the Clinical Excellence Commission. They contributed to the study design and revisions of the manuscript but were not involved in the data assessment and interpretation. We acknowledge the contribution of eHealth NSW to this project.

References

[1] C.J. Czura, "Merinoff symposium 2010: sepsis"-speaking with one voice, *Mol Med* **17** (2011), 2-3.

[2] J.L. Vincent, S.M. Opal, J.C. Marshall, and K.J. Tracey, Sepsis definitions: time for change, *Lancet* **381** (2013), 774-775.
[3] C. Fleischmann, A. Scherag, N.K. Adhikari, C.S. Hartog, T. Tsaganos, P. Schlattmann, D.C. Angus, K. Reinhart, and T. International Forum of Acute Care, Assessment of Global Incidence and Mortality of Hospital-treated Sepsis. Current Estimates and Limitations, *Am J Respir Crit Care Med* **193** (2016), 259-272.
[4] D.C. Angus and T. van der Poll, Severe sepsis and septic shock, *N Engl J Med* **369** (2013), 840-851.
[5] L.A. Despins, Automated Detection of Sepsis Using Electronic Medical Record Data: A Systematic Review, *J Healthc Qual* **39** (2017), 322-333.
[6] R.C. Amland and K.E. Hahn-Cover, Clinical Decision Support for Early Recognition of Sepsis, *Am J Med Qual* **31** (2016), 103-110.
[7] R.C. Amland and B.B. Sutariya, Quick Sequential [Sepsis-Related] Organ Failure Assessment (qSOFA) and St. John Sepsis Surveillance Agent to Detect Patients at Risk of Sepsis: An Observational Cohort Study, *Am J Med Qual* **33** (2018), 50-57.
[8] R.C. Bone, R.A. Balk, F.B. Cerra, R.P. Dellinger, A.M. Fein, W.A. Knaus, R.M. Schein, and W.J. Sibbald, Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine, *Chest* **101** (1992), 1644-1655.
[9] M.M. Levy, M.P. Fink, J.C. Marshall, E. Abraham, D. Angus, D. Cook, J. Cohen, S.M. Opal, J.L. Vincent, G. Ramsay, and C. International Sepsis Definitions, 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference, *Intensive Care Med* **29** (2003), 530-538.
[10] M.M. Churpek, A. Snyder, X. Han, S. Sokol, N. Pettit, M.D. Howell, and D.P. Edelson, Quick Sepsis-related Organ Failure Assessment, Systemic Inflammatory Response Syndrome, and Early Warning Scores for Detecting Clinical Deterioration in Infected Patients outside the Intensive Care Unit, *Am J Respir Crit Care Med* **195** (2017), 906-911.
[11] L. Li, S. Walter, K. Rathnayake, and J. Westbrook, *Evaluation and optimisation of risk identification tools for the early detection of sepsis in adult inpatients*, Macquarie University, 2018.
[12] C.W. Seymour, V.X. Liu, T.J. Iwashyna, F.M. Brunkhorst, T.D. Rea, A. Scherag, G. Rubenfeld, J.M. Kahn, M. Shankar-Hari, M. Singer, C.S. Deutschman, G.J. Escobar, and D.C. Angus, Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3), *JAMA* **315** (2016), 762-774.
[13] M. Singer, C.S. Deutschman, C.W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G.R. Bernard, J.D. Chiche, C.M. Coopersmith, R.S. Hotchkiss, M.M. Levy, J.C. Marshall, G.S. Martin, S.M. Opal, G.D. Rubenfeld, T. van der Poll, J.L. Vincent, and D.C. Angus, The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3), *JAMA* **315** (2016), 801-810.
[14] Clinical Excellence Commission, *SEPSIS KILLS*, retrieved from http://www.cec.health.nsw.gov.au/patient-safety-programs/adult-patient-safety/sepsis-kills
[15] Clinical Excellence Commission, *SEPSIS KILLS - Adult Sepsis Pathway*, retrieved from http://www.cec.health.nsw.gov.au/__data/assets/pdf_file/0005/291803/Adult-Sepsis-Pathway-Sept-2016-with-watermark.pdf
[16] A.R. Burrell, M.L. McLaws, M. Fullick, R.B. Sullivan, and D. Sindhusake, SEPSIS KILLS: early intervention saves lives, *Med J Aust* **204** (2016), 73 e71-77.
[17] T.J. Jackson, J.L. Michel, R.F. Roberts, C.M. Jorm, and J.G. Wakefield, A classification of hospital-acquired diagnoses for use with routine hospital data, *Medical Journal of Australia* **191** (2009), 544-548.
[18] C. Rhee, R. Dantes, L. Epstein, D.J. Murphy, C.W. Seymour, T.J. Iwashyna, S.S. Kadri, D.C. Angus, R.L. Danner, A.E. Fiore, J.A. Jernigan, G.S. Martin, E. Septimus, D.K. Warren, A. Karcz, C. Chan, J.T. Menchaca, R. Wang, S. Gruber, M. Klompas, and C.D.C.P.E. Program, Incidence and Trends of Sepsis in US Hospitals Using Clinical vs Claims Data, 2009-2014, *JAMA* **318** (2017), 1241-1249.
[19] T. Desautels, J. Calvert, J. Hoffman, M. Jay, Y. Kerem, L. Shieh, D. Shimabukuro, U. Chettipally, M.D. Feldman, C. Barton, D.J. Wales, and R. Das, Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach, *JMIR medical informatics* **4** (2016), e28-e28.
[20] A.E. Jones, A. Focht, J.M. Horton, and J.A. Kline, Prospective external validation of the clinical effectiveness of an emergency department-based early goal-directed therapy protocol for severe sepsis and septic shock, *Chest* **132** (2007), 425-432.
[21] P. Bhattacharjee, D.P. Edelson, and M.M. Churpek, Identifying Patients With Sepsis on the Hospital Wards, *Chest* **151** (2017), 898-907.

Address for Correspondence

Corresponding author: Ling Li
Email: ling.li@mq.edu.au
Full mailing address: Level 6, 75 Talavera Road, Macquarie University, NSW 2109, Australia
Tel: +61 2 9850 2423
Fax: +61 2 9850 2499

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190310

Using Machine Learning on Home Health Care Assessments to Predict Fall Risk

Yancy Lo[a], Selah F. Lynch[a], Ryan J. Urbanowicz[a,b], Randal S. Olson[b], Ashley Z. Ritter[c], Christina R. Whitehouse[c,d], Melissa O'Connor[d], Susan K. Keim[c], Margaret McDonald[e], Jason H. Moore[a,b], Kathryn H. Bowles[c,e]

[a] *Clinical Research Informatics Core, Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA, USA*
[b] *Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA, USA*
[c] *NewCourtland Center for Transitions and Health, School of Nursing, University of Pennsylvania, Philadelphia, PA, USA*
[d] *Fitzpatrick College of Nursing, Villanova University, Villanova, PA, USA*
[e] *Center for Home Care Policy & Research, Visiting Nurse Service New York, New York, NY, USA*

Abstract

Falls are the leading cause of injuries among older adults, particularly in the more vulnerable home health care (HHC) population. Existing standardized fall risk assessments often require supplemental data collection and tend to have low specificity. We applied a random forest algorithm on readily available HHC data from the mandated Outcomes and Assessment Information Set (OASIS) with over 100 items from 59,006 HHC patients to identify factors that predict and quantify fall risks. Our ultimate goal is to build clinical decision support for fall prevention. Our model achieves higher precision and balanced accuracy than the commonly used multifactorial Missouri Alliance for Home Care fall risk assessment. This is the first known attempt to determine fall risk factors from the extensive OASIS data from a large sample. Our quantitative prediction of fall risks can aid clinical discussions of risk factors and prevention strategies for lowering fall incidence.

Keywords:

Falls, Health Risk Assessment, Machine Learning

Introduction

Falls are the leading cause of death due to injury in the home, especially for the elderly [1–3]. Falls are typically recurrent; those who fall once are two to three times more likely to fall again [4,5]. In 2015, direct costs related to fatal and non-fatal falls were $637.5 million and $31.3 billion respectively, making it one of the most costly patient safety problems among people aged 65 and older [6,7]. Identifying risk factors for falls is critical to the design of prevention protocols. Many research studies analyzed falls among hospitalized, long-term care, and community-dwelling older adults, yet few studies focus specifically on the frail, homebound population of older adults receiving home health care (HHC) services [8–10].

Medicare and agency policies direct clinicians to screen every HHC patient for fall risk, but only one HHC validated tool exists. The Missouri Alliance for Home Care assessment (MAHC-10) is a 10-item fall risk screening tool [11] with excellent sensitivity (97%), but with poor specificity (13%). In our study cohort, MAHC-10 identified over 93% of the cohort as having high fall risk, but only 5.14% actually had a fall (Figure 1). Therefore, using MAHC-10 as the default fall risk screening tool may increase the cost and burden of healthcare providers by triggering unnecessary provision of fall prevention strategies to almost every HHC patient. Moreover, MAHC-10 provides the clinician a score out of 10 instead of an actionable profile or fall risk as a probability per individual patient, making the design and implementation of personalized fall prevention difficult.

In this study, we devised a machine learning pipeline to explore the utility of existing HHC data containing rich patient information to predict fall risk. This is the first study to analyze large and comprehensive HHC datasets representing the characteristics of vulnerable older adults and the care they received in their homes to create models of effective fall risk prediction. With this larger dataset and additional input features from the electronic health record (EHR), we evaluated whether our models could achieve higher precision and accuracy than the existing risk scoring system.

Methods

Data Description

We leveraged patient information from the Centers for Medicare and Medicaid (CMS)-mandated Outcome and Assessment Information Set (OASIS) [12] data on an ethnically diverse population of nearly 60,000 patients from one large HHC agency in New York City. The OASIS is a mandatory detailed assessment with over 100 items evaluating a patient's clinical, behavioral, cognitive, and environmental conditions. The newer version (OASIS-C) was utilized for the patient population included in this study.

The HHC cohort was also assessed for fall risk using the 10-point Missouri Alliance for Home Care fall risk assessment (MAHC-10) [11]. A score greater than or equal to 4 is clinically regarded as at risk for falls. We took the intersection of patients who had both OASIS-C and MAHC-10 data and were over 65 years old, resulting in a final cohort of 59,028 unique patients. We supplemented the feature set with additional demographic information from the EHR, including language group and borough of residence (New York City).

Data Cleaning

We defined the binary outcome of fall incidence per patient from three sources: two OASIS-C items indicating whether the patient received emergent care or hospitalization due to falling (items M2310 and M2410), and whether a date of last fall was recorded in the EHR. Therefore, the outcome is True (presence of a fall) for a patient if the date of last fall is between the start and end of the HHC episode, or the answer is Yes for either M2310 or M2410. Using this definition, the fall incidence rate in our cohort is 5.14%. An episode of home care can be up to 60 days.

The OASIS-C start of care assessment contains 114 items, the majority being multiple-choice questions. Upon review from

clinical experts, we considered 46 items that are relevant to fall risk prediction. For the multiple-choice OASIS-C items, we categorized the items based on the type of answer choices and applied the data cleaning strategy per category as follows.

For OASIS-C items with ordinal answer choices (for example, a 0-5 scale where 0 denotes never and 5 denotes all the time), we represented these items as continuous features. For OASIS-C items and EHR items with nominal answer choices (for example, gender, language group, type of assistance), we used one-hot encoding to create a binary feature per answer choice. The same encoding scheme applied for OASIS-C items that allowed multiple answers ("select all that applies"). We discarded patients under 65 years old to focus on the older adult population receiving HHC and used age as a continuous feature. Since OASIS-C is a mandatory assessment at the start of HHC, missing data is minimal (< 0.005%). We considered the rare missing entries as a separate answer choice for the corresponding questions.

Two OASIS-C items require listing diagnosis codes and severity of each diagnosis on a 4-point scale. These items were combined to reflect the overall physical well-being of a patient. To incorporate these items meaningfully into our feature set, we constructed summary features of diagnoses including: total number, total severity, and average severity. After data cleaning and feature engineering, our feature space contained 169 features from the OASIS-C and EHR that described the demographics, clinical, behavioral, and environmental characteristics of each HHC patient.

Machine Learning Pipeline

We devised a two-step machine learning pipeline with a feature selection step followed by a falls classification step. We randomly split the data into training (50%) and testing (50%) sets. Given the high-dimensionality and potential multi-collinearity of our feature set, we employed the ReliefF feature selection algorithm [13,14] to rank the features by the ReliefF score, computed based on the discriminative ability of each on the outcome, conditional on its neighboring features in the feature space.

We trained and compared random forest classifiers using three feature sets. First, we considered the OASIS-C features with positive ReliefF scores. We trained a random forest model with 5-fold cross-validation, using 300 estimators and default parameters as specified in the python scikit-learn [15] module *(OASIS model)*. Second, we took the ten items from MAHC-10 scoring as binary features and trained a random forest model using the same parameters as the OASIS model *(MAHC model)*. Finally, we explored if the features from OASIS-C and MAHC-10 together would augment the prediction accuracy of the random forest model. We combined the OASIS-C feature set with the 10 features from MAHC-10 to train a random forest classifier with the same parameter settings and 5-fold cross-validation *(Combined model)*.

Model Assessment

We evaluated the accuracy and precision of predictions from the OASIS model, the MAHC model, and the Combined model against the baseline MAHC-10 total score (hereafter called baseline), where a score greater than or equal to 4 is clinically considered at risk of falls. Since our outcomes are heavily skewed with only about 5% experiencing a fall, we used balanced accuracy as the metric to assess model accuracy. Balanced accuracy is defined as the average accuracy calculated from the two outcome classes. To highlight the performance of each model in correctly predicting the positive outcome, we contrasted the precision-recall curves of our models to that of the baseline. To summarize the precision-recall curves, we computed the average precision (AP), defined as the arithmetic mean of precisions at different recall thresholds.

Furthermore, we computed the area under ROC curves (AUC) for the test data corresponding to each feature set and contrasted them to the AUC of the baseline. To assess for significant differences between ROC curves across models, we performed bootstrapping on the test data to generate a confidence interval for each ROC curve. For each test set, we resampled with replacement for 1,000 times the outcome states and predicted probabilities, then calculated the AUC for each bootstrapped ROC curve. The 95% confidence interval of a ROC curve is given by all curves with AUC between the 2.5 and 97.5 percentile of all bootstrapped AUC values.

To evaluate the clinical significance of the OASIS model, we ranked the input features by their importance scores generated from the random forest classifier. To quantify the contribution of each feature to the accuracy of the model, we computed the gain in balance accuracy per added feature by refitting a random forest classifier iteratively, adding one feature at a time from the ranked feature list. Finally, by performing a hierarchical clustering of pairwise correlations between top-ranked OASIS-C features and the MAHC-10 items, we revealed fall risk factors identified by the OASIS model that were not captured in the baseline MAHC-10 assessment.

Results

We compared the random forest classifiers for three feature sets: 137 out of 169 OASIS- and EHR-derived features with positive ReliefF scores (OASIS model), MAHC-10 items (MAHC model), and OASIS-C plus MAHC-10 items (Combined model), in contrast to the MAHC-10 scoring (baseline). Overall, the MAHC model had comparable performance as the baseline, while the OASIS model and the Combined model had almost identical metrics that outperform the MAHC model and the baseline.

Although the baseline and the MAHC model both had an AUC of 0.6, the balanced accuracy of the MAHC model at 0.58 was slightly higher than that of baseline scoring measured at only 0.51. The OASIS model and Combined model both attained a balanced accuracy of 0.62 and an AUC of 0.67. The 95% bootstrap confidence interval of the OASIS model ROC was (0.66, 0.68), which was completely above the MAHC model ROC 95% confidence interval at (0.59, 0.62).

The precision of all models and the baseline was low, due to the low proportion of cases in our dataset. The average precision (AP) of the baseline was 0.07, and the AP of the MAHC model was 0.08. The OASIS and Combined models had an improved precision at AP=0.10. Consistent with the AP trend, the precision-recall curve of the OASIS model (and that of the Combined model) was above the curve of the MAHC model at all sensible recall thresholds (Figure 2).

Given the above metrics and the rule of parsimony, the OASIS model was the best out of the three models and the baseline. To investigate clinical relevance of the OASIS model, we ranked the input of 137 OASIS-C features by feature importance scores estimated by the classifier. The most important feature was age, with an importance score of 0.05, followed by the average and total severity of home care diagnoses. Frequencies of therapy visit and pain also had high feature importance scores. Balanced accuracy of the random forest classifier increased as features were added to the model in the order of importance; the balanced accuracy converged at around 0.62 after the top 45 features were added. The remaining 92 features had small

contributions to the balanced accuracy of the random forest classifier.

To further evaluate the potential gain of using OASIS-C over MAHC-10 for fall risk prediction, we computed the pairwise correlation between each top-ranked OASIS-C feature among the MAHC-10 items, and performed a hierarchical clustering on the correlations. Four OASIS-C features had an analogous MAHC-10 item, including history of falls, visual impairment, cognitive impairment, and pain (Figure 3), as reflected by the strong positive correlations that were statistically significant (p-value $< 10^{-5}$). However, each of the four MAHC-10 items was also significantly correlated with a broad range of other OASIS-C items, indicating the heterogeneity among patients who scored the same on the MAHC-10 scale. In addition, some OASIS-C items were correlated in opposite directions to different MAHC-10 items, meaning that the effect of these features might be masked in the total MAHC-10 score. In particular, four top-ranked OASIS-C features were weakly correlated in opposite directions to MAHC-10 items but were not correlated to the total MAHC-10 score: frequency of ADL/IADL (activities of daily living/ instrumental activities of daily living) assistance, number of inpatient diagnoses, patient living alone, and patient living with others. These features provide new information on the patient's fall risk that was not available from the baseline scoring system.

Discussion

In patients receiving HHC services, falls rank as the top avoidable event that leads to disability, hospital admission and emergency department care [16,17]. Motivated by the low specificity of the existing fall risk assessment for HHC patients, we investigated the benefit of predicting fall risk using the Centers for Medicare and Medicaid Services (CMS) mandatory OASIS assessment for HHC coupled with supplemental EHR data. OASIS evaluates in detail the clinical, behavioral, cognitive, and environmental properties of a patient upon the start of a HHC episode. To analyze a large sample size of almost 60,000 and a high-dimensional input set comprising over 130 features with positive ReliefF score, we trained a random forest classifier on 50% of the sample and tested the accuracy and precision of the classifier on the remaining 50% of the data. Predictions leveraging this big data show improved accuracy and precision over a simplistic 10-point scoring system. We investigated if the ten items in the MAHC-10 would be more informative when used as features in a random forest model. We found a negligible change in precision and a slight improvement in balance accuracy in the MAHC model.

Furthermore, since all MAHC items are correlated to one or more OASIS items, complementing the OASIS features with the ten MAHC items did not result in detectable improvement over the OASIS-only model. This signals a potential to implement the OASIS model for clinical use, such that clinicians can obtain the per-patient fall risk from the model, once the mandatory OASIS assessment is completed. Clinicians burdened by required assessments and documentation may appreciate the efficiency gained.

Using our random forest classifiers, each patient in the test set gets an estimate of fall risk as a probability. This probability and the ranked list of important features are clinically relevant and potentially of great value when health care providers discuss fall risk with patients, allowing providers to customize and prioritize prevention strategies corresponding to actionable risk factors for each patient. Clinical decision support provided at the point of care may help clinicians target the most effective interventions and help patients recognize their risk and the impact of embracing preventive strategies to decrease their risk.

Limitations

This study used data from one large home care agency in New York City. The precision of the OASIS model is still lower than ideal for clinical use because the two outcome classes in our dataset are heavily imbalanced. Almost 95% of the patients in our cohort were not reported to have a fall. The low fall incidence from our structured data is consistent with the presumption that falls are often underreported. Recent reports in a national sample of older adults revealed 72% failed to report a fall when asked [18]. We hypothesize that a significant proportion of fall cases are recorded in the EHR narrative data instead of the structured data. Therefore, an immediate extension of our work is to use natural language processing techniques to identify additional fall instances from EHR narratives [19]. We also hope to extend the measurement period of falls beyond the home care episode. By having a more accurate estimate of fall incidence, we can fine-tune our classifiers and achieve higher precision.

Conclusions

This is the first known large-scale study to predict fall risk and characterize risk factors in the HHC community using the readily available OASIS assessment. By using machine learning models to analyze the rich feature set in a large cohort, we see promising improvement in the precision and accuracy of fall risk prediction over the MAHC-10 scoring system. Our results suggest that fall risk is a complex trait affected by a large number of risk factors of small effects. The machine learning approach also allows us to predict fall risk as a probability for each patient, and rank the importance of each risk factor. Our model confirms that a broad range of factors including age, clinical diagnoses, daily habits, living environment and hygiene, all contribute to a patient's fall risk. Further study incorporating an expanded feature set from the EHR will improve the estimate of fall incidence and fall risk prediction.

Figures and Graphs

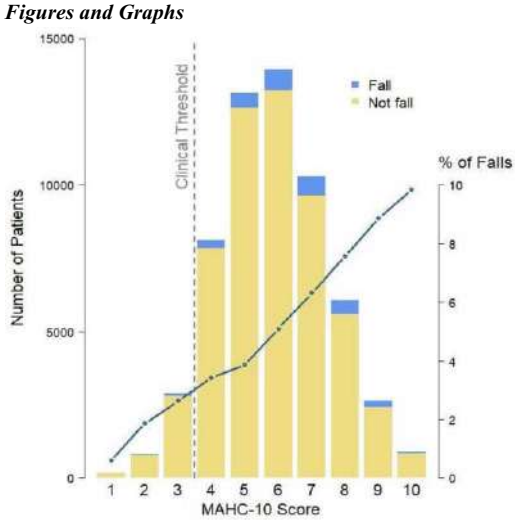


Figure 1. Number of patients per MAHC-10 score. Each bar shows the observed fall cases in blue and the non-cases in yellow. The black line shows the percentage of fall case per score. The clinical threshold of high fall risk is MAHC >=4 (Dashed vertical line).

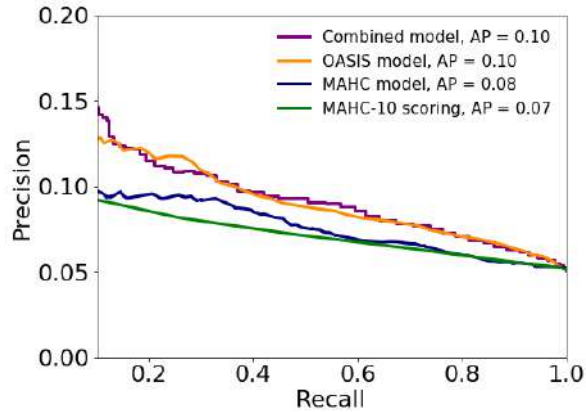


Figure 2. Precision-recall curves of the baseline scoring, MAHC model, OASIS model, and Combined model. For clarity, the panel only shows recall > 0.1 and precision < 0.2.

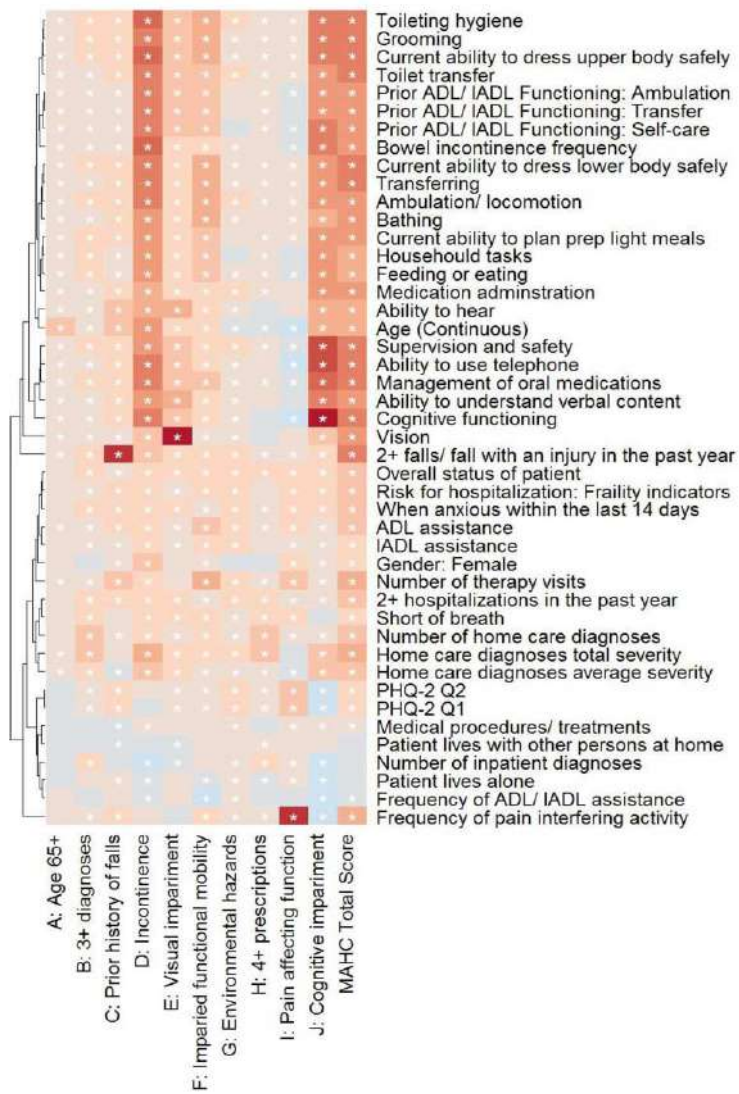


Figure 3. Hierarchical clustering of pairwise correlation (ρ) between top OASIS features (rows) and MAHC-10 items (columns). Blue represents negative correlation (-1 < ρ < 0) and red represents positive correlation (0 < ρ < 1). The asterisk (*) in a cell represents significant correlation after Bonferroni correction for multiple testing ($p < 10^{-5}$).

Acknowledgements

This study was supported by the University of Pennsylvania School of Nursing Frank Morgan Jones Fund. We thank Danielle Mowery, PhD for her valuable comments on the study methods.

References

[1] Centers for Disease Control and Prevention, Important Facts about Falls, (2017). https://www.cdc.gov/homeandrecreationalsafety/falls/adultfalls.html.
[2] K. Cameron, E. Schneider, D. Childress, C. Gilchrist C, Falls Free®: 2015 National Falls Prevention Action Plan, 2015. https://www.ncoa.org/wp-content/uploads/FallsActionPlan_2015-FINAL.pdf.
[3] G. Bergen, M.R. Stevens, and E.R. Burns, Falls and Fall Injuries Among Adults Aged ≥65 Years - United States, 2014, *MMWR Morb. Mortal. Wkly. Rep.* **65** (2016) 993–998.
[4] J.L. O'Loughlin, Y. Robitaille, J.F. Boivin, and S. Suissa, Incidence of and risk factors for falls and injurious falls among the community-dwelling elderly, *Am. J. Epidemiol.* **137** (1993) 342–354.
[5] World Health Organization, WHO global report on falls prevention in older age, (2007). http://www.who.int/ageing/publications/Falls_prevention7March.pdf.
[6] E.R. Burns, J.A. Stevens, and R. Lee, The direct costs of fatal and non-fatal falls among older adults - United States, *J. Safety Res.* **58** (2016) 99–103.
[7] Institute for Healthcare Improvement, No place like home: Advancing the safety of care in the home, (2018). http://www.ihi.org.
[8] C.L. Lewis, M. Moutoux, M. Slaughter, and S.P. Bailey, Characteristics of individuals who fell while receiving home health services, *Phys. Ther.* **84** (2004) 23–32.
[9] T. Yamashita, H. Jeon, A.J. Bailer, I.M. Nelson, and S. Mehdizadeh, Fall risk factors in community-dwelling elderly who receive Medicaid-supported home- and community-based care services, *J. Aging Health*. **23** (2011) 682–703.
[10] A.L. Byers, T. Sheeran, A.E. Mlodzianowski, B.S. Meyers, P. Nassisi, and M.L. Bruce, Depression and risk for adverse falls in older home health care patients, *Res. Gerontol. Nurs.* **1** (2008) 245–251.
[11] MAHC 10 Fall Risk Assessment Tool, Missouri Alliance for Home Care, n.d. https://www.homecaremissouri.org/projects/falls/documents/Oct2012FINALValidatedFallriskassessmenttool.pdf.
[12] Outcome and Assessment Information Set (OASIS), (2018). https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/OASIS/index.html.
[13] R.J. Urbanowicz, R.S. Olson, P. Schmitt, M. Meeker, and J.H. Moore, Benchmarking Relief-Based Feature Selection Methods for Bioinformatics Data Mining, *J. Biomed. Inform.* (2018). doi:10.1016/j.jbi.2018.07.015.
[14] I. Kononenko, Estimating attributes: Analysis and extensions of RELIEF, in: Machine Learning: ECML-94, Springer Berlin Heidelberg, 1994: pp. 171–182.
[15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, Scikit-learn: Machine Learning in Python, *J. Mach. Learn. Res.* **12** (2011) 2825–2830.
[16] Centers for Medicare and Medicaid Services, OASIS-C Based Home Health Agency Patient Outcome, Process and Potentially Avoidable Event Reports, (Updated June 2018). https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/OASIS/09aa_hhareports.html.
[17] M.D. Courtney, H.E. Edwards, A.M. Chang, A.W. Parker, K. Finlayson, and K. Hamilton, A randomised controlled trial to prevent hospital readmissions and loss of functional ability in high risk older adults: a study protocol, *BMC Health Serv. Res.* **11** (2011) 202.
[18] G.J. Hoffman, J. Ha, N.B. Alexander, K.M. Langa, M. Tinetti, and L.C. Min, Underreporting of Fall Injuries of Older Adults: Implications for Wellness Visit Fall Risk Screening, *J. Am. Geriatr. Soc.* **66** (2018) 1195–1200.
[19] V.J. Zhu, T.D. Walker, R.W. Warren, P.B. Jenny, S. Meystre, and L.A. Lenert, Identifying Falls Risk Screenings Not Documented with Administrative Codes Using Natural Language Processing, *AMIA Annu. Symp. Proc.* **2017** (2017) 1923–1930.

Address for Correspondence

Corresponding author: Yancy Lo, yancylo@pennmedicine.upenn.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190311

Renal Biopsy Recommendation Based on Text Understanding

Yang Lu[a], Zheng Jia[a], Xian Zeng[a], Chunyue Feng[b], Xudong Lu[a], Huilong Duan[a], Haomin Li[b]

[a] *College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, Zhejiang, China*
[b] *The Children's Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China.*

Abstract

Due to various etiologies and pathogenesis of kidney diseases, an invasive procedure called renal biopsy may be needed to determine the specific type of kidney disease, its severity, and the best treatment for it. This study aims to detmine if a text understanding technology based on admission records can recommend such an invasive procedure objectively. To understand clinical documents from nephrology, a semi-automatic learning-based lexicon construction method based on CRF and Word2vec was used. We constructed a dictionary of symptom terms for the nephrology department from clinical document, and then extracted patients' symptoms and detected their negation from admission notes. Combined with the preliminary diagnosis given by the doctor, an eigenvector was produced and fed to a machine learning classifier. When compared to the gold standard marked by physicians, the final recommendation achieved 83.5% accuracy, 80.6% precision, 76.6% recall, and 78.6% f1-measure respectively.

Keywords:
Natural Language Processing; Kidney Diseases; Decision Support Systems, Clinical

Introduction

Renal biopsy, also known as kidney biopsy, is a clinical procedure to diagnose multiple suspected renal diseases, especially those unable to be identified alternatively. It is the only definitive route to diagnosis and identify interstitial inflammation including lymphocytes, monocytes, and eosinophils [1].

As an invasive procedure, not all the patients with hematuria and albuminuria need to have a kidney biopsy. This requires comprehensive consideration of patients' conditions. Also, renal biopsy puts patients at risk for bleeding, pain, arteriovenous fistula, and other complications such as hematoma, infection, and hypertension [2]. Physicians should pay special attention to contraindications beforehand.

Artificial intelligence (AI) technologies, such as natural language processing (NLP) and machine learning algorithms, have significantly affected clinical practice and research. AI algorithms usually deal with (transformed) structured data [3]. However, for clinical applications, a large amount of information is concealed in unstructured documents, which is a traditional way for clinicians to record and communicate information.

This study used a corpus of hospital admission notes by the nephrology department. These notes include natural language decscriptions of basic patient information, including history, physical examination findings, reasons why the patient is being admitted for inpatient care, plan of care, etc. This information serves as the basis for further clinical treatments. Even so, different physicians will make different quality choices. The hypothesis of this study is text understanding technologies can evaluate the situation described in free-text documents and provide objective, homogenous decision support.

Methods

In order to solve the above problems, methods were developed acording to the following flow diagram (Figure 1). Traditional knowledge engineering tasks, such as ontology and semantic network for text understanding, are labor-intensive an do not have broad applicability. So, a semi-automatic construction of domain-specific semantic lexicon was designed. Then an NLP pipeline was constructed to extract comprehensive clinical information to feed different clinical decision support models.

CRF-based Semi-automatic Construction of Domain Lexicon

The lack of Chinese medical terminology resources and the complexity of expression in Chinese natural language have led to many difficulties in the automatic acquisition and utilization of information in China.

A previously reported method [4], which iteratively discovers new symptom terms from the clinical corpus using the conditional random field (CRF) algorithm, was used in this study. A general homegrown medical dictionary served as the seed dictionary. The training domain corpus was segmented using the Pangu segmentation tool [5] and reverse maximum matching (RMM) to label based on the seed dictionary. For each iteration, the CRF model output newly identified terms. These new terms update to the seed dictionary and work for the following iterations. After post-processing eliminated false positive terms, an XML file was generated for each test document. The similarity calculation was performed based on the XML file, and the result determined the number of iterations when no more new terms were identified.

The default feature template of the CRF toolkit was used. In the process of calling the CRFSharp API, files named CRFSharp.dll and CRFSharpWrapper.dll were added as references, which provided the core algorithm and low-level and high-level interfaces [6]. They were employed when training the CRF model from training corpus in the program, i.e., model encoding. In order to achieve better results, we modified some of the default parameters to ensure that the extracted symptom terms were more common (occurred at a higher frequency) and that the iterations were more thorough.

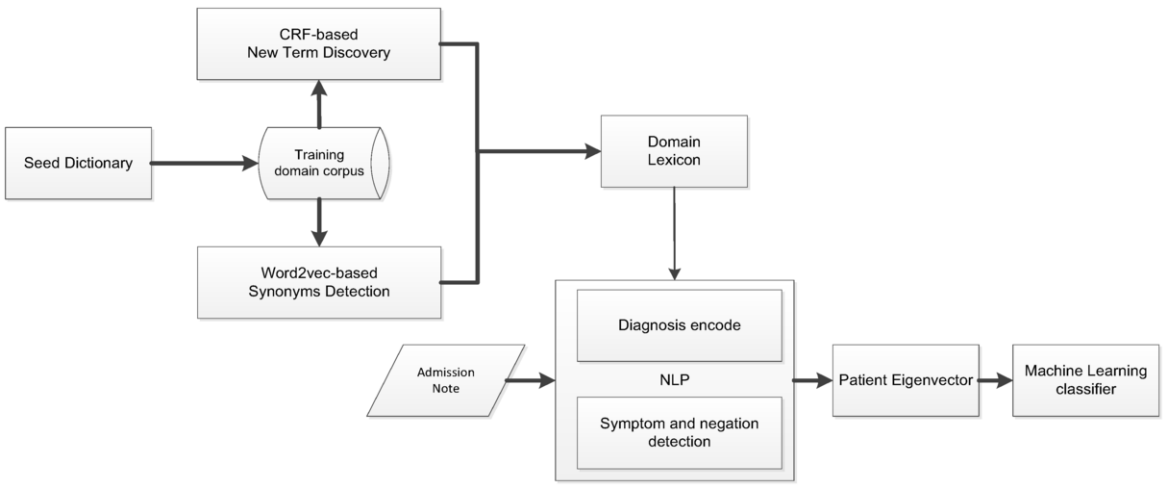


Figure 1–The Flow Diagram of Renal Biopsy Recommendation Based on Text Understanding

After manual review, the accurate new words were integrated with the initial seed dictionary, and we obtained a list of accurate symptom terms. Due to the vague definition of symptom terms, the results might include other signs and diagnoses according to the needs of doctors.

Symptom Synonyms Identification

Due to the different habits of recorders, the same symptoms might have different expressions. Thus, we needed to sort out synonyms to construct the eigenvector, since the integration of the seed dictionary and newoly identified terms might regard different expressions of the same symptom as different symptoms. In order to identify these cases as completely as possible, symptom synonyms detection was required.

Tokenization

Unlike English, Chinese does not separate words by a space or other punctuation [7]. In order to deal with Chinese corpora, we used the "Jieba" Chinese text segmentation, a python-based Chinese word segmentation module. In addition to its built-in dictionary, customized dictionaries were also introduced to improve the accuracy of tokenization, especially for texts in professional fields. After loading the domain specific dictionaries, the Jiaba segmentation generated a Trie tree, divided the text content into phrases, used DAG and dynamic programming to obtain the maximum probability path, and used HMM model and the Viterbi algorithm [8] for phrases not included in the dictionary (i.e., identification of new words). Finally, it built a word generator using Python's yield syntax, returning words one by one.

Word Embedding

Word embedding is actually a manner of word representation, to a form which is suitable for computer processing such as a vector or matrix. In NLP, the representation of words can be divided into two categories: one-hot representation and distributed representation. Since the use of one-hot representation often causes data sparsity [9], people prefer the latter. Word embedding is closely related to the neural language model (NLM) [10]. This process, based on distributed hypothesis, involves of mapping words into real vector spaces [11] and evaluating the similarity between the meaning of two words by their contexts.

Through the Word2vec algorithm, we could represent a word in a lower dimensional space and map the related concepts to similar feature vectors [12].

Synonym Identification

Word2vec comes with a data set called text8, which can train the tokenized texts to obtain a model. We used the generated model to find synonyms for symptom terms based on the similarity of word vectors. By calling the function provided in Word2vec, we were able to calculate the distances between word vectors of other words and the input word. Results were sorted according to their cosine similarity order. The closer the value was to 1, the more similar the position of the words were in the context, and we regarded them as synonyms or words that described different performances of the same symptom. Conversely, the closer the value was to 0, the lower the similarity was between words [13-15].

By manually screening the results, we obtained an accurate list of words covering the symptom terms that had appeared in the nephrology medical records and their synonyms (if any).

Eigenvector Construction

With the list of symptom terms obtained, we matched the text content to terms and then performed negation detection to determine if the patient had the mentioned symptoms or not. After further processing, an eigenvector could be generated for each patient. In the n-dimensional vector, the first column is the patient id, and the last column is reserved for the label, with the remaining columns having a value of 1 or 0, indicating that symptoms and diseases appear or do not occur, respectively. The disease here is the result of a preliminary diagnosis. Therefore, the number of columns of the feature vector is n +2, where n represents the sum of the numbers of the symptoms and diseases (synonyms are considered one symptom).

Reverse Maximum Matching

In the field of Chinese word segmentation, commonly used methods are segmentation algorithms based on a dictionary, understanding, and statistics, among which dictionary-based is the simplest and most practical method [16]. Due to the features of Chinese word formation, dictionary-based string matching often uses reverse maximum matching, which generally

possesses higher segmentation accuracy and encounters fewer ambiguities [17].

Input (Trigger)									
State	symptoms	不† be not 未 do not	无 be without 否认 deny 排除 rule out	有 have 见 see 存在 exist 感到 feel	查出 find 出现 appear 发现 discover 伴有 associate with	闻及 hear 触及 touch 探及 detect	及 (*conj.*) and 及(*v.*) reach	或 or 和 and 与 and 、	Other POS type
0	0	1	2	0			0	0	0
1	0(R)	0(R)	0(R)	2			2	0(R)	0(R)
2	2(M)	0(R)	0(R)	0(R)			2	2	0(R)

Table 1– State-transition Table for FSA of NegDetector
R：Reset NegAff to positive
M：Modify NegAff to negative
†Words in light grey cells are NANegation and corresponding verbs.

The specific steps of the RMM algorithm are as follows: the large segment of text is divided into a plurality of sentences according to the punctuation marks. Each sentence is cyclically read, and the length is n Chinese characters (1 Chinese character is equivalent to 2 characters). Set a maximum length k, and take the last k words as the matching field each time. If the matching fails, remove the first word (k becomes k-1) and continue looping until the matching is successful and gets a word. Repeat the above steps until the remaining length n=0.

Negation Detection

The negation detection method here, an FSA-based algorithm called NegDetector, was derived from an article by Zheng Jia et al. [18]. A concept named NegAff was proposed, whose value was used to describe the affirmation or negation of terms in the electronic medical record (EMR). According to the appearance of sentence structure and negative signals, a state conversion table of FSA was established as Table 1.In the initial state, NegAff defaulted to affirmative, with its value modified according to the FSA state. If we were in state 2 and we saw a term or noun, then the NegAff of the word was modified to negative. If we were going to 0 (R), it meant we need to return to the start state and reset NegAff.

Through the above algorithm, the value of NegAff was defined to determine the contextual polarity of the symptom terms.

Dimension Reduction of Preliminary Diagnoses

The processing object in this section was the disease derived from the preliminary diagnosis of EMR, part of the columns of the eigenvector. The International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) [19] converts words representing diseases and diagnoses into alphanumeric codes, greatly facilitating the storage and retrieval of data. CCD is the Chinese version of Classification and Codes of Diseases (GB / T 14396-2001), expanded and refined on the basis of ICD-10, with the encoding changed from 4-char to 6-char [20].

When dealing with the initial diagnosis of EMR, the use of a detailed classification hierarchy would result in a large number of entries. To this end, in the multilevel disease categories, we had 805 distinct ICD six-char encoding diseases (single-level disease [21; 22], such as I45.102, “完全性右束支传导阻滞”, complete right bundle branch block) classified into subcategories, in which the first four characters were the same (e.g. I45.1, “其他和未明示右束支传导阻滞”, other and unspecified right bundle branch block). It was found that the number of entries could be reduced to 675. That is to say, by reducing the number of coded bits from 6 to 4, the number of distinct diseases could be reduced by 16.1%. Changes in categorization could effectively decrease the number of diseases and control them in a manageable number.

Machine Learning Classifier

The eigenvectors of the cohort were used to train logistic regression models. We randomly selected 70% of the total data as the training set and used the remaining 30% as a test set. Five-fold cross-validation was used on the training set.

Results

Data Description

A total of 3,149 admission notes from the nephrology department were used in this study. We used the recommendations given by physicians in first-day progress note as the gold standard for renal biopsy recommendation. We selected all statements in the document that represent "the patient needs a renal biopsy", such as "Renal puncture examination if necessary, to determine the pathological type" and "Renal puncture examination if necessary.” These were used as the gold standard for providing accurate judgments for subsequent training and test results.

The eigenvector was composed in order of patient ID, 675 4-char encoding diseases, 385 symptoms, and a label representing whether the patient was recommended for renal biopsy. The value 1 stood for “appeared” or “recommended”, and 0 represented the opposite.

Results of Renal Biopsy Recommendation

Results of New Terms Identification

Even if the parameters had been adjusted, there were still low-frequency terms that appeared after multiple iterations. Here we

did not use all newly identified symptoms, and we eliminated symptoms of extremely low frequency. We selected 800 new terms and conducted manual screenings to get 561 new accurate terms. The precision was 70.1%.

In addition, the terminologies identified using machine learning contained combined symptoms, such as "tenderness and rebound tenderness." Among them, "tenderness" and "rebound tenderness" were different symptoms, but might be recognized as a symptom because they were used adjacently in the medical record. In practical application, we regarded the words of the combined symptoms as multiple symptoms, and the degree of refinement of the data became higher so that the granularity became smaller.

Results of Symptom Synonyms Identification

By inputting a word and taking the top ten from the lists that Word2vec retrieved, we could obtain 10 words which appeared in the document that had the closest meaning to the certain word. We chose the symptom terms that had a synonym in the corpus and randomly extracted 10 symptoms and 15 symptoms. After manual proofreading, the calculated precisions were 57.0% and 54.0%, respectively.

Results of Renal Biopsy Recommendation

We performed word segmentation, reverse maximum matching and negation detection on all documents, labeled doctors' recommendations for renal biopsy according to the gold standard, and used scikit-learn to train logistic regression models for binary classification. We used the default parameters of the sklearn toolkit. We used five-fold cross-validation for the training set and calculate the average. The results are shown in Table 2. The ROC curve of the test set obtained by the method of this paper is shown in Figure 2.

Discussion

In this work, we used the accuracy, precision, recall, and F1-measure to evaluate the proposed method. It can be seen that this method works in the recommendation renal biopsy. In addition, a combination of the recommendation system and visualization techniques should be a promising research direction, since it can be used as a support tool. In practical applications, doctors can be assisted while drawing conclusions with a more intuitive view.

Although this method has achieved good results, there are still some inadequacies with it. First, the precision of the word segmentation results has room for improvement. Some terms that were not related to the seed dictionary were not recognized or excluded from screening. Secondly, Word2vec is based on the contextual environment to generate word vectors to judge the similarity of words, which has little to do with the meaning of the words themselves; and some words have few synonyms in the medical records, which lead to high coverage of some synonyms, but low precision. Medical staff input errors (typo and grammatical errors, etc.) can also result in reduced accuracy of term extraction. The above reasons may also lead to prediction failures.

Furthermore, feature selection is performed using the SelectFromModel feature selection method from the scikit-learn library. The features of higher significance and stronger correlation include recurrent and persistent haematuria, nephrotic syndrome, acute tubule-interstitial nephritis, isolated proteinuria, hypersensitivity angiitis, membranous nephropathy, obstructive and reflux uropathy, hyperplastic IgA nephropathy, nephrostomy, hyperplasia of prostate. Further analysis of these clinical meaning of clinical indicators with renal biopsy will help to concrete the expertise to a more explicit form.

We believe that the proposed approach represents an interesting strategy to deal with clinical text data and knowledge fusion for therapy recommendation besides renal biopsy, which may provide advantages in clinical decision supporting applications.

Conclusions

The present paper introduced a renal biopsy recommendation system based on text understanding. We utilized a semi-automatic domain lexicon constructed using CRF and Word2vec for preliminary text analysis, then extracted symptoms and diagnoses to construct an eigenvector with negation detection. Finally, with the help of the machine learning classifier, the system was able to draw a conclusion on whether the patient should perform a renal biopsy based on the given gold standard. To some extent, it provides credible clinical decision support for doctors. Future work could be built relating to method improving and practical application such as visualization. Although clinically it is still necessary to rely on the experience of doctors for kidney biopsy recommendations, this method can provide some help. It is believed that with the accumulation of data and the improvement of the database, the indicators of this method will gradually get better, and expand the application scenarios to assist doctors in various departments to make clinical decisions.

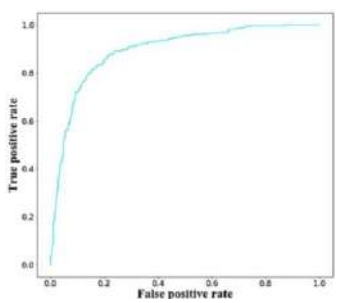


Figure 2-ROC Curve of the Test Set

Table 2– Results of the Test set and Five-fold Cross-validation for the Train set

Index	Test set	Train set				Average
Accuracy	83.5%	83.0%	84.0%	85.2%	85.2%	85.1%	84.5%
Precision	80.6%	81.7%	79.7%	84.6%	80.6%	82.2%	81.8%
Recall	76.6%	73.0%	79.2%	76.1%	81.6%	79.1%	77.8%
F1-measure	78.6%	77.1%	79.5%	80.1%	81.4%	80.6%	79.7%
AUC	0.893	0.894	0.893	0.906	0.911	0.911	0.903

References

[1] R.J. Tan, S.I. Bastacky, and Y. Liu, Molecular Basis of Kidney Disease, in: *Molecular Pathology (Second Edition)*, Elsevier, 2018, pp. 531-553.

[2] Mayo Clinic. Kidney biopsy, 2018.

[3] T. Ruan, J. Gao, D. Feng, X. Qian, T. Wang, and C. Sun, Process and methods of clinical big data mining based on electronic medical records, *Big Data Research* **3** (2017), 83-98.

[4] M. Ju, H. Duan, and H. Li, A CRF-based Method for Automatic Construction of Chinese Symptom Lexicon, in: *International Conference on Information Technology in Medicine & Education*, 2016.

[5] J. Liu, Lucene.Net.Analysis.PanGu, *Github*, 2016.

[6] Z. Fu, CRFSharp, *Github*, 2017.

[7] C. Trim, The Art of Tokenization, 2013.

[8] Johnson0722, The basic principle of Chinese word segmentation and the usage of jieba segmentation, *CSDN*, 2017.

[9] J. Turian, L. Ratinov, and Y. Bengio, Word representations: A simple and general method for semi-supervised learning, in: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, 2010, 384-394.

[10] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, A neural probabilistic language model, *Journal of Machine Learning Research* **3** (2003), 1137-1155.

[11] T. Young, D. Hazarika, S. Poria, and E. Cambria, Recent trends in deep learning based natural language processing, *IEEE Computational Intelligence Magazine* **13** (2018), 55-75.

[12] A. Ng, Word Representations, in: *Coursera*.

[13] F. Xiong, Y. Deng, and X. Tang, The Architecture of Word2vec and its Applications, *Journal of Nanjing Normal University (Engineering and Technology Edition)* (2015), 43-48.

[14] W. Zheng and P. Xu, Research on Chinese Word Clustering with Word2vec, *Computer engineering & Software* **34** (2013), 160-162.

[15] L. Zhou, Exploration of the Working Principle and Application of Word2vec, *Sci-Tech Information Development & Economy* (2015), 145-148.

[16] Y. Liu and W. Li, Research of reverse backtracking matching algorithm for chinese word segmentation, *Journal of Applied Science and Engineering Innovation* **1** (2014).

[17] F. Cai, *Research and Implementation of Portal Information Retrieval System*, Zhejiang University of Technology, 2010.

[18] Z. Jia, H. Li, M. Ju, Y. Zhang, Z. Huang, C. Ge, and H. Duan, A finite-state automata based negation detection algorithm for Chinese clinical documents, in: *Progress in Informatics and Computing (PIC), 2014 International Conference on*, IEEE, 2014, pp. 128-132.

[19] World Health Organization, *International statistical classification of diseases and related health problems*, 2004.

[20] Z. Jia, W. Qin, H. Duan, X. Lv, and H. Li, A Hybrid Method for ICD-10 Auto-Coding of Chinese Diagnoses, *Studies in Health Technology and Informatics* **245** (2017), 427-431.

[21] I. Channel, General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, 2013.

[22] L. Li, W.-Y. Cheng, B.S. Glicksberg, O. Gottesman, R. Tamler, R. Chen, E.P. Bottinger, and J.T. Dudley, Identification of type 2 diabetes subgroups through topological analysis of patient similarity, *Science Translational Medicine* **7** (2015).

Address for correspondence

Haomin Li
hmli@zju.edu.cn

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190312

A Paradigm Shift: Sharing Patient Reported Outcome via a National Infrastructure

Karen Marie Lyng[a], Sanne Jensen[a], Morten Bruun-Rasmussen[b]

[a] *Danish Health Data Authority, Denmark*
[b] *MEDIQ – Medical Informatics and Quality Development, Denmark*

Abstract

Digital solutions transform the way clinical services can be provided and make it possible for patients to participate in decisions concerning their own treatment. With the aim to support a better and more efficient healthcare system in Denmark, it has been agreed among authorities and care providers to establish a national infrastructure for sharing data between hospitals, municipalities, general practitioners and patients and concurrently develop standardized national digital cross-sector questionnaires for the purpose. Sharing data via the national infrastructure enables proactive involvement through patient reported outcomes (PRO). The national infrastructure forms a paradigm shift 1) for collaboration by moving from a baton-passing workflow to sharing-based workflow and 2) for the development of digital cross-sector questionnaires. Cross-sector questionnaire definitions are stored in a national questionnaire repository, and are used in local PRO applications to capture the patients' responses.

Keywords:

Patient Reported Outcomes, Information Dissemination, Questionnaire Design, Questionnaires

Introduction

It is well known that history taking is the most important source of information in healthcare and patients are the best providers of information regarding their own health [1]. Usually a health professional interviewing the patient collects the information from the patient. However, health professionals do not always collect all the relevant information from the patient, besides the task of documenting the patient history is time consuming and patients may even do a better job by filling out a standardized and validated questionnaire [2].

Ask the patient – Patient Reported Outcomes

Across health systems, a wide range of activities have been launched for using Patient Reported Outcomes (PROs). PROs are a general designation for patients' responses to questions about their own state of health. According to the Food and Drug Administration (FDA), PRO is defined as: *"any report of the status of a patient's health condition that comes directly from the patient without interpretation of the patient's response by a clinician or anyone else"* [3]. PROs take the form of carefully designed and validated questionnaires that can be used to capture and quantify the patients' experience, health, and treatment impact. The patients' responses to a questionnaire are denominated as PRO-data. By systematically and actively using PRO-data in the dialogue with the patient, the planning and treatment of the care provided to the patient can be personalized to meet individual needs and concurrently fulfill clinical guidelines. Further PRO-data may also support value-based health management and can be used to perform screening for side effects and reduce the number of unnecessary consultations. At the same time, PRO-data create sound new data for research and quality development. Used correctly, PRO-data are just as essential to the quality of care as clinical data from other more accustomed sources. PRO-data should therefore be an integral part of data useful to clinical practice in the future.

The Danish Health Data Authority has established a permanent PRO secretariat [4] that is responsible for development of standardized cross-sector questionnaires to be used in the healthcare sector nationwide. Standardized questionnaires, for cross sector use, are to be stored and accessed in a national questionnaire repository.

Focus is on using PRO-tools to support active patient participation and involvement. PRO-data are to be used in clinical encounters or as a substitution for outpatient contacts. The aim is to support and strengthen communication between patient and healthcare professional and concurrently ensure efficient processes. PRO-data reported by the patient can be used for clinical assessment before the clinical encounter, or as a dialogue support tool during the encounter; furthermore, PRO-data can be applied to support planning of treatment and as a tool for continuous health monitoring. PRO-data can be shared cross sectors, to all relevant healthcare providers including the patient, via the national infrastructure.

Development of cross-sector questionnaire specifications require involvement of all affected stakeholders including patient representation. It is also crucial that the questionnaires are tested and validated in pilot projects before they are used in daily operations at a national level.

Sharing data via a national infrastructure

To support new flexible ways of cross sector cooperation that include patient involvement, requires concurrent development of IT-technologies as well as well-validated standardized and structured questionnaires. A prerequisite is a generic and standardized IT infrastructure that can support easy interconnection to the many different IT-systems that are used by the various care providers in healthcare.

Messaging has been the main technology applied in Denmark for exchanging information among various healthcare organizations for more than 20 years [5]. Messaging is easy to implement as the focus for the individual message (discharge letter, referral, lab result, etc.) is well-defined. The weakness

in messaging is that information is only shared between the sender and the receiver.

Today, there is a high demand among care providers and patients to request access to relevant clinical data at the point of care independent of location. A common national infrastructure for the healthcare sector has therefore been established in Denmark to make it possible to share relevant eHealth data across all healthcare organizations as well as patients at home. Furthermore, an agreement will be made in the 2020 finance act, which will oblige the parties to start connecting the local IT systems to the national infrastructure with the aim of exchanging PRO-data.

Use of PRO-data will be implemented across regions, municipalities and GPs to ensure sufficiently broad use in and across the healthcare sector based on the common infrastructure and the standardized questionnaires.

Methods

Development of standardized national questionnaires

A standardized method for development of national questionnaires has been developed and evaluated [6]. An overview of the development methodology is illustrated in Figure 1.

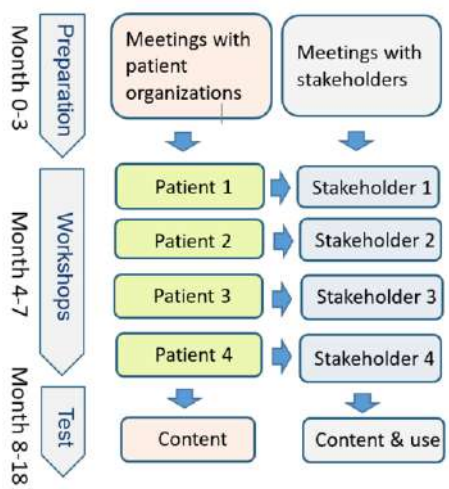


Figure 1 – The Development Methodology for Concerted National Questionnaire Definitions

The design method builds on participatory design principles [7]. A number of meetings and workshops with patient organizations are held to support patient involvement and complement the questionnaire development meetings and workshops with the various stakeholders. The starting point in the development process is always a review of existing PRO material within the area – it is presented to the stakeholders and concurrently a discussion of the aim and practical application is carried out.

By March 2019, six national standardized questionnaires, as shown in Table 1, have been developed.

Table 1 – Fully Developed Areas for PRO in the Program

Area	Status
Apoplexy	Pilot testing
Arthrosis (knee and hip)	Pilot testing
Screening for depression in somatic patients	Pilot testing
Pregnancy and childbirth	Workshops completed
Heart rehabilitation	Workshops completed
Diabetes	Workshops completed

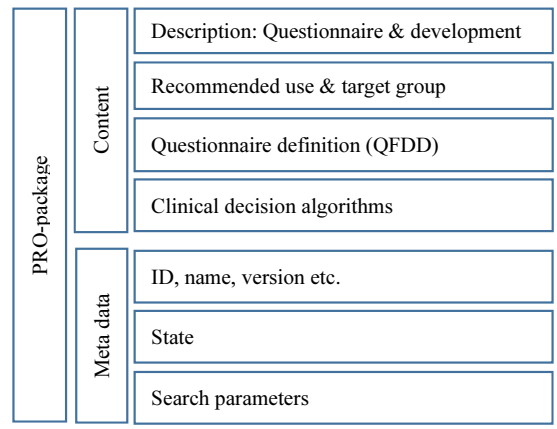


Figure 2 – A PRO-package Contains Content and Metadata

National questionnaire repository

A new national questionnaire repository contains PRO-packages that encompass questionnaires, clinical algorithms, and recommendations for use of PRO questionnaires and PRO-data. A PRO-package contains material as illustrated in Figure 2.

A PRO-package includes a description of the questionnaire and the intended use in several documents that support the technical implementation in the local IT-systems. Metadata are primarily data to be used to search for a specific PRO-package.

The PRO-packages are published via the national questionnaire repository and can be downloaded by healthcare organizations and software providers. However, if a user wants to download a PRO-package, the user must register in the questionnaire repository. As part of the registration, the user provide an e-mail address, that is used to notify the user when the PRO-package is updated.

Development of a national IT infrastructure

The national IT infrastructure builds on the Danish "Reference architecture for collecting health data from citizens" [8]. The reference architecture acts as the common reference for all business areas and IT solutions regarding collection of health data from citizens. The focus in the reference architecture is on data flow from individual citizens to healthcare systems. The data is collected from or provided by the citizen and is communicated to Cross-Enterprise Document Sharing (XDS) repositories in the national infrastructure. Healthcare professionals, that have a "care-relation" with a patient is cleared to look-up relevant data. The use of international standards for communication and content are considered important to ensure semantic interoperability. Furthermore, it

improves quality, as it is the same standardized data set that is applied by all involved care providers.

Further, an important aim of applying an infrastructure based on a standardized reference architecture is to accelerate dissemination and thus implementation in practice.

Sharing PRO-data via the national IT infrastructure focuses on both mono-sectorial and cross-sectorial use of data, thus supporting a smooth cooperation among the various care providers. The national infrastructure is built around a national service platform as shown in Figure 3. The platform among other things holds information on whether there is a 'care-relation' between any given care-provider and the patient.

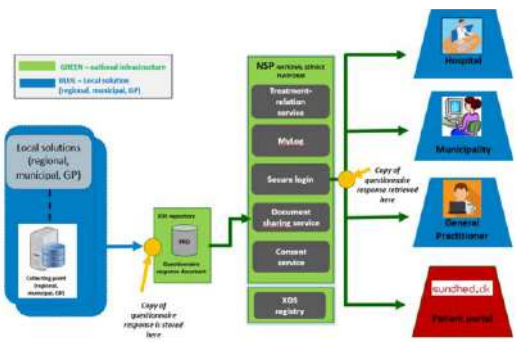


Figure 3 – National Service Platform and Services

Document Sharing Service

The Document Sharing Service (DSS) is a national service for sharing documents among healthcare organizations. Documents can hold any kind of data, but they have to be expressed in an agreed upon standardized format. The DSS is based on the international standard for maintaining and sharing documents among healthcare organizations – Cross-Enterprise Document Sharing (XDS). This is managed through federated document repositories and a document registry to create a longitudinal record of information about a patient within a given clinical affinity domain.

Consent service

According to the Danish healthcare laws, health professionals have by default the right to access the citizens' healthcare data, if it is relevant in relation to actual care situation and the patient has provided oral consent.

In order to restrict access to data a citizen can block:

- Access to all data for a specific period
- Acess to data from a specific healthcare organization
- Access to data for a specific person

The consulting service is linked to the DSS and will automatically filter out documents, if the citizen has blocked access to data.

Standards for questionnaires

Two HL7 CDA v2 standards are used for questionnaires. The Questionnaire Form Definition Document (QFDD) describes how the questionnaire is constructed and organized. The QFDD includes information about author, version, tittle, heading, text for sections and for each question; it is specified how the answer to any question can be given e.g. as free text, numeric values or multiple choice. It is also possible to add conditions where a specific response will open-up a new section with further questions. The QFDD is important to use, when receiving a populated questionnaire in order to understand and interpret the answer.

The Questionnaire Response Document (QRD) is the questionnaire filled by the patient (the PRO-data). The QRD has a link to the QFDD.

The QFDD and the QRD standards ensure that data can be integrated and reused (semantic interoperability) in the numerous existing IT systems, that are used every day in the hospitals, municipalities, and by the general practitioners.

Before the healthcare organizations can begin to share data via the national IT infrastructure, the correct implementation of the QFDD and QRD standards are tested and certified. The tests and certification are mandatory to all users of the IT infrastructure.

Testing and evaluation of the national standardized questionnaires

When a national questionnaire has been developed, a validation is carried out with approximately 20–30 patients, representing a broad diversity of potential responders, e.g. differing in age, sex as well as social and educational backgrounds. Questionnaire algorithms and usability are also tested in small-scale implementation and at several pilot sites. The pilot sites are cross-sectorial according to the relevant cross-sectorial workflows.

A research program has been set up to identify categories of eligible patients, benefits and disadvantages of using PRO-data and the effect application of a system for Patient Reported Information may have on clinical work. From previous experience, we know that the use of PRO-data will empower patients and improve efficiency and effectiveness of care processes in some patient groups [9, 10] and it may even improve outcomes [11].

Results

Dissemination

A new national IT infrastructure based on IHE XDS [14] and HL7 CDA standards has been established and tested. The use of international standards ensures that relevant data can be shared across hospitals, municipalities, and general practitioners independent of which IT-providers and systems are employed. The patients further have access to their own PRO-data as well as other health data via the Danish Public Healthcare Portal [12].

The IT infrastructure is in daily operation and is until now also used for sharing of patient appointments, patient master data, care plans and home monitoring as well as PRO-data.

Maturity the infrastructure

On the Technical Readiness Levels (TRL) scale developed by National Aeronautics and Space Administration (NASA) [13] as a methodology to access the maturity of critical technology solutions, the infrastructure is assessed to be on TRL 8 (in daily operation with real users) by March 2019.

As shown in 4, TRL is based on a scale from one to nine, where nine is the highest maturity level of the technology solution. The use of TRL also provides the basis of

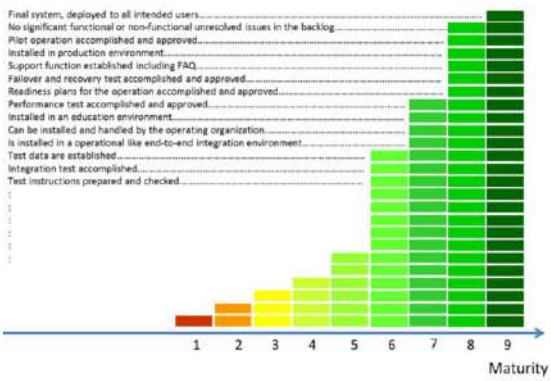


Figure 4 – TRL Scale to Access the Maturity

a consistent description of the maturity level of the national infrastructure.

Questionnaire development

The most important findings during the first rounds of development and evaluation of questionnaire are listed in Table 2.

Table 2 – The Most Important Findings During the Questionnaire Development Process

Area
1. Many issues need to be clarified by the patients before and between stakeholder workshops
2. Most patient pathways are cross sectorial, but cross sectorial workflows are complex and not well described
3. Most previous knowledge on patient reported information is derived from mono-sectorial use in hospitals, as digitalized Patient Reported Information has been only scarcely used in primary care
4. Use of Patient Reported information is a fundamental cultural transformation to health professionals
5. Legal clarification of who may have access to what kind of Patient Reported Information is important
6. Electronic questionnaire systems including algorithms are medical devices and thus must be CE marked in the EU

Discussion

Currently there is a lot of ongoing work in Denmark to develop new IT-solutions that can involve and empower the patient and make the care provision more personalized and efficient. There are several strategies to address this challenge, one is to ensure efficient access to relevant patient data for all care providers in all the various health organizations involved in the care process of a single individual.

Messaging

For many years, messaging has been the preferred approach for communication of health data. Messaging is based on the premise that the sender and the receiver have knowledge of each other; for example, a hospital is sending a discharge letter to the patient's general practitioner. This is a baton workflow, where the diagnosis and treatment are done stepwise by handing over the patient and what is judged as relevant information from one care provider to another care provider. In the baton workflow process, there is a risk of not sending the relevant information or for loss of information, as the patient's data are only communicated as singular messages between the sender and the receiver.

Collaboration via a national infrastructure

Today's tecnology allows the establishment of a national infrastructure where data are shared among many actors. Relevant data can be collected, shared, and applied where and when it is relevant (the services in the National Service Platform ensure that there is legitimate right to access the data).

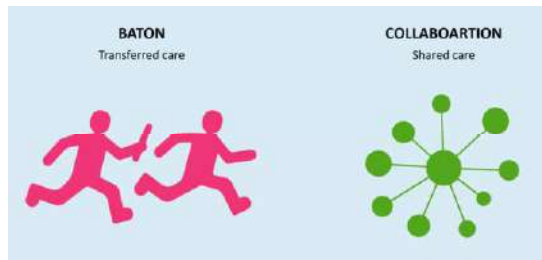


Figure 5 – From Baton to Collaboration

Going from Baton to Collaboration provides a profound shift in the information exchange paradigm as well as in the possible ways to cooperate. Further introducing questionnaires filled by the patients and accessible whenever relevant provide a completely new set of possible ways to cooperate in the healthcare sector.

The national infrastructure is designed as a common generic infrastructure, which can be applied by all care providers, though we have found that it is a challenge that many small healthcare organizations do not have internal IT knowledge and capabilities to establish a smooth integration to the national IT infrastructure. Further legal issues have to be addressed and handled in the way questionnaires are designed and distributed as well as in the design of the infrastructure. It is, however, a benefit that the security issues are taken care of in the design of the IT infrastructure

It is essential that all stakeholders are represented in the process of the development of questionnaires and that all stakeholders actively participate in the dialogue and teamwork. Furthermore, as one of the central aspects in participatory design points out, it is essential that stakeholders collaborate on equal terms [7]. We have managed to support this aspect by conducting dedicated workshops for patients and for clinicians from the primary care sector as we found that they needed more support in the process. Clinicians from the hospitals have an advantage of their prior experience in development and use of questionnaires with patient reported information.

Conclusions

Based on our findings, we are convinced that the increased use of patient reported information involving patients and clinicians in the development, evaluation, and implementation processes will be of great benefit to both the individual patients as well as the entire healthcare system [11]. However, it requires a thorough and exhaustive development program including tracks for technical infrastructure, questionnaire development, evaluation, and application as well as a research program.

Sharing data versus messaging is a major paradigm shift concerning technology, use of eHealth standards, clinical workflow, security, legal aspects, and privacy.

Acknowledgements

The Danish program on Patient Reported Information is a cooperation between state, regions, municipalities, and patient organizations in Denmark.

References

[1] J.R. Hampton, et al., Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients*, Br Med J,* **2(5969)** (1975), 486-9.

[2] J.M Gumpel. and A.M. Mason, Self-administered clinical questionnaire for outpatients. *Br Med J*. **2(5912)** (1974), 209-12.

[3] Food and Drug Administration, U.S. Department of Health and Human Services: *Guidance for Industry. Patient-Reported Outcome Measures: use in medicinal product development to support labeling claims*. 2009. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf

[4] Danish PRO secretariat, www.pro-danmark.dk

[5] H. Bjerregaard, and C. Duedal Pedersen, MedCom: Danish Health Care Network*,* Technology and Informatics **100** (2004), 59-65

[6] S. Jensen, and K.M. Lyng, Establishing a Nation Wide Infrastructure for Systematic Use of Patient Reported Information*, Studies in Health Technology and Informatics* **247** (2018), 226-230.

[7] K.M Lyng. and B.S. Pedersen, Participatory Design for Computerization of Clinical Practice Guidelines. *Journal of Biomedical Informatics* **44** (2011), 909-918

[8] Reference architecture for collecting health data from citizens. Danish *National eHealth Authority (2013)*

[9] K.M Lyng. and H.K. Hegaard, Empowering of pregnant women by making them report their own primary clinical record in a web-based system *Presentation at ISQA* 2013, (2013).

[10] C. Mejdahl, et al., Use of patient-reported outcomes in outpatient settings as a means of patient involvement and self-management support, *European Journal for Person Centered Healthcare* **4(2) (**2016) 359-367.

[11] E. Basch*, Patient-Reported Outcomes - Harnessing Patients' Voices to Improve Clinical Care. N Engl J Med*. **376(2)** (2017), 105-108.

[12] Danish Healthcare Portal, www.sundhed.dk/borger/service/om-sundheddk/ehealth-in-denmark/

[13] Scandurra, I., M. Hägglund, and S. Koch, From user needs to system specifications: Multi-disciplinary thematic seminars as a collaborative design method for development of health information systems. *Journal of Biomedical Informatics* **41(4)** (2008), 557-569.

[14] Integration the Health Care Enterprise*. IHE IT Infrastructure. Volume 2b. (Revision 15.0 Lyly 24, 2018)*

Address for correspondence

Karen Marie Lyng, Head of Department, Department of Data Quality and Registries, The Danish Health Data Authority, Ørestads Boulevard 5, DK-2300 Copenhagen S; Email: kmly@sundhedsdata.dk (preferred method of contact)

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190313

Acceptability of Telemedicine to Help African American Women Manage Anxiety and Depression

Terika McCall[a], Todd Schwartz[b,c], Saif Khairat[a,b]

[a] *Carolina Health Informatics Program, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*
[b] *School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*
[c] *Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC*

Abstract

African American women experience rates of mental illness comparable to the general population, however they significantly underutilize mental health services. Past studies have shown that telemedicine is effective, and can be used to increase access to services. To assess the acceptability of using mobile video calls, a web-based survey was launched to solicit response from African American women. The results of this exploratory study (N=102) showed high acceptance of the use of video calls to communicate with a professional to receive help to manage anxiety and depression (> 70% endorsed). Statistically significant associations were found between age and agreement with the use of video calls, with younger women (< 50 years) more likely to indicate acceptance (p's < .05). The findings of this study demonstrated the need for additional research into the use of telemedicine to provide African American women with more accessible and convenient options for mental health care.

Keywords:

Telemedicine, women, mental health

Introduction

In 2017, an estimated 5.1 million African American adults experienced mental illness in the last year [1]. Women comprised approximately 54% of this population [1]. The prevalence of mental illness among African American women (18.6%) was comparable to the general population (18.9%) [1]. However, there is significant racial disparity in the use of mental health services among African American women compared to their white counterparts [1]. Telemedicine can be used to increase access to mental health care in this population, and reduce disparities.

Anxiety and mood disorders are among the most common mental illnesses in the U.S. [2]. Women are nearly twice as likely as men to experience an anxiety (e.g. generalized anxiety disorder) or mood disorder (e.g. major depressive disorder) in their lifetime [2]. Results from the 2008 *National Health Interview Survey* found that 15.7% of non-Hispanic black women reported having generalized anxiety in their lifetime [3]. Furthermore, 27.4% of non-Hispanic black women reported experiencing depression in their lifetime [3]. Mental illness is underreported in the African American community; therefore, the prevalence estimates may actually be substantially lower than the true burden.

Although there is significant burden of mental illness among African American women, mental health services are underutilized. Results of the 2017 *National Survey on Drug Use and Health* revealed that African American women use mental health services at less than half the rate of their white counterparts (10.6% compared to 23.4%) [1]. Among the African American women that reported experiencing mental illness in the last year, 64.2% did not receive any mental health treatment during that time [1]. There are many reasons why African American women may not seek mental health services when needed. Barriers such as stigmatization of mental illness [4,5], less access to treatment, no or inadequate health insurance, mistrust of providers, and low health literacy prevent traditionally marginalized populations from seeking care [5].

Actual and perceived racism and sexism may cause undue stress and lead to increased anxiety or depression in African American women. A study published by Ward & Heidrich [6] found that African American women's preferred coping strategies included praying and seeking medical and mental health care. Taking this into consideration, how do we make mental health services more accessible to this population? The ideal solution would incorporate technology they already have access to, and require minimal education on its use. Telemedicine may be a viable solution. Specifically, the use of mobile video calls to communicate with a professional to receive help for managing anxiety or depression.

Eighty percent of African American women own smartphones [7]. Most smartphones are equipped with the capability to complete video calls. To our knowledge, there have been no studies that have examined the use of mobile video call use to manage anxiety or depression in African American women. However, evidence from past studies showed that videoconferencing has been effective in helping participants reduce anxiety or depressive symptoms [8–10].

Due to the paucity of published peer-reviewed literature on the use of video calls for therapy, which also include a significant representation of African American women in the study sample, more investigation into the use of this technology with African American women is needed. Given that current mental health services are underutilized by this population, the population should be surveyed to determine the acceptability of this approach to treatment. The aim of this exploratory study is to gauge the acceptability of telemedicine, using mobile video calls, to deliver mental health services to help African American women manage anxiety and depression.

Methods

Study Design and Recruitment

The self-administered web-based survey was launched in June and closed August 2018. Eligible participants were African American women (≥ 18 years of age) that reside in the U.S.A. Participants were recruited through convenience sampling.

Recruitment methods included sending an invitation to take the survey via a direct email from the first author, or email distributed through listservs whose membership is primarily African American women. Participants were also recruited via social media (e.g. Facebook) posts and direct messages. A link to the research information sheet about the study was provided in the email text and social media posts. Following a snowball sampling technique, respondents were encouraged to share the link to the survey with their networks (e.g. family, friends, professional organizations, etc.). No remuneration was offered for participation. The study received notification from the institutional review board (IRB) of exemption from further review.

Measures

Due to the sensitive nature of the questions and for convenience, the computer-assisted web interviewing (CAWI) data collection technique was used to administer the survey instead of in-person interviewing. Respondents were provided an anonymous link to the online questionnaire. No personally identifiable information (PII) was collected in the survey. The survey consisted of 53 questions, and was administered using Qualtrics software. Survey domains included sociodemographic characteristics, attitudes toward seeking professional psychological help, mobile phone use, and acceptability of using a mobile phone to receive mental health care.

Sociodemographic data such as the respondent's race, ethnicity, age, gender, and highest level of education attained were captured at the beginning of the survey. The race, age, and gender questions were also used as screener questions to determine eligibility to complete the survey. If the respondent did not self-identify as a Black/African American (or biracial, Black/African American and another race) female age 18 years or older, they were routed to the end of the survey.

Attitudes Toward Seeking Professional Psychological Help

In the survey, the term "professional" referred to individuals who have been trained to deal with mental health problems (e.g., psychologists, psychiatrists, social workers, and family physicians). Respondents attitudes toward seeking professional psychological help was measured using questions from an adapted version of the validated *Inventory of Attitudes Toward Seeking Mental Health Services (IASMHS)* [11]. The IASMHS consists of 24 questions and the following factors: psychological openness, help-seeking propensity, and indifference to stigma. Response options to the survey items were on a 5-point Likert-type scale from disagree (0) to agree (4).

In order to collect data specifically about attitudes toward seeking professional help for anxiety and depression, six questions in the inventory were revised. The words 'psychological problems' and 'mental disorder' were substituted with 'anxiety' and 'depression'. For example, Item #3 in the IASMHS reads, "I would not want my significant other (spouse, partner, etc.) to know if I were suffering from psychological problems." The revised survey question states, "I would not want my significant other (spouse, partner, etc.) to know if I were suffering from **anxiety**." To prevent asking double-barreled questions, respondents were asked about anxiety and depression independently for the six questions. This increased the total number of questions in the inventory to 30.

Mobile Phone Use

Mobile phone use was ascertained with the following items: (1) current mobile phone ownership (yes/no); (2) frequency of sending text messages (never, less than 1 time per week, 1-6 times per week, 1-3 times per day, 4 or more times per day); (3) ability to complete video calls on mobile phone (yes/no); and (4) frequency of using mobile phone to complete video calls (never, less than 1 time per week, 1-6 times per week, 1-3 times per day, 4 or more times per day).

Acceptability of Mobile Phone Use for Mental Health Care

Acceptability of using a mobile phone to receive mental health care to manage anxiety or depression was measured by the following items: (1) comfortability communicating with a professional through text messaging to receive help for managing anxiety/depression; (2) comfortability communicating with a profession through video call to receive help for managing anxiety/depression; (3) whether having the option to use a video call to complete an appointment with a professional would increase access to mental health services; (4) convenience of using video call to complete an appointment (e.g. less travel time); (5) convenience of using video call to complete an appointment compared to an in-person visit; and (6) perceived helpfulness of having the option to use text messaging to communicate with a professional if feeling anxious/depressed. Response options to the survey items were on a 5-point Likert-type scale from disagree (0) to agree (4).

Before completing the survey, respondents were asked two questions, "Do you have any concerns about using a video call to complete an appointment with a professional?" and "Do you have any concerns about using text messaging to communicate with a professional?" If they answered 'yes' to either question they were presented with an open-ended question asking them to note their concerns in the multi-line text field provided.

Data Analysis

Statistical analyses were conducted using SPSS version 25 software. Descriptive statistics were calculated for sample characteristics and responses to video call questions as mean, standard deviation, and range for continuous variables, and as frequencies and percentages for categorical variables. Age was segmented into two groups (under 50 years and 50+), education was collapsed into three levels (< Bachelor's, Bachelor's degree, Graduate degree), and response options were collapsed into three categories (Agree/Somewhat agree, Undecided, and Disagree/Somewhat disagree). Fisher's exact test was used to determine whether an association exists between the response to each video call question with age group and highest education level, respectively. Statistical significance was determined at the two-sided $p < .05$ level.

Results

Of the 113 respondents that started, 102 (90%) completed the survey. The characteristics of the completers are presented in Table 1. Participants ranged in age from 19 to 80 years, and all identified as either Black/African American or biracial (i.e. Black/African American and another race) and female. Most respondents (98%) identified as non-Hispanic.

Approximately 15% had less than a bachelor's degree, 23% obtained a bachelor's degree, and 62% had a graduate degree. Most participants (96%) reported having video call capability on their mobile phone, and 39% reported completing a video call at least once per week.

Acceptability of Video Call Use

Communicating with a Professional

The majority of respondents (> 70%) indicated agreement (i.e. selected the 'Agree' or 'Somewhat agree' response options) with feeling comfortable using video call to communicate with a professional. Figure 1 shows 70.6% agreement (43.1% agree,

27.5% somewhat agree) with the statement, "I would feel comfortable communicating with a professional through a video call to receive help for managing **anxiety.**" Similarly, 71.6% (40.2% agree, 31.4% somewhat agree) of respondents showed agreement with the statement, "I would feel comfortable communicating with a professional through a video call to receive help for managing **depression.**" See Figure 2.

Access and Convenience

The survey results revealed high levels of agreement with the access and convenience statements. Figure 3 shows 83.3% agreement (69.6% agree, 13.7% somewhat agree) with the statement, "Having the option to use a video call to complete an appointment with a professional increases my access to mental health services." Likewise, 88.2% (72.5% agree, 15.7% somewhat agree) of respondents indicated agreement with the statement, "Using a video call to complete an appointment would save me time traveling to a professional's office." See Figure 4. Furthermore, Figure 5 shows 71.6% (47.1% agree, 24.5% somewhat agree) of respondents indicated agreement with the statement, "Using a video call to complete an appointment with a professional would be more convenient for me than an in-person appointment."

Table 1– Characteristics of Study Participants

Participant characteristics		(N=102)
Age in years, mean (SD)		38.8 (13.1)
Age group in years, n (%)		
	Under 50	81 (79.4)
	50 and older	21 (20.6)
Race, n (%)		
	Black or African American	100 (98.0)
	Biracial[a]	2 (2.0)
Ethnicity, n (%)		
	Hispanic	2 (2.0)
	Non-Hispanic	100 (98.0)
Education, n (%)		
	< Bachelor's degree	15 (14.7)
	Bachelor's degree	24 (23.5)
	Graduate degree	63 (61.8)
Mobile phone video call capability, n (%)		
	Yes	98 (96.1)
	No	4 (3.9)
Frequency of video call use[b], n (%)		
	Never	12 (11.8)
	Less than 1 time per week	46 (45.1)
	1-6 times per week	29 (28.4)
	1-3 times per day	9 (8.8)
	4 or more times per day	2 (2.0)

[a]Biracial defined as identifying as Black or African American and another race.
[b]Participants that indicated that their phone did not have video call capability (n=4) were not presented for the frequency of use question.

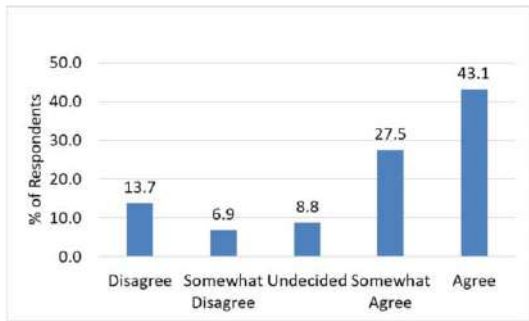


Figure 1–Sample percentages for response to the statement, "I would feel comfortable communicating with a professional through a video call to receive help for managing anxiety"

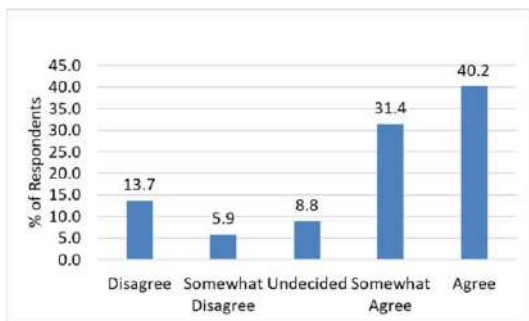


Figure 2–Sample percentages for response to the statement, "I would feel comfortable communicating with a professional through a video call to receive help for managing depression"

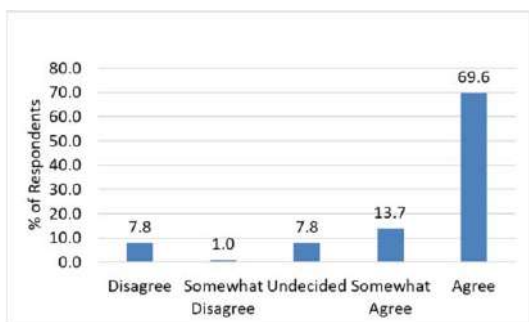


Figure 3–Sample percentages for response to the statement, "Having the option to use a video call to complete an appointment with a professional increases my access to mental health services"

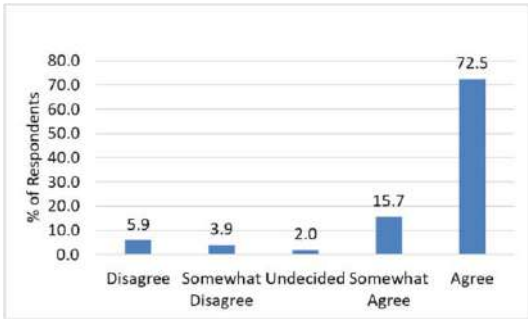


Figure 4–Sample percentages for response to the statement, "Using a video call to complete an appointment would save me time traveling to a professional's office"

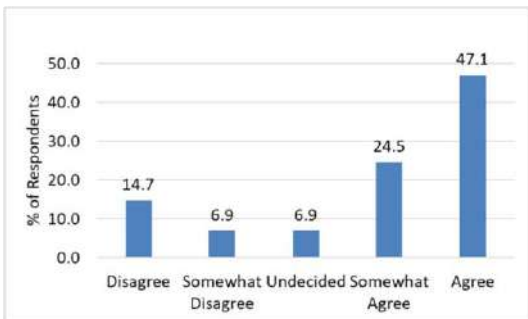


Figure 5–Sample percentages for response to the statement, "Using a video call to complete an appointment with a professional would be more convenient for me than an in-person appointment"

Concerns about Video Call Use

Although overall response to using video calls was positive, approximately 43% of respondents indicated having concerns about using a video call to complete an appointment with a professional. Most of the concerns centered around privacy, confidentiality, and maintaining the video call connection. Regarding privacy, one respondent stated, "I would also be concerned about how private the call is, like is the professional alone, how would I know?" Another respondent voiced connectivity concerns stating, "…the call could get disconnected. That breaks the rhythm of the conversation."

Acceptability of Video Call Use by Age and Education

Statistically significant associations were found between age and level of agreement with the following video call statements: "I would feel comfortable communicating with a professional through a video call to receive help for managing **anxiety**" (highly significant negative association, $p = .002$); "I would feel comfortable communicating with a professional through a video call to receive help for managing **depression**" (highly significant negative association, $p = .003$); "Using a video call to complete an appointment would save me time traveling to a professional's office" (highly significant negative association, $p = .001$); and "Do you have any concerns about using a video call to complete an appointment with a professional?" (positive association, $p = .025$). No statistically significant associations were observed between level of agreement and education (all $p > .05$).

Discussion

To our knowledge, this preliminary study was the first to measure the acceptability of telemedicine, using mobile video call, to deliver mental health services to African American women. The results of this study showed high acceptance. A majority of the women responded that they agree/somewhat agree with the statements regarding feeling comfortable using a video call to communicate with a professional to receive help to manage anxiety and depression (> 70%). Most respondents also believed that the use of video calls will increase their access to mental health services (> 80%), and that it is a convenient option for care (> 70%). Younger women (< 50 years) were more likely to indicate acceptance. While older women (≥ 50 years) were more likely to have concerns about using a video call.

These findings are consistent with literature on the use of telehealth modalities. Although a positive overall reaction was observed in adults, in general, there is greater acceptance among younger adults for the use of technology [12]. Previous studies have also found that older adults generally have many concerns about the use of technology [13].

The concerns of privacy, confidentiality, and connectivity must be addressed when considering use of video calls. Transparency is key to reducing medical mistrust and adoption of video calls as a sustainable mode to deliver care to this population. Clients should also be advised to connect to Wi-Fi or ensure they have a strong cellular signal and sufficient mobile data for high quality video transmission and to prevent disconnection.

Given the stigma of mental illness in the African American community, the option to complete an appointment with a professional in a setting of their choosing may appeal to African American women. Completing a therapy session in a familiar location (e.g. home) may provide a relaxing environment that is viewed as more discrete, and promotes open communication. Furthermore, increasing the accessibility and convenience of therapy helps to overcome some of the barriers to utilizing mental health services.

The main limitations of this preliminary study were recruitment method and sample size. Participants were recruited through convenience sampling, and encouraged to share the survey email or social media posts with their networks. Although no PII was collected in the survey and respondents accessed the survey through an anonymous link, social desirability bias could have resulted if the respondent personally knew the first author. The sample size of 102 respondents is small for this cross-sectional survey, and consisted of mostly younger (under 50 years) and highly educated women (over 85% had at least a bachelor's degree).

This limits our ability to broadly generalize the findings to all African American women. Although stigma may continue to be a barrier for highly educated African American women, access to mental health services, insurance coverage, and health literacy may be less of an issue for this group. However, due to the shortage of African American psychologists (< 5% of active psychologists), access to a professional that meets the ethnicity and gender preferences of the patient may remain an issue [14]. While some patients may have no preference, African Americans who indicate higher levels of mistrust of whites are more likely to discontinue therapy before treatment goals are reached if they are seen by a white counselor [15].

Future work will include relaunching this survey to a larger and more diverse sample of African American women. Questions from the Patient Health Questionnaire (PHQ-9) and Generalized Anxiety Disorder (GAD-7) screeners will be added to capture data on the presence and severity of depression and

anxiety, respectively. In addition, we will explore the acceptability of using text messages to communicate with a professional to receive help for managing anxiety and depression.

Conclusions

The use of telemedicine is a viable solution to help African American women manage anxiety and depression. The study showed high acceptance among this population. Furthermore, over 95% of respondents had the capability to complete video calls on their mobile phone. The use of video calls is an underutilized technology that has the potential to help reduce the disparity in mental health service utilization and improve health outcomes for this population.

Acknowledgements

The first author is supported by funding from the National Library of Medicine's Institutional Training Grant for Research Training in Biomedical Informatics and Data Science (T15-LM012500).

References

[1] Substance Abuse and Mental Health Services Administration, *2017 National Survey on Drug Use and Health* [data file], RTI International [distributor], Research Triangle Park, NC, Retrieved 13 November 2018 from https://pdas.samhsa.gov/#/survey/NSDUH-2017-DS0001.

[2] R. C. Kessler, P. Berglund, O. Demler, R. Jin, K. R. Merikangas, and E. E. Walters, Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication, *Archives of General Psychiatry* 62(6) (2005), 593–602.

[3] Health Resources and Services Administration, *Women's Health USA 2010*, U.S. Department of Health and Human Services, Rockville, Maryland, 2010.

[4] V.L. Thompson-Sanders, A. Bazile, and M. Akbar, African Americans' perceptions of psychotherapy and psychotherapists, *Professional Psychology: Research and Practice*, 35(1) (2004), 19–26.

[5] National Alliance on Mental Illness, *Multicultural Mental Health Facts*, Retrieved 13 November 2018 from https://www.nami.org/Learn-More/Mental-Health-By-the-Numbers.

[6] E. C. Ward and S. M. Heidrich, African American women's beliefs about mental illness, stigma, and preferred coping behaviors, *Research in nursing & health* 32(5) (2009), 480–492.

[7] The Nielsen Company, *Reaching Black Women Across Media Platforms,* Retrieved 14 November 2018 from www.nielsen.com/us/en/insights/news/2017/reaching-black-women-across-media-platforms.html.

[8] N. G. Choi, M. T. Hegel, N. Marti, M. L. Marinucci, L. Sirrianni, and M. L. Bruce, Telehealth problem-solving therapy for depressed low-income homebound older adults, *The American Journal of Geriatric Psychiatry* 22(3) (2014), 263–271.

[9] D. A. Dunstan and S. M. Tooth, Treatment via videoconferencing: a pilot study of delivery by clinical psychology trainees, *The Australian Journal of Rural Health* 20(2) (2012), 88–94.

[10] E. K. Yuen, J. D. Herbert, E. M. Forman, E. M. Goetter, A. S. Juarascio, S. Rabin, C. Goodwin, and S. Bouchard, Acceptance based behavior therapy for social anxiety disorder through videoconferencing, *Journal of Anxiety Disorders* 27(4) (2013), 389–397.

[11] C. S. Mackenzie, V. J. Knox, W. L. Gekoski, and H. L. Macaulay, An adaptation and extension of the attitudes toward seeking professional psychological help scale, *Journal of applied social psychology* 34(11) (2004), 2410–2433.

[12] T. Newlin, T. McCall, P. Ottmar, B. Welch, and S. Khairat, Assessing the satisfaction of citizens using teleconsent in clinical research, *Studies in Health Technology and Informatics* 247 (2018), 685–689.

[13] S. T. M. Peek, E.J.M. Wouters, J. van Hoof, K.G. Luijkx, H.R. Boeije, and H.J.M. Vrijhoef, Factors influencing acceptance of technology for aging in place: a systematic review, *International Journal of Medical Informatics* 83(4) (2014), 235–248.

[14] American Psychological Association, *Demographics of the U.S. psychology workforce: Findings from the 2007-16 American Community Survey*, L. Lin, K. Stamm, and P. Christidis, Washington, DC, 2018.

[15] F. Terrell and S. Terrell, Race of counselor, client sex, cultural mistrust level, and premature termination from counseling among Black clients, *Journal of Counseling Psychology* 31(3) (1984), 371–375.

Address for correspondence

Corresponding Author: Saif Khairat, PhD, MPH, FAMIA, 428 Carrington Hall, Chapel Hill, NC, USA 27514; Email: saif@unc.edu.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190314

Evaluation of a Clinical Decision Support System for the Prescription of Genetic Tests in the Gynecological Cancer Risk

Jesús Moreno-Conde[ab], Celia Alvarez-Romero[ab], Cristina Suárez-Mejías[ab], María Ángeles Martínez-Maestre[b], José Manuel Silvan-Alfaro[b] and Carlos Luis Parra-Calderón[ab]

[a] *Biomedical Informatics, Biomedical Engineering and Health Economics, Institute of Biomedicine of Seville, IBIS / Virgen del Rocío University Hospital / CSIC / University of Seville. Seville, Spain*
[b] *Technological Innovation Group, Virgen del Rocio University Hospital, Seville, Spain*
[c] *Gynecology Department, Virgen del Rocío University Hospital, Seville, Spain*

Abstract

Clinical Decision Support System (CDSS) has been implemented to support physicians about the medical prescription of genetic testing. CDSS is based on open source software. A CDSS for prescribing these genetic tests in BRCA1 and BRCA2 and preventing gynecological cancer risks has been designed and performed in the 'Virgen del Rocio' University Hospital. Clinical evidence demonstrates that BRCA1 and BRCA2 mutations can develop gynecological cancer, but genetic testing has a high cost to the healthcare system.

The developed technological architecture integrates open source tools like Mirth Connect and OpenClinica. The system allows general practitioners and gynecologists to classify patients as low risk (they do not require a specific treatment) or high risk (they should be attended by the Genetic Council), According to their genetic risk, recommending the prescription of genetic tests. The aim main of this paper is the evaluation of the developed CDSS, getting positive outcomes.

Keywords:

Clinical Decision Support Systems, Breast Neoplasms, Ovarian Neoplasms.

Introduction

Breast Cancer (BC) is the most common type of cancer in women in the western world. Besides, Ovarian Cancer (OC) ranks fifth in cancer death in women. However, difficulties in their diagnosis and therapeutic treatment imply high mortality, greater than 50% at five years from diagnosis. The International Agency for Research on Cancer estimated a worldwide incidence of 1.67 million new BC cases per year and over 0.23 million of OC [1].

The appearance of tumors of BC and OC is usually sporadic, but around 10-15% of diagnosed cases are heritable. BRCA1 and BRCA2 genes describe germinal mutations. They are inherent dominantly and with high penetrance in 7% of BC and around 11-15% of OC [2]. Therefore, BRCA1 and BRCA2 mutations increase the risk of BC and OC. Concretely, people with BRCA1 mutations have a 57% risk of developing BC and around 40% of OC. People with BRCA2 mutations have a 49% risk of developing BC and approximately 18% of OC [3-5].

Hereditary gynecologic cancer usually starts in younger people while this trend is not usual for sporadic cancer. Also, it has a more invasive histopathological pattern. However, its diagnosis is important because the result could be positive with determined treatments [6].

In this sense, to know the risk of having a hereditary cancer is essential for people with family history. Moreover, patients' anxiety and concern may be provoked if the risk stays unknown. The treatments for individual patients are different, depending on the BRCA1 and BRCA2 gene mutation. If the mutation is positive, patients will receive monitoring or preventative measures. Otherwise, patients can be relieved of anxiety. In this sense, there are some studies which demonstrated that patients are benefited by the genetic testing results [7-9]. These studies found out a significant decrease in patient concern and anxiety about developing cancer.

In clinical practice, general practitioners and gynecologists are the first in the attendance of the patient. However, the genetic tests to analyze BRCA1 and BRCA2 gene mutation have a high cost to the healthcare system. Previous studies also indicate that performing these genetic tests to the population is not cost-efficient [10]. In this sense, criteria were defined to identify patients with high risk for developing those mutations. They were agreed between the scientific society and the official organization [11]. Genetic testing is only recommended for patients with previous family and personal history. The CDSS was developed for this purpose and it will optimize the prescription and the care process avoids unnecessary patient referrals.

Besides, a recent study has assessed data privacy, security protection and health-promoting role modeling in the technology acceptance model. After controlling for several covariates, perceptions of usefulness, data privacy and security protection, and health-promoting role were all statistically significant factors that influenced the use of electronic personal health records. Those findings suggest that electronic health records users feel more protected and less concerned about privacy and security when their providers use electronic health records [12].

In this paper, this CDSS is functionally and clinically evaluated with preliminary results. Concretely, static rules defined by the Spanish Society of Clinical Oncology (SEOM) will be validated. The dimensions to be assessed are: Perceived Usefulness, Perceived Ease of Use, Social Norms, Facilitating Conditions and Intention to Use.

Methods

Specific questionnaires based on Technology Acceptance Model (TAM) were designed to measure clinical staffs' acceptance of the technology. The model was developed by David (Davis, 1989) and David et al. (David et al., 1989) and it's effective and proven in predicting the use of information and communication technologies. Besides, it has been demonstrated the leading information and communication technology (ICT) application areas for the TAM in health services: telemedicine, electronic health records, and mobile applications. The original TAM has been extended to fit dynamic health service environments by integration of components such as the theory of planned behavior and unified theory of acceptance and use of technology. Those variables frequently reflect the concepts of subjective norm and self-efficacy, but also compatibility, experience, training, anxiety, habit, and facilitators are considered [13].

In another study [14], it was found that the original TAM constructs had a significant impact on the staffs' behavioral intention to adopt HIS in paraclinical departments. The user behavior factors are essential for successful usage of the system and should be considered. It provides valuable information for hospital system providers and policy makers in understanding the adoption challenges as well as practical guidance for the successful implementation of information systems in paraclinical departments.

The TAM model is used to predict the use of technologies, based on two main characteristics: Perceived Usefulness and Perceived Ease of Use. Perceived Usefulness refers to the degree to which a person believes that using a system will improve the performance of a given job. The Perceived Ease of Use aims to assess to what degree a person believes, using a particular system, makes less effort to perform their tasks. The TAM aims to explain the causes of the acceptance of technologies by users. In that sense, it is considered that a person's perceptions of the Perceived Usefulness and Perceived Ease of Use of a system are conclusive in determining their Intention to Use it. According to this model, there are external variables that have a direct influence on the Perceived Usefulness and Perceived Ease of Use. In our study, different variants of TAM are included to evaluate different dimensions.

In order to evaluate the Clinical Decision Support System for the prescription of the BRCA1, BRA2 genetic tests in the prevention of hereditary breast and ovarian cancer, a personalized questionnaire was designed for evaluation by Gynecology based on TAM. The dimensions to be assessed in this questionnaire were: Perceived Usefulness, Perceived Ease of Use, Social Norms, Facilitating Conditions and Intention to Use.

Specifically, the items of the questionnaires assessed for ech dimension are as follows:

Perceived Usefulness

1.1. I believe that the developed system will facilitate coordination with primary care professionals in the integrated assessment of the risk of Hereditary Breast and Ovarian Cancer.

1.2. I believe that the system developed will facilitate coordination with genetics professionals in the integrated assessment of the risk of Hereditary Breast and Ovarian Cancer.

1.3. I believe that the system developed will provide professionals with a useful Clinical Decision Support System for prescribing genetic tests, unifying the criteria for information and referral, where appropriate, to genetic counseling units.

1.4. I believe that the system developed will allow people who express to their family doctor/gynecologist their concern about suffering/being able to suffer in the future from hereditary family breast/ovarian cancer because of their personal or family history, resolve doubts, diminish their worries and their level of anxiety.

Perceived Ease of Use

2.1. I think it would be easy for me to learn how to use the developed system.

2.2. I think it would be easy to acquire the necessary skills to use the developed system.

2.3. Overall, I think the developed system will be easy to use.

Social Norms

3.1. My specialty colleagues would like me to use the developed system.

3.2. My superiors would like me to use the developed system.

3.3. Genetic colleagues would like me to use the developed system.

3.4. Primary care colleagues would like me to use the developed system.

Facilitating Conditions

4.1. I think I will have the technical assistance available to solve problems associated with the developed system.

4.2. I think I will have the necessary resources to use the developed system.

Intention to Use

5.1. I intend to use the system developed for the Integrated Risk Assessment of Hereditary Breast and Ovarian Cancer.

5.2. I intend to use the system developed to manage and visualize primary care patients who are referred to as Genetic Counseling.

5.3. I intend to use the developed system to be able to visualize and know the results of the genetic tests of the patients registered in the developed system on which they decide to perform the BRCA1 BRCA2 genetic tests.

The questionnaire was designed by exchanging questions of different dimensions. Next, the correct use of the application in the departments of Gynecology and Mammary Pathology and Oncology of the HUVR was verified. Individually, in the consultations of Breast Pathology, Gynecological Oncology and in the Department of Maternal-Fetal Medicine, Genetics and Reproduction of the HUVR and the consultations of Gynecology of the Specialties Centre. Dr. Fleming from Seville.

The correct use of the application in the different computers had been checked. A training session was held with the 16

gynecologists. User manuals of the application were also provided based on the professions of the users: primary care physicians, gynecologists and geneticists. After the training, the system was validated with the defined TAM questionnaire and the results are shown in the next section.

Results

16 users completed the questionnaires. Each item is rated from 1 to 5, where 1 represents the most negative, and 5 the most positive. and NK/NA in the case of not knowing or not answering. The results are shown below:

Table 1 – Results of questionnaires

Dimensions	Item	Mean	Typical Deviation
	1.1	4,19	1,33
	1.2	4,20	1,15
Perceived	1.3	4,53	1,06
Usefulness	1.4	4,37	0,88
	Mean	4,32	0,98
	2.1	3,87	1,26
Perceived	2.2	3,85	1,21
Ease of Use	2.3	3,62	1,36
	Mean	3,71	1,23
	3.1	4,06	1,34
Social Norms	3.2	4,87	0,34
	3.3	4,25	1,36
	3.4	4,21	1,25
	Mean	4,34	0,78
Facilitating	4.1	3,47	1,50
Conditions	4.2	3,47	1,50
	Mean	3,47	1,44
	5.1	4,27	1,33
Intention to	5.2	4,60	0,91
Use	5.3	4,62	0,80
	Mean	**4,48**	0,81

The table shows the average score of each item, grouped by dimensions. It is observed that the highest score, and therefore the one that obtained the best rating from health professionals, was the Intention to Use.

On the other hand, the dimension where the professionals detected the most inconveniences was in the Facilitating Conditions. This assessment can be attributed to the learning curve of the application, given that the questionnaire was administered after training. The clinicians carried out specific questionnaires to measure acceptance, using the TAM Model, at the end of the training sessions. The results of these questionnaires are described below.

Within the Perceived Usefulness dimension, it stands out as the most valued item "*I believe that the system developed will provide professionals with a useful Clinical Decision Support System for prescribing genetic tests, unifying the criteria for information and referral, where appropriate, to genetic counseling units*" with an average score of 4,53 while the least valued was "*I believe that the developed system will facilitate coordination with primary care professionals in the integrated assessment of the risk of Hereditary Breast and Ovarian Cancer*" with an average score of 4,19. However, this item was also rated positively, taking into account that the maximum value of the scale is 5.

Within the Perceived Ease of Use dimension, it stands out as the most valued item "*I think it would be easy for me to learn how to use the developed system*" with an average score of 3,87 while the least valued was "*Overall, I think the developed system will be easy to use*" with an average score of 3,62.

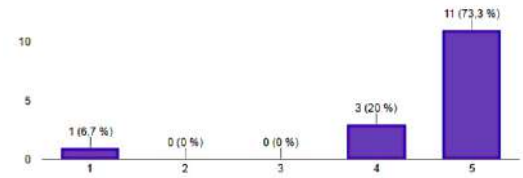


Figure 1 – "I believe that the system developed will provide professionals with a useful Clinical Decision Support System for prescribing genetic tests, unifying the criteria for information and referral, where appropriate, to genetic counseling units" (NK/NA: 1 user)

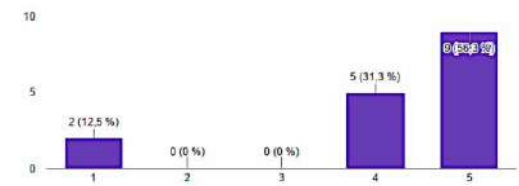


Figure 2 – "I believe that the developed system will facilitate coordination with primary care professionals in the integrated assessment of the risk of Hereditary Breast and Ovarian Cancer"

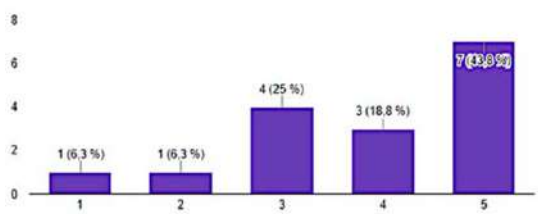


Figure 3 – "I think it would be easy for me to learn how to use the developed system"

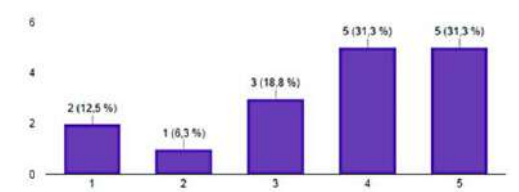


Figure 4 –" Overall, I think the developed system will be easy to use"

Within the Social Norms dimension, it stands out as the most valued item "*My superiors would like me to use the developed system*" with an average score of 4,87, while the least valued item was "*My specialty colleagues would like me to use the developed system*" with an average score of 4,06. However, this item was also rated positively, taking into account that the maximum value of the scale is 5.

Within the Facilitating Conditions dimension, the two items scored an average score of 3,47.

Finally, within the Intention to Use dimension, the most valued item is "*I intend to use the developed system to be able to visualize and know the results of the genetic tests of the patients registered in the developed system on which they decide to perform the BRCA1 BRCA2 genetic tests*" with an average score of 4.62 while the least valued was "I intend to use the system developed for the Integrated Risk Assessment of Hereditary Breast and Ovarian Cancer" with an average score of 4.27. However, this item was also rated positively, taking into account that the maximum value of the scale is 5.

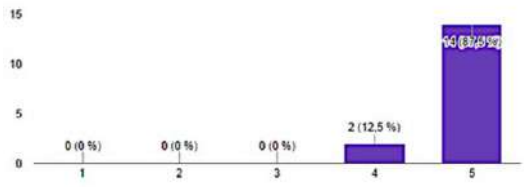


Figure 5– "My superiors would like me to use the developed system"

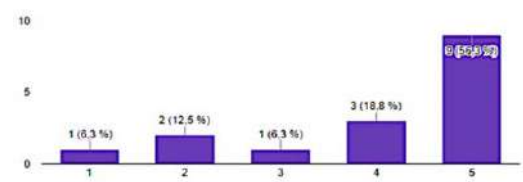


Figure 6 – "My specialty colleagues would like me to use the developed system"

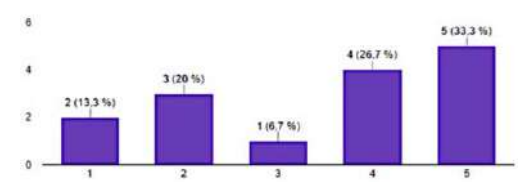


Figure 7 –" I think I will have the technical assistance available to solve problems associated with the developed system" (NK/NA: 1 user)

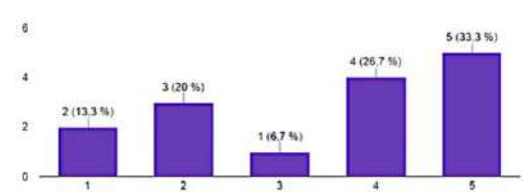


Figure 8 –" I think I will have the necessary resources to use the developed system" (NK/NA: 1 user.)

Discussion

The dimension of the questionnaires that obtained the best rating from health professionals was the Intention to Use. Therefore, this score is relevant for predicting the future use of the developed tool in this project. Within the Intention to Use dimension, the most valued item was "*I intend to use the developed system to be able to visualize and know the results of the genetic tests of the patients registered in the developed system on which they decide to perform the BRCA1 BRCA2 genetic tests*" with an average score of 4.62, meaning other items of this dimension also rated positively.

Besides, the dimension where the professionals detected the most inconveniences was in the Facilitating Conditions, due to the learning curve of the application, given that the questionnaire was administered after training. The two items scored an average score of 3,47, and these scores could be improved. For instance, guaranteeing and providing the required technical assistance and the necessary resources to use the developed system.

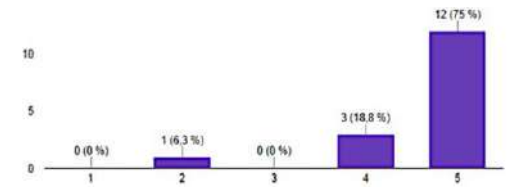


Figure 9–"I intend to use the developed system to be able to visualize and know the results of the genetic tests of the patients registered in the developed system on which they decide to perform the BRCA1 BRCA2 genetic tests"

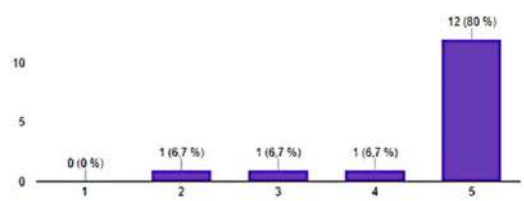


Figure 10 – "I intend to use the system developed for the Integrated Risk Assessment of Hereditary Breast and Ovarian Cancer" (NK/NA: 1 user)

Finally, other dimension with a low score is Perceived Ease of Use, where stands out as the most valued item "*I think it would be easy for me to learn how to use the developed system*" with an average score of 3,87 while the least valued was "*Overall, I think the developed system will be easy to use*" with an average score of 3,62. So it is essential to promote the realization of training sessions and the use of user manuals. Besides, these items rated positively, taking into account that the maximum value of the scale is 5.

In general, the results of the questionnaires show favorable and positive feedbacks. Therefore the use of the developed system is recommended to clinicians for supporting the prescription of genetic tests in the gynecological cancer risk.

According to analysis of the feedbacks, the developed CDSS will be useful for clinical practice. As a result, a tool adapted to the needs and preferences of clinicians, who have also formed part of the project's research team for the definition of requirements, the design of the tool and the implementation of clinical knowledge in the CDSS, has been achieved.

We are working on the development of Artificial Intelligence and Data Mining algorithms. So that through these predictive algorithms can potentially improve the recommendations following the guidelines of the SEOM for the prescripting genetic tests, since currently there are a high percentage of patients, who were prescribed the test and the result turned out

negative. In this way, the system will evolve into a learning healthcare system.

Predictive models are using information collected in electronic medical records, in addition to that referred to in clinical guidelines, to develop more efficient algorithms, with greater precision in prescribing the genetic tests. In this way, we are analyzing the set of variables relevant to this prescription, as well as their association with the target variable.

Conclusions

In this study, a CDSS for prescribing the genetic tests in BRCA1 and BRCA2 and preventing gynecological cancer risks was designed and evaluated in the Virgen del Rocío University Hospital. The developed technological architecture allows general practitioners and gynecologists to classify patients as low risk (who are not required a specific treatment) or high risk (who should be attended by the Genetic Council). Also, it provides recommending related to prescription of genetic tests and their genetic risk of suffering gynecological cancer.

An evaluation of the developed CDSS, for usefulness perceived by clinicians, the perceived ease of use and the intention to use, was also carried out and the results showed positive feedbacks.

Acknowledgements

This research was funded in part by the Ministry of Health of the Government of Andalusia, Proyectos de Investigación en Salud 2016, 'Herramienta de Soporte a la Decisión a la Valoración Integrada del Riesgo de Cáncer de Mama y Ovario Heredofamiliar' project (PIN-0212-2016) and by the Platform for Innovation in Medical Technologies and Health (ITEMAS Platform, code PT17/0005/0036), co-funded by FEDER from European Regional Development Funds (European Union). Many thanks to the Project Coordinator MD Maria Angeles Martinez Maestre and to all clinical researchers and biomedical informatics researchers from this project, for his contributions and suggestions, who helped us to develop this work.

We are especially grateful to the Ethics Committee for the favorable evaluation. The president was Mr. Víctor Sánchez Margalet. The Ethics Committee was made up of 22 members, who considered that the research study meets the requirements for the suitability of the protocol about the objectives of the study and conforms to the ethical principles applicable to this type of study.

References

[1] GLOBOCAN 2012, in, International Agency for Research on Cancer.

[2] G. Llort, I. Chirivella, R. Morales, R. Serrano, A.B. Sanchez, A. Teulé, E. Lastra, J. Brunet, J. Balmaña, and B. Graña, SEOM clinical guidelines in Hereditary Breast and ovarian cancer, *Clinical and Translational Oncology* **17** (2015), 956-961.

[3] T.R. Rebbeck, N. Mitra, F. Wan, O.M. Sinilnikova, S. Healey, L. McGuffog, S. Mazoyer, G. Chenevix-Trench, D.F. Easton, and A.C. Antoniou, Association of type and location of BRCA1 and BRCA2 mutations with risk of breast and ovarian cancer, *Jama* **313** (2015), 1347-1361.

[4] S. Narod, Modifiers of risk of hereditary breast cancer, *Oncogene* **25** (2006), 5832.

[5] N. Andrieu, D.E. Goldgar, D.F. Easton, M. Rookus, R. Brohet, A.C. Antoniou, S. Peock, G. Evans, D. Eccles, and F. Douglas, Pregnancies, breast-feeding, and breast cancer risk in the International BRCA1/2 Carrier Cohort Study (IBCCS), *Journal of the National Cancer Institute* **98** (2006), 535-544.

[6] J.L. Profato and B.K. Arun, Genetic risk assessment for breast and gynecological malignancies, *Current Opinion in Obstetrics and Gynecology* **27** (2015), 1-5.

[7] S. Pruthi, B.S. Gostout, and N.M. Lindor, Identification and management of women with BRCA mutations or hereditary predisposition for breast and ovarian cancer, in: *Mayo Clinic Proceedings*, Elsevier, 2010, pp. 1111-1120.

[8] C. Foster, D. Evans, R. Eeles, D. Eccles, S. Ashley, L. Brooks, R. Davidson, J. Mackay, P. Morrison, and M. Watson, Predictive testing for BRCA1/2: attributes, risk perception and management in a multi-centre clinical cohort, *British journal of cancer* **86** (2002), 1209.

[9] M. Watson, C. Foster, R. Eeles, D. Eccles, S. Ashley, R. Davidson, J. Mackay, P. Morrison, P. Hopwood, and D.G.R. Evans, Psychosocial impact of breast/ovarian (BRCA 1/2) cancer-predictive genetic testing in a UK multi-centre clinical cohort, *British journal of cancer* **91** (2004), 1787.

[10] M.J. Esplen, J. Hunter, M. Leszcz, E. Warner, S. Narod, K. Metcalfe, G. Glendon, K. Butler, A. Liede, and M.A. Young, A multicenter study of supportive-expressive group therapy for women with BRCA1/BRCA2 mutations, *Cancer: Interdisciplinary International Journal of the American Cancer Society* **101** (2004), 2327-2340.

[11] I. Ejarque, M. García-Ribes, J.V. Sorlí, E. Arenas, and V. Martín, Papel de la atención primaria ante el cáncer hereditario, *Atención primaria* **40** (2008), 525-529.

[12] K. Gartrell, A. Trinkoff, C. Storr, M. Wilson, and A. Gurses, Testing the electronic personal health record acceptance model by nurses for managing their own health, *Applied clinical informatics* **6** (2015), 224-247.

[13] B. Rahimi, H. Nadri, H.L. Afshar, and T. Timpka, A systematic review of the technology acceptance model in health informatics, *Applied clinical informatics* **9** (2018), 604-634.

[14] H. Nadri, B. Rahimi, H.L. Afshar, M. Samadbeik, and A. Garavand, Factors affecting acceptance of hospital information systems based on extended technology acceptance model: a case study in three paraclinical departments, *Applied clinical informatics* **9** (2018), 238-247.

Address for correspondence

Cristina Suárez Mejías.

Address: Avda. Manuel Siurot s/n 41013 Sevilla, Spain.

Email: cristina.suarez.exts@juntadeandalucia.es

Phone: +34 955 013 662

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190315

Using Health Information Exchange: Usage and Perceived Usefulness in Primary Care

Aude Motulsky [a, b], Claude Sicotte [b], Marie-Pierre Moreault [b], Tibor Schuster [c], Nadyne Girard [d], David Buckeridge [d], Marie-Pierre Gagnon [e], Robyn Tamblyn [d]

[a] Research Center, Centre Hospitalier de l'Université de Montréal, Montreal, Quebec, Canada
[b] Department of Management, Evaluation and Health Policy, Université de Montréal, Montreal, Quebec, Canada
[c] Department of Family Medicine, McGill University, Montreal, Quebec, Canada
[d] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada
[e] Faculty of Nursing Sciences, Université Laval, Quebec City, Quebec, Canada

Abstract

Health information exchange (HIE) is seen as an essential technology for improving health care quality and efficiency by allowing exchange of patient-centered data over time and across organizations. The objective of this study was to evaluate the usage and the perceived usefulness of a nationwide HIE in a centralized model that was implemented in 2013 in the province of Quebec, Canada. A mixed-method study was conducted with a longitudinal descriptive analysis of usage data combined with in-depth comparative case study in four selected primary care organizations and two emergency departments. Perceived benefits were reported by users across all dimensions of care performance, including accessibility, efficiency, quality and safety, and patient experience; however, the experience of users was very heterogeneous and strongly associated with the commercial electronic record system available in their work place and the implementation strategy.

Keywords:

Health Information Exchange; Electronic Health Records; Quality of Health Care

Introduction

Allowing for information exchange along the care continuum is seen as an essential component of a high quality and efficient health care system.[1,2] Around the world, many jurisdictions have implemented standards or systems to allow for this electronic exchange to be possible.[3] In short, two types of models are emerging: on one hand, the centralized model, with national or regional shared records that are available to authorized providers and/or patients in various fashions, such as in Finland and Sweden, and on the other hand, the federated model, where individual organizations share information based on exchange protocols, as is done in the United States.[4] The information available for exchange is highly variable from one jurisdiction to another and ranges from imaging information to hospital discharge summaries.[5] In primary care, where care is by definition anchored in the patient's life, information continuity is particularly challenging, and the promised benefits of HIE are especially important. For example, a study conducted in the state of New York demonstrated that when a general practitioner accessed the regional HIE in the 30 days following the hospitalization of one of his or her patients, the risk of readmission for the same problem as the original hospitalization dropped by 57%.[6]

In Canada, Quebec was the first province to implement a province-wide HIE in a centralized model, starting with medication, laboratory, and clinical imaging information. The objective of this study was to evaluate the usage and the usefulness perceived by users in primary care with this system, 2 years after its full implementation.

Design of the System and Available Features

The Ministry of Health manages the HIE in Quebec by for the entire healthcare system. It was built as a pull model, with central data warehouses, where every authorized provider is allowed to access and retrieve data depending on local certifications. All residents of the province (estimated population 8.2 million) are identified using their unique health insurance numbers. A secure connection between retail pharmacies, medical clinics, laboratories and diagnostic imaging centers allows for information to be transmitted to central warehouses.(for a detailed description of the medication domain, see a previous study [7]) A number of options are available to clinicians to access this information depending on the tools available at their work sites. Clinicians can use their electronic medical record (EMR) application if it was certified by the Ministry for interoperability. Ten different EMRs were certified in 2018, nine for primary care providers (outpatient EMR, O-EMR) and one for hospital-based providers. Alternatively, a web-based Viewer application can be used that was developed by the Ministry of Health and enables data access without EMR application. To access an individual patient's information through the Viewer, users need a certificate of security that was provided by the Ministry of Health on a USB device, combined with a password. As of Jan 2018, devices were distributed to 53,000 individuals in the province.

Users who access data through their EMRs are able to both view the patient's information, and import it, in a more or less structured format depending on the type of information and the commercial EMR they are using; however, users accessing data through the web-based Viewer only have the ability to view and print the patient's information.

Methods

Study Design and Data Collection

A mixed-method concurrent study was conducted, from January 2016 until December 2017. First, a descriptive analysis of usage was performed, using audit trails on accesses to the HIE obtained from the Ministry of Health. Information available for each access to the HIE included user ID, date, user role (e.g. physician, nurse, pharmacist), and which tool was accessed the HIE (e.g. Viewer, EMR1, EMR2). Second, an in-depth comparative case study was conducted, where a case was defined as a medical clinic or an emergencey department, where

clinicians declared using the HIE regularly. Purposeful selection of cases was performed in an exploratory phase, where the criteria for selection was clinicians' regular system usage that was assessed using usage data combined with exploratory interviews with stakeholders in the province. In each targeted organization, the manager was invited to participate and to invite clinicians to an interview and an observation session. All clinicians (nurse, physician and pharmacist) were invited per site. Table 1 describes the cases selected, as well as the participants in each case. After consent was obtained, semi-structured interviews were conducted with users (24 general practitioners, eight nurses, nine pharmacists) following an interview guide with open questions about their usage of the HIE, their satisfaction, and their perception of the usefulness of the system in relation to their daily activities. Observation sessions were also organized at each site to complete interviews and describe usage practices.

Data Analysis

Adoption was estimated using the number of actual users compared to the potential number of users, per role. A user was defined as an individual who accessed the HIE at least once in October 2017 (this month was selected as a typical month of usage because individual users could not be followed during the whole study period due to replacement of user IDs without mapping between old and new IDs). Descriptive statistics on the proportion of potentially authorized users who actually accessed the HIE was calculated, according to the users' roles. The proportion of users accessing the HIE by tool was also calculated in Oct 2017 to describe usage. Finally, to describe the level of use for each data domain, the number of accesses to each domain was calculated by user for each week in 2017.

All audio files were transcribed verbatim, and thematic content analysis was performed using the framework proposed by Lau and colleagues (quality of the information and the system, usage patterns, perceived usefulness) [8]. Emerging codes were also allowed until saturation of the findings. Preliminary reports were shared with the participants at each site, and their comments were included in the final analysis. Adoption and level of use were combined with case study results to analyse the users' experience. The ethics committee of the Centre Hospitalier de l'Université (CHU) de Québec and McGill University approved this project.

Table 1 – Organizations Selected and Participants at Each Site (GP=General Practitioner; phm = Pharmacist; nurse = Clinical Nurse or Nurse Practitioner, ED = Emergency Department)

Site	Tool available for HIE usage	Participants
Family health team 1 - Urban region A	O-EMR 1 + Viewer	9 GPs, 1 nurse, 1 phms
Family health team 2 - Rural region	O-EMR2 + Viewer	3 GPs, 2 nurses, 1 phms
Family health team 3 - Urban region B	O-EMR3 + Viewer	3 GPs, 1 nurse, 1 phms
Family Health Team 4 - Urban region A	O-EMR4 + Viewer	3 GPs, 3 nurses, 1 phms
ED Academic Health Center 1 Urban region A	Viewer	2 GPs, 3 phms
ED Academic health center 2 Urban region A	Viewer	4 GPs, 2 phms, 1 nurse

Results

Usage of the HIE

Adoption

Table 2 presents the number of users compared to the potentially authorized users per role. Overall, adoption was higher in primary care, where the vast majority of pharmacists (78%), general practitioners (74%) and nurse practitioners (72%) in the province accessed the HIE in October 2017. Interestingly, only 26% of specialists accessed the tool. The Viewer was used by 89% of users in October 2017, while only 17% of users accessed the HIE using an interoperable EMR (Table 3). While more than 90% of GPs in the province have access to an interoperable EMR in their workplace, only 50% of GP users accessed the HIE using their O-EMR, while 74% of GP users accessed the HIE using the Viewer, indicating that a rudimentary adoption of the HIE is high, while the adoption of more advanced features linked to HIE integration still needs to be improved. Finally, only 12 % of specialist users accessed the HIE using an integrated EMR.

Table 2 – Number of Users by Role in Oct 2017

Role	N potential users[1]	N Actual users (%[2])
Physicians (total)	**20,052**	**9,612 (48%)**
General practitioners	9,503	6,992 (74%)
Specialists	10,239	2,620 (26%)
Pharmacists	**9,212**	**7,162 (78%)**
Nurses (total)	**74,469**	**8,762 (12%)**
Nurse practitioners	413	297 (72%)
All roles	NA	**31,915**

[1]Data obtained from the annual reports of the respective professional associations; [2]actual users / potential users

Table 3 – Number of Users by Tool, Selected Roles in Oct 2017

Role	N users[1] (%[2])	
	EMR	Viewer
General practitioners	3,496 (50%)	5,187 (74%)
Specialists	309 (12%)	2,428 (92%)
Pharmacists	102 (1%)	6,790 (94%)
Nurse practitioners	146 (49%)	232 (78%)
All roles	**5,571 (17%)**	**28,395 (89%)**

[1]Users may have accessed with more than one tool (sum >100%) Access with a pharmacy management system is not presented; [2]=N users with this tool / N actual users with this role (table 2)

Level of Use

The medication domain was the one with the highest number of weekly accesses by physician users, in comparison to the lab and imaging domains, with an average of the mean number of weekly accesses per user of 21 for meds, six for labs and seven for images. Figure 1 presents detailed weekly accesses by tools, per domain. For the medication domain (Fig 1A), the number of weekly accesses using an outpatient EMR (O-EMR) was higher when compared to the Viewer and an EMR in the acute care setting (mean ± SD : 35 ± 4 vs 17 ±1 vs 10 ±1 respectively). This was aligned with the available feature of the O-EMR confirmed with the case study, where only these tools in outpatient settings had the availability of the most advanced feature for importing granular medication data from the HIE, to be reused for clinical activities, such as generating a new electronic prescription. In the lab domain (Fig 1B), the level of use was similar for O-EMR and Viewer, and lower for EMRs in the acute care setting. For images (Fig 1C), the level of use

was higher using the Viewer application. This might be related to the fact that the image was only accessible using the Viewer application, while O-EMR and EMR only provided access to the report.

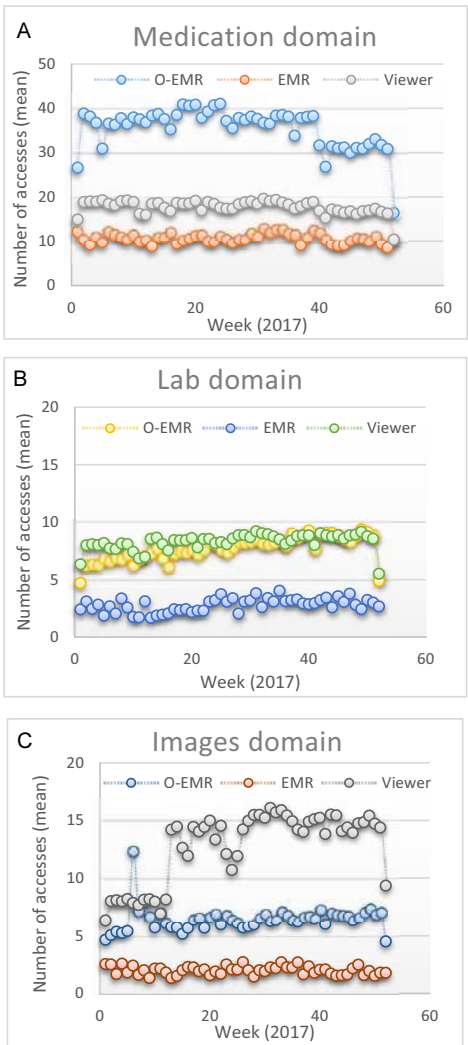


Figure 1. Level of Use of Each Data Domain of the HIE by Physician Users in 2017, by Tool Used to Access the HIE (O-EMR = Outpatient Electronic Medical Record; EMR = Electronic Medical Record) (Scale for Graph A is Different)

Perceived Benefits in Primary Care

Benefits associated with HIE usage were identified across all cases, by a comparative analysis, highlighting differences and similarities among and between cases. Interestingly, perceived benefits were similar across cases, and rapidly saturated in terms of how using the HIE impacted care quality, productivity, and the experience of patients. The diversity of perceived benefits was great and aligned with what is already documented with such clinical information sharing in other settings.[9–11] They were also aligned with the expectations associated with the implementation of the HIE.

Overall, the added value seemed greater for medication information, while this was the only source of information in all cases visited. On the other hand, in the lab and imaging domains, the added value depended on the regional organization and available tools in the workplace. Indeed, other systems in place in some regions for electronically sharing this information were making a difference.

To be precise, perceived benefits were reported across all dimensions of care performance : productivity, quality and safety, accessiblity, and experience of patients (Table 4). According to users, these benefits are actualized, because the information accessible through the HIE is of greater quality: more complete, more reliable, more valid and available when needed. For productivity, clinicians perceived that using the HIE reduces delays related to clinical information management (e.g. receive active medication list, lab results or imaging reports by fax), thus reducing delays for patients, reducing duplication of exams, and dedicated (avoidable) resources. Then, the users perceived that the quality and safety of care was improved, thanks to a better informed clinician, who is more confident about the clinical judgement based on a more accurate portrait of the patient, thus reducing errors and improving continuity. Similarly, the participants perceived that access to care was improved, in particular for general practitioners, who mentioned that they were less reluctant to register complex or vulnerable patients who became less difficult to follow with this kind of tool facilitating care coordination through clinical information sharing. Finally, some clinicians reported that the experience of their patients was improved, particularly in relation to timely access to information and reduced unnecessary visits.

Table 4. Perceived Benefits Associated with HIE Usage

Productivity
↓ Delay in getting information
↓ Delay in receiving results
↓ Duration of visits
↓ Duration of care episodes
↓ Duplicate exam
↓ Avoidable visit
Quality and security
↑ Confidence in decision making
↓ Errors
↑ Appropriateness of care
↑ Continuity and coordination of care between the team members and different heathcare organizations
Accessibility
↑ Volume of patients
↓ Waiting time
↑ Management of vulnerable and complex patients who consume healthcare services
Patient experience
↑ Relationship with clinicians and teams
↑ Comprehension and involvement
↓ Unnecessary visits and wait times
↑ Satisfaction

Legend : ↑ = Improve or Increase; ↓ = Decrease

A Very Heterogeneous Journey

However, the experience of users, a mediating factor in the actual level of use and persistence over time, was highly heteronegeous and closely associated with the commercial tool available on the work site, as well as the implementation strategy. While in some organizations (family health team 1 (FHT1) and FHT2), using the HIE was reported as easy and was highly integrated in the daily activities, in some other organizations (FHT3, FHT4, ED1, ED2), using the HIE was

only motivated by certain specific situations, where the added value was high enough to compensate for the efforts and irritants reported by users to access and integrate the clinical information from the HIE in routine clinical activities. The main barriers were related to the completeness of the information, the usability of the applications, the performance of the system (e.g. some long periods of shut downs), and the process for access (a pull system requiring a USB device on a physical desktop) not well aligned with the fluid nature of care.

When is Using HIE Worthwhile?

In the ED, the added value was clear and very high, for almost every patient but especially when the patient was unconscious, unable to communicate for various reasons, or when the clinician had a doubt about the veracity of the story from the patient; however, both EDs had access to the HIE only through the Viewer application, with limited integration into their work flow and basic features of viewing information. In medical clinics (family health teams), the added value was particularly high for patients with chronic conditions, navigating through the health care system with a long trajectory requiring clinical information sharing or for walk-in clinics when patients were not known.

Discussion

To our knowledge, this is the first study conducted with HIE users in Canada. While evidence of HIE positive outcomes are growing [12,13], it is important to deepen our understanding of the different systems, and associated users' experience across different jurisdictions. The actual experience of users, and thus the potential outcomes, are highly related to the availability of given features, and their levels of use by many health care professionals, in a given setting.[14] Our results highlight the fact that one dimension needs to be deepened when studying HIE, namely the integration within the main tool used by healthcare professionals, i.e. the commercial electronic record available in their work place. While a recent systematic review conducted by Menachemi and colleagues [13] highlighted the fact that most studies with positive outcomes were associated with American community HIEs, one should consider what it means in terms of newly available information, and the actual features available for a given clinician, in a particular setting. As already described by Opoku and colleagues in the US, this is strongly associated with the health care system organization, and underlying systems [15]. Nonetheless, what this means in other jurisdictions is poorly described. This is where this study makes a contribution by deepening the description of an HIE in Canada.

First, adoption was higher in primary care, with almost three quarters of general practitioners, pharmacists and nurse practioners in the province accessing the HIE. Moreover, perceived usefulness by clinicians in family health teams and emergency departments was important, with a diversity of benefits identified by users. The added value seemed significant for clinicians in these cases, even if some barriers were mentioned as to the quality of information and of the system. This relates to previous studies in Finland [9], New York state [6,8–10], and Midwestern states [16], highlighting the benefits of HIE especially for primary care providers, and emergency department physicians.

On the other hand, a lower adoption rate by specialists was observed, which may be related to many factors. First, most of these physicians work in hospitals, where the availability of an integrated EMR with an HIE is scarce, and most clinicians only have access to the Viewer application to access the HIE. In fact, most hospital centers in the province do not have an advanced EMR available, and paper charts are still the norm as the basis for the integration of clinical information. Moreover, the added value of the HIE is lower in acute care centers, because what you can do with the Viewer application is very basic : view and print clinical information. While very useful in the emergency department, to complete the medication list, and review lab results and imaging reports from other organizations, as a way to get a quick overview of a patient and his or her recent care episode, it does not seem to be enough for specialists to access the system in their usual patient care. Furthermore, the performance of the system for viewing images is not satisfactory for diagnostic purposes, and the PACS (picture and archiving communication system) already available on a regional basis in the province reduced the added value of the HIE.

Interestingly, the HIE in the province of Quebec was the first clinical information system integrating information between primary care and acute care settings, and integrating information between private and public organizations. In Quebec, most hospitals are publicly owned, and clinicians practicing within these organizations have access to some form of an integrated system while all pharmacies are privately owned, and about 30% of lab centers and imaging centers are private organizations.

Overall, it is important to note that HIE is a generic term to describe clinical information sharing, but what is shared, and how it is integrated into clinical and cognitive workflows, are crucial to benefit actualization. This study is one step ahead on this road in that it highlights the fact that the **added value** associated with HIE needs to be considered **in comparison** to other available systems in a given setting. In other words, two main questions need to be considered around such systems: 1) what it adds to what is already known about a given patient (quality of the information); 2) how easy and quick it is to use the clinical information from the HIE versus the other available systems (quality of the system).

This study has a few limitations. First, only a few cases were selected in three regions of the Province (out of 18), thus limiting the external validity of the findings. While triangulation of data sources and types allowed us to present interesting findings, further research is needed to deepen our understanding of the actual practices in a more diverse sample of work sites. Moreover, this study was conducted in the early phases of the implementation of the integrated EMR/HIE features. Some users were not very familiar with these features when we visited their work site, and further research should investigate routine practices when users are more experienced with their interfaces.

Conclusions

This study was able to describe the level of use of the HIE at the scale of the whole system in the province of Quebec, and characterize the perceived usefulness for clinicians using the system. Interestingly, the adoption is higher in primary care and at the emergency department, where the added value of the system is higher. Further research is needed to deepen our understanding of the heterogeneity of usage practices of the HIE, particularly with different commercial EMR and their specific integration features, as well as the various factors associated with perceived usefulness and outcomes actualization.

Acknowledgements

Canada Health Infoway funded this study. Aude Motulsky was financially supported by the Centre de recherche du CHUM and the FRQ-Santé. Tibor Schuster is funded through a CIHR Canada Research Chair (Tier II). We thank all participants for their time and willingness to contribute to this study.

References

[1] J.R. Vest, Health information exchange: National and international approaches, *Adv Health Care Manag*. **12** (2012) 3–24.

[2] J.R. Vest, L.M. Kern, T.R. Campion, M.D. Silver, and R. Kaushal, Association between use of a health information exchange system and hospital admissions, *Appl Clin Inform*. **5** (2014) 219–231. doi:10.4338/ACI-2013-10-RA-0083.

[3] L. Ohno-Machado, Electronic health records and health information exchange, *Journal of the American Medical Informatics Association*. **25** (2018) 617–617. doi:10.1093/jamia/ocy057.

[4] S. Klapman, E. Sher, and J. Adler-Milstein, A snapshot of health information exchange across five nations: An investigation of frontline clinician experiences in emergency care, *Journal of the American Medical Informatics Association*. **25** (2018) 686–693. doi:10/gcx2fh.

[5] F. Sadoughi, S. Nasiri, and H. Ahmadi, The impact of health information exchange on healthcare quality and cost-effectiveness: A systematic literature review, *Computer Methods and Programs in Biomedicine*. **161** (2018) 209–232. doi:10/gdvnw2.

[6] J.R. Vest, L.M. Kern, M.D. Silver, R. Kaushal, and HITEC investigators, The potential for community-based health information exchange systems to reduce hospital readmissions, *J Am Med Inform Assoc*. **22** (2015) 435–442.

[7] A. Motulsky, D.L. Weir, I. Couture, C. Sicotte, M.-P. Gagnon, D.L. Buckeridge, and R. Tamblyn, Usage and accuracy of medication data from nationwide health information exchange in Quebec, Canada, *Journal of the American Medical Informatics Association*. **25** (2018) 722–729.

[8] F. Lau, M. Price, and K. Keshavjee, From benefits evaluation to clinical adoption: Making sense of health information system success in Canada, *Healthc Q*. **14** (2011) 39–45.

[9] T. Maenpaa, P. Asikainen, and T. Suominen, Views of patient, healthcare professionals and administrative staff on flow of information and collaboration in a regional health information exchange: A qualitative study, *Scand. J. Caring Sci.* **31** (2017) 939–947.

[10] W.R. Hersh, A.M. Totten, K.B. Eden, B. Devine, P. Gorman, S.Z. Kassakian, S.S. Woods, M. Daeges, M. Pappas, and M.S. McDonagh, Outcomes From Health Information Exchange: Systematic review and future research needs, *JMIR Med Inform*. **3** (2015) e39.

[11] R.S. Rudin, A. Motala, C.L. Goldzweig, and P.G. Shekelle, Usage and effect of health information exchange: A systematic review, *Ann. Intern. Med.* **161** (2014) 803–811.

[12] D.M. Walker, Does participation in health information exchange improve hospital efficiency?, *Health Care Manag Sci*. **21** (2018) 426–438.

[13] N. Menachemi, S. Rahurkar, C.A. Harle, and J.R. Vest, The benefits of health information exchange: An updated systematic review, *J. Am. Med. Inf. Assoc.* **25** (2018) 1259–1265.

[14] M. Price, and F. Lau, The clinical adoption meta-model: A temporal meta-model describing the clinical adoption of health information systems, *BMC Medical Informatics and Decision Making*. **14** (2014) 43. doi:10.1186/1472-6947-14-43.

[15] W. Opoku-Agyeman, and N. Menachemi, Are there differences in health information exchange by health system type?, *Health Care Management Review*. **41** (2016) 325–333.

[16] B.D. Gordon, K. Bernard, J. Salzman, and R.R. Whitebird, Impact of Health Information Exchange on emergency medicine clinical decision making, *West J Emerg Med*. **16** (2015) 1047–1051.

Address for correspondence

Aude Motulsky
aude.motulsky@umontreal.ca
+1 (514) 343-5631

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190316

Evaluation of a Nationwide e-Prescribing System

Aude Motulsky [a, b], ManQing Liang[a], Marie-Pierre Moreault[b], Elizabeth Borycki[c], Andre Kushniruk[c], Claude Sicotte[b]

[a] *Research Center, Centre hospitalier de l'Université de Montréal, Montreal, Canada*
[b] *Department of Management, Evaluation and Health Policy, Université de Montréal, Montreal, Canada*
[c] *School of Health Information Science, University of Victoria, Victoria, Canada*

Abstract

Electronic prescribing, defined as the electronic generation and transmission of a medication order for community-dwelling patients, is presented as an essential technology to improve medication use. The objective of this study was to evaluate a nationwide e-prescribing system in Quebec, Canada. A mixed-method study was conducted from July 2017 until June 2018. A descriptive analysis of e-prescription usage was performed using aggregated usage data, combined with an exploratory descriptive analysis of the e-prescribing system from the perspective of users of two electronic health records (EHR) and pharmacy management systems (PMS) (n=9 prescribers; 8 pharmacy technicians and 11 pharmacists). Overall, the adoption of the system was low, with only 2% of prescriptions being electronically transmitted and retrieved during the study period. Alignment problems were identified on the prescriber's and receiver's side, generating safety issues, and hindering the potential for benefits realization.

Keywords:

Electronic Prescribing, Medication Adherence, Health Information Technology

Introduction

Improving safety and quality of medication use in primary care is a priority, both for clinical and economical reasons. For more than two decades, electronic prescribing (e-prescribing) has been presented as an essential technology to support clinicians and patients towards this goal[1]. While the technology may differ by jurisdictions, the terms e-prescribing usually refers to any computerized system used to generate and communicate information related to medication prescriptions for community-dwelling patients[2]. The opioid epidemic, off-label use, polypharmacy and potentially inappropriate medications are examples of clinical issues that are promised to be resolved, or reduced, with e-prescribing[3.4]. Many jurisdictions around the world have implemented nationwide e-prescribing systems, including the generation of a prescription using a computer system (with or without a clinical decision support system), and the electronic transmission of the prescription to the dispensing pharmacy. European countries are leading the way, with Finland, Denmark and Sweden at almost 100% of the outpatient prescriptions being transmitted electronically[5–7].

While it is already known that e-prescribing can support the decision making processes of clinicians, and reduce legibility and transcription problems, it is also known that it can create new problems and errors at all steps of the medication management process[8–10]. Precisely, issues with the design of the e-prescribing feature have already been described, leading to e-prescriptions of highly variable quality[11], while the transmission and reception models are heterogeneous in different jurisdictions, and pose various issues for the pharmacy work processes[12,13].

In Quebec, a nationwide e-prescribing system was implemented in 2013, in a central pull model, connecting all primary care electronic record systems with pharmacy management systems for electronic transmission of the prescriptions. The objective of this study was to evaluate the system after its full implementation, focusing on the adoption of the system in the province, and its quality for improving the prescribing and dispensing processes in primary care.

Methods

Description of the e-Prescribing System

The e-prescribing system is managed by the Ministry of Health, and is constituted of a central repository of e-prescriptions that are generated from certified electronic health record (EHR) systems, and then accessible to certified pharmacy management systems (PMS) in the province for importation and execution in pharmacies. All prescribers (e.g. general physicians, nurse practitioner) using a certified EHR system can use their local e-prescribing feature to generate an e-prescription, that is then validated when transmitted to the central e-prescription repository. Pharmacists and their team can then log-in to the central repository for a given patient, and import e-prescriptions in their local PMS to dispense the medication. The system uses a national index registry for patient identification, and Drug Identification Numbers (DIN) as the index for medication identification. DINs are managed by Health Canada and issued for every medication that receives approval to be marketed in Canada. They uniquely identify the product name, the active ingredient, the manufacturer, the strength, the pharmaceutical form, and the route of administration. At the time of the study, only EHR and PMS in the outpatient setting were certified for connection with the e-prescription repository. No feature was designed for the patients. Details on the system have been described elsewhere[14].

Data Collection and Analysis

A mixed-method study was conducted in parallel, from July 2017 until June 2018. First, a descriptive analysis of usage of the e-prescribing system was performed, using aggregated usage data provided by the Ministry of Health. Information available included the number of prescriptions dispensed by all retail pharmacies in the province, to all citizens with a health insurance number (mandatory), the number of electronic prescriptions (eRx) sent by prescribers using a certified EHR system, and the number of eRx retrieved by pharmacists, by region and in the whole province (population approximately 8 million inhabitants). Adoption was estimated by calculating the proportion of eRx compared to the total number of prescriptions dispensed, per month. To triangulate these observations, one pharmacy was visited to manually gather all prescriptions

executed within a typical week, and classify them per type (manuscript vs electronic). The proportion of prescriptions of each type was calculated per day.

Second, an exploratory descriptive study was conducted by interviewing and observing users of the e-prescribing system, both on the prescribing and receiving sides. Purposeful selection of high users was performed based on the declared regular usage of the system, specifically in regions targeted for their high adoption of e-prescribing (based on usage data). Two commercial EHR systems and 5 PMS were analyzed (details on participants are presented in Table 1).

On the prescriber side, frequent e-prescribing users were invited to participate to an interview and an observation session using think aloud protocols around defined prescribing scenarios. Semi-structured interview guide was elaborated to describe their usage of the system, their work process, and their experience with using the system on a daily basis. Moreover, typical scenarios were designed from previous studies[15], to identify problems related to the e-prescribing feature, including all steps of the process (medication review and reconciliation, medication selection, validation and transmission of the e-prescription[16]. Users were encouraged to "think aloud" and verbalize their thoughts as they were completing the scenarios. Their screen and voice were then recorded. Seven physicians and 2 nurses participated (See Table 1).

On the pharmacy side, all users (pharmacists and pharmacy technicians) of the PMS in a given pharmacy were invited to participate. A convenient sample of prescriptions was executed while the screen and the voice of the user were recorded. Semi-structured interviews were also conducted with users, including open questions about their usage of eRx, their work process, and their experience with using the system on a daily basis. Overall, 11 pharmacists and 8 pharmacy technicians participated (see Table 1).

Audio files were transcribed. Verbatim and thematic content analysis was performed to describe the flow of each step of the process, and identify alignment problems per step of the process. Videos were used to confirm the problem identified as described previously[17]. Emerging codes were allowed, until saturation of the findings. This project was approved by the ethics committee of the Université de Montréal, and all participants consented before their participation.

Table 1 – Data Gathered

Systems	Participants
PMS A	1 pharmacist 4 technical assistants
PMS B	4 pharmacists 4 technical assistants
PMS C	3 pharmacists
PMS D + E	3 pharmacists
EHR A	4 physicians
EHR B	3 physicians 1 nurse practitioner 1 nurse

Legend: PMS = Pharmacy management system; EHR = Electronic Health Record

Results

Adoption

The number of eRx sent and retrieved by all prescribers and pharmacies in the province are presented in Figure 1, and compared to the total number of prescriptions dispensed in retail pharmacy during the study period (target for eligible prescriptions). The total number of eRx sent represented on average 13% of all prescriptions dispensed during the study period (10% in July 2017, and 14% in March 2018). Hence, the adoption was low on the prescriber side, with a total number of individual clinicians sending eRx varying from 2,397 clinicians in July 2017, to 3,946 clinicians in March 2018 (while the potential is more than 10,000 clinicians, including general practitioners and nurse practitioners).

In terms of pharmacy, this observation is reflected in the fact that the vast majority of prescriptions that were received were not electronically transmitted (Table 2). When analyzing the characteristics of all prescriptions dispensed in a typical pharmacy, we observed that while 55% of all new prescriptions were created through an EHR, only 35% were actually transmitted electronically. This means that 20% of prescriptions were generated by using an e-prescribing system, but were printed instead of electronically transmitted. Interestingly, almost one fourth of all prescriptions (23%) were still manuscript in this pharmacy located in the region with the highest rate of adoption of eRx in the Province. This suggests that even if the adoption of EHR in primary care in Quebec has increased, the adoption of the e-prescribing feature is lagging.

Moreover, the level of use is low in pharmacy, where only 2% of prescriptions were actually received electronically on average during the study period. When compared to the eRx that were sent, only 16% were retrieved by pharmacists, leading to the vast majority of eRx "sleeping" in the central e-prescribing repository (which will be deleted after two years).

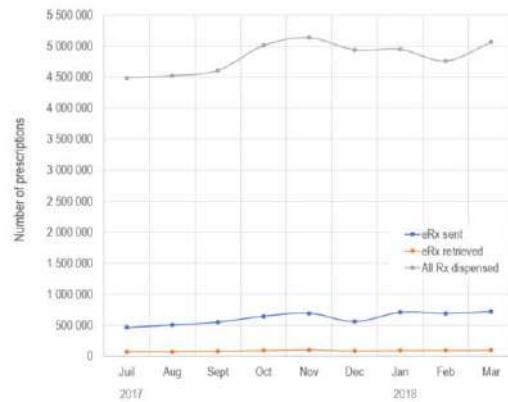


Figure 1 – Level of Use of e-Prescribing in Quebec

Problems on the Prescribers' Side

The main problems that were identified on the prescriber's side were related to the following features: a) design of the medication order; b) the absence of clinical decision support including information about the dose and the characteristics of patients (except allergies); c) the absence of a feature for electronic prescription requests; d) the systematic printing of a paper copy of the prescription.

Table 2 – Proportion of Prescriptions Executed in a Typical Pharmacy During Five Consecutive Weekdays, per type (in June, 2018)

Type of prescriptions	Total	Mean (SD) per day	%
Manuscript	99	20 (7)	23%
Printed form	27	5 (3)	6%
Fax	63	13 (4)	15%
EHR-printed	84	17 (5)	20%
EHR-electronically	149	30 (14)	35%
Verbal	8	2 (1)	2%
Total	430	86	100%

a) First, the design of the medication order were all based on a product-based design using DINs (Fig 2A). Here, the product refers to what you can take in your hand as a patient (e.g. tablet of 500 mg of acetaminophen made by Apotex). The product includes information about the brand, the pharmaceutical form (e.g. tablet, liquid, inhaler), the molecule that is aimed to be administered (e.g. acetaminophen) and the amount of the molecule within this product (or the strength, e.g. 500 mg). This type of design requires the user to select the product (e.g. Apo-acetaminophen comp. 500 mg), and the instructions are built using the number of "unit" the patient has to take, based on the formulation of the product (e.g. 2 tabs). In contrast, the molecule-based medication ordering (Fig. 2B) would require the prescriber to select a molecule and its route, following by the dose and the frequency. The problems we identified with the product-based design were the error-prone selection of the medication, because the list of products was long, and was not up to date. For example, all clinicians were wrongly able to validate the order of a product that is not on the market anymore in Canada (Lasix™) (Fig 3A1).

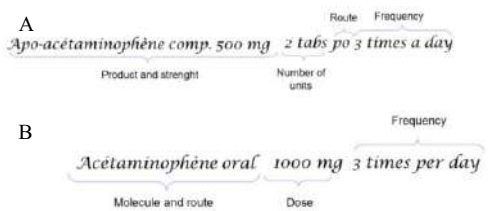


Figure 2 – Product-Based Design of the Medication Order (A) Observed, Compared to the Molecule-Based Design (B)

b) Second, no system had clinical decision support integrating information related to the dose, or the characteristics of the patients (such as age or diagnostic). Consequently, all clinicians were wrongly able to validate a prescription at a lethal dose, for example if they thought it was for an adult and it was actually for a child. The absence of structured and standard information required for intelligent alerts was limiting the utility of the decision support feature from the point of view of prescribers (Fig 3A2).

c) Third, the feature for requesting prescription repeats, from pharmacists or from patients, was not available. Patients would need to call their prescribers, and pharmacists would need to send paper requests by fax to get repeats of an ongoing prescription. This was seen as a major irritant by most prescribers and pharmacists interviewed, given the volume of transactions it generated per day. Moreover, a safety issue was associated with this situation, because most prescribers would not add this paper prescription to their electronic record, and would simply manually sign the request, and send it back to the pharmacy. The electronic record of patients would then become incomplete (Fig 3A3).

d) Finally, the last problem was due to systematic printing of a paper copy of the prescription when the electronic transmission was validated. Consequently, it was not infrequent that prescribers would manually modify the paper copy or a writable PDF form that was created before printing, while the electronic copy would remain unchanged (Fig 3A4). This creates a major safety issue because two copies of the same prescription would then exist. An unclear legal status for the electronic prescription seemed to have led to this situation, where vendors and prescribers were being told that they had to print a paper copy, and pharmacists, that they have to wait for the paper prescription to be allowed to retrieve the eRx.

Problems on the Receivers' Side

The main problem in pharmacy was related to the fact that the execution of a prescription always began with a paper copy of the prescription (Fig 3B1). Pharmacy staff were not informed when an electronic prescription was available for one of their patients, given the design of the system that was developed without a feature for allowing a push or an alert to an assigned pharmacy. The only way for pharmacists to know that an electronic prescription was available was a sign (a logo or a number depending on the EHR) on the paper copy of the prescription. Moreover, in three (out of 5) pharmacy systems, viewing eRx for a given patient would require the staff to execute a request, while in two other pharmacy systems, eRxs for a given patient would be visible from the summary page of the patient record without a specific action. Moreover, because most prescriptions were not electronic, pharmacy staff struggled to adapt their process to a relatively "rare" event.

At the second step of the process, when the eRx repository was accessed, the first problem was associated with the absence of visual aid or code to target medications associated with the same order (Fig 3B2.1). For pharmacy staff, this would mean that they had to be carefully reviewing the list of medications in the repository (that can be long given the low adoption on the pharmacy side) to select precisely the ones that they would need to import. Because orders are chronologically ordered in the repository, this created confusion for some participants. Even though batch importation of many medications was possible with some systems, it was not the case in all of them, constraining the execution process at the pharmacy (Fig 3B2.2).

Once all medications would have been imported in the local pharmacy system, the auto-population of the different fields of the prescriptions was problematic (Fig. 3B3). For all prescriptions, at least one field had to be manually modified, namely the instructions, that were free text (even if initially structured in the prescriber's EHR system). A manual copy-paste of the instructions, from the eRx to the pharmacy system, was possible in some PMS, but not all. For other fields, such as medication ID, prescriber ID, quantity, refills, and duration, manual modifications were also frequent, indicating a lack of standardization. Moreover, it seemed that manual modifications were more frequent with some pharmaceutical forms or some type of prescribers: while capsules and tablets were usually correctly auto-populated in the PMS, other pharmaceutical forms (inhalers, injectables, drops, creams) seemed to create problems more frequently. Similarly, while physicians as a prescriber were generally correctly auto-populated in the PMS, residents, nurse practitioner and nurses seemed to generate more manual modifications in the local pharmacy system for the prescriber ID field.

Finally, the last problem identified for the validation in the PMS was the absence of a visual representation of the original eRx, that is required by pharmacist in the final step of the validation process (Fig 3B4). More problematic, one PMS had no

inalterable version of the eRx available, leading to an impossibility for the pharmacist to know if the staff had manually changed any field during their execution process, a shortfall creating an important risk of error.

Overall, the execution of a prescription using the paper copy was generally perceived as quicker and easier for pharmacy staff, while they recognized that a printed (electronic) prescription was easier to read than a manuscript prescription.

Discussion

To our knowledge, this is the first study to evaluate an e-prescribing system fully implemented in Canada, both from the prescriber and the receiver's sides. While the potential for positive outcomes associated with computerized provider order entry (CPOE) for medications have been demonstrated in some settings, many safety issues have been documented[16,18,19]. Our study adds to this literature by highlighting some specific challenges hindering adoption and increasing the risk of technology-induced errors in the medication management process.

First, our results suggest that the misalignment between the system and the prescribing and dispensing processes were numerous, with no simple workarounds for users. Second, the implementation of the transmission feature was still not completed, hindering the potential for benefits at various steps. Third, the quality of e-prescription needs to be improved, and safety issues were identified both by prescribers and pharmacists using the system with their respective EHR or PMS. Consequently, the level of use of the system was low, with only 2% of prescriptions being electronically transmitted and retrieved in pharmacy. Overall, the low adoption of e-prescribing that was observed in this study is probably essentially related to the poor quality of the system from the user's point of view, both on the prescriber's and the receiver's side.

On the Incomplete Implementation of the System

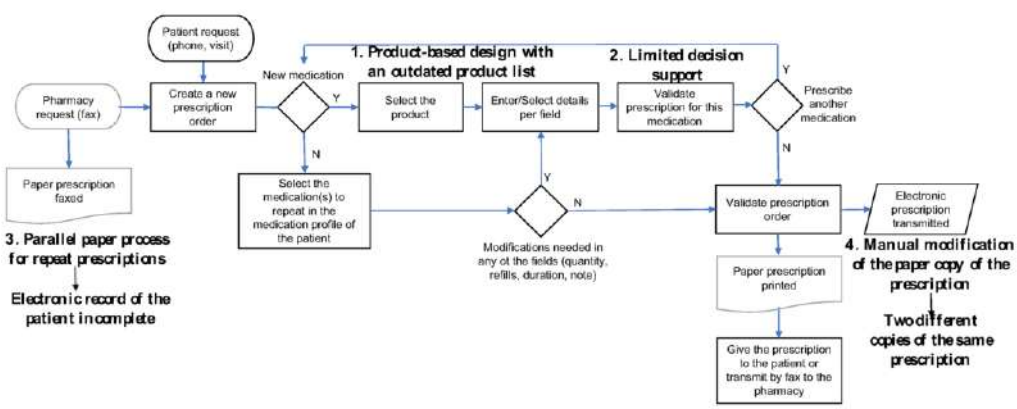
E-prescribing consists of different features that all needs to be coherently implemented for benefits to be actualized, from the request, to the creation & validation of the order, and finally its transmission. While the adoption of EHR with an e-prescribing feature for creation of orders has increased drastically in the past decade in Quebec, our results suggest that the implementation is still incomplete. First, only 2% of prescriptions were electronically transmitted and retrieved during the study period. While many prescribers, such as clinicians in acute care centers, do not have access to a certified EHR including the e-prescribing feature in Quebec, our results also suggest that many prescribers have access to an EHR to create their medication orders, but then only print it (or they may electronically fax it to the pharmacy). It is thus important that further research on e-prescribing considers this distinction between the creation of the order and its transmission, where one step of the process can be electronic, and not the other.

Moreover, even when prescribers were using the transmission feature, we have observed a systematic printing of the paper copy of the prescription, sometimes associated with manual modifications of the paper copy as a work-around for the design of the feature. This creates a major safety issue, with an increased risk of error because two different copies of the same prescriptions are created. While e-prescribing is being presented as a technology designed specifically to decrease the risk of falsification, by ensuring the integrity of prescriptions, this situation undermines the credibility of the system, and reduces its potential for benefits at this step.

On the Quality of e-Prescription

Our results suggest that the quality of e-prescription transmitted through the system could be improved. Specifically, two elements were problematic: 1) the limited decision support, with no alert including dose or patient-related characteristics; 2) the product-based design of the e-prescription, not well aligned with the cognitive process of prescribing medications, because it includes information about the packaging and the format that was not useful for clinicians. A recent study by Quist and colleagues have already described the high variability in the display of medication names in different CPOE systems in US (both inpatient and outpatient)[20]. Our study suggests that the structure of the prescription order, and specifically the logic behind the selection of the medication (product-based or dose-based) might be more important than the name of the medication itself. Precisely, caution attention needs to be dedicated to the fields, their format and content, and their reference terminology (if any). This is not yet standardized in an electronic format in the Canadian setting, as observed in this study. Consequently, on the pharmacy side, the auto-generation of the various fields was associated with problems, where manual manipulations were required, decreasing the potential for benefits realization here again. While this lack of standardization was already described during the pilot phase of the technology, no changes to the system were made in the past 5 years[14]. This first study since the full implementation of the system highlights the need for further research on the usability of this e-prescribing system, including a more diverse sample of commercial systems and users.

A



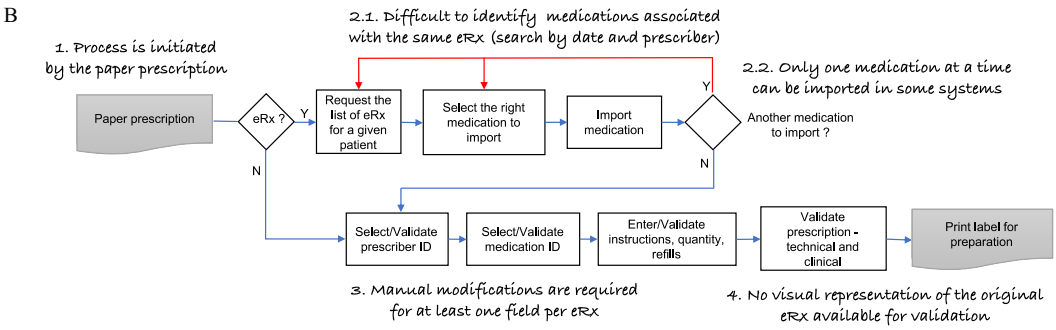


Figure 3 – Problems Identified on the Prescriber's Side (A) and Pharmacy's Side (B)

Conclusions

The adoption of a nationwide e-prescribing system was impeded by the low quality of the system, and its incomplete implementation, where the dispensing processes were based on the paper copy of the electronic prescription. Overall, this study highlights the need for improved certification mechanisms of EMR and PMS related to e-prescribing feature at the level of the province, as well as a proactive implementation strategy addressing the identified issues. Further research should also analyze the design of the prescription order, and adopt a standardized way to describe different types of design for medication prescription to improve our ability to compare systems and their effectiveness to improve the medication management processes.

Acknowledgements

Canada Health Infoway funded this study. Aude Motulsky was financially supported by the Centre de recherche du CHUM and the FRQ-Santé. We thank all participants for their time and willingness to contribute to this study.

References

[1] D.S. Bell, and M.A. Friedman, E-prescribing and the medicare modernization act of 2003. Paving the on-ramp to fully integrated health information technology?, *Health Aff (Millwood)*. **24** (2005) 1159–69.

[2] A. Motulsky, L. Lamothe, and C. Sicotte, Impacts of second-generation electronic prescriptions on the medication management process in primary care: a systematic review, *International Journal of Medical Informatics*. **82** (2013) 473–91. doi:10.1016/j.ijmedinf.2013.01.012.

[3] J. Wong, A. Motulsky, M. Abrahamowicz, T. & al., Off-label indications for antidepressants in primary care: descriptive study of prescriptions from an indication based electronic prescribing system, *BMJ*. **356** (2017) j603.

[4] J.A. Cooper, C.A. Cadogan, S.M. Patterson & al., Interventions to improve the appropriate use of polypharmacy in older people: a Cochrane systematic review, *BMJ Open*. **5** (2015) e009235.

[5] Health information and quality authority, Scotland, EPrescribing and Electronic Transfer of Prescriptions: an International Review | hiqa.ie, n.d. http://www.hiqa.ie/publications/eprescribing-and-electronic-transfer-prescriptions-international-review (accessed June 21, 2014).

[6] J.E. Brennan, Adoption of national electronic prescribing services in primary care in Europe - policy lessons for Ireland, (2015). https://ulir.ul.ie/handle/10344/4827 (accessed January 11, 2018).

[7] J. Brennan, A. McElligott, and N. Power, National health models and the adoption of eHealth and ePrescribing in primary care - New evidence from Europe, *Journal of Innovation in Health Informatics*. **22** (2015) 399–408.

[8] O.K. Odukoya, and M.A. Chui, E-prescribing: a focused review and new approach to addressing safety in pharmacies and primary care, *Research in Social & Administrative Pharmacy : RSAP*. **9** (2013) 996–1003.

[9] L.J. Schleiden, O.K. Odukoya, and M.A. Chui, Older Adults' Perceptions of E-Prescribing: Impact on Patient Care, *Perspectives in Health Information Management*. **12** (2015) 1d.

[10] J.M. Grossman, D.A. Cross, E.R. Boukus, and G.R. Cohen, Transmitting and processing electronic prescriptions: experiences of physician practices and pharmacies, *Journal of the American Medical Informatics Association*. (2011).

[11] Y. Yang, S. Ward-Charlerie, A.A. Dhavle, M.T. Rupp, and J. Green, Quality and Variability of Patient Directions in Electronic Prescriptions in the Ambulatory Care Setting, *J Manag Care Spec Pharm*. (2018) 1–9.

[12] O. Odukoya, and M. Chui, Relationship between e-prescriptions and community pharmacy workflow, *Journal of the American Pharmaceutical Association*. **52** (2012) e168–e174.

[13] O. Odukoya, and M. Chui, Retail pharmacy staff perceptions of design strengths and weaknesses of electronic prescriptions, *Journal of the American Medical Informatics Association*. **19** (2012) 1059–1065.

[14] A. Motulsky, C. Sicotte, M.-P. Gagnon & al. Challenges to the implementation of a nationwide electronic prescribing network in primary care: a qualitative study of users' perceptions, *Journal of the American Medical Informatics Association*. **22** (2015) 838–848.

[15] S.P. Slight, T. Eguale, M.G. Amato, A.C. Seger, D.L. Whitney, D.W. Bates, and G.D. Schiff, The vulnerabilities of computerized physician order entry systems: a qualitative study, *J Am Med Inform Assoc*. **23** (2016) 311–316. doi:10.1093/jamia/ocv135.

[16] Y. Lee, M.-Y. Jung, G.W. Shin, S. Bahn, T. Park, I. Cho, and J.-H. Lee, Safety and Usability Guidelines of Clinical Information Systems Integrating Clinical Workflow: A Systematic Review, *Healthc Inform Res*. **24** (2018) 157–169.

[17] A. Kushniruk, E. Borycki, S. Kuwata, and J. Kannry, Predicting changes in workflow resulting from healthcare information systems: ensuring the safety of healthcare, *Healthc Q*. **9 Spec No** (2006) 114–118.

[18] C.L. Brown, H.L. Mulcaster, K.L. Triffitt, & al., A systematic review of the types and causes of prescribing errors generated from using computerized provider order entry systems in primary and secondary care, *JAMIA* **24** (2017) 432–440.

[19] G.D. Schiff, M.G. Amato, T. Eguale & al. Computerised physician order entry-related medication errors: analysis of reported errors and vulnerability testing of current systems, *BMJ Qual Saf*. **24** (2015) 264–271.

[20] A.J.L. Quist, T.-T.T. Hickman, M.G. Amato & al. Analysis of variations in the display of drug names in computerized prescriber-order-entry systems, *Am J Health Syst Pharm*. **74** (2017) 499–509.

Address for correspondence

Aude Motulsky, aude.motulsky@umontreal.ca; +1 (514) 343-5631

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190317

Citizen Perspectives on Cross-Border eHealth Data Exchange: A European Survey

Pantelis Natsiavas[a], Christine Kakalou[a], Kostas Votis[b], Dimitrios Tzovaras[b], Vassilis Koutkias[a]

[a] *Institute of Applied Biosciences, Centre for Research & Technology Hellas, Thermi, Thessaloniki, Greece*
[b] *Information Technologies Institute, Centre for Research & Technology Hellas, Thermi, Thessaloniki, Greece*

Abstract

Efficient and secure cross-border eHealth data exchange has been recently identified by the European Commission as one of the top-three priorities for the digital transformation of health and care in the European Union. To this end, various organizational, legal, ethical, and technical challenges, related to citizens' privacy and health data security arise. This paper discusses an online survey that was conducted with the participation of European citizens, aiming to identify how they feel about exchanging their health data with healthcare professionals or eHealth service providers and to what extent they are aware of the privacy, legal, security, and technology acceptance issues (e.g. use of biometrics, mobile apps, etc.). The survey rationale, structure, and results are presented, while potential barriers and facilitators regarding cross-border health data exchange and the adoption of eHealth solutions at large are discussed.

Keywords:

Health Information Exchange, Privacy, Computer Security, Surveys and Questionnaires

Introduction

Recent advances in digital health technologies outline a paradigm shift for healthcare delivery [1, 2]. While technologies such as the Internet of Things (IoT), big data analytics, Artificial Intelligence (AI), robotics, the Future Internet offer new opportunities for more effective and personalised healthcare delivery, they also introduce significant challenges that have to be addressed [3]. In particular, due to the increased connectivity and the underlying technical complexity of these novel technological artifacts, complex cybersecurity risks have to be tackled. To this end, cybersecurity threats could prove a significant barrier for their adoption in the healthcare sector. These concerns are further reinforced by various alarming reports purporting approximately 90% of healthcare institutions as victims of security breaches and cyberattacks [4]. Since 2010, cyberattacks have increased up to 125% and have been the main culprit of health data security breaches. As a result, patients/citizens, healthcare providers (HCPs) as well as policy makers tend to be reluctant about digital health services [5]. Investors also express scepticism to fund such activities, thus significantly affecting the acceptance of these new technologies as a part of the provided healthcare services.

As the number of travellers for work, education, and tourism constantly increases in the European Union (EU), cross-border health data exchange becomes a natural need to support proper healthcare services and continuity of care. However, people, especially those suffering from chronic diseases, are facing obstacles in travelling outside their country of residence, due to the lack of an established framework for health data exchange among healthcare organizations across the EU. To this end, one of the top three priorities of the European Commission regarding the digital transformation of health and care in the Digital Single Market constitutes citizens' secure access to their health data, including across borders of the EU [6].

To-date, the core effort in the EU for enabling cross-border health data exchange has been focusing on interoperability aspects, with projects such as epSOS (European Partners – Smart Open Services) [7], OpenNCP (Open-source and reference version of the NCP software [8] - the software implementation of epSOS), and lately the Trillium project, which focuses on EU-US cooperation and particularly on exchanging patient summary data. However, limited focus has been given to the cybersecurity aspects that are entailed in cross-border health data exchange.

Aiming to address this challenge, the EU-funded H2020 KONFIDO (Secure and Trusted Paradigm for Interoperable eHealth Services) project [9] that develops a toolset to facilitate secure cross-border exchange, storage, and overall handling of health data [10]. The toolset leverages various novel technologies, such as homomorphic encryption [11], photonic Physical Unclonable Functions (p-PUF) [12], a Security Information and Event Management (SIEM) system, [13] and blockchain-based auditing [14]. In addition, it builds upon existing frameworks, mainly OpenNCP and eIDAS (electronic IDentification, Authentication, and trust Services) [15]. OpenNCP offers a set of interoperability services to enable national and regional eHealth platforms to conduct cross-border health data exchange, while eIDAS implements the EU regulation regarding electronic identification and trust services for electronic transactions in the internal market, which includes eHealth applications among others.

As part of the KONFIDO user requirements engineering phase [16], we conducted an online survey to gain useful insights for the technical development of the envisaged toolset. The main goal of the survey was to identify how patients/citizens feel about exchanging their health data with healthcare professionals or eHealth service providers, as well as to what extent patients/citizens are aware of the entailed privacy and security issues. The survey also aimed to investigate technology acceptance issues, like the use of mobile health apps and the potential use of biometrics based on input collected from citizens across Europe.

In this paper, we present the overall survey methodology and the results, and conclude by consolidating these outcomes in terms of key barriers and facilitators regarding the acceptance and ultimately the adoption of digital health technologies focusing on cybersecurity and interoperability issues.

Methods

The survey was designed and implemented upon key principles of human psychology [17]. The main steps involved in this methodology are:

1. Deciding what information should be collected
2. Deciding how to conduct the survey
3. Constructing a draft of the respective questionnaire
4. Revising the draft questionnaire
5. Pre-testing the questionnaire
6. Revising the questionnaire and its use procedures

Several online sources were investigated prior to the questionnaire design. These included relevant surveys conducted by other organizations, reports, and scientific papers. In addition, the survey was designed and deployed, incorporating sophisticated features, such as:

- conditional workflow of questions based on the answers submitted on earlier questions, so that only questions relevant to the responder appear;
- input validation to avoid erroneous data entries;
- use of control questions (or "trap questions") to verify response quality; and
- export of the collected responses in a format that facilitates further data analysis

The respective questionnaire was built in an iterative fashion to validate its alignment with the survey scope and comprehension. The final version of the questionnaire was published in seven European languages, namely, Danish, Dutch, English, French, Greek, Italian, and Spanish.

The questions (35 in total) were organized in six sections:

7. *Awareness regarding Information Technology (IT) risks*: Focused on identifying the responder's awareness level regarding the risks entailed (both explicitly and implicitly) in using digital health tools
8. *Legislation*: Aimed at identifying the responder's familiarity with relevant legislation artifacts
9. *Cross-border medical treatment*: Aimed to provide insights on whether the responder was medically treated or hospitalized abroad, detailing the circumstances under this event
10. *Cross-border health data exchange*: Concerned with the responder's opinion regarding the need of cross-border health data exchange
11. *Barriers and facilitators*: Aimed at identifying key issues that facilitate or discourage cross-border health data exchange from a patient's/citizen's viewpoint
12. *Demographics*: Aimed to identify some key information about the responder, in order to facilitate the statistical analysis of the obtained data

The online survey was disseminated publicly via relevant forums,, mailing lists, and social media (i.e. KONFIDO project Twitter and Facebook accounts), targeting citizen groups that could be related with the subject of cross-border health data exchange (for example, chronic patient associations, immigrant groups, medical tourism groups) and also the general public. No exclusion criteria were applied aiming to increase participation of people who were not necessarily aware of the recent advances in the eHealth domain and its security aspects..

Before conducting the survey, it was approved by the Bioethics Committee of the Centre of Research and Technology Hellas (CERTH), as CERTH was responsible for the data collection and control of the study. The survey did not require the disclosure of the responder's identity.

Results

The survey was available online for three weeks and collected a total of 437 responses, out of which about 30% of responders contributed to the online questionnaire but did not complete it (124 incomplete responses out of 437 in total). This is a typical behavior in online surveys, as the responder might quit the process for any reason, therefore, the incompletely taken surveys were still taken into account. More specifically, regarding the demographics of the responders, their average age was 43.96 years, their gender distribution was 38.54% females and 57.96% males, and they were distributed in 14 European countries (most of the responders declared that they were from Greece, Germany, or Denmark). Since the survey distribution was conducted via public Internet communication channels, we cannot confirm the number of people reached through the online invitation campaign and, therefore, the response ratio cannot be calculated. We summarize the main findings in the following section.

Awareness regarding IT risks

The key results from this questionnaire section can be summarized as follows:

a) 11.96% did not thought about possible health data risks.

b) Only 36.41% felt informed about these risks.

c) 66.21% of the responders did not read the respective applications' "Terms and Conditions", with more than 30% declaring that they did not feel it is worthy, given the time required to read them and 19.79% declaring indifference towards them (Figure 1).

Figure 1 - Answers to Question "Why haven't you read the terms and conditions?" (Only Responders Who Answered the Previous Question that They Did Not Read the "Terms and Conditions" Were Asked)

Furthermore, the responders expressed confusion and lack of confidence on the subject. Only 26.09% of the responders felt confident regarding their electronic health data privacy, 38.04% felt concerned but helpless about, and 16.3% stated that they avoid using eHealth services due to the lack of confidence regarding their data handling.

An interesting remark is that the responders clearly preferred to entrust their health data to national and state organizations than to private ones, despite the rather high-rate (35.33%) declared in using applications exploiting personal health data.

These applications are typically provided and operated by private companies, mostly belonging to the category of lifestyle/wellbeing monitoring.

It should also be noted that only 20% of the responders felt that their privacy was fully covered in the "Term and Conditions" of the applications. The findings regarding the reasons that led responders to not read the "Terms and Conditions" were directly linked to legislation complexity, legislation misalignment between countries and the need for usability – most of the responders felt that too much time is required to read them, and 12.50% declared that they do not understand them.

Legislation

The need for raising awareness was also evident from the findings of this section (Figure 2 and Figure 3). While almost 80% of the responders declared being familiar with the EU General Data Protection Regulation (GDPR) [18], more than 50% declared that they were familiar with legislation items that do not really exist (as captured by a control question). In particular, about 27.37% declared being aware of data concepts/initiatives that do not really exist. The need for raising awareness was also supported by the fact that 24.93% of the responders expressed no opinion on whether the current legislation effectively protects them.

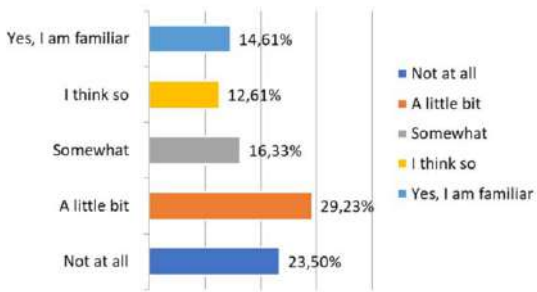


Figure 2 – Answers to the Question "Are you familiar with legislation that concerns the use of health data?"

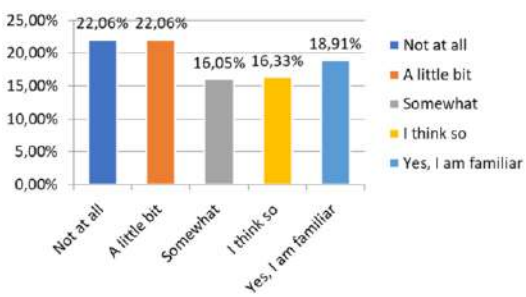


Figure 3 – Answers to the Question "Are you familiar with the data ownership concept?"

Regarding the effectiveness of the current legislation, only 27.79% of the responders felt protected, while 37.54% felt either defenceless, or insufficiently protected (Figure 4). As shown in Figure 5, citizens argued that legislation alignment among EU countries is a necessity. Furthermore, dissemination regarding legal issues and the level of legal protection provided were also identified as important items.

Cross-border Medical Treatment

28.16% of the responders were medically treated abroad. The reasons for their hospitalization or medical treatment abroad

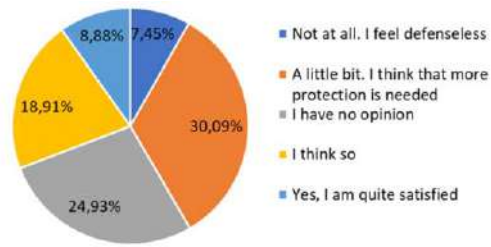


Figure 4 – Answers to question "Are you satisfied with the level of protection provided by current legislation?".

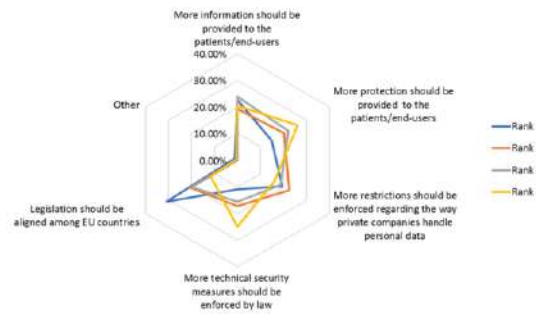


Figure 5 – Ranking of reasons for question "Regarding legislation, please rank the most important things to be improved regarding cross-border health data exchange".

were clearly depicted: 44.90% of the responders reported a sudden incident while travelling (e.g. a car accident), 37.76% were immigrants, and 16.33% referred to other reasons (e.g. studying abroad).

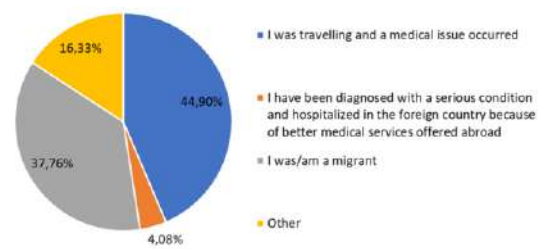


Figure 6 – Answers to the Question "In what circumstances have you used medical services abroad?"

These findings clearly highlighted that unscheduled access to healthcare services abroad, such as accidents/incidents while travelling, were critical for European citizens.

Cross-border Medical Data Exchange

The responders were highly in favor of cross-border data exchange, since only 9.28% expressed a negative opinion (Figure 7). Among the responders who viewed this issue as a critical one, the main concern involved technical issues and, particularly, information security aspects. Furthermore, a high percentage of responders (71.56%), would consent in sharing medical data with foreign medical personnel in case of an emergency, with the rest of them mostly worrying about the technical issues.

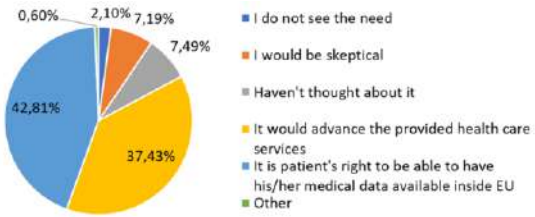


Figure 7 – Answers to the Question "How do you feel about medical data shared with foreign healthcare professionals /institutions to facilitate treatment while being abroad?"

It is also clear that the responders would prefer to use a European authentication card, rather than using biometric characteristics to authenticate themselves for cross-border eHealth data exchange. In addition, the need for a detailed description of the underlying context for using health data was evident, given that 53.29% of the responders wanted to be thoroughly informed before consenting to their health data usage.

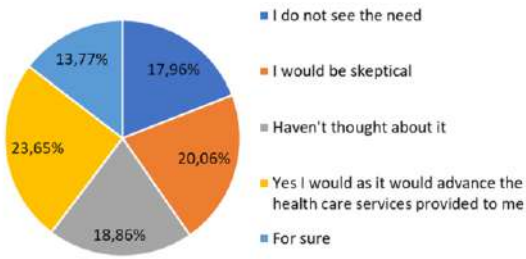


Figure 8 – Answers to the Question "Would you prefer using a biometric characteristic (e.g. fingerprints) instead of an ID card to facilitate cross-border medical data exchange?"

An interesting remark is that the vast majority of the responders (72.46%) declared that they were in favor of sharing personal data for research purposes, at least under the terms of anonymization. Finally, it should be noted that the responders who were sceptical towards cross-border health data exchange did not focus on the cross-border data exchange, but rather on the security challenges that have to do with data sharing, regardless of whether this sharing had to do with foreigners or not.

Barriers and Facilitators

The responders identified the following key factors for the acceptance of cross-border health data exchange:

1. A common legislation among EU Member States
2. Better control of data management practices applied by companies
3. More information on the processing of citizen health data

These key points clearly support the application of new European regulation, GDPR. Providing consent is one key aspect of the overall data sharing process: it could, on one hand, facilitate the process and, on the other hand, act as a barrier. The results clearly indicate the need for a flexible consent process, as 73.44% of the responders supported that "*Patient consent should be actively enforced. However, in some special cases (e.g. when the patient is unconscious), it could be skipped in favour of the provided medical services*" (Figure 9).

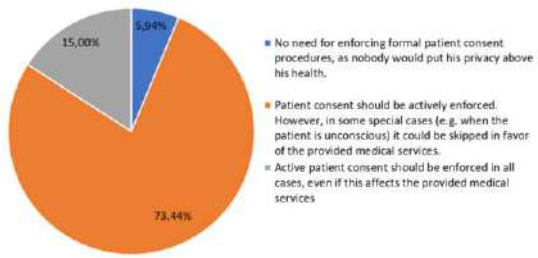


Figure 9 – Answers to the question "Please select in your opinion the level of required patient consent, in cases of cross-border data exchange"

Finally, the most important barriers regarding the acceptance of cross-border health data exchange are:

1 Lack of trust regarding the intentions of data collection

2. Lack of suitable legislation

3. The risks of interlinking these data with other personal information already available and traceable in the Internet (e.g. posts in social media platforms)

Consolidated Outcomes

The main outcomes of the survey focusing on how these could contribute to the development of cybersecurity solutions for health data exchange at large, and in cross-border scenarios in particular, are as follows:

- Raise awareness among patients/citizens and other stakeholders. More specifically, awareness should be raised regarding: (a) the risks of using IT for health data management (b) citizens' confidence with respect to their personal data handling (c) the need to simplify the "Terms and Conditions" for using eHealth services/applications (d) the need for a flexible and comprehensive consent process, and (e) the need for a clear and aligned legislation across EU Member States.
- Incidents during travel seem to be of high value. However, other use cases (such as immigrants living temporarily abroad) should be considered.
- The main control of cross-border health data exchange should be built upon national infrastructures (such as the national Electronic Health Record), as the patients/citizens tend to trust them more.
- Technical solutions should focus on using a Europe-wide authentication method such as one based on eIDAS and avoid the use of biometric characteristics.
- Provide a simple and comprehensible consent process, while being flexible in cases where the patient is unconscious (emergency scenarios).

Discussion and Conclusion

Citizens are a key stakeholder in eHealth data management, as cross-border health data exchange becomes a necessity across the EU. Recent regulation and legislation activities such as the GDPR set a framework for legal, ethical, and practical issues related to health data management and personal data protection. The KONFIDO project develops a technology toolset,

aiming to enhance information security for cross-border health data exchange, building upon emerging European frameworks such as OpenNCP and eIDAS.

To this end, the project relied on an intensive end-user engagement strategy [16], aiming to obtain feedback related to the current landscape and the practical issues that health data exchange entails through digital health solutions. Various activities were conducted, including a survey with health IT professionals, eHealth companies, and health policy makers [19], to identify digital health acceptance barriers and facilitators, using the survey presented in the current paper.

We presented the main results of the survey focusing on European citizens. A list of key issues was identified and a number of challenges were consolidated. These results could be used as a beacon for the development of new technical solutions in the context of health data exchange and health data management at large, including cross-border health data exchange. We thus argue that our findings provide useful insights for stakeholders of the European eHealth ecosystem, encouraging them to adapt their services and reinforce the acceptance and consequently the adoption of their solutions by the targeted end-users.

The main limitations of the current study relate to the risk of bias due to the following reasons: (a) the specific questions could be considered as "leading" responders to specific answers based on the reader's subjective judgement, and (b) the non-uniform distribution of the responders across European countries. The lack of detailed responders' demographic information is because of the fact that all the questions were optional and as the demographics section was put last, only a small portion of the responders answered them. As part of the presented methodological approach, all project partners revised the questionnaire to avoid responder "guidance" and also tried to disseminate the survey as widely as possible. Despite these limitations, we consider the results valuable and be able to provide useful insights. In order to reduce the effect of bias and further explore the collected data, the collected responses will be analyzed in combination and in comparison with the relevant studies and surveys conducted by other organizations such as the Healthcare Information and Management Systems Society (HIMMS).

Acknowledgements

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 727528 (KONFIDO - Secure and Trusted Paradigm for Interoperable eHealth Services). This paper reflects only the authors' views and the Commission is not liable for any use that may be made of the information contained therein.

The authors would like to thank the KONFIDO partners for their help on the survey dissemination and the questionnaire translation to the supported European languages.

References

[1] M. Herrmann, P. Boehme, T. Mondritzki, J.P. Ehlers, S. Kavadias, and H. Truebel, Digital Transformation and Disruption of the Health Care Sector: Internet-Based Observational Study, *J Med Internet Res.* **20** (2018) e104.

[2] B. Meskó, et al., Digital health is a cultural transformation of traditional healthcare, *MHealth*. **3** (2017) 38.

[3] A. Sharma, et al., Using Digital Health Technology to Better Generate Evidence and Deliver Evidence-Based Care, *J Am Coll Cardiol.* **71** (2018) 2680–90.

[4] Cyber Attacks and Negligence Lead to Rise in Medical Data Breaches, (n.d.). https://www.nbcnews.com/tech/tech-news/cyber-attacks-negligence-lead-rise-medical-data-breaches-n575471 (accessed November 25, 2018).

[5] M.P. Jarrett, Cybersecurity - A Serious Patient Care Concern, *JAMA*. **318** (2017) 1319.

[6] European Commission, Communication to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society, Brussels, 25.4.2018 COM(2018) 233 final, Available at: http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51628.

[7] epSOS project web site, (n.d.). http://www.epsos.eu/ (accessed November 16, 2018).

[8] M. Fonseca, K. Karkaletsis, I.A. Cruz, A. Berler, and I.C. Oliveira, OpenNCP: a novel framework to foster cross-border e-Health services., *Stud Health Technol Inform*. **210** (2015) 617–21.

[9] KONFIDO project website, (n.d.). http://www.konfido-project.eu/konfido/ (accessed Nov. 15, 2018).

[10] M. Staffa, et al., An OpenNCP-based Solution for Secure eHealth Data Exchange, *J Netw Comput Appl.* **116** (2018) 65–85.

[11] X. Yi, R. Paulet, and E. Bertino, Homomorphic Encryption, in: *Homomorphic Encryption Appl.*, Springer, Cham, 2014: pp. 27–46.

[12] C. Mesaritakis, et al., Physical Unclonable Function based on a Multi-Mode Optical Waveguide, *Sci Rep.* **8** (2018) 9653.

[13] S. Bhatt, P.K. Manadhata, and L. Zomlot, The Operational Role of Security Information and Event Management Systems, *IEEE Secur Priv.* **12** (2014) 35–41.

[14] A. Theodouli, S. Arakliotis, K. Moschou, K. Votis, and D. Tzovaras, On the Design of a Blockchain-Based System to Facilitate Healthcare Data Sharing, in: 2018 17th IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. 12th IEEE Int. Conf. Big Data Sci. Eng., IEEE, 2018: pp. 1374–1379.

[15] eIDAS web site, (n.d.). https://www.eid.as/home/ (accessed November 16, 2018).

[16] P. Natsiavas, et al., Comprehensive user requirements engineering methodology for secure and interoperable health data exchange, *BMC Med Inform Decis Mak.* **18** (2018) 85.

[17] J.J. Shaughnessy, et al., Research methods in psychology, n.d., 10th Edition, McGraw-Hill, 2014

[18] EU General Data Protection Regulation (GDPR), (n.d.). https://eugdpr.org/ (accessed November 25, 2018).

[19] P. Natsiavas, C. Kakalou, K. Votis, D. Tzovaras, N. Maglaveras, I. Komnios, and V. Koutkias, Identification of Barriers and Facilitators for eHealth Acceptance: The KONFIDO Study, in: Springer, Singapore, 2018: pp. 81–85.

Address for correspondence

Dr Vassilis Koutkias, Institute of Applied Biosciences, Centre for Research & Technology Hellas, 6th Km. Charilaou-Thermi Road, P.O. BOX 60361 GR – 57001, Thermi, Thessaloniki, Greece.

Email: vkoutkias@certh.gr.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190318

Implementation of Clinical Decision Support Services to Detect Potential Drug-Drug Interaction Using Clinical Quality Language

Binh-Phi Nguyen[a], Thomas Reese[c], Stefen Decker[a,b], Daniel Malone[d], Richard D. Boyce[e], Oya Beyan[a,b]

[a] *Informatik 5, RWTH Aachen University, Aachen, Germany,*
[b] *Fraunhofer FIT, Sankt Augustin, Germany*
[c] *Department of Biomedical Informatics, University of Utah, Salt Lake City, Utah, USA*
[d] *College of Pharmacy, University of Arizona, Tucson, Arizona, USA*
[e] *Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA*

Abstract

Potential drug-drug interactions (PDDI) rules are currently represented without any common standard making them difficult to update, maintain, and exchange. The PDDI minimum information model developed by the Semantic Web in the Healthcare and Life Sciences Community Group describes PDDI knowledge in an actionable format. In this paper, we report implementation and evaluation of CDS Services which represent PDDI knowledge with Clinical Quality Language (CQL). The suggested solution is based on emerging standards including CDS Hooks, FHIR, and CQL. Two use cases are selected, implemented with CQL rules and tested at the Connectathon held at the 32nd Annual Plenary & Working Group Meeting of HL7.

Keywords:

Potential Drug-Drug Interaction, Clinical decision support systems, Electronic health records.

Introduction

Drug-drug interactions (DDI) are biological processes that result in a clinically meaningful change to the response of at least one co-administrated drug [1]. Identifying potential DDIs during the care process is important to ensure patient safety and health care quality. DDIs can often be predicted and mitigated at the point of care. However, a large number of drugs, lack of knowledge of DDIs [2] hinders the ability of physicians to preemptively identify all potential DDIs [3]. Therefore it is important to guide care providers with well design computerized alerting systems during the medication prescribe process [4].

In the last two decades, numerous clinical decision support (CDS) systems have been developed to support physician by presenting alerts in real time to the clinician about the current medication's potential impact to patients [5]. Ideally, CDS should provide clinicians with the relevant reference information such as knowledge or suggestions, intelligently filtered and presented at appropriate times [6]. Currently, there are many CDS systems that provide drug interaction alerts [4]. Most of the systems curate their own knowledge base since there is no complete source of potential DDI (PDDI) knowledge [7]. Also, there are currently no broadly accepted standards to guide implementers in the organization and presentation of PDDI information [8]. Most EHR vendors either develop their own internal rule engines based on their custom standards, or contract with a third party. Since these rules are written in a way just for those systems, it is difficult to exchange, review and rewrite the rules. Lack of agreement on standards hinders the reuse of the rules by third parties, creating additional burden to rewrite rules at each side.

The minimum information models help to describe the PDDI information in a detailed, actionable and contextualizable format. Standardized PDDI information can be exchanged via broadly applicable formats. In this research, we will demonstrate the use of the PDDI minimum information model developed by Semantic Web in Healthcare and Life Sciences Community Group [9], by implementing them with CQL language and extend CQF Ruler[1] to provide decision support.

The aim of this work is to present an example implementation of representing PDDI logic in CQL and execution of them by using the FHIR Clinical Reasoning module. We selected Digoxin-Cyclosporine and Warfarin-NSAIDs as use cases because they are non-trivial PDDIs for which alerts can be contextualized to specific patient cases. The developed prototype detects a PDDI and provides alerts using Clinical Decision Support Hooks to Electronic Health Record systems that subscribe to the CDS services.

The following sections provide an overview of our methods, prototype architecture, implementation of PDDI with CQL, and evaluation.

Methods

PDDI: The PDDI Minimum Information Model [9] [10] is a standard proposed by the W3C Semantic Web in Healthcare and Life Sciences Community Group. It is aimed to help the clinicians to keep up with the PDDI evidence base, document and share PDDI information by summarizing PDDI evidence from primary sources using the information elements from the PDDI minimal information model.

The information model contains 10 core information items that should be used to describe every PDDI CDS knowledge artifact as follows: (i) Clinical consequences; (ii) Contextual information/modifying factors; (iii) Drugs involved, (iv) Evidence; (v) Frequency of exposure to the interacting drug pair; (vi) Frequency of harm for persons who have been exposed to the interacting drug pair; (vii) Mechanism of the interaction, (viii) Recommended actions; (ix) Seriousness rating; and (x) Operational classification of the interaction.

CQL: Clinical Quality Language (CQL)[2] is a Health Level Seven International (HL7) authoring language standard that can

[1] https://github.com/DBCG/cqf-ruler

[2] http://cql.hl7.org

be used with both CDS and electronic Clinical Quality Measures to represent the PDDI information. The CQL syntax provides a clinical focused, author-friendly and human-readable language to clinical domain experts, since it allows for rich, modular and flexible expression of the logic, and supports different data models including FHIR. CQL is a query language built up by combining rules, namely statements, to describe the available data in terms of a data model. For decision support, the rules will be evaluated in the context of a specific patient to produce a response at some specific point in a workflow.

CDS Services: CDS Service [11] is a role which provides real-time clinical decision support as a remote service. It enables a consumer to ask for clinical decision support based on his current context. The consumer gives relevant contextual information as part of the request and receives clinically relevant suggestion describing potential actions to be taken. The service implementation must comply with CDS Hooks specification.

CDS-Hooks[3] is an emerging standard gained considerable interest from EHR vendors. It is a "hook" based pattern designed to provide a simple way to initiate requests for CDS, from any point in a clinical workflow. It specified the basic actions of registering for CDS services, calling those services, and then receiving the CDS service response in form of simple cards providing appropriate information within the context of the EHR.

Concept

Architecture: As shown in the Figure 1, the PDDI CDS Implementation was designed as a self-contained service, with three main components including HAPI FHIR server[4], CDS Services, and CQL engine[5]. The first component is the HAPI FHIR server which is an option to store the definition of services, called PlanDefinition, supported by the system and knowledge base of CQL rules, called Library. The second component is the CDS services, which serve as a core logic to map the incoming request to the corresponding service defined by PlanDefinition, execute the CQL logic of the service defined in Library following the instruction from the PlanDefinition, and generate response cards. By this way, the new service can be introduced into the system without changing the core CDS services and the configuration. The third component is the CQL engine which provides CQL rule execution for the core CDS services.

As described in detail in the current draft of the PDDI Implementation Guideline (IG) [12], the CDS Discovery service is hosted at a stable endpoint, i.e. {baseUrl}/cds-services, and allows EHRs to discover the list of available supported CDS Service, e.g. {baseUrl}/cds-services/warfarin-nsaids-cds. The list of CDS Services contains information such as a description of the CDS Service, when it should be invoked, and any data that is requested to be prefetched.

PDDI CDS Implementation Workflow: The PDDI IG envisions two hooks which are medication-prescribe at order authorization, and medication-select at the time of selecting a medication and prior to the order authorization. For this paper, we focus only on medication-prescribe in the Level 1 Implementation.

Firstly, the EHR invokes the "medication-prescribe" hook and PDDI CDS service is called by sending an HTTP POST request, namely CDS Hooks Request, containing JSON to the service endpoint (e.g. {baseUrl}/cds-services/warfarin-nsaids-cds).

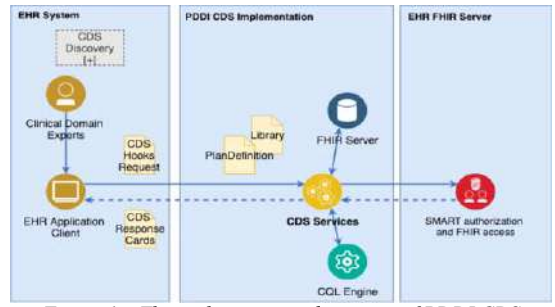


Figure 1 – The architecture and process of PDDI CDS Implementation

The CDS Hooks request contains specific information for the hook that was triggered including hook name FHIR server url, context data that the service will need, and prefetch data that is pre-queried data based on the CDS discovery. If the PDDI CDS service does not receive prefetch data in the request, it will query the EHR FHIR Server via network call with the authentication given by the EHR Application Client.

When CDS Service processes the request, the corresponding PlanDefinition and Library resources are loaded from FHIR server. Once the resources are loaded, the CQL logic in Library resource is decoded and evaluated by CQL engine with the data received either in the prefetch or from EHR FHIR server. The CQL engine evaluates the CQL logic following the guidance specified in the PlanDefinition resource.

After the evaluation is done, CDS Response cards are generated and returned to the client. Each Card has specified attributes including summary, detail, indicator, and list of suggestions providing actionable information. The specified attributes map to the core elements of the minimum information model (e.g. summary = Drugs Involved, detail = Clinical consequences, Seriousness, Mechanism of Interaction, and Evidence). The Card indicator element dictates how the EHR presents the alert (e.g. indicator = "hard-stop" could be a modal alert).

CDS Services Workflow: When the PlanDefinition is loaded by PDDI CDS services for the corresponding request, the services will load the CQL library defined in the PlanDefinition and decode it based on base64format.

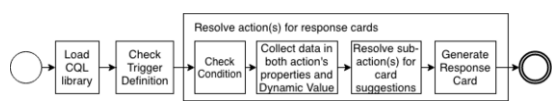


Figure 2 – The flow of processing Plan Definition

After the CQL library is loaded, the services check the hook event specified in the incoming request with the trigger definition defined in the PlanDefinition. If the hook event meets the trigger definition, the services begin to resolve the actions by evaluating the condition defined in CQL library to see whether or not the definition specified in the Action is to be applied. Then, the services collect the data in the action's properties and evaluate the dynamic value specified in the CQL library for the customizable properties. For the suggestions, the services resolve the sub-actions as the same approach as resolving actions.

[3] https://cds-hooks.org
[4] http://hapifhir.io
[5] https://github.com/DBCG/cql_engine

When the data is ready, the corresponding response card is generated and the services continue to resolve the next action until no actions left. Once finished, the services return the response cards as a result to the client.

Results

The implementation focuses on two PDDI CDS use cases which are Digoxin + Cyclosporine and Warfarin + NSAIDs. This section only discusses the Digoxin + Cyclosporine use case. Please refer to our repository (http://doi.org/10.5281/zenodo.1481220) for the full implementation and the PDDI IG [12] for the detail including PDDI model of the two use cases.

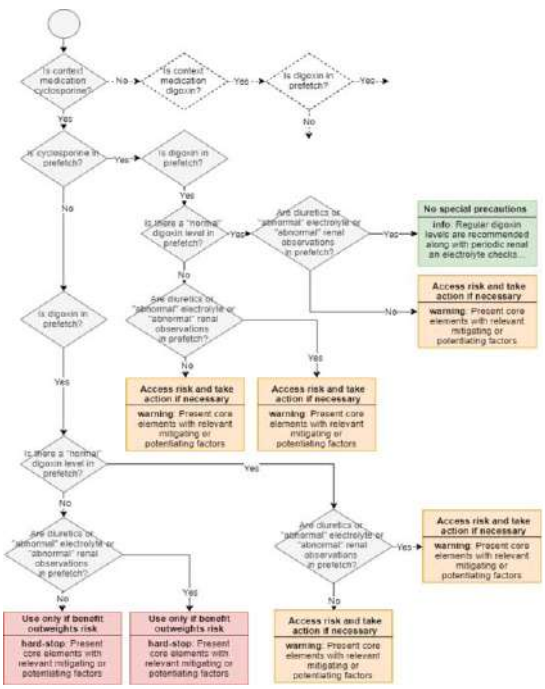


Figure 3 – The partial of decision tree for Digoxin Cyclosporine use case [12]

Use Case

The Digoxin + Cyclosporine exemplar artifact has two main decision blocks. Whether the patient is taking digoxin and/or cyclosporine at the time of the current order for digoxin or cyclosporine, and whether the patient has risk factors that may potentiate the risk of digitalis toxicity.

As shown in the Figure 2, each decision block has a certain suggestion in rectangle shape which presents core elements of PDDI model, necessary actions and level of indicator (i.e. "info" – green, "hard-stop" – red, and "warning" – orange).

Implementation of Plan Definition: PlanDefinition is a definition of the service served as a guide for the incoming request. PlanDefinition contains a set of definition including Library, Trigger Definition, Condition, Action, and Dynamic Value (Figure 3).

The Library element defines the reference to the logic used by the PlanDefinition. An example of digoxin-cyclosporine-cds PlanDefinition is the reference to Library/digoxin-cyclosporine-cds Library. This Library contains the CQL logic library encoded in base64 format as a string used by the PlanDefinition.

The Trigger Definition uses the Name Event, which allows triggering of an event opposed to a scheduled or fixed event. As an example, by specifying medication-prescribe as an event, the service only serves for the incoming request sent from medication-prescribe hook.

The condition element is used to determine whether or not the CDS logic specified in the Library is to be applied. By specifying "Inclusion Criteria" as a CQL statement for the condition, an action(s) is initiated once the condition is satisfied (i.e., true or false).

The Action element defines a list of response card needed to be generated. The subaction element indicates a list of suggestion which recommends a set of changes in the context of the current activity.

The Dynamic Value enables customization of the response card's properties by collecting the dynamic data defined in CQL statements. For the action, we define three dynamic values, such as "Get Summary" statement for the card title, "Get Detail" for the card description and "Get Indicators" for the card level. Since each decision block for PDDIs has one or more individualized information components, integrating patient-specific and product-specific data into Card elements is facilitated by the Dynamic Value element.

Implementation of CQL: All artifact logics of CQL for the Digoxin Cyclosporine use case are wrapped in the Digoxin_Cyclosporine_CDS library. A set of declarations, including data model, included libraries, valuesets, parameters, and context, provide information about the library.

The PDDI CDS uses FHIR model, version 3.0.0 as the primary data model to support for the FHIR resources.

A common library, namely PDDICDSCommon, contains all supported statements. The PDDI IG provides a number of valueset used in this library. Each valueset describes RxNorm codes drawn from one or more code systems for certain drugs. For example, the "Digoxin" valueset has this number of codes which is 197604 for Digoxin 0.125 MG Oral Tablet, 245273 for Digoxin 0.0625 MG Oral Tablet and other RxNorm codes. CQL evaluate the rules using all these RxNorm codes.

A ContextPrescription parameter refers to the list of Medication Request prescribed by the clinician specified in the incoming request. The Patient context restricts the information within a scope of single patient.

As specified in the condition element of the PlanDefinition, the "Inclusion Criteria" statement is used to check whether the patient of the incoming request is taking digoxin and/or cyclosporine at the time of the current order for digoxin or cyclosporine.

```
define "Inclusion Criteria":
  ( "Is Context medication cyclosporine"
      and "Is digoxin in prefetch"
  ) or (
    "Is Context medication digoxin"
      and "Is cyclosporine in prefetch"
  )
```

To express "Is Context medication cyclosporine" criterion, we need to check the existence of the list of Medication Request containing Cyclosporine specified in the ContextPrescription. The "Cyclosporine Prescription" statement requires that all codes from medication of the ContextPrescription parameter belong to the valueset identified by "Cyclosporine".

```
define "Is Context medication cyclosporine":
  exists ("Cyclosporine Prescription")
```

```
define "Cyclosporine Prescription":
  ContextPrescriptions P
    where Common.ToCode(P.medication.coding[0]) in
"Cyclosporine"
```

Next, the second criterion is "Is digoxin in prefetch". This criteria requires that the code of MedicationRequest is the code in the valueset identified by "Digoxin". Since we are in the Patient context, this query retrieves all Medication Request for only current Patient in the context. To filter the results which authored within the 100-day look-back period, we had to construct an interval from the 100-day look-back to infinitive.

```
define "Is digoxin in prefetch":
  exists ("Digoxin Rx")

define "Digoxin Rx":
  [MedicationRequest: "Digoxin"] MR
    where MR.authoredOn.value in Interval[Today()-100 days,null]
```

We use the same approach for the remaining criteria.

As defined in the dynamicValue element of the PlanDefinition for the "summary" property of the response card, the "Get Base Summary" statement provides the short description of the drugs involved in this interaction.

```
define "Get Base Summary":
  'Potential Drug-Drug Interaction between digoxin ('
    + (if "Is Context medication digoxin"
        then Common.GetDrugNames("Digoxin Prescription")
      else Common.GetDrugNames("Digoxin Rx")
    )
    + ') and cyclosporine ('
    + (if "Is Context medication cyclosporine"
        then Common.GetDrugNames("Cyclosporine Prescription")
      else Common.GetDrugNames("Cyclosporine Rx")
    )
    + ')'
```

For the "detail" property of the response card, the "Get Base Detail" statement provides detail of PDDI minimum information.

```
define "Get Base Detail":
  'Increased risk of digoxin toxicity...'
```

For the "indicator" property of the response card, the "Get Base Indicator" statement specifies the importance of what this card conveys.

```
define "Get Base Indicator":
  if "Is Context medication cyclosporine" then
    if "Is cyclosporine in prefetch" then
      if "Is there a normal digoxin level in prefetch" then
        if "Are diuretics or abnormal electrolyte or abnormal
renal observations in prefetch"
          then 'info'
        else 'warning'
      else 'warning'
    else
      if "Is there a normal digoxin level in prefetch" then
        then 'warning'
      else 'hard-stop'
  else
    if "Is digoxin in prefetch" then
      if "Is there a normal digoxin level in prefetch" then
        if "Are diuretics or abnormal electrolyte or abnormal
renal observations in prefetch"
          then 'info'
        else 'warning'
      else 'warning'
    else 'warning'
```

The first criterion of the "Get Base Indicator" statement is "Is there a normal digoxin level in prefetch". This criterion requires that all codes of Observation are in the valueset identified by "Digoxin LOINC". The normal digoxin observation is identified by measure the number of quantity less than 0.9ng/mL. To filter the results which take effects within the 30-day look-back period, we had to construct an interval from the 30-day look-back and infinitive. Because the most recent of Observation is important for the check, we select the latest result after ordered by the effective date.

```
define "Is there a normal digoxin level in prefetch":
  exists ("Normal Digoxin Observation")

define "Normal Digoxin Observation":
  Last (
    [Observation: "Digoxin LOINC"] O
      where O.effective.value in Interval[Today()-30 days, null]
        and Common.ToQuantity(O.value) < 0.9 'ng/mL'
      sort by effective.value
  )
```

The second criterion of the "Get Base Indicator" statement is "Are diuretics or abnormal electrolyte or abnormal renal observations in prefetch". This criteria contains three main components which are the diuretics medication requests involved, the abnormal electrolyte observations involved, and the abnormal renal observations involved.

We apply the same approach to detect the diuretics medication requests involved by checking the existence of Aldosterone Antagonists and Loop Diuretics in the Medication Request. For the remaining components, a similar approach for the abnormal electrolyte observations and renal observations has been used to detect the abnormal level of Potassium, Magnesium, Calcium, and Renal.

Table 1 – The variety of resources tested in two use cases

Warfarin-NSAIDs	Digoxin-Cyclosporine
MedicationRequest	MedicationRequest
MedicationDispense	MedicationDispense
MedicationStatement	MedicationStatement
MedicationAdministration	MedicationAdministration
Patient	Patient
Encounter	Encounter
Condition	Observation

Evaluation

We prepared FHIR resources (Table 1) to cover all cases of two decision trees and tested around 170 FHIR resources in STU3 version for 21 different patients on 14 cases of Digoxin-Cyclosporine and 7 cases of Warfarin-NSAIDs decision tree to support the draft IG. We also performed the evaluation at the Connectathon held at the 32nd Annual Plenary & Working Group Meeting of HL7 held in September 2018.

The evaluation was done using Postman[6] and CDS Hooks Sandbox[7], a tool developed by CDS Hooks team demonstrate how CDS Hooks would work with an EHR system. In brief, the clinician enters the medication, e.g., Ketorolac Tromethamine 10 MG Oral Tablet, for the specific treatment, and the CDS Hooks Sandbox then invokes the "medication-prescribe" hook to send the request to the PDDI-CDS services endpoint (e.g. {baseUrl}/cds-services/warfarin-nsaids-cds). After the PDDI CDS service processes the request, the CDS Response cards are returned. Finally, the CDS Response cards is presented by the CDS Hooks Sandbox.

Discussion

We were able to implement a PDDI CDS service using CQL, FHIR, and CDS Hooks. This shows the feasibility and that CQL was sufficiently expressive to cover to realistic use cases. We suggested enhancements to CDS Hooks that would enable it to support PDDI CDS using the minimum information model, and the combination of tools indeed performed PDDI CDS as a

[6] https://www.getpostman.com

[7] https://sandbox.cds-hooks.org

Figure 4– The screenshot of CDS Hooks Sandbox tool to test our PDDI-CDS services through CDS Developer Panel

service for two patient use cases. One of the limitations is a lack of tools supported for CQL because the community who is using CQL and CQL itself is entirely new. Therefore, it takes a lot of time in debugging to find out the problem.

As described in the Concept section, we need two components including FHIR server and CQL engine to implement CDS services. There are many FHIR server implementations including HAPI FHIR, FHIR .NET API[8] and others. However, HAPI FHIR is the most stable and popular implementation. It has a well-written documentation and full FHIR version support. Apart from that, there are less choices for CQL engine and we chose CQL-Evaluation-Engine from Database Consulting Group. This implementation is based on CQF-Ruler open source developed by Database Consulting Group, and Level 1 implementation of the draft PDDI-CDS Implementation Guide developed by HL7 Clinical Decision Support Work Group.

The Level 1 Implementation uses a single CDS service call and response with the medication-prescribe hook. As a future work, we plan to implement Level 2 implementation which we are working with the CDS Hooks developers to define the new medication-select hook and clarify how DetectedIssues will be returned in the card responses.

To create the CQL artifacts, the CDS Authoring Tool[9] is a promising candidate which is a component of the CDS Connect project funded by the Agency for Healthcare Research and Quality (AHRQ). The CDS Authoring Tool introduces an interface for composing CDS logic step by step using simple forms and exporting it as CQL artifacts using the HL7 FHIR DSTU 2 data model.

Conclusions

Based on the PDDI-CDS IG, we successfully implemented PDDI CDS services supported the Level 1 implementation of PDDI-CDS IG. The implementation follows strictly the CDS Hooks specification which enables EHR system to interoperate and exchange the data by using FHIR STU3 standard to enhance the better clinical diagnostic decision making. In the future, we will focus on the Level 2 implementation to help advance the standards including the new mediation-select.

Acknowledgments

This work funded in part by grants from the United States Agency for Healthcare Research and Quality R01 LM011838, R21 HS023826 and R01 HS025984; and National Library of Medicine grants T15 LM007124 and R01LM011838.

References

[1] OBO Technical WG, Drug-drug Interaction and Drug-drug Interaction Evidence Ontology, (2018). http://www.obofoundry.org/ontology/dideo.html (accessed Nov 10, 2018).

[2] H. van der Sijs, J. Aarts, and et al., Turning Off Frequently Overridden Drug Alerts: Limited Opportunities for Doing It Safely, *J Am Med Inform Assoc*, **15** (2008), 439-448.

[3] R.T. Scheife, L.E. Hines, and et al., Consensus recommendations for systematic evaluation of drug-drug interaction evidence for clinical decision support, *Drug Saf,* **38** (2015), 197-206.

[4] T.H. Payne, L.E. Hines, and et al., Recommendations to improve the usability of drug-drug interaction clinical decision support alerts, *J Am Med Inform Assoc* **22** (2015), 1243-50.

[5] J.I. Wolfstadt, J.H. Gurwitz, and et al., The Effect of Computerized Physician Order Entry with Clinical Decision Support on the Rates of Adverse Drug Events: A Systematic Review, *J Gen Intern Med* **23** (2008), 451–458.

[6] A.M. Sirajuddin, J.A. Osheroff, et al., Implementation pearls from a new guidebook on improving medication use and outcomes with clinical decision support. Effective CDS is essential for addressing healthcare performance improvement imperatives, *J Healthc Inf Manag* **23** (2009), 38-45.

[7] K.W. Fung, J. Kapusnik-Uner, and et al., Comparison of three commercial knowledge bases for detection of drug-drug interactions in clinical decision support, *J Am Med Inform Assoc* **24** (2017), 806-812.

[8] M.Z. Herrero, I.B. Segura, and P. Martínez, Conceptual models of drug-drug interactions: A summary of recent efforts, *Knowledge-Based Systems* **114** (2016), 99-107.

[9] W3C Group, HCLS Drug-Drug Interaction, 2018. https://github.com/w3c/hcls-drug-drug-interaction/ (accessed Nov 10, 2018).

[10] M. Brochhausen, M. Herrero-Zazo, and et al., MPIO - A novel Minimum Potential drug-drug interaction Information Ontology implemented in OWL, *AMIA Informatics Summit*, San Francisco, CA. USA, 2018.

[11] HL7, HL7 IG: Decision Support Service, Release 1, (2018). http://www.hl7.org/implement/standards/product_brief.cfm?product_id=334 (accessed Nov 10, 2018).

[12] HL7, PDDI CDS Implementation Guide, (2018). http://hl7.org/fhir/uv/pddi/2018Sep (accessed Nov 10, 2018).

Address for correspondence
Oya Beyan, beyan@fit.fraunhofer.de.

[8] https://github.com/FirelyTeam/fhir-net-api

[9] https://cds.ahrq.gov/authoring/

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190319

A Decision Support System for Pathology Test Result Reviews in an Emergency Department to Support Patient Safety and Increase Efficiency

Anthony Nguyen[a], Hamed Hassanzadeh[a], Yushi Zhang[a], John O'Dwyer[a], David Conlan[a], Michael Lawley[a], Jim Steel[a], Kylynn Loi[a], Peter Rizzo[b]

[a] *The Australian e-Health Research Centre, CSIRO, Brisbane, Queensland, Australia,*
[b] *The Prince Charles Hospital, Department of Health, Brisbane, Queensland, Australia*

Abstract

The review of pathology test results for missed diagnoses in Emergency Departments is time-consuming, laborious, and can be inaccurate. An automated solution, with text mining and clinical terminology semantic capabilities, was developed to provide clinical decision support. The system focused on the review of microbiology test results that contained information on culture strains and their antibiotic sensitivities, both of which can have a significant impact on ongoing patient safety and clinical care. The system was highly effective at identifying abnormal test results, reducing the number of test results for review by 92%. Furthermore, the system reconciled antibiotic sensitivities with documented antibiotic prescriptions in discharge summaries to identify patient follow-ups with a 91% F-measure – allowing for the accurate prioritization of cases for review. The system dramatically increases accuracy, efficiency, and supports patient safety by ensuring important diagnoses are recognized and correct antibiotics are prescribed.

Keywords:

Clinical Decision Support Systems, Data Mining, Emergency Medicine

Introduction

The failure to follow-up on pending test results when patients have been discharged from hospitals is a potential threat to patient safety [1][3]. Poor test result follow-up can significantly impact patient safety and clinical care, including missed diagnoses and suboptimal patient outcomes [3]. The Emergency Care Research Institute (ECRI) underscored the problem of "Test result reporting and follow-up" as one of the top 10 patient safety concerns in healthcare organizations [4].

This is especially problematic in Emergency Departments (EDs) where there can be a significant number of pending test results when patients have been discharged from the hospital. The current manual process in EDs for checking test results for abnormalities, then ensuring patients with abnormal results are appropriately followed-up is sub-optimal [5][6]. The process is labor-intensive, un-prioritized by its nature, and has negative impacts on patient care and staff workloads. Specifically, the current process 1) potentially delays the follow-up of critical abnormal cases because all test results need to be reviewed then reconciled against the patient's disposition recorded in the ED; 2) consumes valuable clinical hours that could be better spent engaged in direct patient care; and 3) is prone to errors due to the time intensive nature of work trying to find the "needle in the haystack" of results, in addition to ongoing pressures of clinician workloads.

Furthermore, the process of checking test results in EDs has not kept pace with the rapid growth and expansion of hospital services due to increasing population, aging, and chronic diseases. This has placed increased pressures on "systems" and staff workloads, leading to reduced efficiencies and reduced direct clinical contact hours with patients potentially compromising patient safety and care.

Proposed is an automated solution that changes the way EDs detect and report abnormal test results. The system initially focused on the review of microbiology test results as these can take up to several days to be reported upon, by which time a patient will have been discharged from the ED. These test results often contain information on the results of culture strains present and their antibiotic sensitivities, both of which make a significant impact on ongoing patient safety and clinical care.

In particular, contributions include 1) 'trigger' algorithm to identify abnormal microbiology test results and related antibiotic sensitivities from pathology reports; 2) extraction of antibiotic prescriptions documented in ED discharge summaries using text mining based 'trigger' keywords derived from clinical terminology semantics based on SNOMED CT[1] to reconcile against the antibiotic sensitivities; and 3) automate a protocol with a prioritized listing of cases to support the test result follow-up checking process. The solution overcomes the labor-intensive and error-prone nature of test result reviews to enhance patient safety and efficiency.

Background

Clinical decision support (CDS) systems have the potential to improve clinical efficiency and patient safety [7][11]. These systems have been applied across diverse settings and range from information management systems through to clinical alerts, and diagnosis and/or treatment recommendations. Despite the significant body of literature on CDS systems, automation in the context of test result reviewing has been limited.

Health system processes and test result management systems have been implemented with the aim to improve test result review processes [3]. However, physicians are still unsatisfied with how they manage test results [5][6]. For example, an end-to-end workflow is much desired along with the ability to filter normal (irrelevant) results to help prioritize the workflow. This

[1] Systematized nomenclature of medicine – clinical terms

suggests that current systems do not include robust CDS features or functions such as prioritization and/or filtering to the end user [8].

Automated notifications of abnormal test results can significantly reduce the number of results for review. Abnormality detection from pathology and imaging test results range from 'trigger' algorithms that use clinical logic on structured electronic health record (EHR) data to identify abnormalities [12][15] through to advanced computational approaches such as machine learning to identify abnormalities from unstructured narrative EHR data [16][18].

Furthermore, the abnormality task can be extended to also reconcile or link the abnormal findings in radiology reports with the patient's disposition recorded in ED information system to provide decision support to the manual review process [16].

Here, the proposed solution for microbiology test result reviews leverages previous work on 'trigger' algorithms and reconciliation to develop an end-to-end test result reconciliation solution. Microbiology test results most frequently report on bacterial antimicrobial (or antibiotic) susceptibility. Antibiotics have been used to treat infections or diseases caused by bacteria and have saved many lives since their introduction. However, antibiotics may not be effective when overused due to antimicrobial resistance. This "antimicrobial stewardship" problem is a global problem and like "Test result reporting and follow-up," it was also identified as being within the top 10 patient safety concerns in healthcare organizations [4]. This end-to-end solution can aid in ensuring that appropriate antibiotic prescription continues through the treatment of the infection thus fulfilling the needs for antimicrobial stewardship.

Methods

The workflow for identification and reconciliation of abnormal microbiology test results with ED discharge summaries will be presented along with the proposed methodology for automating it. The workflow was adapted from an actual test result review process within an ED in Brisbane, Australia.

Data

The dataset was obtained from The Prince Charles Hospital (TPCH), Brisbane, Australia[2]. The dataset comprised 31,787 ED encounters (pertaining to 15,916 unique patients) from July 2013 to December 2014 (18-month period). A separate dataset of microbiology test results, ordered from the same hospital and ED, was obtained from a state-wide pathology information system and comprised 29,503 microbiology pathology HL7 messages (pertaining to 18,560 patients) from a wider 4-year time period. The ED encounters and microbiology test results were matched based on their unique patient identifier and test result order date is within the patient's ED admission and discharge date/time. A total of 16,867 ED encounters had matching microbiology test results over the dataset time period.

A subset of abnormal test results and matching ED discharge summaries (142 cases) was manually reviewed by ED senior medical officers to determine their follow-up requirement. Eighty percent of the cases (113 cases) were used for system development while the remaining cases were used for testing (29 cases). An additional 282 cases were subsequently obtained to assess the generalisability and robustness of the system via a pilot study. Table 1 presents a summary of the gold standard dataset.

Table 1 – Gold standard dataset for abnormal microbiology test result reconciliation

Dataset	Follow-up	No follow-up
Development set	78 (69%)	35 (31%)
Test set	20 (69%)	9 (31%)
Pilot set	211 (75%)	71 (25%)

Abnormal test result identification and filtering

The TPCH ED, on a daily basis (including weekends), would print out the microbiology test results and place them in a dedicated area for sorting. Results must be sorted by hand to identify abnormal results requiring follow-up from those that do not. This process was not without errors. A typical day may find the sorting pile overlooked or half completed due to clinical demands. Other errors include failure to recognize abnormal results from distraction or incorrect interpretation of results.

The proposed system addressed the abnormal test result identification issue by utilizing the concept of the trigger algorithm [12][15]. Triggers were applied to the microbiology test results to identify abnormal results and related antibiotic sensitivities. The presence or absence of antibiotic sensitivity results in pathology HL7 messages was used as the trigger.

The set of abnormal test results could be further filtered based on the patient's discharge destination. Patients who were admitted into the hospital as an in-patient would not be required to be followed-up as they would be appropriately followed-up by other clinicians in the hospital. To achieve this, the system again applied triggers to the discharge destination field in the ED information system to identify patients who were not admitted to any of the hospital wards. A list of all possible ED discharge destinations with their follow-up requirement was provided for the discharge destination filtering.

The processing and filtering of pathology reports resulted in a significant reduction in the number of test results requiring review by the ED.

Abnormal test result reconciliation

The next stage of the review process was abnormal test result reconciliation whereby abnormal test results must be correlated against the clinical record. The ED clinician searches for patients with abnormal test results, one by one, in their ED medical record to determine if patients were required to be followed-up due to an inappropriate diagnosis or treatment. This correlation process was also not without errors. For example, appropriate action may not have been taken due to failure to recognize the misdiagnosis or incorrect treatment.

This stage involved the application of text mining and clinical terminology semantics for the extraction of antibiotic prescriptions documented in ED discharge summaries for reconciling against antibiotic sensitivities in pathology test results. To extract antibiotic prescriptions in discharge summaries, a 'trigger' keyword list containing antibiotic names was compiled. This was used to match occurrences of these keywords in discharge summaries.

A baseline list of antibiotic 'trigger' keywords was derived from the list of possible antibiotic sensitivities identified in microbiology test results (e.g. Trimethoprim, Co-trimoxazole,

[2] Research ethics was obtained from the Metro North Hospital and Health Services Human Research Ethics Committee.

and Di(Flu)cloxacillin). Antibiotic names and their expanded forms (e.g. dicloxacillin and flucloxacillin for Di(Flu)cloxacillin) would form candidate antibiotics to use as 'trigger' keywords.

An extended 'trigger' keyword list was compiled by supplementing the antibiotic names with also their brand names. Here, the Australian Medicines Terminology (AMT)[3] and the SNOMED CT expression constraint language (ECL)[4] was used to generate the list of trade names. Ontoserver [19], a clinical terminology server with support for SNOMED CT and AMT as well as SNOMED CT's ECL, was used to generate the list of antibiotic trade names.

ECL templates were devised to return a list of trade names given 1) a single active antibiotic ingredient, and 2) more than one active antibiotic ingredients. An example ECL for a single antibiotic ingredient `2691011000036102|Trimethoprim|` is as follows:

```
^ 929360021000036102|Trade product reference set|
AND >> (
  ( ^ 929360031000036100|Trade product unit of use
reference set| : (
      700000081000036101|has intended active
ingredient| = 2691011000036102|Trimethoprim|
      ) AND [1..1] 700000081000036101|has intended
active ingredient| = * )
)
```

Ontoserver provides an ECL high-level reference[5] to aid in the interpretation of the above expression. In brief, the ECL returns a list of trade names (from `Trade product reference set`) that contains medications (from `Trade product unit of use reference set`) that have the specified antibiotic as its active ingredient. The `[1..1]` constrains the results to medications with only a single ingredient.

If multiple antibiotic ingredients were applicable to a certain antibiotic sensitivity test, then the above ECL can be adapted to include additional antibiotics using the `OR` operator. An example ECL extract for 'Di(Flu)cloxacillin' would be as follows:

```
...

700000081000036101|has intended active ingredient| =
  ( 2018011000036104|dicloxacillin| OR
    2115011000036102|flucloxacillin| )

...
```

For cases where an antibiotic has more than one active ingredient such as 'Co-trimoxazole' which contains both `2605011000036103|sulfamethoxazole|` and `2691011000036102|trimethoprim|`, then the following ECL was applied:

```
^ 929360021000036102|Trade product reference set|
AND >> (
  ( ^ 929360031000036100|Trade product unit of use
reference set| : (
      700000081000036101|has intended active
ingredient| = 2605011000036103|sulfamethoxazole|,
      700000081000036101|has intended active
ingredient| = 2691011000036102|trimethoprim|) )
)
```

The resulting list of antibiotic 'trigger' keywords was used to extract antibiotics documented in discharge summaries.

Discharge summaries often document the antibiotic treatments administered during the ED encounter. Intravenous (IV) medications for certain antibiotics can only be administered as IV-only. These IV-only antibiotics are generally not the full course of antibiotics and thus not relevant for reconciling against antibiotic sensitivities in test results. A list of all IV-only antibiotics was provided for antibiotic filtering. The resultant non-IV-only antibiotics would be used for the next stage of reconciliation.

The reconciliation of antibiotic prescriptions extracted from discharge summaries against antibiotic sensitivities from test results applied the following rule-based logic. The patient would not be followed-up if they were prescribed with an antibiotic that had a corresponding antibiotic sensitivity in their test result of 'sensitive'. However, the patient would require follow-up if any of the following reconciliation events occur:

- The patient has not been prescribed any antibiotics.
- The prescribed antibiotic was not tested (and thus its antibiotic sensitivity was unknown).
- The prescribed antibiotic resulted in an antibiotic sensitivity of 'resistant.'

The logic would be applied to each culture strain (or bacterial organism) identified in the test result.

Evaluation Measures

The efficiency of the system was evaluated based on the resultant number of test results identified by the system for clinical review (normalized on a weekly basis) compared to the full set of test results that would have been manually reviewed.

The reconciliation phase to determine whether or not a patient required follow-up was evaluated against the gold standard. The effectiveness of the system was measured using sensitivity (or recall) and positive predictive value (PPV or precision). To provide a single, overall evaluation measure, precision and recall were combined into a third evaluation measure, F-measure.

Results

A total of 16,867 ED encounters had matching microbiology test results over the dataset time period. This averages to ~216 test results per week, which an ED clinician would need to sort and review the test results.

Abnormal test result identification and filtering

The efficiency of the system in filtering irrelevant test results is tabulated in Table 2.

The proposed trigger to filter normal (irrelevant) test results from abnormal test results resulted in 2,605 abnormal test results – an average of 33 reports per week requiring review. The trigger was confirmed by ED clinicians to be accurate in identifying abnormal from normal test results.

[3] AMT is a subset of SNOMED CT-AU (Australian extension) for medicines commonly used in Australia.

[4] ECL is a formal language for defining bounded sets of clinical meanings represented by pre-coordinated or post-coordinated expressions.

[5] https://ontoserver.csiro.au/shrimp/ecl_help.html

Table 2 – System efficiency results in terms of reducing the number of abnormal test results review

	Number of test results	Weekly number of test results
Full manual review	16,867	216
+ Abnormality identification	2,605	33 (↓ 85%)
+ Discharge destination filtering	1,379	18 (↓ 92%)

When the abnormal test results were filtered based on the ED discharge destination, the number of test results that actually required clinical review was reduced to 1,379 (18 reports per week).

Abnormal test result reconciliation

To further provide clinical decision support to the test result review process, the reconciliation of antibiotic prescriptions extracted from discharge summaries against antibiotic sensitivities in test results allows for the prioritization of cases for clinical review. Table 3 presents the classification effectiveness of the proposed reconciliation approach.

Table 3 – System effectiveness results in classifying the follow-up requirement

Dataset	PPV	Sensitivity	F-measure
Development set	0.802	0.936	0.869
Test set	0.905	0.950	0.900
Pilot set	0.858	0.943	0.898

Discussion

The automated identification and prioritization of abnormal microbiological pathology reports for clinical review will bring about benefits in cost efficiencies, quality of care, and performance.

Noteworthy, was that a simple solution based on triggers was able to substantially reduce the number test results for review. Results show a 92% reduction in microbiology test results that required review.

Furthermore, the system was able to accurately identify the follow-up requirements (F-measure of 90%) for the prioritization of cases for clinical review. The very limited differences in performances between the development, test and pilot dataset show the generalisability and robustness of the proposed antibiotic extraction approach using 'trigger' keywords and a rule-based reconciliation logic.

Error analysis on the development dataset revealed four error categories: 1) antibiotic misspellings in discharge summaries (4 cases); 2) context of antibiotic mentions (3 cases; e.g., previous prescriptions and non-IV-only administered antibiotics); 3) extrapolation of prescribed antibiotics with sensitivity results (4 cases; e.g., cephalexin prescription considered a correct prescription for an organism sensitive to Cefazolin test result); 4) missed identification of antibiotic mentions in discharge summaries (1 case); and 5) gold standard inaccuracies confirmed to be human errors (7 cases). These error categories form avenues for future work to improve system performances.

Conclusions

The proposed IT solution combines text mining and decision support technologies for the novel identification and reconciliation of abnormal microbiology test results in EDs. It has the potential to dramatically increase the accuracy and efficiency of microbiology test result review to support patient safety by ensuring important diagnoses are recognized and correct antibiotics have been prescribed. The increased efficiency will allow significantly more clinical hours devoted to the direct treatment of patients presenting to hospital EDs to increase their quality of care.

Next steps include the planning and development of end-user software that would allow for appropriate presentation without impeding on clinical workflow [7][20]. The clinical decision support system would then be implemented and trialed by clinicians in actual practice.

The proposed 'triggers' and reconciliation logic is applicable to any software system using HL7 communications. This would allow changes to be made among different pathology systems or EHRs to provide the notification and reconciliation of abnormal test results. It can also be applied to other areas of pathology as well as the reporting of radiology test results, which are similar in processes.

Acknowledgements

We thank Dr. Mike Watson of TPCH ED for assisting with the initial project management and data extraction; Sally Taylor of TPCH Pharmacy Department, Dr Robert Horvath, Dr Alex Chaudhuri and Dr John McNamara of TPCH Infectious Disease Department for their domain expertise and advice; and Doreen Lelieveld for assisting Dr Peter Rizzo with the review of microbiology test results and discharge summaries for patient follow-ups.

References

[1] Roy CL, Poon EG, Karson AS, Ladak-Merchant Z, Johnson RE, Maviglia SM, Gandhi TK. Patient safety concerns arising from test results that return after hospital discharge. Ann Intern Med. 2005 Jul 19;143(2):121-8.

[2] Walz SE, Smith M, Cox E, Sattin J, Kind AJ. Pending laboratory tests and the hospital discharge summary in patients discharged to sub-acute care. J Gen Intern Med. 2011 Apr 1;26(4):393-8.

[3] Callen J, Georgiou A, Li J, Westbrook JI. The impact for patient outcomes of failure to follow up on test results. How can we do better? EJIFCC. 2015 Jan;26(1):38.

[4] ECRI Institute. Top 10 patient safety concerns for healthcare organizations. 2017.

[5] Poon EG, Gandhi TK, Sequist TD, Murff HJ, Karson AS, Bates DW. "I wish I had seen this test result earlier!": dissatisfaction with test result management systems in primary care. Arch Intern Med. 2004 Nov 8;164(20):2223-8.

[6] Dalal AK, Poon EG, Karson AS, Gandhi TK, Roy CL. Lessons learned from implementation of a computerized application for pending tests at hospital discharge. J Hosp Med. 2011 Jan;6(1):16-21.

[7] Musen MA, Middleton B, Greenes RA. Clinical decision-support systems. In Biomedical informatics 2014 (pp. 643-674). Springer, London.

[8] Sittig DF, Wright A, Osheroff JA, Middleton B, Teich JM, Ash JS, Campbell E, Bates DW. Grand challenges in clinical decision support. J Biomed Inf. 2008 Apr 1;41(2):387-92.

[9] Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, Samsa G, Hasselblad V, Williams JW, Musty MD, Wing L. Effect of clinical decision-support systems: a systematic review. Ann. Intern. Med. 2012 Jul 3;157(1):29-43.

[10] Miller A, Moon B, Anders S, Walden R, Brown S, Montella D. Integrating computerized clinical decision support systems into clinical work: a meta-synthesis of qualitative research. Int J Med Inform. 2015 Dec 1;84(12):1009-18.

[11] Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? J Biomed Inf. 2009 Oct 1;42(5):760-72.

[12] Rosen AK, Mull HJ, Kaafarani H, Nebeker J, Shimada S, Helwig A, Nordberg B, Long B, Savitz LA, Shanahan CW, Itani K. Applying trigger tools to detect adverse events associated with outpatient surgery. J Patient Saf. 2011 Mar 1;7(1):45-59.

[13] Murphy DR, Laxmisan A, Reis BA, Thomas EJ, Esquivel A, Forjuoh SN, Parikh R, Khan MM, Singh H. Electronic health record-based triggers to detect potential delays in cancer diagnosis. BMJ Qual Saf. 2014 Jan 1;23(1):8-16.

[14] Murphy DR, Meyer AN, Bhise V, Russo E, Sittig DF, Wei L, Wu L, Singh H. Computerized triggers of big data to detect delays in follow-up of chest imaging results. Chest. 2016 Sep 1;150(3):613-20.

[15] Singh H, Thomas EJ, Sittig DF, Wilson L, Espadas D, Khan MM, Petersen LA. Notification of abnormal lab test results in an electronic medical record: do any safety concerns remain? Am J Med. 2010 Mar 1;123(3):238-44.

[16] Koopman B, Zuccon G, Wagholikar A, Chu K, O'Dwyer J, Nguyen A, Keijzers G. Automated reconciliation of radiology reports and discharge summaries. AMIA 2015 (Vol. 2015, pp. 775-784).

[17] Hassanzadeh H, Nguyen A, Karimi S, Chu K. Transferability of artificial neural networks for clinical document classification across hospitals: A case study on abnormality detection from radiology reports. J Biomed Inform. 2018 Sep 1;85:68-79.

[18] Hassanzadeh H, Kholghi M, Nguyen A, Chu K. Clinical Document Classification Using Labeled and Unlabeled Data Across Hospitals. AMIA 2018 (Vol. 2018, pp. 545-554.

[19] Metke-Jimenez A, Steel J, Hansen D, Lawley M. Ontoserver: a syndicated terminology server. J Biomed Semantics. 2018 Dec;9(1):24.

[20] Horsky J, Schiff GD, Johnston D, Mercincavage L, Bell D, Middleton B. Interface design principles for usable decision support: a targeted review of best practices for clinical prescribing interventions. J Biomed Inf. 2012 Dec 1;45(6):1202-16.

Address for correspondence

Anthony Nguyen, Email: Anthony.Nguyen@csiro.au.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190320

A Model Driven Approach to the Design of a Gamified e-Learning System for Clinical Guidelines

Job N. Nyameino[a,b], Fazle Rabbi[c], Khalid A. Mughal[a], Martin C. Were[d,e,b], Yngve Lamo[c]

[a] *Department of Informatics, University of Bergen, Bergen, Norway,*
[b] *Institute of Biomedical Informatics, College of Health Sciences, Moi University, Eldoret, Kenya*
[c] *Department of Computing, Mathematics, and Physics, Western Norway University of Applied Sciences, Bergen Norway.*
[d] *Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA*
[e] *Vanderbilt Institute for Global Health, Vanderbilt University Medical Center, Nashville, TN, USA*

Abstract

Clinical practice guidelines (CPGs) are indispensable in the practice of evidence-based medicine. However, the cost of effective CPG dissemination strategies is prohibitive and not cost-effective. Therefore, scalable strategies using available technology are needed. We describe a formal model-driven approach to design a gamified e-learning system for clinical guidelines. We employ gamification to increase user motivation and engagement in the training of guideline content. Our approach involves the use of models for different aspects of the system, an entity model for the clinical domain, a workflow model for the clinical processes and a game model to manage the training sessions. A game engine instantiates a training session by coupling the workflow and entity models to automatically generate questions based on the data in the model instances. Our approach is flexible and adaptive as it allows for easy updates of the guidelines, integration with different device interfaces and representation of any guideline.

Keywords:

Practice Guideline; Statistical Model; Computer Games

Introduction

Over the past three decades, the concept of evidence-based medicine has become the pre-eminent paradigm in informing clinical decision making. Evidence-based medicine is the conscientious, explicit, and judicious use of best evidence in making decisions about the care of individual patients [1]. The volume of published medical literature is immense and it is practically impossible for a clinician to read and appraise all the available evidence. This has led to efforts to systematically appraise the literature and provide summaries of the evidence in the form of clinical guidelines that can be more easily consumed by healthcare workers.

Clinical guideline are statements that include recommendations intended to optimize patient care. They are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options [2]. Guidelines serve a variety of functions such as improving effectiveness and quality of care, decreasing variations in clinical practice, and decreasing costly and preventable mistakes and adverse events [3].

Development of clinical guidelines is not by itself enough. For a health system to derive value from available guidelines, effective dissemination and implementation strategies have to be employed to facilitate the integration of the guideline recommendations in clinical practice. Dissemination refers to the methods by which guidelines are made available to potential users, while implementation refers to ensuring that the users act on the recommendations [4]. Currently, there is no consensus on the most effective dissemination strategies although the use of a multifaceted dissemination approach has been found to be more effective than single interventions [5,6]. Further, the passive distribution of printed educational material on care guidelines was found to be less effective than active approaches such as educational outreaches [7].

Despite the observed effectiveness of educational outreaches as a dissemination strategy, these often require significant financial and human resource commitment that might not be readily available in low resource settings. For example, the Emergency, Triage, Assessment and Treatment plus Admission care (ETAT+) course is a 5.5 day course developed in Kenya to train healthcare workers on the guidelines for emergency paediatric care. During its initial implementation, the cost of the whole training for 32 participants using 5 facilitators cost $5000 for facilitators and meals, not to mention the cost of having providers away from work and the often associated additional costs for trainings provided at a conference facility [8]. This is neither scalable nor sustainable.

To address these dissemination challenges, new, innovative and cost-effective strategies that leverage contemporary technologies need to be employed [9]. A study conducted in Kenya showed that ownership of mobile phones among healthcare workers was nearly universal with 98.6% of respondents reporting mobile phone ownership and 75% reporting using their mobile phones to access work-related information [10]. This provides an opportunity to use the mobile electronic platforms to conduct educational outreach in a scalable, cost effective and automated fashion.

In this paper, we describe a formal model driven approach to the design of a gamified elearning system for clinical guidelines. Gamification is the use of game design elements in non-game contexts [11]. It uses game-based mechanics, aesthetics and thinking to engage people, motivate action, promote learning and solve problems [12]. The term "Gamification" is relatively new and has been used to describe the use of game-based concepts and techniques outside recreational activities, with the goal of increasing the motivation and engagement of the participants and improving the results. The benefits of gamification are an increase in motivation and engagement, which can be applied in education or work-related contexts.

To be effective, gamification of guidelines would require flexibility and adaptiveness as the users have various learning styles. In this paper, we describe an innovative model-driven approach for dissemination of a gamified eLearning system for clinical guidelines, and the resultant prototype of this system.

Methods

Model Driven Engineering

To design our gamified training system, we aim to use a model driven engineering (MDE) approach. MDE is a system development paradigm that promotes the use of models as the primary artefacts that drives the whole development process [13]. An MDE approach provides several advantages that we leveraged for our purposes. First, it improves communication by exploiting abstraction and domain-specificity which can target different audiences. Secondly, MDE facilitates separation of the specification of system functionality from the specification of the implementation of that functionality on a specific technology platform. This is especially significant since technology platforms are in a constant state of flux, changing frequently in response to business needs and technological development. By adopting MDE, the business logic of the system and its application technologies can evolve independently of each other. Implementations execution is mad easier by abstraction on different platforms while maintaining the structure and behaviour of the system.

In MDE, models are specified using modelling languages. A modelling language is defined by a metamodel (a model of models) and is a set of all possible models that conform to this metamodel [14]. There are two types of modelling languages, namely: (1) General Purpose Modelling Languages (GPLs) and (2) Domain-Specific Modelling Languages (DSLs). GPLs, as the name suggests, tend to be general and with poor support for domain-specific notation. Conversely, DSLs are tailored to a specific application domain that offers appropriate notations and abstractions. DSLs are thus more expressive and easier to use with concomitant gains in productivity and maintenance costs [14]. The abstract syntax of the DSL is a conceptualization of all the concepts, abstractions and relations underlying the domain represented as a model which acts as the metamodel of the modelling language.

A metamodel architecture introduces a generic pattern of metamodelling hierarchy in which models at each level are specified by a modelling language at the level above it and conform to the corresponding metamodel of the language. Figure 1 illustrates a metamodelling hierarchy where a model M_i at a certain level *i* (e.g. M_0, M_1 etc) conforms to a metamodel M_{i+1} at the level above until a model M_j has itself as metamodel, called a reflexive model.

However there are several aspects to design in a software model which requires the co-ordination of multiple models. These aspects include the design of information systems that have several sub-systems that communicate with each other. Rabbi et al. proposed to use an integration of multiple metamodeling hierarchies to co-ordinate various aspects of a system [15]. Our approach is also based on the idea of coordination of multi-metamodeling hierarchy where we integrate entity models, workflow models and models that represents the gaming aspects of a training program. Below, we describe each of these models and their integration within our model.

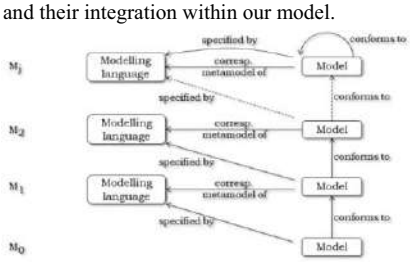


Figure 1:Generic pattern: modelling languages and metamodels [16]

Entity Model

In order to conceptualize the clinical encounters, we define an entity model relevant to the medical domain. This involves identification of the significant concepts in the clinical encounter, with a clear delineation of their attributes and relations (Figure 2). The figure shows a constraint [preCondition] implied over the relations on patient, diagnosis, and treatment. The semantic of the constraint specifies that "for every treatment, patients' diagnosis needs to be confirmed".

This clinically-relevant entity model forms the data model from which the game engine populates instances in the workflow model (see below).

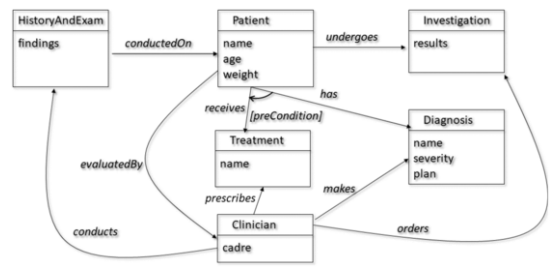


Figure 2: A simplified entity model of the clinical encounter domain

Workflow Model

Once the entity model is defined, we define the workflow model. Clinical practice guidelines can run into tens or hundreds of pages and are usually summarized in algorithmic flow charts that show the process of treatment for a given scenario. Our approach leverages the flowchart structure of the guideline summaries to design our workflow model.

Figure 3 demonstrates the use of the metamodelling hierarchy to depict the flow of tasks during a clinical encounter using an use case of asthma. At level M_2, we have a meta-metamodel showing the abstraction of how tasks flow from one to the next. Level M_1 shows a generic abstraction of the process of a clinical encounter from the initial assessment, diagnosis, treatment and evaluation of the treatment, after which the treatment can be repeated or the patient reassessed afresh to review the diagnosis. Finally, at level M_0, we show an instance of the recommended flow of tasks when treating a child with asthma.

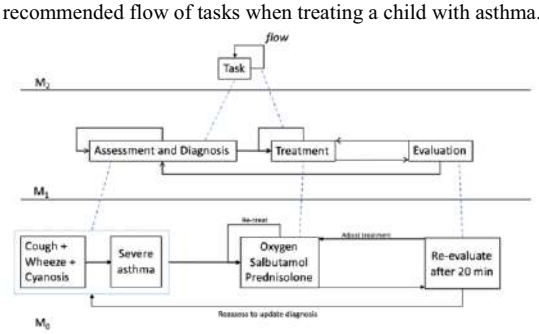


Figure 3: Workflow model showing the flow of tasks during a clinical encounter

Game Model

The design of gamified e-learning systems should be undertaken in view of the core concepts of games i.e. goal-oriented activities with reward mechanisms and progress tracking [17]. To become familiar with guideline content,

learners need to know how to treat the different aspects of a disease condition as outlined within the guideline. Game models use reward mechanisms and progress tracking aids in order to increase the users engagement and motivation.

The game engine in our model automatically generates questions from the entity and workflow models to instantiate a training module. The questions are categorized according to the learner's skill level (beginner, intermediate, advanced). The model also specifies a *learner profile* that tracks the learner's activities. Figure 4 shows the model for the game engine.

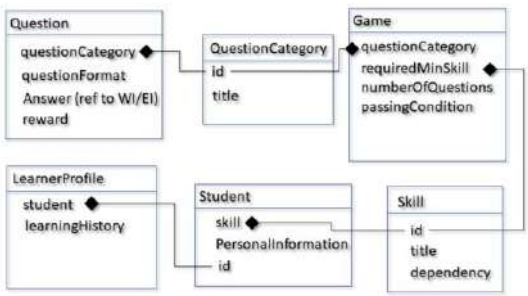


Figure 4: Game model

Integrated Training Model

The training model is built by coupling and coordinating the entity and workflow models discussed above based on the principles introduced by Rabbi et al [15]. The states of the training module (*TM*) are defined by a set of elements that include a pair of workflow instance (*WI*) and an entity instance (*EI*): $TM_i = <EI_i, WI_i>$. This coupling of models is illustrated in Figure 5. Here, we show a part of the entity model with values from a given scenario where based on the *History & Examination* findings, a *Diagnosis* of Severe Asthma is made and its *Treatment* specified. The flow of how this process should happen is shown in the workflow model.

The game engine instantiates a training session by generating questions based on the entity model and workflow model. For example, it could initially generate a scenario based on the patient details and history and examination findings and ask what the diagnosis is. If answered correctly, it will move on to the next task and ask about the treatment. A training session is composed of a sequence of training modules and is evolved from the initial state of a training flow and progresses based on the answer provided by the user.

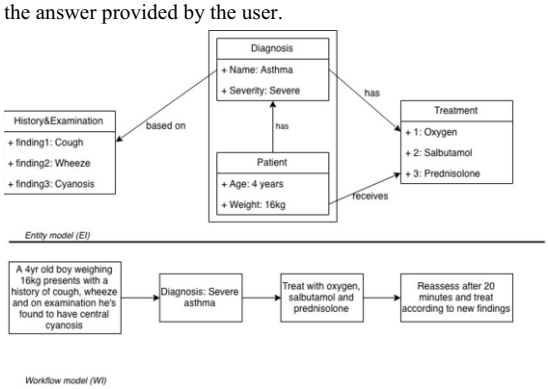


Figure 5: Coupled entity and workflow models

Figure 6 illustrates the idea of the progression of the states of training session. Depending on the answer given by the user a game engine consults with the training flow and evolves the state of the training session. We use the diagram predicate approach for representing the status of our training modules and their transformation [18]. Two annotations **< Enabled >** & **< Disabled >** are used to represent the current status of the training modules. A training module TM_0 when annotated with the **< Enabled >** predicate indicates that the training module is currently active and is being considered for training. Once TM_0 is completed successfully, it is disabled and TM_1 is enabled. This continues until the training session is over. If a wrong answer is given the training module is repeated with hints until the correct answer is given.

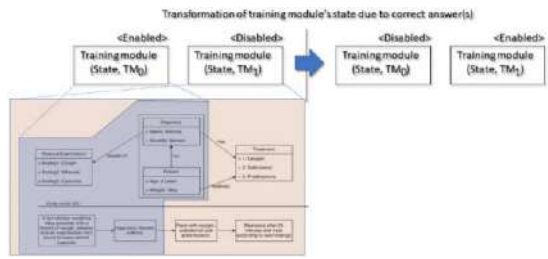


Figure 6: Progression of the states of the training session

Model Based Analysis

Since our approach is based on models, there are lot of opportunities to make sure the guideline scenarios being represented by the models are correct. We use model transformation rules to control the flow of training sessions. Transformation rules are often used in an MDE approach to encode the knowledge of a transformation that can be analyzed. This allows us to perform reasoning over the models. This means anylizing the models for validity, semantic consistencies and inconsistencies. For example we can perform reasoning over the models to answer the following questions:

- Are there sufficient number of questions to run a training session?
- Classify the questions over different stages e.g., identify questions that belong to the 'Diagnosis and Assessment' stage of a guideline training.
- Can we use a computer system such as that one by Alloy to randomly produce a scenario and generate questions [19]? How can we make sure that the scenarios are valid instance of a guideline?
- If there are certain number of questions to be asked in a session, what informaiton are common among the questions?
- How can we find out an order of the questions that complies with the order of the guideline?

We have applied DPF constraint checking [18] to find answer to some of the reasoning questions mentioned above. The fundamental idea of DPF constraint checking is based on the principle of category theory. Several techniques have been studied in [20] to produce valid DPF instances that satisfies the constraints in the metamodel. However, the technique has not been applied for multilevel metamodeling. In future, we will enhance the technique presented in [20] and use Alloy to randomly produce valid guideline scenarios.

Modular Architecture

The vision of this approach is to develop general components that can be utilized in the development of training programs for different guidelines. Figure 7 shows the overall architecture of our system. The training management module contains reusable elements that control the flow of questions in a training session

and keeps usage logs. The entity models are the elements that should differ from one training program to another, however their metamodel should remain the same. The game engine components of the architecture is responsible for integrating with other systems. As an example, in this architecture, we show how the game engine is being interfaced with the dialogflow framework of Google assistant and the React framework [21] [22].

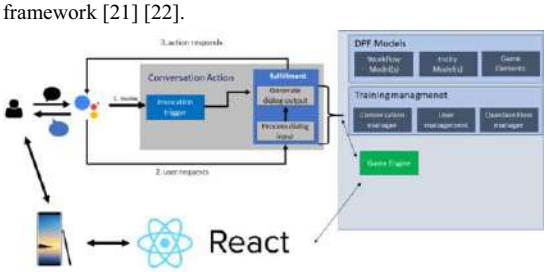


Figure 7: Overall architecture of the MDE approach

Learning Analytics

In our implementation, we utilize learning analytics for providing better training environment for the user. The training program records the history of training sessions. Therefore, we can perform some analysis over the training data such as whether the questions are too easy or too difficult for learners. While presenting questions to the user, the system does not only know how many questions the user has correctly answered, but also the time taken by the user to finish a training session. This allows the trainers to adjust the difficulty of the training sessions. Learning analytics are to identify the part of a training session that students often struggle with and the outcome of a session. A dashboard for the trainers with visualization can then be used to show the progress of the trainees. The goal is to make the training programs adaptable for various levels of difficulty to increase learners' interest to the sytem for practicing use of guidelines. This adaptability cannot be achieved without use of the learning analytics. The application of MDE enables achievement of this requirement as all the training programs share an abstraction. Building a visualization is therefore achievable in general across different training modules. With this model, machine learning techniques can also be applied to automatically adjust the difficulty of training sessions.

Results

The approach we describe in this paper has been materialized through a prototype tool for the gamification of CPGs. The specific use-case we employ is based on asthma care guidelines [23]. The model based approach for game engine allows us to develop and/or customize a new training program by making changes in the model. The MDE approach for modularizing different components of the system and its seperation from the user interaction allows us to integrate with various frameworks of user interaction such as Google's Dialogflow and React Framework [21] [22]. Since the model for the guideline and gamification remains the same, adding various techniques such as mobile-based interaction and voice-based interaction requires very small effort. With the prototype tool we show the potential of using MDE approach supporting different learning style of students. With the proposed idea of applying learning analytics for providing adaptive training to the trainee will provide better training results.

One indirect result of our approach is the support for innovating technologies. The generalized game engine can be integrated with new platforms such as Amazon Echo, integration with IoT devices that are used in clinical setup e.g., robot doll patients.

Another result is the support for ICT research in the training of practitioners. Data science and machine learning techniques can be used to understand the learning pattern of students and adapt them for better learning outcome.

Discussion

This work presents a model-driven approach to the design of an elearning system for clinical guidelines. Most current research on the use of gamification in medical education is focused on the effect of the educational games in knowledge and skill acquisition and on their acceptability. Akl et al present an educational game for teaching clinical practice guidelines to internal medicine residents [24]. In their description, domain experts developed multiple choice questions based on clinical guidelines which were then uploaded into the system via a question editor. This approach is inefficient in both the manual development of questions, and in the static nature of available questions. Our model-based approach allows for questions to be generated automatically, and the learner analytics platform helping to adapt the questions to the strengths and weaknesses of the learner. The automation of different aspects of the system means that once it is set up, it will require minimal resources to maintain, hence saving costs.

Our modular approach provides several other advantages. It makes it easier to update guidelines as only parts of the entity and workflow models change while the rest of the system remains the same. It is flexible also enough to allow for integration with various devices supporting different means of user-interaction and the abstraction of the clinical processes allows for the representation of any guideline. Further, the use of gamification will potentially enhance the learning experience by increasing student engagement with the learning material as reported in several studies [11,17,25].

There are a number of limitations to the gamified elearning system we describe in this work. First, full training of guideline content that requires the learning of some physical skills - such as performing cardiopulmonary resuscitation (CPR) – cannot be fully performed using our system as our system can only train on guideline content that do not require hands on training. Secondly, dissemination and implementation of clinical guidelines is a continuum that cannot be separated. Our system only addresses the dissemination half of that continuum and is not enough to bridge the gap between recommended and actual practice. Finally, there are several non-technological barriers to adherence to clinical practice guidelines, [26] and our system only addresses two of these barriers: lack of awareness and lack of familiarity with the guidelines.

In the near future, we plan to enhance our prototype tool for the development of other clinical guidelines. We wil also evaluate the acceptability and effectiveness of the proposed technique as a dissemination strategy for clinical guidelines within resource-limited settings.

Conclusions

We set out to design a gamified e-learning system for the training of clinical practice guideline content. Our approach allows for a flexible, adaptive and potentially cost-effective e-learning system.

Acknowledgements

This work was supported in part by the NORHED program (Norad: Project QZA-0484). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Norwegian Agency for Development Cooperation. All authors report no conflict of interest.

References

[1] D.L. Sackett, Evidence-based medicine, *Seminars in Perinatology*. **21** (1997) 3–5. doi:10.1016/S0146-0005(97)80013-4.

[2] Institute of Medicine (US) Committee on Standards for Developing Trustworthy Clinical Practice Guidelines, Clinical Practice Guidelines We Can Trust, National Academies Press (US), Washington (DC), 2011. http://www.ncbi.nlm.nih.gov/books/NBK209539/ (accessed August 24, 2017).

[3] T. Kredo, S. Bernhardsson, S. Machingaidze, T. Young, Q. Louw, E. Ochodo, and K. Grimmer, Guide to clinical practice guidelines: the current state of play, *Int J Qual Health Care*. **28** (2016) 122–128. doi:10.1093/intqhc/mzv115.

[4] L. Thomas, Clinical practice guidelines, *Evid Based Nurs*. **2** (1999) 38–39. doi:10.1136/ebn.2.2.38.

[5] J.M. Grimshaw, H.J. Schünemann, J. Burgers, A.A. Cruz, J. Heffner, M. Metersky, and D. Cook, Disseminating and Implementing Guidelines, *Proc Am Thorac Soc*. **9** (2012) 298–303. doi:10.1513/pats.201208-066ST.

[6] J.M. Grimshaw, R.E. Thomas, G. MacLennan, C. Fraser, C.R. Ramsay, L. Vale, P. Whitty, M.P. Eccles, L. Matowe, L. Shirran, M. Wensing, R. Dijkstra, and C. Donaldson, Effectiveness and efficiency of guideline dissemination and implementation strategies, *Health Technol Assess*. **8** (2004) iii–iv, 1–72.

[7] J.M. Grimshaw, L. Shirran, R. Thomas, G. Mowatt, C. Fraser, L. Bero, R. Grilli, E. Harvey, A. Oxman, and M.A. O'Brien, Changing provider behavior: an overview of systematic reviews of interventions, *Med Care*. **39** (2001) II2-45.

[8] G. Irimu, A. Wamae, A. Wasunna, F. Were, S. Ntoburi, N. Opiyo, P. Ayieko, N. Peshu, and M. English, Developing and introducing evidence based clinical practice guidelines for serious illness in Kenya, *Archives of Disease in Childhood*. **93** (2008) 799–804. doi:10.1136/adc.2007.126508.

[9] B. Djulbegovic, and G.H. Guyatt, Progress in evidence-based medicine: a quarter century on, *The Lancet*. **390** (2017) 415–423. doi:10.1016/S0140-6736(16)31592-6.

[10] N. Muinga, B. Sen, P. Ayieko, J. Todd, and M. English, Access to and value of information to support good practice for staff in Kenyan hospitals, *Glob Health Action*. **8** (2015). doi:10.3402/gha.v8.26559.

[11] S. Deterding, M. Sicart, L. Nacke, K. O'Hara, and D. Dixon, Gamification. Using Game-design Elements in Non-gaming Contexts, in: CHI '11 Extended Abstracts on Human Factors in Computing Systems, ACM, New York, NY, USA, 2011: pp. 2425–2428. doi:10.1145/1979742.1979575.

[12] K.M. Kapp, The gamification of learning and instruction: game-based methods and strategies for training and education, John Wiley & Sons, 2012.

[13] I. Kurtev, J. Bézivin, F. Jouault, and P. Valduriez, Model-based DSL Frameworks, in: Companion to the 21st ACM SIGPLAN Symposium on Object-Oriented Programming Systems, Languages, and Applications, ACM, New York, NY, USA, 2006: pp. 602–616. doi:10.1145/1176617.1176632.

[14] A. Rodrigues da Silva, Model-driven engineering: A survey supported by the unified conceptual model, *Computer Languages, Systems & Structures*. **43** (2015) 139–155. doi:10.1016/j.cl.2015.06.001.

[15] F. Rabbi, Y. Lamo, and W. MacCaull, Co-ordination of Multiple Metamodels, with Application to Healthcare Systems, *Procedia Computer Science*. **37** (2014) 473–480. doi:10.1016/j.procs.2014.08.071.

[16] A. Rutle, Diagram predicate framework: A formal approach to MDE, The University of Bergen, 2010. https://bora.uib.no/handle/1956/4469 (accessed November 7, 2016).

[17] D. Strmečki, A. Bernik, and D. Radošević, Gamification in E-Learning: Introducing Gamified Design Elements into E-Learning Systems, *Journal of Computer Science*. **11** (2016) 1108–1117. doi:10.3844/jcssp.2015.1108.1117.

[18] F. Rabbi, Y. Lamo, I.C. Yu, and L.M. Kristensen, WebDPF: A web-based metamodelling and model transformation environment, in: 2016 4th International Conference on Model-Driven Engineering and Software Development (MODELSWARD), 2016: pp. 87–98.

[19] D. Jackson, Alloy: a lightweight object modelling notation, *ACM Transactions on Software Engineering and Methodology*. **11** (2002) 256–290. doi:10.1145/505145.505149.

[20] X. Wang, Towards Correct Modelling and Model Transformation in DPF, (2016).

[21] Google, Dialogflow, *Dialogflow*. (n.d.). https://dialogflow.com/ (accessed November 23, 2018).

[22] React – A JavaScript library for building user interfaces, (n.d.). https://reactjs.org/index.html (accessed November 23, 2018).

[23] MoH, Kenya, Basic Paediatric Protocols, 2016.

[24] E.A. Akl, R. Mustafa, T. Slomka, A. Alawneh, A. Vedavalli, and H.J. Schünemann, An educational game for teaching clinical practice guidelines to Internal Medicine residents: development, feasibility and acceptability, *BMC Medical Education*. **8** (2008) 50. doi:10.1186/1472-6920-8-50.

[25] F. Faiella, and M. Ricciardi, Gamification and learning: A review of issues and research, *Journal of E-Learning and Knowledge Society*. **11** (2015) 13–21.

[26] M.D. Cabana, C.S. Rand, N.R. Powe, A.W. Wu, M.H. Wilson, P.-A.C. Abboud, and H.R. Rubin, Why Don't Physicians Follow Clinical Practice Guidelines?: A Framework for Improvement, *JAMA*. **282** (1999) 1458–1465. doi:10.1001/jama.282.15.1458.

Address for correspondence

Job N. Nyameino, MBChB, MSc. Email: jbngena@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190321

Governance and Sustainability of an Open Source Electronic Health Record: An Interpretive Case Study of OpenDolphin in Japan

Placide Poba-Nzaou[a], Naoto Kume[b], Shinji Kobayashi[b]

[a] *Department of Human Resource and Organization, University of Quebec in Montreal, Montreal, Quebec, Canada*
[b] *Department of EHR, Graduate School of Medicine, Kyoto University, Kyoto, Japan*

Abstract

Electronic Health Records (EHRs) are at the heart of reforms aiming for improving the efficiency and quality of citizens healthcare services. Although there is still some skepticism, open source (OS) EHR is a growing phenomenon in health informatics. Given the widespread adoption of OS software (OSS) in several domains, including operating systems, and enterprise systems, the repeated shortfalls faced by healthcare organizations with dominant proprietary EHRs create an opportunity for other alternatives, such as OSS to demonstrate their abilities in addressing these well-documented problems, including inflexibility, high costs, and low interoperability. However, scholars have expressed extensive concerns about the sustainability of OS EHR. Recognizing that OSS project sustainability relies on their governance arrangements, this case study reports on the evolution of the governance and sustainability of a Japanese OS EHR project and provides rich insights to other open source EHR initiative stakeholders, including physicians, developers, researchers, and policy-makers.

Keywords:

Electronic Health Records, Japan, Medical Informatics

Introduction

In order to deal with the unsustainable increasing cost of health care, aging population, and chronic diseases burdening their population, all industrialized countries are investing in Electronic Health Record (EHR)[1]. However, Health Care Organizations (HCOs) have been slow to implement EHR in most OECD countries because of their high cost. In addition, the large majority of HCOs have adopted proprietary EHRs and are facing repeated shortfalls with those tools, including dissatisfaction with costs and interoperability issue [2]. This situation creates an opportunity for Open Source Software (OSS) to demonstrate its abilities in addressing those challenges. In the context of Japan, the Japanese Medical Association (JMA) estimated that EHR implementation costs in all medical providers would amount to US $180 billion over a 10-year period, which is not affordable for Japan without a significant reduction in costs. JMA also suggested the implementation of OSS as one of the options to overcome the obstacle [3]. However, although open source projects have been launched at a rapid pace in a growing number of medical clinics and hospitals, scholars expresse extensive concerns about the sustainability problem [4]. Thus, it's paramount to understand the factors that may impede wide adoption of the open-source medical informatics tools.

Recognizing that OSS projects governance affects the project sustainability [5], we report on a case study of the governance and sustainability of OpenDolphin, a Japanese open source EHR project. The source code can be found on GitHub: https://github.com/dolphin-dev/OpenDolphin. A screenshot of the OpenDolphin interface is presented in Figure 1.

Figure 1— OpenDolphin Interface

This study was guided by the following questions: (1) how did the governance of OpenDolphin EHR evolve over time?; (2) how did the governance of the project enable or constrain OpenDolphin EHR sustainability?

Conceptual Background

The framework is adapted from de Laat [6] three stages (see Figure 2) and Markus [7] governance dimensions (e.g., vision and goals, ownership of the assets, etc.).

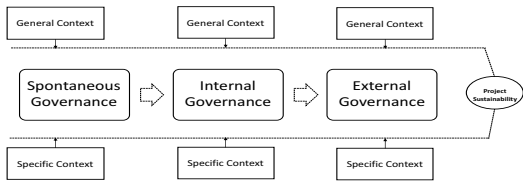


Figure 2— Research Framework

General Context of Healthcare in Japan

Over the past decades, the Japanese healthcare system has been cost efficient and achieved significant progress in terms of life expectancy, infant mortality, and the eradication of communicable diseases [8]. Moreover, whole system reform has been undertaken in order to preserve its in the future [9], and the promotion of Medical ICT is one of the pivotal measures in the growth strategy of the Abe Government known as "Abenomics" [10].

Specific Context of Medical Technology in Japan

In order to deal with the challenges associated with the countrywide plan of health IT diffusion, the JMA declared its intent to develop OSS as their "infostructure" in 2000. This momentum towards OSS has influenced the development of OSS projects in the medical field in Japan, including the OpenDolphin project. Overall, the adoption of EMR by medical

providers in Japan is imbalanced. Whereas the adoption rate is about 70% for large hospitals (>400 beds), it is below 14% for small hospitals (<200 beds) [11]. Thus, Japan lags behind when compared with other OECD countries. The prevalence of electronic health record in clinics in Japan is low and is estimated to be 16.5% [12] then about 30% [11]. Overall, the lack of financial resources is one of the most important barriers to EHR adoption [11].

Methods

OpenDolphin EHR was developed within the Dolphin project. It's an open-source client system for regional health information systems as a response to the recognition of the inability of the Japanese marketplace to sustain the objective of 60% of diffusion of EHR in Japan within 5 years, as announced by the government [13]. The main contributor to the development of OpenDolphin is OS-Service Corp, a for-profit organization; and as of 2017, the system was used by over 300 clinics in Japan.

We relied on five sources of evidence: semi-structured interviews, written documents, a short questionnaire, and field notes from researchers. In addition, all research team members participated in a conference organized by the JMA and related to open health informatics in Tokyo.

In total 7 key informants from the OpenDolphin community were interviewed (see Table 1). In-depth semi-structured interviews were deemed appropriate for this study because it focuses on perceptions and meanings of informants [14]. The same interview grid was used for all interviews and covered questions related to the theoretical framework. Each interview lasted on average one and a half hour. During the interviews, the research team took notes that were later developed into detailed field notes. A short questionnaire was used to collect demographic and factual information.

Table 1—Informants Characteristics

Name	Role/Education	ODolph	OSS	HIT
Alpha	(1), (2), (3), (4)/B	a	a	c
Beta	(1), (2)/B	c	c	c
Akyo	(1), (2), (3), (4)/B	c	c	c
Gama	(5)/H	a	a	c
Delta	(1), (2), (3), (4)/D	c	c	c
Epsilon	(1), (2)/B	a	b	b
Zeta	(2)/B	a	a	a

Role: (1) project manager (2) developer (3) document creation (4) system administrator (5) sales representative
Experience (years): 0–5 (a), 6–10 (b), over 10 (c)
Education: H: high school, B: bachelor degrees, D: medical doctorate

Data analysis started as soon as data collection began and was conducted in an iterative manner, mainly based on the fundamental principle of the hermeneutic cycle [15]. Given that we aimed to understand the governance and its implication for the sustainability of OpenDolphin 'through the meanings that informants assign to them' [14], we read to get familiar with the collected materials. More specifically, interview transcripts were read by moving back and forth between the whole and parts of the studied phenomenon. In the same manner, interview transcripts were read between informant statements and the research framework. After several iterations, we applied data display techniques [16]. Then, a narrative approach was employed to describe the case in the form of a 'narrative report' [17] that helped us to develop an in-depth and global view of the OpenDolpin history, governance, and sustainability.

Results

We draw on the conceptual framework to present the results.

Phases 1 and 2: Spontaneous and internal governance

The development of OpenDolphin project began in 2001, following a tender launch by the Japan Ministry of Economy, Trade and Industry (METI) and targeting regional medical associations. In Japan, medical associations are built upon a three-layer model corresponding to the architecture of the country's administration [18]: the regional medical associations at the regional level, the prefectural medical associations at prefectural level, and the Japan Medical Association at the national level. All these organizations operate cooperatively, yet each and everyone has an autonomous structure.

The tender was awarded to the "Dolphin project" led by a consortium formed by three universities (Miyazaki University, Kumamoto University, and Kyoto University) and one regional medical association (Tokyo Medical Association). They received a fund of 60 million yen—about $570,000 US. One of the key reasons why the consortium won was that it proposed to adopt the Medical Markup Language (MML) for the Dolphin project, so as to facilitate the exchange of data between EHR systems within regions and beyond. In fact, MML was developed in order to generate a set of standards through which medical data, within Japan and hopefully internationally, can be stored, accessed, and exchanged among different physical locations [19]. The MML was created in Japan from research and funded by the Japanese Ministry of Health and Welfare [20].

Vision and Goals. From the outset, the OpenDolphin project's goal was to electronically emulate the user experience of the paper medical record at a low cost. However, instead of trying to simply achieve a high level of functional coverage, the goal was to create good user experience emphasizing simplicity, ease of use, and usability. In addition, OpenDolphin was envisioned to be compatible with different operating systems used by clinics in Japan. A statement from Akyo (a fictitious pseudonym given to the main contributor to the initial project) illustrates the vision: we wanted to have a "*clinical user-oriented EHR with UX [user experience] resembling that of a notebook*".

Ownership of the Assets. The ownership of the asset was sealed by the grant awarded by the METI, which requires the research output and the software, to comply with a Bayh-Dole Act-like regulation adopted by Japan in 1999. The Bayh-Dole Act is a US federal legislation enacted in 1980 and co-sponsored by Senators Birch Bayh and Robert Dole. It allows universities, nonprofit research institutions, and small businesses to own, patent, and market any inventions resulting from federally funded research programs within their organizations [21]. Hence, the GNU General Public License (GPL) version 2 was the *de facto* choice.

Software Development Processes. The project started in 2000, following a requirement specification stage that took about half a year, which was followed by a software development stage with two full-time developers, including one senior (here identified as Akyo) and one junior and lasting about a year. Afterward, about six months later, the junior developer left the project and from then on, the senior developer took charge of all the software development tasks. The software was delivered to the research team in 2001 and was released to the public the same year. Another principle that guided the development of

OpenDolphin was "*DIY (Do It Yourself)*". This principle stemmed from the fact that, according to Akyo, "*there are various doctors with different needs, so the project needs to work as a platform on which DIY (Do It Yourself) offers solutions to the needs not answered by the project as-is when created*".

Community Management, Leadership, and Finance. During this first phase, Akyo was the only contributor for OpenDolphin. However, due to the newness of OpenDolphin EHR and the absence of a community of users and developers, it was adopted by 23 clinics. The revenue generated from the services related to adoption and the use was not enough to support Akyo as a professional software developer. He left and started to conduct small business activity. At that time, the fees for services related to OpenDolphin was entirely at Akyo's discretion. Meanwhile, four medical doctors from clinics that adopted OpenDolphin started to contribute to software development at a very high level of expertise, including developing new functionalities.

Use of Information and Tools. At this stage, the following tools were used: Repository: GitHub (dolphin-dev/OpenDolphin); Development tools: NetBeans, SourceTree, SublimeText, TeraTerms, BitBucket (only for certain parts of the software).

Given the small size of the community, only Twitter and Email were used to communicate with other actors. According to Akyo, during this stage, he was the only person who could answer the coming questions by email.

Technologies and Translation Language. The selection of technologies was guided by key principles, including high interoperability, freedom, and low cost. Hence, the following technologies were selected: database (PostgreSQL), development language (Java), application server platform Wildfly (JBOSS), and client application (Apache MAVEN for XML). At this stage, OpenDolphin was available only in Japanese.

During this stage, the OpenDolphin project accomplished some achievements that confirmed its position as a promising open source EHR: (1) the successful connection with ORCA (Online Receipt Computer Advanced) system in 2001; (2) the release of a MacOS version in 2006; (3) the launch of an ASP service in 2007; (4) the achieved compatibility with the iPhone and the iPod Touch in 2009.

Phase 3: External governance

As of 2009, OpenDolphin has a small community of developers, users, and support providers. The same year, OS-Service Corp, one of the support providers who have been involved with OpenDolphin for few years, suggested cooperating with Akyo for the aim of "scaling up the services related to OpenDolphin". Thus, Akyo joined OS-Service Corp and the OpenDolphin project moved from a community-managed to a corporate-managed governance model. Since assuming responsibility for OpenDolphin formal governance, OS-Service Corp has been taking or planning to take new initiatives aimed at developing and sustaining the OpenDolphin project and products.

Vision and Goals. OS-Service Corp is planning to create an EHR with a combined hospital management system (HMS). According to Akyo, historically, HMS has been the foundation upon which the EHR system is installed. He indicated that they are considering the opposite, in other words, to have the EHR as the foundation system, and implement the HMS functionality on top of that.

Ownership of the Assets. In the meantime, the license had been moved from *GNU General Public License (GPL) version 2 to GPL version 3*. In order to develop its business, OS-Service Corp created a new version of OpenDolphin named OpenDolphinPro, a more commercial version that is also distributed under GNU GPL license. OpenDolphiPro is packaged with services provided by OS-Service Corp.

Software Development Processes. With the arrival of OS-Service, the software development team increased to eight with four internal developers within OS-Service Corp and four medical doctor users who contribute actively at a very high level of expertise.

In order to improve the sustainability of the OpenDolphin project, OS-Service Corp is moving from JBoss AS (Application Server) 6 to server 7. In fact, the development of JBoss AS 6 has been stopped. JBoss AS 7, now known as WildFly, is a complete rewritten server that is faster and easier to configure. With the intention of reducing the risk related to this technological transition, OS-Service Corp has partnered with Red Hat Corporation in Japan in order to use their deployment methods and development environments as references. In addition, OS-Service Corp has introduced an agile development method to improve the efficiency, as stated by OS-Service Corp manager: "We are employing an agile development method where the project manager goes through the details of every request, and uses them to assign required members to do the development. Because of this, we do a much faster implementation than development, based on the waterfall model".

OS-Service Corp has also implemented a formal process for handling requests from clients. There is now a dedicated customer support division that collects detailed requests from the users, and passes the requests on to the development division after deciding on the design through meetings. The new functionalities ensuing from these endeavors are generally implemented in OpenDolphinPro and released as minor updates. Thereafter, these functionalities are integrated into the regular OpenDolphin version.

Base on their experience, the OS-Service Corp team stated that "having the software always up to date is ideal. However, given that OpenDolphin differs from web applications (in the sense that clinics have source code updated at different levels), there is a need to consider the special circumstances from each individual clinic". In order to properly manage this situation, the decision to update the software for a clinic is made by the customer support division based on the assessment of each clinic's environment and structure.

Community Management, Leadership, and Finance. At this stage, the number of resellers nationwide is increased to 27. The OpenDolphiPro sold by reseller is packaged with services provided by OS-Service Corp. Since the initial release, OpenDolphin has always been sold by salespeople who directly visit the target clinics. However, according to OS-Service Corp, from this point on, the company will put more emphasis on the cloud version of OpenDolphin. Accordingly, the company is swiching to a digital strategy and expecting to sell faster than the past. In fact, between 2009–2016, with one sales executive, OS-Service Corp was selling at a rate of 30–40 clinics per year through clinic visits and sometimes 50 clinics a year. The sales target of the cloud version is set at 100 per year. It's worth mentioning that the sales executive has left the company and since then, OS-Service Corp is struggling to attract new clients.

At this stage, it important to mention that, OS-Service Corp is just above the break-even point and the management team is taking various initiatives with the expectation to depart from that. The break-even point is reached when a company's total costs equal its total revenue.

The management of OS-Service Corp has come to realize the need to create a partnership with various actors in order to

increase its community at large. One of the initiatives planned is to connect with the Medical Open Source Software (MOSS) meeting held twice a year in Japan which brings together medical informatics experts, advocates, and supporters.

Use of Information, Communication, and Tools. The tools in use at this stage are the same as the previous one. Repository: GitHub; Development tools: NetBeans, SourceTree, SublimeText, TeraTerms. Whereas the OpenDolphin version is managed on GitHub, the OpenDolphinPro is managed on OS-Service Corp premises.

Technologies and Development Language. As stated earlier, OpenDolphin client was written with Java but the client rewritten in JavaScript. This initiative stems from the fact that JavaScript is primarily a client-side scripting language designed to run in the internet browser without having to be compiled like Java. In addition, OS-Service Corp is planning to add artificial intelligence (AI) components in OpenDolphin. Regarding to language translation, OS-Service Corp is planning to develop an English version of the software so as to move into the international market and extend its market reach.

During this stage, the OpenDolphin project made some achievements that confirmed its position as a promising open source EHR: (1) the release of a comprehensive documentation named OpenDolphin Perfect Guide in 2016; (2) the release of three iOS apps (DolphinPro, VisitTouch, Super EHRTouch); (3) the launch of authentication based on SSO (Single Sign-On); (4) the establishment of compatibility with iPads in 2010; (5) obtaining the status of certified partner solution for IBM Japan in 2011; (5) the completion of connection between ORCA and OpenDolphin cloud solution in 2013; (6) the celebration of the 10th anniversary OpenDolphin cloud ZERO (pay-as-you-go system) in 2014.

Among the most important achievements is the fact that, even if OpenDolphin EHR was at first targeting medical clinics, it has successfully expanded its market share from small hospitals.

Discussion

Open source software is recognized as a growing phenomenon with a promising potential in the medical informatics field. Nonetheless, scholars have expressed extensive concerns about open source software sustainability [4]. Although the body of knowledge related to open source in health informatics has kept growing, there have been few empirical investigations of open source clinical information systems in the context of industrialized countries outside of the western world. Our study reports on an empirical instance of open source EHR that is developed and used in an industrial country outside the western world. It contributes to deepening our understanding of the governance evolution of OS EHRs and how it enables or constrains the sustainability of the OS projects. Our results reveal that the model provided by de Laat [6] is useful in explaining the evolution of the governance of OpenDolphin. However, we are able to match only two phases out of the three suggested by the model, as the first two phases are difficult to separate in the case of OpenDolphin.

Our data reveal that the turning point from the "spontaneous and internal" governance phase to the "external governance" was triggered by the concern about the sustainability of OpenDolphin. In fact, by the time Akyo received OS-Service Corp's offer, he had already noticed the ineffectiveness of the governance model, even though the main concern of OpenDolphin during the first phase was related to the inefficiency of the business model.

Watson and Boudreau [22] distinguished four different business models of OSS production or distribution: open community, corporate distribution, sponsored OSS, and second-generation OSS. Open community is a model for which the development and support of the software mainly rely on volunteers with limited commercial interests, while corporate distribution is a model that takes advantage of quality products developed by open community models, "improving distribution methods for these products, and providing complementary services in order to make these OSS products more accessible to a broader market." [22]. Examples of such models include RedHat and SpikeSource. Sponsored Open Source is models of OSS projects sponsored by corporations or foundations or both. Examples of such models include Apache Web Server with Apache Foundation and Eclipse with IBM. The second-generation model is also known as professional open source—is composed of firms that are considered hybrid because their models are between a corporate distribution and sponsored OSS. Of note is the fact that second-generation firms "typically own or tightly control the software code and can exploit their intimate knowledge of the code to provide higher-quality service that could potentially competing service providers" [22]. Examples of such open source projects include MySQL and JBoss.

The deal with OS-Service Corp had an impact that went beyond the business model and *de facto* induced new governance arrangements. The analysis also reveals a high degree of dependency between the business model and the governance model. It also suggests that, during the first phase, OpenDolphin was a technological success but not an economic success. Of note is the fact that, during the first phase of governance, the community of software developers has never really grown beyond Akyo. One possible explanation might be that there were few initiatives undertaken to attract contributors. OS-Service Corp brought the needed resources to sustain technological success and convert it to economic success. Our data reveals that the involvement of OS-Service Corp has altered the "business model" of the OpenDolphin project, moving it from "open community" to "sponsored OSS project" as explained below.

At the beginning of the third phase, OpenDolphin has experienced rapid growth by attracting a large number of clinics each year (30–40 per year through clinic visits). However, since the departure of the sales executive who was in charge of sales, OpenDolphin growth has been stagnant, despite sales activities being managed by OS-Service management. Even if the stagnation can be partially attributed to the departure of the sale executive, OS-Service management feels the need to make some changes in the way they deal with their main market segment formed by a medical clinic. In fact, OpenDolphin situation is surprising for at least five reasons: (1) the OpenDolphin EHR seems to be appreciated by medical doctors who are using it; (2) the market has a great potential for growth due to the low rate of adoption of EHR by medical clinics in Japan (about 30%); (3) the rate of adoption of EHR by newly opened clinics is high (about 80%); (4) the choice of OSS by the Japan Medical Association as their "infostructure" in 2000 and the release of ORCA, an open-source software used by medical clinics in Japan, has created a fertile ground for OS to expand at least in primary care organizations; (5) OpenDolphin was created as part of a project led by regional medical associations, which gives it a certain level of proximity to the medical field in Japan.

Going back to the stagnation of the OpenDolphin growth, our informants identified a characteristic of the local market that may explain why OpenDolphin, a product that meets the needs of medical clinics and does so at a low cost, is struggling to

attract newly opened medical clinics, knowing that the cost is one of the main barriers to the adoption of EHR by small hospitals and clinics. In Japan, a consultant on behalf of a medical doctor usually handles the process of opening a new clinic. Since consultants' fees are calculated as a percentage of the total costs of the project, including the costs associated with the acquisition of an EHR, consultants may have the tendency not to recommend OpenDolphin because of its low cost.

In order to strengthen the sustainability of the OpenDolphin project, the management at OS-Service Corp has decided to put more emphasis on the OpenDolphin cloud and digital marketing, with the aim of attracting more medical clinics. The management team seems to recognize the limitation of both the current business and the governance models. The initiatives they are undertaking may substantially alter the governance and the business models, hence the question arises as to whether other options may exist, in terms of governance models and other business models from which OpenDolphin can choose. When contrasting OpenDolphin characteristics with the four open source business models suggested by Watson and Boudreau [21], it's illuminating to recall that OpenDolphin goes through two of them (the community model and Corporate Distribution model). Given that the Second-Generation allows the firm to own or tightly control the code source, this option does not apply to OpenDolphin because of the characteristics of the GPL license. Hence, within our theoretical background, the only remaining option is the Sponsored model. If OpenDolphin has to become a sponsored project, OS-Service Corp may become the corporate sponsor in the first case or be associated with a foundation in the second case. Each of these options will allow OS-Service Corp to share the costs of developing new functionalities.

Overall, the governance of the OpenDolphin project has evolved as a consequence of the alteration of its business model so as to strengthen its sustainability. However, the results appear to be mixed. It does not seem easy for OS-Service Corp to handle, by itself, all the challenges faced by OpenDolphin. Our study reveals that it is not only the management of OS-Service feeling the needs, but also some of the elements are already in place waiting to be highlighted and combined in order to create an OpenDolphin ecosystem that will help strengthen its sustainability. This ecosystem may include OSS experts, advocates, and supporters among both health IT decision makers and Health IT policymakers.

Conclusions

Our results indicate that governance does enable or constrain the sustainability of OpenDolphin but, considering the business model also provides a broader understanding of the sustainability of the OSS project. One of the main limitations of this study is related to the methods, this is an interpretive single case study. Thus, caution should be exercised in interpreting these results in other contexts.

References

[1] OECD. improving Health sector Efficiency. The Role of Information Technology Communication: OECD Publishing, 2010.
[2] Friedberg M, Crosson FJ, Tutty M. Physicians' concerns about electronic health records: implications and steps towards solutions. Health Affairs blog 2014;11:963
[3] Kobayashi S, Yahata K, Goudge M, Okada M, Nakahara T, Ishihara K. Open source software in medicine and its implementation in Japan. *J Inf Technol Healthc* 2009;7(2):95-101
[4] Alsaffar M, Yellowlees P, Odor A, Hogarth M. The State of Open Source Electronic Health Record Projects: A Software Anthropology Study. *JMIR Med Inform* 5 (2017), e6.
[5] De Noni I, Ganzaroli A, Orsi L. The governance of open source software communities: An exploratory analysis. *Journal of Business Systems, Governance & Ethics* 2011;6(1):1-18
[6] de Laat PB. Governance of open source software: state of the art. *Journal of Management & Governance* 2007;11(2):165-77
[7] Markus ML. The governance of free/open source software projects: monolithic, multidimensional, or configurational? *Journal of Management & Governance* 2007;11(2):151-63
[8] Uetsuka Y. Characteristics of Japan's Healthcare Systems and the Problems. *Japan Medical Association journal: JMAJ* 2012;55(4):330-33
[9] Statistics Japan. Statistical Handbook of Japan 2017, 2017.
[10] Japan Government. About Abenomics. Secondary About Abenomics2018. https://www.japan.go.jp/abenomics/ [accessed on May 20, 2018].
[11] Tanaka H. Medical ICT in Japan. 2016. http://ngbrc.com/img/file11.pdf [accessed on May 15, 2018].
[12] Takabayashi K, Suzuki T. Japanese EMRs and IT in medicine: expansion, integration, and reuse of data. Healthcare informatics research 2011;17(3):178-83
[13] Takada A, Guo J, Tanaka K, et al. Dolphin Project. *Journal of Medical Systems 29 (2005), 391-400.*
[14] Walsham G. Doing interpretive research. European journal of information systems 2006;15(3):320-30
[15] Klein HK, Myers MD. A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS quarterly* 1999:67-93
[16] Miles MB, Huberman AM. Qualitative data analysis: An expanded sourcebook: sage, 1994.
[17] Langley A. Strategies for theorizing from process data. *Academy of Management review* 1999;**24**(4):691-710
[18] Sahara H. Role of regional medical associations in large-scale disasters. Japan Med Assoc J 2012;55:273-76
[19] Araki K, Ohashi K, Yamazaki S, et al. MML for XML-based hospital information interchange. *Journal of medical systems* 2000;24(3):195-211
[20] Kimura M, Ohe K, Yoshihara H, et al. MERIT-9: a patient information exchange guideline using MML, HL7 and DICOM. *International journal of medical informatics* 1998;51(1):59-68
[21] Public Law 96-517. 96th Congress, 2nd session (12); 1980.
[22] Watson RT, Boudreau M-C, York PT, et al. The business of open source. *Comm. of the ACM* 2008;51(4):41-46

Address for correspondence

Placide Poba-Nzaou

email: poba-nzaou.placide@uqam.ca

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190322

Fit Between Individuals, Tasks, Technology, and Environment (FITTE) Framework: A Proposed Extension of FITT to Evaluate and Optimise Health Information Technology Use

Mirela Prgomet[a], Andrew Georgiou[a], Joanne Callen[a], Johanna Westbrook[a]

[a] Centre for Health Systems and Safety Research, Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

Abstract

Evaluating and optimising 'fit' between technology and clinical work is critical to ensure the intended benefits of technology implementations are achieved. Using a mixed method approach (structured observation, interviews, field notes) we collected data regarding users, tasks, technology, and factors impeding technology use from a sample of 38 clinicians on two wards at an Australian hospital. We used the FITT framework to assess the relationships between users, tasks, and technology.

Our findings showed that even when adequate fit between users, tasks, and technology was attained additional factors related to the environment (including the temporal rhythms of a ward, infection control rooms, or space limitations) ultimately affected technology use. Thus, we propose the fit between individuals, task, technology and environment (FITTE) framework as a means to evaluate and optimise technology use by explicating the relationships between users, tasks, technology, and the environment in which they operate.

Keywords:

Computers; Health services research; Models, theoretical

Introduction

Health information technologies (HITs) undoubtedly have the potential to make care delivery more efficient and effective. However, HITs can also result in unintended consequences, including: unfavourable workflow issues; paper persistence; changes in work practices (such as altering of the pace and sequencing of clinical activities); and only partial support for the information needs and work practices of clinicians [1; 2]. Researchers indicate that the 'fit' between technology and clinical work is what leads to HITs being either used and incorporated into routine, worked around, or rejected [3]. Where there is a mismatch between technology and work practices, the intended benefits of the HIT may not be achieved.

Research investigating fit between technology and clinical work has, by and large, focused on software applications with limited consideration regarding hardware computing devices. In order to achieve optimal beneficial outcomes from HITs, it is also essential to ensure the right fit of hardware devices. Fixed computing devices, such as desktop computers, may constrain the work practices of clinicians by mandating where information can be accessed or documented. Mobile computing devices, on the other hand, are not tied to one location but rely on battery and wireless connectivity, and have different screen sizes and data input mechanisms (keyboard, stylus, and/or touchscreen), all of which impact their usability. Considering that computers play an integral part in the success of software implementations, how fixed and mobile devices fit into and support clinical work practices warrants attention.

Theoretical Frameworks

The Technology Acceptance Model (TAM) and Task-Technology Fit (TTF) model are two prominent theories that describe potential means by which to explore the notion of fit between technology and work practices. TAM has been widely used in health care and employs the constructs of perceived usefulness and perceived ease of use as determinants of: attitudes towards using technology; intention to use technology; and the actual use of technology [4]. Findings from Holden and Karsh's seminal review of TAM, and its application in health care literature, underscored TAM's value as a theoretical framework to assess the relationship between users and technology [3]. However, the review also pointed to the need for additional constructs to allow for further relationships to be explored. Dishaw and Strong, likewise, highlight that a shortcoming of TAM in aiding understanding of the use of technology may lie in the absence of an explicit construct examining tasks [5]. They indicate that, unlike TAM, the TTF model provides explicit inclusion of a task–technology construct.

The TTF model looks at task characteristics and technology characteristics, which together determine task–technology fit and influence technology utilisation [6]. The premise of the model is that technology will be used if it adequately supports the demands of a task. In applying the TTF model to assess the fit of a picture archiving and communications system (PACS), Lepanto et al. concluded that "TTF is a valid tool to assess perceived benefits, but it is important to take into account the characteristics of users" [7]. Others have similarly indicated that a limitation of TTF is that it does not explicitly include a construct that examines user characteristics [5].

Ammenwerth et al.'s propose Fit between Individuals, Tasks, and Technology (FITT) as a framework that encompasses the interactions of users and technology (i.e. TAM concepts) and tasks and technology (i.e. TFF concepts), and the interaction between users and tasks [8]. FITT was developed specifically for the health care domain and was based on an analysis of literature on technology adoption. FITT posits that the use of technology is dependent on the fit between the attributes of the individuals (users), attributes of the tasks, and attributes of the technology.

The framework defines individuals as either an individual user or a user group. Examples of user attributes include: knowledge of the technology; motivation to execute tasks; openness to new ways of working; cooperation within a team; and organisational context. Tasks comprise whole tasks or work processes (such as, documentation or order entry), the attributes of which can

include: organisation of tasks; activities and their interdependence; and task complexity. Technology is defined as any tool required to execute a task; encompassing both computer-based and paper-based tools. Examples of technology attributes include: usability of the tool; functionality of the tool to support a given task; integration of tools; and availability of tools. Where fit between attributes of the users, tasks, or technology is lacking, problems with the adoption of technology arise [8].

The article where FITT was introduced was focused on a retrospective analysis of a case study that assessed the adoption of a nursing documentation system in three wards to test and validate FITT [8]. Application of FITT was shown to facilitate understanding of the relationships between users, tasks, and technology, and the factors leading to either the failure, or the successful adoption, of technology in each ward. For example, a factor that affected overall fit in the paediatric ward was identified as the fit between tasks and technology dimension, whereby the unavailability of mobile computing devices or fixed computing devices located at the patient bedside disrupted the workflow of nurses who were accustomed to undertaking documentation at the patient bedside. By pinpointing the issues affecting the use of technology, FITT helped to determine areas where changes could be introduced in order to optimise fit.

The aim of our research was to (i) evaluate clinicians' use of fixed and mobile computing devices using the FITT framework and (ii) identify factors affecting the optimal use of devices.

Methods

A mixed method approach comprising structured observation, interviews, and field notes was used for data collection. Observation can be particularly valuable for determining whether technology is used in expected or unexpected ways, while interviews can complement observation by providing clarification about what was seen in the field.

Data collection was conducted on two wards (a surgical and a general medical ward) at a 320-bed teaching hospital in Sydney, Australia. The hospital had several HITs, including: an electronic medical record (for documenting discharge summaries); computerised provider order entry (for test ordering and results viewing); an electronic medications management system; and PACS. Tasks such as recording vital sign observations and progress notes were paper-based. Both fixed (desktop computers) and mobile computing devices (computers on wheels (COWs)) were available on the wards. Some doctors also had tablet computers that were connected to the hospital's information system.

A paper-based data collection form was developed and used to capture variables regarding users, tasks, and technology, including the: activity being conducted (i.e., ward round, medication round, or outside of round); task being performed (e.g., administer medication, order medication, order test, document progress notes); technology used to perform the task (e.g., desktop computer, COW, paper medical record); and factors impeding the use of technology. The data collection form was also used to document free text field notes and interviews. Definitions for the observed work tasks were based on classifications used in previous observational studies [9; 10].

A sample of 38 clinicians (26 nurses (19 female) and 12 doctors (6 female)) were observed for 90 hours and 45 minutes. Medication administration rounds accounted for 45 hours and 10 minutes, ward rounds 28 hours and 50 minutes, and outside of rounds 16 hours and 45 minutes of the observation time. In total, 4,423 clinical tasks were recorded: 2,321 during medication administration rounds, 1,444 during ward rounds, and 658 outside of rounds. Twenty-seven clinicians also participated in informal interviews, providing explanations regarding observed events.

Clinicians were observed in the course of their daily work (between 7am and 5pm, Monday to Friday). Each observation session lasted a maximum of two hours. There were no set criteria for the selection of observed clinicians; they were chosen at random from those that were on the study ward on any given day that observations were conducted. Informal interviews were carried out opportunistically when clarification was needed and the situation allowed. Ethics approval for the research was obtained from the hospital Human Research Ethics Committee.

The quantitative data were analysed in SPSS using descriptive statistics to calculate frequencies of the collected variables, including: tasks conducted by doctors and nurses; devices used to conduct tasks; and locations in which tasks were conducted. Qualitative data from interviews and field notes were analysed for common themes, particularly regarding the factors affecting device use. FITT was applied in order to assess the relationships between users, tasks, and technology.

Results

FITT: Doctors, Tasks, and Technology

Doctors and Tasks

Doctors' work differed on ward rounds and outside of rounds. Ward rounds were conducted in teams, with doctors moving from one bedside to the next as they reviewed each patient. Ward round tasks occurred across several locations, including the patient bedside, the corridors, and while doctors were in transit between locations.

The main types of tasks doctors conducted on ward rounds included reviewing the patient record, reviewing test results, documenting progress notes, ordering medications, and ordering tests. Outside of rounds, tasks were largely conducted independently and doctors were observed conducting 93.8% (n=212) in a stationary location. The main tasks observed outside of rounds included reviewing the patient record, documenting discharge summaries, ordering medications and ordering tests.

Tasks and Technology

The hospital's hybrid information system required doctors to have access to both computing devices and paper-based medical records. For ward rounds, mobile devices fit the mobile nature in which tasks were conducted by meeting doctors' information needs as they moved throughout the ward. Of the tasks completed on a computer on ward rounds (n=762), almost all were completed with the use of mobile devices (n=759; 99.6%). The mobile cart design of the COWs also provided a convenient means to store items, allowing doctors to transport several paper-based medical records at a time.Outside of rounds, desktop computers provided a better fit. Of the computerised tasks conducted outside of rounds (n=120), the substantial majority were completed on desktop computers (n=116; 96.7%). The desks on which the computers were stationed provided space to set opened paper-based medical records so that paper-based and electronic information could be viewed at the same time.

Doctors and Technology

While conducting ward rounds doctors had one COW to use amongst the team. One of the junior doctors used the COW on the medical ward, while the senior doctor leading the ward round used the COW on the surgical ward. Thus, on the surgical ward, the junior doctors were observed using paper to document

details regarding the test orders, medication orders, or medication modifications and the information was entered electronically after the ward round.

Several doctors expressed a preference for electronically available information ,as paper-based records were often misplaced and time had to be spent searching for them. While searching for a printed pathology form that he ultimately failed to locate, *Doctor 4* stated that a benefit of information being computerised was that the pathology form could have just reprinted. Doctors also reported a preference for accessing electronic information via mobile devices, with two doctors further explaining that they needed mobile devices for ward rounds as their work practices are mobile *(Doctor 1 and 7)*. While doctors said they liked the COWs, many conveyed a desire for tablet computers: *Doctor 3* stated it would be easier to conduct his work with a tablet computer than with the COWs, which were bulky and had battery issues.

FITT: Nurses, Tasks, and Technology

Nurses and Tasks

Nurses were observed to largely undertake their work independely during medication administration rounds and outside of rounds. Nurses on medication rounds were constantly on-the-move throughout the ward, often going from a patient's bedside, to the medication room to obtain medications, and back to the bedside.

Nurses' main types of tasks on medication rounds included accessing information to prepare medications, documenting administration of medications and reviewing patient records. When conducting tasks outside of rounds nurses tended to have a base location (at a desk or parked COW) where they completed most tasks (n=382; 88.4%). The main tasks observed outside of rounds included reviewing the patient record, documenting notes and patient observations (such as vital signs).

Tasks and Technology

As with doctors, the hybrid information system required nurses to use both computing devices and paper-based medical records. Mobile devices suited mobile tasks of medication round by providing nurses access to information while moving around the ward. Of the computerised tasks on medication administration rounds (n=1,966), the substantial majority were completed using COWs (n=1,885; 95.9%). Desktop computers were considered not to be conducive to medication round tasks *(Nurse 2 and 9)* and were rarely used.

Outside of rounds both mobile and fixed computing devices appeared suited to undertaking tasks. Despite the availability of several desktop computers throughout the ward most of the computerised tasks conducted outside of rounds (n=92) were completed on COWs (n=58; 63%),. As tasks conducted outside of rounds were predominantly non-mobile, the COWs were largely observed being used in a stationary manner. Nurses were often observed in the corridor using the COW as a bench to complete paper-based records.

Nurses and Technology

Several nurses perceived that it was quicker and easier to complete tasks using paper *(Nurse 4, Nurse 5, Nurse 8, and Nurse 11)*. Nurses sometimes used a computing device to access information regarding medications, which they transcribed onto paper and then used the information to prepare the necessary medication. Nurses explained that when they only needed to prepare one or two medications (particularly for medications such as paracetamol or vitamins) they found it quicker and easier not to wheel a COW around with them *(Nurse 12, Nurse 20, and Nurse 21)*. Nurses were also observed using printed handover sheets and scrap pieces of paper to document notes, such as self-reminders, or to temporarily document vital sign observations that they later transcribed into the paper-based medical record.

The general consensus, however, was that the computer system, particularly mobile devices, provided several benefits over paper-based medical records. A key benefit identified by the nurses was the ease of access to patient information and clinical information when the need for it arose *(Nurse 12, 13, 15, 16, and 19)*. Not having to carry around several paper-based records at a time, being able to stow other necessary items in the COW (such as medications and wound dressings), and not needing to search paper-based textbooks were also seen as benefits *(Nurse 2, 12, 13, and 15)*. *Nurse 2* explained that if there were special instructions on how to administer a medication then that information would appear on the system next to the medication order, and saved time in not having to look in a textbook. The nurse also liked having a COW to conduct medication round tasks, and preferred to sit down and use the desktop computer for tasks outside of rounds.

Other Factors Affecting FITT: Environment

Temporal Rhythms

The timing of medication rounds and ward rounds and the number of available mobile devices was found to influence device use. On the surgical ward eight nurses concurrently undertook the morning medication administration rounds. As there were eight COWs on the ward, nurses explained that there was competition for the use of COWs for morning medication rounds, as doctors were conducting their morning ward round at the same time. *Nurse 8* referred to this competition as a "battle", which doctors often won. *Nurse 2* reported that when no COWs were available she would use a desktop computer and transcribe details onto paper about patients required medications so that she could take the information with her. She would then prepare the medications, administer them to her patients, and locate an available computing device to document the adminstration of the medications. Doctors similarly conveyed it was sometimes a struggle to access a COW when ward rounds occurred at the same time as medication administration rounds *(Doctor 2, 5 and 7)*.

Space Limitations

During medication rounds, nurses were generally observed positioning the COW beside the patient bedside so that they were in reach of the patient's bedside drawers (where most medications corresponding to the patient's specific needs were kept). However, several instances were observed where lack of space directly beside the patient bedside was an issue and the COW had to be positioned elsewhere in the room. Similarly, doctors on ward rounds would usually use the COW at the foot of the patients' bed but when lack of space directly at the bedside was an issue doctors had to find adequate space to use the COW elsewhere within the patient room.

Infection Control

In instances where the patient was quarantined in an infection control room both doctors and nurses were observed having to use COWs in the corridor just outside the patient's room. This often required them to walk back and forth between the patient bedside and the COW when needing to access or document information.

Low Battery

On occasions when low battery was an issue, nurses plugged the COW into an available power outlet either in the patient room or in the corridor, and moved between the patient bedside and the COW when they needed to access to document information. Doctors were observed plugging the COW into a power outlet in the corridor while they engaged in discussions

in between visiting patients. When moving on from the corridor to visit their next patient, doctors were observed unplugging the COW from the power outlet in order to take the device with them to the bedside despite the low battery.

Discussion

We found that clinicians' use of devices could be largely attributed to a relationship between attributes of the tasks and attributes of the technology. The mobile nature of ward rounds and medication administration rounds suited the mobile nature of COWs, while the substantially less mobile nature of tasks conducted outside of rounds suited the stationary nature of desktop computers. However, this relationship alone did not account for all the device use behaviours that were observed.

Ammenwerth et al.'s fit between individuals, task, and technology (FITT) framework, posits that the optimal use of technology is dependent on the interaction between three key dimensions: attributes of users, tasks, and technology [8]. A distinctive feature of the FITT framework, compared to other theoretical frameworks aimed at understanding technology use, is the emphasis on the interaction between users and tasks and the subsequent impact that this interaction has on the use of technology. Application of the FITT framework in this present study aided in the identification of distinct differences between the attributes of the observed user groups and how they conduct tasks. Nurses were found to largely conduct their work independently and, hence, could select the computing device that they perceived provided the best fit for their tasks. Doctors, on the other hand, worked in teams during ward rounds, thus device use amongst the team was influenced by the team leader. When junior doctors were not the primary users of the COW it meant they had to document patient treatment decisions on paper, which they then had to enter electronically after the ward round. These findings highlight that, irrespective of a congruent relationship between mobile computing devices and the mobile nature of ward rounds, user attributes affected optimal fit. Thus, validating Ammenwerth et al.'s argument about the importance of the user dimension when examining the use of technology.

Nonetheless, even when adequate fit between the attributes of users, tasks, and technology was attained additional factors related to the environment were found to affect the use of technology. Environmental attributes included: department type (levels and timing of ward activities); physical environment (space, layout, power outlet locations); or organisational policies and procedures (infection control requirements).

Environment Factors Affecting Use of Technology

One of the key factors found to affect the use of devices was the temporal rhythms of the ward. When the timing of ward rounds and medication administration rounds coincided it resulted in more clinicians requiring the concurrent use of COWs than were available. Nurses that were not able to access a COW reported instead having to use a desktop computer. As desktop computers were not available at the patient bedside, where information was largely needed during medication administration rounds, nurses would transcribe information from the desktop computer onto paper. Although transcribing allowed information to be taken to the bedside it also introduced the potential for errors, as well as negatively impacting efficiency as a result of the additional documentation.

The presence of infection control rooms on a ward was also found to affect the use of devices. In cases where a patient was isolated, clinicians could not take the COW into the room. Instead they had to leave the COW outside the room and walk between the patient bedside and the COW when needing to access or document information. Similarly, battery issues and lack of space at the bedside, often due to the presence of other medical equipment or furniture, impacted clinicians' ability to use COWs at the bedside. Andersen et al., who observed clinicians' use of devices on hospital wards, similarly found that lack of space was a critical factor preventing the use of COWs at the patient bedside [11]. A survey of nurses reported that a lack of space resulted in the need to undertake double documentation: using paper to document information at the bedside and then copying the data onto the COW [12].

The commonality amongst the identified environmental factors is that they restricted the ability of clinicians to use COWs at the patient bedside and, hence, impacted on the use and optimal fit of computing devices. This meant that, not only were the benefits associated with having a mobile device at the bedside, such as ease of access to information, subsequently lost but the potential for errors was introduced due to clinicians having to work around the restraints imposed by these factors. Often temporary paper resources, such as nurses' handover sheets or scrap pieces of paper were used as an interim means by which clinicians overcame factors affecting the use of computing devices. Examples included nurses transcribing information from desktop computers onto paper when COWs were unavailable or junior doctors' documentation of treatment decisions when senior doctors were using the COW. Temporary paper resources used in such instances have been described as "transitional artefacts" which are used to bridge a gap between clinical workflow needs and formal electronic documentation [13]. The persistence of temporary paper resources in such cases could potentially be decreased, or even eliminated, by addressing environmental factors hindering direct electronic information access or input. For example, evaluating ward activities to identify peak periods of demand for technology and either ensuring sufficient device availability or adjusting the timing of activities. Similarly, providing a dedicated device, that can be sanitised, within each infection control room.

Extending the FITT Framework: FITTE

The above findings highlight the importance of examining environmental factors as an entity in and of themselves and suggest the need for an extension to the FITT framework. While the dimensions of individuals, tasks, and technology were found to be critical in assessing fit, ultimately it was factors within the environment, such as the temporal rhythms of a ward, the presence of infection control rooms, or space limitations, which influenced the optimal use of technology. Presently, the FITT framework enmeshes factors related to the environment (or context) of a setting as an intrinsic part of the user attribute. Yet, context is recognised to be a critically important factor affecting the use of technology [14; 15].

The addition of a separate and overarching "environment" dimension to the FITT framework would aid in the assessment of factors related to the context in which users, tasks, and technology operate. Thus, we propose the FITTE framework: whereby optimal adoption and use of technology is determined by the fit between individuals, tasks, technology and environment (Figure 1). The distinction of environment as a separate dimension is necessary as it is likely that this is where the key differences between different sites and settings lie. As such, an environment dimension may help to explain why a technology that works in one setting does not show the same success in another setting. Future research could look at using a multi-dimensional work observation tool, such as the Work Observation Method By Activity Timing (WOMBAT) [16], to obtain data about users, tasks, technology and environment that could be evaluated through the lens of the FITTE framework.

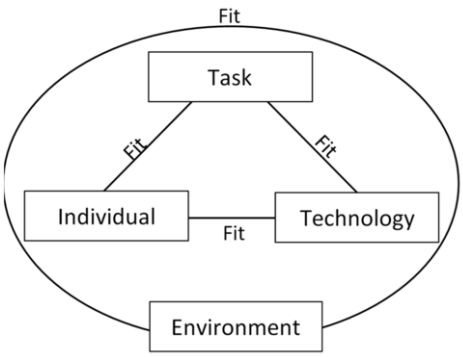


Figure 1 – FITTE Framework: Optimal Adoption and Use of Technology Depends on the Fit between Individuals, Tasks, Technology and Environment.

Limitations

As with any observational research there is a possibility of introducing the Hawthorne effect, where participants modify their behaviour in the presence of the researcher. While it cannot be known whether participants changed their behaviour, given that the focus of the study was examining how computing devices fit clinicians' work and that no assessment of quality was being made, any magnitude of behavioural change is likely to have had minimal influence on the study findings.

Conclusions

Due consideration needs to be given to all the factors that may affect device use, as technology can significantly impact the efficiency and effectiveness of clinical work practices, both in intended and unintended ways. In particular, it is important to identify the environmental nuances that may affect the ideal fit of computing devices. The FITTE framework provides a means to evaluate the use of technology by explicating the relationships between users, tasks, and technology, and the environment in which they operate to identify factors leading to either the failure of or the successful adoption of technology.

References

[1] D.F. Sittig, A. Wright, J. Ash, and H. Singh, New unintended adverse consequences of electronic health records, *Yearb Med Inform* **25** (2016), 7-12.

[2] F. Magrabi, E. Ammenwerth, H. Hyppönen, N. de Keizer, P. Nykänen, M. Rigby, P. Scott, J. Talmon, and A. Georgiou, Improving evaluation to address the unintended consequences of health information technology: a position paper from the working group on technology assessment & quality development, *Yearb Med Inform* **25** (2016), 61-69.

[3] R.J. Holden and B.T. Karsh, The technology acceptance model: its past and its future in health care, *J Biomed Inform* **43** (2010), 159-172.

[4] F.D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS Quart* **13** (1989), 319-340.

[5] M.T. Dishaw and D.M. Strong, Extending the technology acceptance model with task–technology fit constructs, *Inform Manage* **36** (1999), 9-21.

[6] D.L. Goodhue and R.L. Thompson, Task-technology fit and individual performance, *MIS Quart* **19** (1995), 213-236.

[7] L. Lepanto, C. Sicotte, and P. Lehoux, Assessing task-technology fit in a PACS upgrade: do users' and developers' appraisals converge? *J Digit Imaging* **24** (2011), 951-958.

[8] E. Ammenwerth, C. Iller, and C. Mahler, IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study, *BMC Med Inform Decis* **6** (2006), 3.

[9] J.I. Westbrook, A. Ampt, L. Kearney, and M.I. Rob, All in a day's work: an observational study to quantify how and with whom doctors on hospital wards spend their time, *Med J Australia* **188** (2008), 506-509.

[10] J.I. Westbrook, L. Li, A. Georgiou, R. Paoloni, and J. Cullen, Impact of an electronic medication management system on hospital doctors' and nurses' work: a controlled pre–post, time and motion study, *J Am Med Inform Assn* **20** (2013), 1150-1158.

[11] P. Andersen, A.-M. Lindgaard, M. Prgomet, N. Creswick, and J. Westbrook, Mobile and fixed computer use by doctors and nurses on hospital wards: multi-method study on the relationships between clinician role, clinical task, and device choice, *J Med Internet Res* **11** (2009), e32.

[12] L.E. Moody, E. Slocumb, B. Berg, and D. Jackson, Electronic health records documentation in nursing: nurses' perceptions, attitudes, and preferences, *CIN: Comput Inform Nu* **22** (2004), 337-344.

[13] Y. Chen, Documenting transitional information in EMR, in: *CHI 2010: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Atlanta, Georgia, USA, 2010, pp. 1787-1796.

[14] J.L. Callen, J. Braithwaite, and J.I. Westbrook, Contextual implementation model: a framework for assisting clinical information system implementations, *J Am Med Inform Assn* **15** (2008), 255-262.

[15] R. Pawson, T. Greenhalgh, G. Harvey, and K. Walshe, Realist review - a new method of systematic review designed for complex policy interventions, *J Health Serv Res Policy* **10 Suppl 1** (2005), 21-34.

[16] E.C. Lehnbom, M.T. Baysari, and J.I. Westbrook, Quantifying health professionals' patterns of work and communication and the impact of health information technology: a workshop on how to design and use the Work Observation Method by Activity Timing (WOMBAT) tool, in: *Medinfo 2013: The 14th World Congress on Medical & Health Informatics*, IOS Press, Copenhagen, Denmark, 2013, p. 1251.

Address for correspondence

Dr Mirela Prgomet: mirela.prgomet@mq.edu.au

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190323

A Chinese Survey of Women's Use and Expectation of Pregnancy Applications

Li Qing[a], Shan Weiying [b]

[a] *College of nursing, Chengde Medical University, Chengde, Hebei, China,*
[b] *Chengde Vocational Technical Schools, Chengde, Hebei, China*

Abstract

This study aimed to explore the use and expectation of smartphone applications about pregnancy among pregnant women using a survey method. A total of five hundreds pregnant women from 3 hospitals in the city of Chengde in hebei province in China were randomly selected to conduct a questionnaire survey. The study found that 61.48% of pregnant women used at least one pregnancy apps, mainly for fetal development information (83.3%) and maternal changes information (65.3%); only 8% of pregnant women worried about the security of personal information very much. The main expected functions of pregnancy apps are prenatal examination reminder (65%) and fetal monitoring (54.7%). Also, 77.3% of pregnant women hoped that information provided pregnancy apps should be recognized by relevant authorities. It indicated that pregnancy apps in the market in China was not able to meet the needs of pregnant women and needed further improvement in future.

Keywords:

Pregnant women; Smartphone; Mobile Applications

Introduction

In 2013, WHO listed "using mobile health applications (apps) to provide health education and behavior promotion for patients" as one of the twelve functions of mobile health [1]. In 2015, China government put forward "Internet plus" action plan and in 2016 "healthy China 2030" program outline also led the medical, nursing and other industries to use Internet information technology to promote health reform and improve public health literacy.

In recent years, with the implementation of the "comprehensive two-child" policy and the popularization of the concept of eugenics in China, the large group of pregnant and lying-in women in China will have a strong demand for health care professional guidance in the next years. As information technology play more and more roles in people's daily life, Pregnancy related apps has become one of the convenient choices for pregnant women to acquire health knowledge and promote health behavior.

Researches on apps in smart phone applied to maternal health management has been increasing all over the world since Apple launched AppStore in 2008, but most Chinese studies focused on the introduction, comparison and analysis of the function of these apps [2-4]. While a few studies involve quantitative evaluation of the application effect of such apps [5,6],The studies considering pregnant women's need and using experience are limited.

In view of that, this study investigated the pregnant women's status quo and demand for pregnancy apps, aiming to understand its current situation and existing problems, which can provide reference for future similar researches or further development of high-quality, professional pregnancy apps.

Methods

Sample

In this study, a simple random sampling method was used to select 500 pregnant women who were from 3 hospitals in Chengde from May 2017 to October, and 488 questionnaires were collected effectively, with an effective recovery rate of 97.6%. The inclusion criteria of respondents are: (1) pregnant women visiting obstetric outpatient or inpatient; (2) using smart phones; (3) understanding the contents of apps in smart phones; (4) volunteered to participate in the survey. The human ethics committee of Chengde Medical University Approved the study.

Measurements

Anonymous questionnaire was used to investigate pregnant women's usage and expectation for apps in pregnancy management. The questionnaire was designed on the basis of referring to relevant literatures, and consulting obstetric experts and software engineers. Then, after revising the questionnaire and finishing the test of reliability and validity through pre-investigation, this questionnaire could be applied to the formal investigation. The internal consistency of the questionnaire (Cronbach's alpha coefficient) was 0.781, and the content validity value (CVI) was 0.825. The main contents of the questionnaire include: (1) the general demographic characteristics of the respondents, such as age, occupation, education level, place of residence, pregnancy history, gestational weeks and so on; (2) the usage of apps in pregnancy management; (3) the expectation of apps in pregnancy management.

Data Analysis

SPSS 19.0 statistical software was used to input and analyze data. The statistical analysis methods are mainly general descriptive analysis and chi-square test.

Results

Demographic characteristics

Among 488 respondents, 300 pregnant women (61.48%) had installed and used at least one pregnancy app; therefore, this study only investigated and analyzed the use and demand of pregnancy apps among 300 pregnant women. The demographic characteristics of the total of the 300 respondents are presented

in Table 1. The age of the respondents was 26.74±0.81 years old, between 18 and 38. And 66% of pregnant women had their occupations, while 34% of them were housewives. Outpatient pregnant women accounted for 82.3%, pregnant women from ward accounted for 17.7%.

Table 1. Respondent demographics (n=300)

Age (n,%)	Residence (n,%)	Education (n,%)	gravidity (n,%)	parity (n,%)	Gestational age (n,%)
18-23: 38(12.7)	Rural: 52(17.3)	<Final high school: 36(12.0)	once: 161(53.7)	no: 184(61.3)	<12 weeks: 64(21.3)
24-38: 117(39.0)	Suburbs: 51(17.0)	High school: 69(23.0)	twice: 80(26.7)	1 time: 101(33.7)	13-28 weeks: 110(36.7)
29-33: 97(32.3)	Town: 197(65.7)	Undergraduate: 178(59.3)	3 times: 41(13.7)	2 times: 15(5.0)	29-40 weeks: 126(42.0)
34-38: 48(16.0)		graduate/above: 17(5.7)	4 times and more : 18(6.0)		

Use of pregnancy apps

The Survey showed that 216 (72.0%) of 300 pregnant women who had used pregnancy apps installed one pregnancy app and 84 (28.0%) installed two or more pregnancy apps. The frequency of the use of pregnancy apps by pregnant women was: 60.0% of these respondents had used pregnancy apps every day; 20.3% had used every week or so; 4.6% every month or so; 3.3% only once or twice; and 11.7% of them cannot remember how often they used apps.

Table 2 shows the respondents' reasons for their used of pregnancy apps. The main reason for using an app was to understand the information about fetal development (83.3%) and about changes in women's bodies during pregnancy (65.3%).

Table 2.Reasons for using pregnancy apps (respondents could nominate more than one reason)

Reason for use	n	%
Information about fetal development	250	83.3%
Information about changes in body in pregnancy	196	65.3
Tracking fetus	154	51.3
Tracking own body	108	36.0
Keeping tracking of appointments/ medical information	90	30.0
Continuous recording of prenatal examination results	54	18.0
Online discussions with other pregnant women	51	17.0
Uploading/storing pregnancy photos	5	1.7
Uploading / storing fetal ultrasound images	4	1.3

In addition, 29.67% of the respondents had unloaded pregnancy management apps after installation, and 12.7% of them did not continue to install the new apps. After reinstallation, 87.3% pregnant women still used their apps. The reason why pregnant women unloaded apps were mainly finding better apps (29.2%), fatigue after usage (15.7%), and the anxious or worried content in apps (13.5%). And 11.2% and 10.1% of pregnant women believed that the information provided by pregnancy apps was inaccurate and useless respectively.

The survey found that 40.7% of pregnant women never checked the source and accuracy of information provided by pregnancy apps. When asked whether they worried about their personal information leakage (such as name, identity card number, phone number, address, etc.) during the use of apps

Table 3. The comparison on the degree of concern of pregnant women with different demographic characteristics about the personal information leakage in pregnancy apps (n)

demographics		the degree of concern of pregnant women about information leakage					X^2	P
		never	a bit	medium	very	not sure		
Age	18~23	16	15	3	1	3	14.931	0.245
	24~28	28	53	22	5	9		
	29~33	24	42	15	12	4		
	34-38	12	21	7	6	2		
Residence	rural	24	14	7	2	5	19.087	0.014*
	suburbs	8	27	9	3	4		
	town	48	90	31	19	9		
Educational level	<high school	14	13	3	2	4	18.615	0.017*
	High school	28	24	8	6	3		
	>Undergraduate	38	94	36	16	11		
Job	in employment	43	89	35	20	11	10.476	0.033*
	housewives	37	42	12	4	7		
parity	primipara	45	89	24	16	10	5.913	0.206
	multigravida	35	42	23	8	8		

* P<0.05, there was a statistically significant difference.

or not, 26.7% of pregnant women replied "never worry about it"; 43.7% felt "a little worried", 15.7% "worried generally", only 8% "very concerned", and 6% of them were not sure about this problem. Moreover, there was a statistically significant difference in the degree of concern about the safety of personal information when using apps in different places of residence, educational level and job conditions, as shown in Table 3.

Expectation for a pregnancy app

According to this survey, 96% of pregnant women would like to install a completely free pregnancy app in their phone, and only 4% of them could accept a paid high-quality version of app. For the interface characteristics of app, 62.7% of pregnant women hoped that the interface should be simple and easy to operate, and 36.3% of respondents also desired no-implanted-advertisements apps. Table 4 showed ideal feature in a pregnancy app.

Table 4. Ideal features in a pregnancy app (respondents could nominate more than one feature)

Feature	n	%
Keeping tracking of appointments/ medical information	195	65.0
Fetal movement monitoring	164	54.7
Common symptoms and mitigation methods during pregnancy	137	45.7
making an appointment for registration	131	43.7
Health care guidance for pregnancy, diet, exercise and rest	129	43.0
Information about changes in body in pregnancy and fetal development	112	37.3
Monitoring of uterine contraction	106	35.3
Weight monitoring	97	32.2
Recommending a doctor	64	21.3
Communication platform with other expectant mothers / new mothers	49	16.3
Timely feedback and guidance to users' questions	46	15.3
Entertainment (Mother and baby mall, games)	35	11.7
Name the baby	35	11.7
Predict baby image	27	9.0

In addition, 61.7% of pregnant women did not want the using app involved upload or share personal information, 29.3% of them felt that they were willing to upload or share personal data if the security of user information can be enough ensured. And 77.3% of the respondents hoped that the information provided by pregnancy apps should be approved by the relevant authorities while 16% of the pregnant women did not take this for granted. Moreover, there is no statistically significant difference among pregnant women with different age, educational level and residence on the demand of security, privacy and authoritative information about pregnancy apps.

Discussion

This study showed that utilization rate of apps for pregnant women in Chengde in China was 61.48%, which was lower than that of apps (73%) in Australia in 2016 [7], but higher than that of pregnant women (55.4%) in Korea in 2015 [8] and the utilization rate of pregnant women (59%) in Ireland in 2015 (59%) [9]. In this study, 60% of pregnant women used pregnancy apps every day, while only 26% of pregnant women in Australia used it every day [7]. A qualitative study in Germany showed that 60% of pregnant women used apps (1 times / week) regularly [10]. This may be related to the using habits of mobile phones in different cultures and social backgrounds.

The popular reasons for pregnancy apps was to acquire the information of fetal development (83.3%) and changes of body during pregnancy (65.3%), and 30% of pregnant women used apps to remind prenatal examination during pregnancy Keeping and track appointments, which was similar to the results of survey made by Lupton [7]. The study of Yeonkyu [8] in South Korea indicated that the most frequent use of pregnancy apps among pregnant women tended to be the risk of pregnancy and the symptoms of pregnancy. A study in India also showed that the frequency of mobile phone app used by mothers in families with higher or lower family income was higher than that in families with middle or lower income. It may be due to the fact that high income families were not willing to share their experience with others through mobile apps [11].

Table 2 showed that only 8% of pregnant women were very concerned the safety of personal information when using apps. Moreover, the degree of worry of rural women was lower than that of rural and urban residents. Therefore more attention of pregnant women's privacy and information safety in mobile health needs to be paid, especially in pregnant women with low educational level, full-time mothers and rural women.

In the study, 29.2% of pregnant women had to unload former app if they found better similar app, and 11.2% of pregnant women believed that the information provided by some apps was inaccurate, and 40.7% of pregnant women never checked the source and accuracy of information provided by pregnancy apps. All of these hinted that the quality of apps in Chinese market was not uniform, the users' information security literacy was relatively low, and the lack of relevant apps supervision, management and evaluation mechanism. In 2015, BinDhim [12] and other studies also found that although the US Food and Drug Administration (FDA) and the British National Health Service (NHS) had established a certain health related apps evaluation mechanism, 77% of users did not evaluate the credibility of apps developers, which further indicated the importance and urgency of improving the public's e-health literacy.

Most pregnant women wanted to install free apps for pregnancy management, which embodied the public's need for the accessibility and popularity of mobile health. 62.7% of pregnant women hoped that the interface is simple and easy to operate, and 36.3% of pregnant women also desired that apps would not be implanted with advertisements. A survey in Germany also showed that all respondents preferred an easy-to-use interface in Wed-based applications and did not want to be held up with time-consuming technical issues [10]. A South Korean survey on the needs of pregnant women's dietary guidelines app found that the right age pregnant women had more ability to access to information, higher demand for app's design, convenience and content diversity than those with advanced age [13].

As for the main feature of pregnancy apps, 65% and 54.7% of pregnant women would like to have the function of antenatal examination and fetal monitoring, which was not consistent with the main reasons of apps in the study. This indicated that pregnancy apps in Chinese market had not been yet able to meet the actual needs of pregnant women, present pregnancy apps needed to be further improved. Some similar apps products in

other countries which include gestational apps with increased fetal movement monitoring, uterine contraction monitoring, weight and blood pressure monitoring, and gestational diabetes diet management as well as providing health education for pregnancy.

Yeonkyu's research showed that pregnant women had more needs for expert opinions on diet and drug management in pregnancy apps during gestation period [8]. Two Australian researches also showed that pregnant women had a strong desire for health care information provided by health care professionals and gynecologist experts [14]; and hoped that apps could help them to assess their weight, nutrition and health status and it should be linked to reliable websites, which could provide short answers to their daily concerns [15]. Molly's study showed that 89% of pregnant women had a clear desire to use smart phone app to guide them to maintain healthy weight during pregnancy in united states [16]. According to Maren' study of in Germany, 87% of pregnant women wanted applications to be implemented in routine pregnancy care in order to detect and prevent serious pregnancy conditions already at an early age; besides, 23% of women requested sharing data among health care professionals(ie, physician and midwives) or even health insurance companies through standardized networks, in that case they would not have to fill out another form each time [10].

Furthermore, 61.7% of the respondents suggested applications should not be involved in uploading or sharing personal information. And 77.3% of pregnant women hoped that the information provided by pregnancy apps needed the approval of the relevant authorities. So health care organization and authoritative agency in China should pay more attention to the supervision and legislation about mobile health field, in view of the specific requirements for the design, function, supervision and evaluation system of apps in future.

Conclusions

The usage rate of pregnancy apps among pregnant women is acceptable, pregnant women's consciousness for information safety is low, and the demand for professional and authoritative apps is high. There are differences between the reasons and needs of apps use. It indicates that pregnancy apps in the market in China is not able to meet the needs of pregnant women, so the function of pregnant apps needs further improvement.

This study chose only pregnant women who took routinely antenatal care in 3 top hospitals in city of Chengde, not involving in those who had no regular prenatal appointment and those who lived in the remote areas in this city due to the limit of time, manpower and material resources. However, these groups of pregnant women are more likely to suffer from some adverse pregnancy outcomes, such as low birth weight newborn, gestational complications, neonatal death, etc. Under the background of the rapid development of information technology and mobile health globally, maternal and child health in impoverished and remote regions should be paid more attention and benefited a lot. Therefore, further researches will be required to explore suitable pregnancy applications for pregnant women in consideration of different demographic characteristics as well as their using experience and expectations.

Acknowledgements

Funding for this study was provided by Educational Department in Hebei Province in China.

References

[1] Labrique AB, Vasudevan L, Kochi E, et al. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract. (2013) 1:160-171.

[2] Wu Qianlan, Wang Jing, He Xiuyu, et al. Application of mobile medical app in maternal and child health management. Chinese Journal of health information management (2015) 12: 217-220.

[3] Wang Fang, Zhu Bing Hua, Xu Xinfen. Problems in the application of app in "palm pregnant school" and coping strategies. nursing and rehabilitation (2016) 15: 1067-1069.

[4] Ding Lachun, Wang, Shi Lei, et al. The development and application of maternal and child mobile medical health platform "glebo". Journal of medical informatics (2016) 37: 40-43.

[5] Zhu Xu, Yu Kai, Tao Jing. Maternal and child health manual app as a carrier to achieve personalized health education for pregnant women. health research (2017), 37: 241-243.

[6] Qu Xin Yun, Wang Hongying, Zhou run, et al. Application analysis of health assistant App in maternal health management. Chinese digital medicine (2017) 12: 118-120.

[7]Deborah Lupton,Sarah Pedersen..An Australian survey of women's use of pregnancy and parenting apps.Women and Birth(2016)29:368–375.

[8] Yeonkyu Lee, Mikyung Moon. Utilization and Content Evaluation of Mobile Applications for Pregnancy, Birth, and Child Care. Healthc Inform Res (2016) 22:73-80.

[9] A.O'Higgins, O.Murphy, A.Egan,et al.The use of digital media by women using the maternity services in a developed country.Ir Med J(2015)5:213-222

[10]Maren Goetz,Mitho Müller, Lina Maria Matthies,et al. Perceptions of Patient Engagement Applications During Pregnancy: A Qualitative Assessment of the Patient's Perspective. JMIR Mhealth Uhealth(2017) 5: e73-85.

[11] Jayaseelan R, Pichandy C, Rushandramani D. et al. Usage of Smartphone Apps by Women on their Maternal Life. J Mass Communicat Journalism (2015) 5:267-273.

[12] BinDhim, Hawkey,Trevena. A systematic review of quality assessment methods for smartphone health apps. Telemedicine and e-Health (2015) 21:1-8.

[13] S. B. Kim, J.WKim, M.H Kim,et al.A study on Consumer's Needs for Development of Diet Guide Application for Pregnant Women.Korean J Community(2013)8:588-598.

[14] Deborah Lupton. The use and value of digital media for information about pregnancy and early motherhood: a focus group study. Pregnancy and Childbirth (2016) 16:171-181.

[15] L.Hearn, M. Miller, L. Lester.Reaching perinatal women online: the Healthy You, Healthy Baby website and app. J Obesity(2014)10:57-68.

[16] Molly EWaring,Tiffany A Moore Simas , Rui S Xiao.Pregnant women's interest in a website or mobile application for healthy gestational weight gain. Sexual & Reproductive Healthcare (2014) 5:182–184.

Address for correspondence

Corresponding author : Shanweiying , email: chengdeuyn@163.com
Phone number: 86-13831468160

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190324

How and in what Contexts Does Networked Health IT Improve Patient Safety? Elicitation of Theories from the Literature

Rebecca Randell[a], Maysam Abdulwahid[b], Joanne Greenhalgh[c], Natalie King[b], Judy M. Wright[b], Justin Keen[b]

[a] School of Healthcare, University of Leeds, Leeds, UK,
[b] Leeds Institute of Health Sciences, University of Leeds, Leeds, UK,
[c] School of Sociology & Social Policy, University of Leeds, Leeds, UK

Abstract

Healthcare systems worldwide are investing in networked health IT systems that link healthcare providers across multiple organisations. Much of the policy arguments in favour of such investment rely on the assumption that networked health IT will lead to improved patient safety. As part of the first stage of a realist review to determine how and in what contexts networked, inter-organisational health IT does lead to improved patient safety, we elicited stakeholders' theories from the literature that reveal possible answers to this question. A key mechanism appears to be that the information provided supports improved decision making. Greatest benefits are likely to be found in relation to medication information, in scenarios where the patient is less able to provide accurate information about their medications themselves. However, access and use of this information depends on ease of access, clinicians' perception of the likelihood that the desired information will be available, and clinicians' trust in the information.

Keywords:

Health Information Technology; Health Information Exchange; Patient Safety.

Introduction

Healthcare systems worldwide are investing in networked health IT (HIT) systems that link healthcare providers across multiple organisations. For example, large-scale shared electronic health record projects have been undertaken in the United Kingdom (UK), the United States (US), Canada, Australia, Sweden, Estonia, Singapore, and Hong Kong [1; 2]. Much of the policy arguments in favour of such investment rely on the belief that networked, inter-organisational HIT will lead to improved patient safety [3-7], defined by the World Health Organisation as 'the prevention of errors and adverse events to patients associated with healthcare' [8].

How networked, inter-organisational HIT will lead to such improvements is rarely explicated in the policy literature, beyond the assumption that if clinicians have access to more information they will access and use that information, which in turn will result in better decisions and safer patient care. For example, in 2012 the Department of Health in England published a document entitled 'The Power of Information: Putting all of us in control of the health and care information we need', which set out a ten-year framework for transforming information for health and care and relies on the notion that 'not sharing information has the potential to do more harm than sharing it' [3]. Similarly, a report by the US Department of Health suggests that health information exchange (HIE) can improve safety 'by improving the timeliness and completeness of important patient health information' [5].

At present, there is a lack of evidence to support these claims [1]. Others have previously noted that networked, inter-organisational HIT is a complex intervention [9], meaning that it is aimed at producing change in the delivery and organisation of healthcare services and comprises a number of separate components that may act both independently and interdependently [10; 11]. These components are not only technological but also organisational and social, and they can all impact the extent to which the technology is successfully introduced and subsequent process and patient outcomes. It could be argued that networked, inter-organisational HIT is more complicated than a complex intervention because it spans several settings, with distinct organisational and social cultures and norms in each one. Previous research has revealed that there is significant variation in the use of information provided by networked, inter-organisational HIT, in terms of the amount and type of information that is accessed [9]. Given the complexity of the intervention, such variation is to be expected and raises the question: how and in what contexts does networked, inter-organisational HIT lead to improved patient safety?

We are currently undertaking a review of the literature with the purpose of answering this question. Using the methodology of realist reviews [12; 13], we will elicit, test, and ultimately refine theories on this topic. In this paper, we report on findings from the first stage of the review, the theory elicitation stage. These theories will be tested, using evidence from empirical studies, in subsequent stages of the review.

Methods

Realist review is an approach to synthesising evidence that represents a divergence from traditional systematic review methodology. Realist reviews identify theories of how an intervention is intended to work, for whom, and in what circumstances, and then test and refine those theories through consideration of primary studies [12]. For realists, interventions themselves do not produce outcomes. Rather, interventions offer resources; outcomes depend on how recipients respond to those resources, which is highly dependent on context. Realist theories, referred to as Context Mechanism Outcome (CMO) configurations, explain how different contexts trigger particular mechanisms (the reasoning and responses of recipients) which, in turn, give rise to a particular pattern of outcomes, where C + M = O. For example, from a realist perspective, networked, inter-organisational HIT in and of itself will not result in

improved patient safety. Rather, it is how clinicians respond to and make use of (or not) the resources that networked, inter-organisational HIT provides that will determine the impact on patient safety and how they will respond is likely to vary according to context, such that a doctor in a busy emergency department may respond differently than a nurse in an outpatient clinic. Realist approaches have much to offer the health informatics community, providing a means to not only determine if HIT interventions provide benefit in terms of outcomes, but to understand why and in what contexts such benefits may occur [14].

A realist review involves several stages. An important initial stage in a realist review is 'theory elicitation', where reviewers explore the literature with the explicit purpose of identifying theories [13]. It is only once the theories have been identified that identification of primary studies takes place. Searching should be purposive and iterative, driven not by the intervention but by the theories. This can provide particular benefit when undertaking a review on a topic where there is limited evidence, as is the case with networked, inter-organisational HIT, because the reviewer can draw on evidence from other domains where the intervention is different but the underlying theory remains the same. For example, networked IT systems to support the exchange of data between organisations have been introduced in a range of industries, such as government, manufacturing, and banking, for the purpose of process improvement, which may be based on similar theories of how networked, inter-organisational IT can lead to benefit [15].

Here we report findings from the theory elicitation stage of the review. Three main searches were undertaken for this purpose, one focusing on government policies and official reports, one focusing on opinion leaders in the area of HIT and patient safety, and one focusing on academic and practitioner literature concerned with networked, inter-organisational HIT and patient safety.

Search Strategy

Searches were conducted using synonyms for HIT, e.g. medical records, combined with synonyms for networked IT, e.g. computer networks; Health Information Exchange (HIE), defined as "the electronic movement of health-related information among organizations according to nationally recognized standards" [16]; and interoperability, defined as "the ability of different information technology systems and software applications to communicate, exchange data, and use the information that has been exchanged" [17]. In some searches, these were combined with synonyms for patient safety, e.g. adverse events, errors. These searches were conducted in February and March 2018 on the following databases: Medline (1946 to present), Embase (1996 to present), and Health Management Information Consortium (1983 to present). Full details of planned searches are available via PROSPERO (CRD42017073004).

Inclusion Criteria

Records were first screened based on title and abstract and then, where available, full papers of potentially relevant records were retrieved and screened. We aimed to identify papers that described stakeholders' theories about how and in what circumstances introduction and use of networked, inter-organisational HIT leads to improved patient safety. We did not restrict our attention to a particular form of networked, inter-organisational HIT such as HIE. No restrictions were placed on the type of healthcare setting.

Data Extraction, Analysis, and Synthesis

To provide an overview of the relevant articles, a short description of each was presented and summarised in a table format. Furthermore, we abstracted out any theories or assumptions, or fragments of theories, concerning the mechanisms through which networked, inter-organisational HIT improves patient safety and/or the contexts in which this may occur. Given the focus of this phase of the review on eliciting theories rather than testing them, when considering empirical studies we focused on the discussion sections, in an attempt to identify authors' theories about why networked, inter-organisational HIT did or did not result in the intended outcomes. In an attempt to construct initial theories, similar theories or theory fragments were grouped together.

Results

The searches reported here identified 375 records, of which 34 records were found relevant. For two of these, only abstracts were available, leaving 32 articles. Stakeholders' theories are likely to be found in editorials, letters, commentaries, and news articles and so these are often the focus of the theory elicitation stage of a realist review [13]. This was the case with our review; the majority of publications were editorials [18], letters to the editor [19-22], commentaries [23-25], and news articles [26]. However, the publications also included original research studies that sought the opinions of HIT policy and opinion leaders [27] and clinicians [28; 29], reports on experiences and lessons learned from the introduction of networked, inter-organisational HIT [30-33], and two systematic reviews [1; 34]. The publications covered a range of networked HIT, including shared EHRs [1; 18; 22-24] and networked picture archiving and communications systems [19].

The included articles discussed barriers to the introduction of networked, inter-organisational HIT – e.g. patient consent to sharing, cost, incompatibility of systems, information held within paper records – as well as drivers for it, such as financial incentives and patient expectations [1; 15; 18-20; 24; 25; 29; 31-35]. However, our concern was not with what constrains or leads to the introduction of networked, inter-organisational HIT but, once it is in place, the contextual factors that support and constrain its use and subsequent impact.

While some articles considered potential risks to patient safety that may be introduced by use of networked, inter-organisational HIT [22; 31], largely the articles reflected the same belief in the potential for improved patient safety that is promoted within the policy arena. Similar to the policy literature, how this would be achieved was often not explicated. For example, an interview study with Canadian HIT policy and opinion leaders reported that:

'clinical data sharing across the continuum of care was believed to be critical for improving safety and effectiveness, especially electronic prescribing and drug management in the near term.' [27]

Only one of the articles referred to a theoretical model that might explain the impact of networked, inter-organisational HIT. Bowden & Coiera [1], in their systematic review of the impact of accessing primary care records during unscheduled care, refer to information value theory [36], which would suggest that networked, inter-organisational HIT can only have impacts on care when the information it provides to clinicians triggers a change in a decision that has the potential for a better (higher value) outcome.

Despite the lack of explicit theory within the remaining articles, we were able to identify two key mechanisms through which authors anticipated that networked, inter-organisational HIT

would lead to improved patient safety: through clinicians making use of the information provided by networked, inter-organisational HIT to inform their decisions about patient care and through clinicians making use of networked, inter-organisational HIT to better coordinate patient care. We consider these two mechanisms in further detail below.

Improved Decision Making

A key anticipated mechanism is that clinicians will respond to the provision of accurate patient information by using that information in their decision making, resulting in improved decision making – although what is meant by 'improved' is rarely articulated – and consequently increased patient safety. For example, Alvarez [23] states:

'providing access to reliable electronic patient encounter data will result in improved diagnostic capability for providers and consequently more appropriate treatment.' (p.34)

A context where networked, inter-organisational HIT was considered to be particularly beneficial for decision making was the emergency department [1; 37; 38], due to the lack of up-to-date medical records at the point of care [31].

In terms of the information to be accessed, a patient's medication history was considered to be particularly important [23; 29; 31]. This was especially the case for patients with mental health issues, where information regarding mental health medications was perceived not only to be critical for decision making but often difficult to obtain accurately from patients [31]. Similarly, information on medications for elderly patients was seen as important, again due to anticipated difficulties in obtaining accurate information from the patient themselves. However, what information will be accessed is likely to depend on the stage in the patient journey, with information on medications, allergies, and diagnoses being the focus during triage and immediate treatment, while access to the full patient record is potentially useful later in the patient's care [1].

Other contextual factors that appear to determine whether information will be accessed and used include the ease of accessing patient information and the clinician's perception of the likelihood that the desired information will be available [31]. Where ease of access is not achieved, this may be overcome by having other staff, such as those in training, searching for information. Related to ease of access is the extent to which the networked HIT is integrated into existing workflows [26; 31]. To use the information, clinicians have to be confident that the information is accurate and up to date [22; 31; 34].

Experience of individuals may also influence the likelihood of clinicians accessing information via networked, inter-organisational HIT, with those with experience of using networked HIT typically being more positive than those without [34]. The benefits to be obtained may also vary according to levels of experience and specialism. For example, Alvarez [23] suggests that sharing of radiology images will benefit smaller hospitals by providing them with timely access to high-quality interpretations by radiology specialists.

Improved Coordination of Care

When reviewing the literature retrieved using the search term 'interoperability', an additional mechanism was identified, whereby the ability to share information provided by networked, inter-organisational HIT is used as a means of communication, leading to improved patient safety through increased coordination of care. While we were only able to elicit theory fragments, we report it here because it represents an alternative theory to the one concerning improved decision making that underlies much of the policy literature.

The e-Health Stakeholder group, a multidisciplinary group established in 2012 with the aim of discussing and contributing to the development of HIT policy at EU level, published a report entitled 'Perspectives and Recommendations on Interoperability' [2]. The report suggests that faster access to patient health records not only enables better decision making but also improved care coordination between multiple clinicians. Because of the fragmented nature of healthcare, where a patient's journey can involve multiple clinicians, there is the potential for miscommunication or error, with communication breakdown or failures in healthcare being one of the most frequent causes of adverse events [30; 39]. Networked, inter-organisational HIT can facilitate communication between clinicians working in different organisations through, for example, the transfer of hospital discharge reports to a patient's general practitioner (GP) or requests from the GP for a hospital appointment, to improve the coordination of care [30].

Beyond the scenarios described above, we were unable to elicit much from the literature regarding the contexts in which this mechanism would be triggered. However, in contrast to the decision making mechanism described above, it appears that this mechanism has less relevance to unplanned care and greater relevance to longer term care, such as palliative care and management of long term conditions such as diabetes [30]. We can also anticipate that some of the contextual factors identified in relation to the decision making mechanism, regarding ease of access of information and trust in that information, also apply here, in order for the clinician receiving the information to incorporate it into care planning for the patient.

Discussion

We have undertaken the first part of a realist review to identify stakeholders' theories regarding how and in what contexts networked, inter-organisational HIT may result in improved patient safety. The findings reveal two possible mechanisms through which improved patient safety may be achieved, one concerned with decision making and one concerned with care coordination, the relevance of which depends on the care context. Drawing together the theory fragments from the literature, two initial theories, formulated as CMO configurations, are presented in Table 1. Given realist evaluation's concern with identifying what works, for whom, in what circumstances, these theories describe what is needed to produce a positive outcome. The implication is that, in the absence of the necessary contextual factors, the mechanism that produces the desired outcome will not be triggered.

Recommendations for Future Work

Evaluation of complex interventions requires a strong theoretical foundation [40]. Bowden & Coiera [1] argue that future evaluations of networked, inter-organisational HIT need to based on appropriate theory, something that is absent in previous studies. We would agree with this and add that, ideally, not only the evaluation but also the introduction of networked, inter-organisational HIT, and HIT more generally, should be based on appropriate theory that explicates how the intended benefits are expected to be achieved. Doing so provides a way for knowledge, in terms of what works and how, to cumulate; if we become explicit about the theories that underlie the introduction of HIT, we can then test those theories, using the refined theories to inform future implementations.

Table 1 – Initial theories

Context		Mechanism			Outcome
		Resource	Response		
Emergency care Patient is unable to provide accurate medication information Information is easy to access, accurate and up to date	+	Access to medication information	Clinicians access medication information and, trusting that information, use it to inform their decision making	=	Improved decision making Reduced medication errors Increased patient safety
Long term care provided by clinicians spread across multiple organisations Information is easy to access, accurate and up to date	+	Ability to share information	On receiving information, clinicians access it and, trusting that information, incorporate it into their care planning for the patient	=	Improved coordination of care Increased patient safety

Strengths and Limitations

A strength of this work is that we have demonstrated how, when the introduction of HIT is not based on explicit theory concerning how the intended benefits will be achieved, the theory elicitation stage of a realist review provides a means of unearthing stakeholders' theories. The resulting theories can be tested and refined through the use of primary studies, as we will do, or they can be tested and refined through the collection of empirical data [41].

Nonetheless, what is presented here is only the first stage of a realist review and so we can make no claims about the truth of the theories that we have elicited from the literature. However, while the initial theories do not necessarily reflect our views, they do reflect commonly held views in one or more academic and practitioner communities.

Conclusions

Worldwide, there are efforts to introduce networked, inter-organisational HIT. While such HIT promises many benefits in terms of patient safety, these are not always achieved. We have undertaken a realist review to identify stakeholders' theories regarding how and in what contexts networked, inter-organisational HIT may result in improved patient safety. One of the key mechanisms identified in the literature is that access to 'additional' information available through networked inter-organisational IT systems can support enhanced decision making. This mechanism was more likely to yield benefits in relation to medication information, particularly in scenarios where the patient is less capable to provide accurate information themselves. However, different factors can determine the clinician's decision to access and use these systems, such as ease of accessibility, perceived usefulness of the information provided, and their trust that the information is available, accurate, and up-to-date.

Acknowledgements

This research is funded by the National Institute for Health Research (NIHR) Health Services and Delivery Research (HS&DR) Programme (project number 16/53/03). The views and opinions expressed are those of the authors and do not necessarily reflect those of the HS&DR Programme, NIHR, NHS or the Department of Health.

References

[1] T. Bowden and E. Coiera, The role and benefits of accessing primary care patient records during unscheduled care: a systematic review, *BMC Medical Informatics & Decision Making* **17** (2017), 138.
[2] E-health Stakeholder Group, Perspectives and Recommendations on Interoperability in: A. Eadie, ed., E-health Stakeholder Group, 2014 p. 33.
[3] Department of Health, *The Power of Information: Putting all of us in control of the health and care information we need*, Department of Health, London, 2012.
[4] HIQA, Overview of Healthcare Interoperability Standards, (2013).
[5] D.J. Mark L. Graber, Robert Bailey, Report of the Evidence on Health IT Safety and Interventions, in, 2016.
[6] Institute of Medicine, *Health IT and patient safety: Building Safer Systems for Better Care*, National Academies Press (US), Washington (DC), 2011.
[7] National Information Board, Personalised health and care 2020 : patient, carers and service user vision, in, Department of Health, London, 2017.
[8] WHO, The World Health Organisation. Definition of Patient Safety, in, 2018.
[9] J.R. Vest and E. Abramson, Organizational Uses of Health Information Exchange to Change Cost and Utilization Outcomes: A Typology from a Multi-Site Qualitative Analysis, *AMIA ... Annual Symposium proceedings. AMIA Symposium* **2015** (2015), 1260-1268.
[10] M. Campbell, R. Fitzpatrick, A. Haines, A.L. Kinmonth, P. Sandercock, D. Spiegelhalter, and P. Tyrer, Framework for design and evaluation of complex interventions to improve health, *BMJ* **321** (2000), 694-696.
[11] Medical Research Council, A framework for development and evaluation of RCTs for complex interventions to improve health, in, Medical Research Council, London, 2000.
[12] R. Pawson, *Evidence-based policy: A realist perspective*, SAGE, London, 2006.
[13] R. Pawson, T. Greenhalgh, G. Harvey, and K. Walshe, Realist review – a new method of systematic review designed for complex policy interventions, *Journal of Health Services Research and Policy* **10** (2005), 21-34.
[14] R. Randell, J. Greenhalgh, and D. Dowding, Using realist reviews to understand how health IT works, for whom,

and in what circumstances, *Journal of the American Medical Informatics Association* (2015).
[15] J.P. Glaser and H.G. Lo, Concepts for building inter-organizational systems in healthcare: lessons from other industries, *Journal of Healthcare Information Management* **20** (2006), 54-62.
[16] N.A.f.H.I. Technology, Report to the Office of the National Coordinator for Health Information Technology on defining key health information technology terms, in, Office of the National Coordinator for Health Information Technology, US Department of Health and Human Services Washington, DC, 2008.
[17] Healthcare Information and Management Systems Society, Definition of Interoperability, in, 2013.
[18] E. Rynning, Public trust and privacy in shared electronic health records, *European Journal of Health Law* **14** (2007), 105-112.
[19] R. Chelvarajah, S. Bavetta, and J. Benjamin, Patient safety and image transfer between referring hospitals and neurosciences centres: an alternate best practice?, *British Journal of Neurosurgery* **25** (2011), 144-145.
[20] J. Tooker, Health information technology: Improving quality and value of patient care, *MedGenMed Medscape General Medicine* **7 (2) (no pagination)** (2005).
[21] P.G. Wood, Reader responds to "Could computerization harm patient safety?", *MedGenMed Medscape General Medicine* **8 (3) (no pagination)** (2006).
[22] M. Hawking, The Single Shared Electronic Patient Record (SSEPR): problems with functionality and governance, *Informatics in Primary Care* **16** (2008), 157-158; discussion 159.
[23] R. Alvarez, The electronic health record: a leap forward in patient safety, *Healthcarepapers* **5** (2004), 33-36; discussion 82-34.
[24] C.M. Cotter, Making the case for a clinical information system: the chief information officer view, *Journal of Critical Care* **22** (2007), 56-65.
[25] D. Hillblom, A. Schueth, S.M. Robertson, L. Topor, and G. Low, The impact of information technology on managed care pharmacy: Today and tomorrow, *Journal of Managed Care and Specialty Pharmacy* **20** (2014), 1073-1079.
[26] K. Traynor, Solid data lacking on HIT and patient safety, *American Journal of Health-System Pharmacy* **69** (2012), 91-92.
[27] E. Zimlichman, R. Rozenblum, C.A. Salzberg, Y. Jang, M. Tamblyn, R. Tamblyn, and D.W. Bates, Lessons from the Canadian national health information technology plan for the United States: Opinions of key Canadian experts, *Journal of the American Medical Informatics Association* **19** (2012), 453-459.
[28] K. Ishikawa, H. Ohmichi, Y. Umesato, H. Terasaki, H. Tsukuma, N. Iwata, T. Tanaka, A. Kawamura, K. Sakata, T. Sainohara, M. Sugimura, N. Konishi, R. Umemoto, S. Mase, S. Takesue, and M. Tooya, The guideline of the personal health data structure to secure safety healthcare. The balance between use and protection to satisfy the patients' needs, *International Journal of Medical Informatics* **76** (2007), 412-418.
[29] P. Fontaine, T. Zink, R.G. Boyle, and J. Kralewski, Health information exchange: participation by Minnesota primary care practices, *Archives of Internal Medicine* **170** (2010), 622-629.
[30] D. Font, J. Escarrabill, M. Gomez, R. Ruiz, B. Enfedaque, and X. Altimiras, Integrated Health Care Barcelona Esquerra (Ais-Be): A Global View of Organisational Development, Re-Engineering of Processes and Improvement of the Information Systems. The Role of the Tertiary University Hospital in the Transformation, *International Journal of Integrated Care [Electronic Resource]* **16** (2016), 8.
[31] L.K. Gottlieb, E.M. Stone, D. Stone, L.A. Dunbrack, and J. Calladine, Regulatory and policy barriers to effective clinical data exchange: lessons learned from MedsInfo-ED, *Health Affairs* **24** (2005), 1197-1204.
[32] A.H. Goroll, S.R. Simon, M. Tripathi, C. Ascenzo, and D.W. Bates, Community-wide implementation of health information technology: the Massachusetts eHealth Collaborative experience, *Journal of the American Medical Informatics Association* **16** (2009), 132-139.
[33] A. Berler, A. Tagaris, and C. Chronaki, European Patient Summary Guideline: Focus on Greece, *Studies in Health Technology & Informatics* **224** (2016), 1-6.
[34] Y.M. Hopf, C. Bond, J. Francis, J. Haughney, and P.J. Helms, Views of healthcare professionals to linkage of routinely collected healthcare data: a systematic literature review, *Journal of the American Medical Informatics Association* **21** (2014), e6-10.
[35] Committee on Patient Safety Quality & Improvement and Committee on Practice Management, Committee opinion no. 621: Patient safety and health information technology, *Obstetrics & Gynecology* **125** (2015), 282-283.
[36] R.A. Howard, Information Value Theory, *IEEE Transactions on Systems Science and Cybernetics* **2** (1966), 22-26.
[37] C. Chronaki, A. Estelrich, G. Cangioli, M. Melgara, D. Kalra, Z. Gonzaga, L. Garber, E. Blechman, J. Ferguson, and S. Kay, Interoperability standards enabling cross-border patient summary exchange, *Studies in Health Technology and Informatics* **205** (2014), 256-260.
[38] J.A. Jacob, On the Road to Interoperability, Public and Private Organizations Work to Connect Health Care Data, *JAMA* **314** (2015), 1213-1215.
[39] C. Foisey, 4 Ways Technology Is Improving Patient Safety, (2017).
[40] E. Murray, S. Treweek, C. Pope, A. MacFarlane, L. Ballini, C. Dowrick, T. Finch, A. Kennedy, F. Mair, C. O'Donnell, B. Ong, T. Rapley, A. Rogers, and C. May, Normalisation process theory: a framework for developing, evaluating and implementing complex interventions, *BMC Medicine* **8** (2010), 63.
[41] R. Randell, S. Honey, N. Alvarado, A. Pearman, J. Greenhalgh, A. Long, P. Gardner, A. Gill, D. Jayne, and D. Dowding, Embedding robotic surgery into routine practice and impacts on communication and decision making: a review of the experience of surgical teams, *Cognition, Technology & Work* **18** (2016), 423-437.

Address for Correspondence

Rebecca Randell, School of Healthcare, University of Leeds, Leeds LS2 9UT, UK. Email: r.randell@leeds.ac.uk

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190325

An Ontology-Based Personalized Decision Support System for Use in the Complex Chronically Ill Patient

E. Román-Villarán[a], F.P. Pérez-Leon[a], G. A. Escobar-Rodriguez[a], A. Martínez-García[a], C. Álvarez-Romero[a], C.L. Parra-Calderón[a, b]

[a]*Group for Research and Innovation in Biomedical Informatics, Biomedical Engineering, and Health Economy. Institute of Biomedicine of Seville, IBiS/"Virgen del Rocío" University Hospital /CSIC/University of Seville, Seville, Spain.*
[b]*Head of Innovation Technology, "Virgen del Rocío" University Hospital, Seville, Spain.*

Abstract

Management of the complex chronically ill patient is highly problematic. This is due to the need to complement recommendations in clinical guidelines with recommendations based on treatments performed on a representative set of patients. It is necessary to identify conflicts between the recommendations of different guidelines designed for handling specific chronic diseases. The PITeS-TIiSS project (code PI15/01213) aims to overcome this problem by designing and deploying an ontology-based personalized clinical decision support tool. This helps to improve personalized decisions and reduces variability in clinical practice in an integrated care environment. This paper describes the methodology followed in developing the ontology used to infer clinical knowledge and to develop and implement the support tool. The tool will access the information provided by the Electronic Health Record of Andalusian Health Public Service, the main healthcare provider in a region in Spain with more than eight million inhabitants.

Keywords:

Polypharmacy, Decision Support Systems, Clinical, Chronic Disease

Introduction

The high prevalence of chronic diseases in the population has led to an increase in the number of people living with complex chronic disease or multiple diseases. They are considered to be the leading cause of death worldwide (60% of the total) [1].Complex Chronically ill Patients (CCP) are patients with multimorbidity [2]. Drug use in CCP and elderly persons often requires specific knowledge of clinical, functional, and care areas. The complexity of patients with chronic diseases makes them a risk group that is more susceptible to errors and, therefore, adverse reactions [3].

A study of 12 clinical practice guidelines (CPGs) conducted by the National Institute of Health and Care Excellence (NICE) in the U.K. highlights multiple potentially dangerous interactions between the recommendations of these guidelines. The results show the need to adopt innovative and interactive mechanisms for the production and dissemination of clinical guidelines to facilitate informed decision making in the treatment of CCPs [4]. They also suggest the need to advance the interaction between electronic clinical guidelines and basic medical knowledge in order to allow the guides to be adapted to a specific patient. This results in adapted treatment of the disease [5].

A personalized clinical decision support system (PCDSS) was designed to overcome these problems. The incorporation of this type of tool into clinician workflows is one of the main strategies to prevent errors in decision-making in integrated care environments [6]. The proposed tool focuses on the reduction of variability and on fair and personalized clinical practice in the field of integrated care of CCPs. In this case, an ontology was used to infer specific clinical knowledge [7].

An ontology is a formal and explicit specification of a shared conceptualization. It helps to model by first identifying the relevance of concepts and their explicit definition and related constraints. In addition the ontology can be read by a machine and can capture consensual knowledge [8].

The CDSS are handly tools in clinical practice. However, in the area of pluripathological patients they are not as flexible as they should be. This is why it was decided to use ontological reasoning for tool modeling. Galopin et al. have found that there are fewer treatment conflicts between the different diseases with ontological reasoning [9].

Velickovski et al. [10] proposed a CDSS for CCPs that focused on the diagnosis and prevention of chronic obstructive pulmonary disease. The key difference with the present proposal is that it does not use ontology as the primary basis for the development of the decision support tool.

Böttiger et al. [11], on the other hand, proposed a tool focused on drugs and adverse reactions. Its main aim was to prevent and avoid possible interactions and adverse effects on polymedicated patients. Again, these researchers did not use the ontology for the development of their proposed tool.

A CDSS in which an ontology was used as the central element in development was proposed by Zhang et al. [12], which focused on the diagnosis, evaluation, and treatment of patients with diabetes mellitus 2. The key difference between this proposal and ours is that the former focuses on complex chronic diseases, while the present proposal focuses on the different drugs used to treat complex chronic diseases. Also, the recommendations provided in the present research are justified because they show the items selected initially by healthcare professionals.

Methods

Objectives

This study is part of the PITeS-TIiSS project [13], funded by the Spanish Carlos III Institute of Health and whose specific objectives include the following:

- Define a methodology to represent and implement evidence-based integrated care plans. These would incorporate recommendations and personalized

treatment for CCPs based on the most appropriate CPGs and the evaluation of the results of previously applied therapies.

- Applying the methodology described above in the setting of CCPs. The aim is to produce a set of recommendations, clinical pathways and specific protocols.
- Avoid increasing the time spent using the tool during medical care.

The aim is to integrate the tool into the workflow of the health professional. This facilitates access to information provided by the Electronic Health Record (EHR), known as DIRAYA [14], of the Andalusian Health Service, the main health services provider to Andalusia (a region in southern Spain with a population of more than eight million inhabitants). This is one of the largest registration systems in Europe. The project will benefit from a service-oriented architecture based on an internal communication bus (using open source Mirth Connect software). This bus will incorporate new modules and improve performance by reusing integrations with many EHR services (e.g., patient identification, clinical authentication and the option to query relevant health data sets). This design reduces development cost by reusing common components and enables research on implementation of Integrating the Healthcare Enterprise (IHE) [15] and ISO 13606 [16] interoperability standards in several projects.

Study of the Knowledge Bases

Initially, within the framework of the PITeS-TIiSS project, a generic scenario focused on the integrated care of CCPs was defined. In this scenario, the sphere of care was shared by two centers: "Virgen Del Rocío" University Hospital (HUVR) and a primary care center in the HUVR area.

Given the disease prevalence, the information available in the EHR and the availability of patients, the prescription decision scenarios defined for the validation of the technological developments in this project were the following:

- Scenario 1: Atrial fibrillation. Treatment with anticoagulant drugs.
- Scenario 2: Secondary prevention of cardiovascular disease.Treatment with statins.

Subsequently, together with clinical researchers, we identified and exhaustively analyzed a series of knowledge bases related to the selected scenarios. To this end, the clinical researchers identified the guidelines and pathways used in clinical practice, as well as other cross-sectional knowledge tools to support decision making in these diseases. The knowledge bases, currently in use in conventional integrated care, were identified and collected in the protocols of the HUVR and its affiliated centers. These bases were as follows:

- STOPP/START Criteria: This tool is used to detect an elderly patient's prescription medication. Also to indicate which, according to their individual conditions, is the most correct pharmaceutical prescription for a patient. It has been used in other parts of Europe [17].
- LESS/CHRON Rules: This tool is used to guide the prescription of medicines in patients with chronic diseases. It is the result of a recent consensus on the medicine prescription [18].
- PROFUND Index: A widely used outcome prediction scalein Spain [19].
- CPGs: Guides developed systematically to assist healthcare professionals and their patients in maing decisions about appropriate care under specific clinical circumstances [20]. In order to obtain the most appropriate information on the defined decision-making scenarios, the clinical researchers finally recommended the use of two specific CPGs. Scenario 1 was studied using the CPG on Diagnosis and Treatment of Atrial Fibrillation [21]; Scenario 2 was studied using the CPG on dyslipidemias [22].

Given the complexity of the knowledge bases, some of conventions have been defined for this paper. A clinical statement is a text that makes a recommendation in the knowledge bases (e.g., warfarin should be prescribed if the patient has atrial fibrillation and his/her glomerular filtration rate is greater than or equal to 15 ml/min). Also, within this clinical statement, we can identify two premises (e.g., having a glomerular filtration rate greater than or equal to 15 ml/min) that compare a series of clinical concepts (e.g., glomerular filtrate) with numerical values, Boolean values, or other concepts. Once the resources to be analyzed were identified and the conventions defined, a set of tabulated information was extracted using a methodology based on the following steps: (1) The clinical statements are extracted from the different knowledge bases; (2) the clinical concepts associated with these clinical statements are identified and, once identified, (3) the questions to be shown to the healthcare professional for each of the previous concepts are established in the tool. Subsequently, (4) the premises associated with the clinical statement are identified, (5) the recommendation to be given by the tool when makng the specific clinical statement (e.g., prescribing warfarin) is noted, and (6) the existing relational logic between the concepts contained in a premise is identified and designed (by means of a propositional or Boolean logic system). Finally, (7) the logical relations are established between the premises that fulfill all (AND), only one (OR), etc. The whole process is carried out by the research team in a systematic and recurrent way until a consensus is reached among the clinical researchers.

In this case, the recommendations corresponded to the prescription and/or deprescription of the medicines in the two defined scenarios. They are shown in the tool developed according to the introduced patient information.

Before its implementation, the project was approved by an ethics committee made up of 21 members and chaired by Mr. Víctor Sánchez Margalet.

Information Modeling and Semantic Standards Mapping

Ontology was defined from the previous tabulated document and developed on the knowledge extracted from the knowledge bases. Ontology modeling has been selected as methodology since it allows a precise classification of a patient in diagnosis, care management and translational research. Compared to terminologies, the big difference is that ontologies define the relationships between concepts in a way that allows logical and computational reasoning. In addition, it allows conclusions to be drawn from related claims. Ontologies can help organize and analyze large amounts of data that are too large for a healthcare professional to handle [23].

Protégé was used to implement clinical statements, premises, concepts, and the relationships between them. The concepts were added to the ontology and, through the use of annotations in the ontology, the extracted concepts were mapped. To facilitate communication exchange, each concept was mapped with a semantic interoperability standard. This allowed all information to be semantically communicable. For this purpose, it was decided to apply FHIR [24] of HL7. Also, the concepts were terminologically mapped with SNOMED CT [25] to ensure semantic interoperability at the terminological level (See Figure 1). For those terms that were not assigned

codes in FHIR or SNOMED CT, they were assigned our non-standard code in order to map them in the ontology.

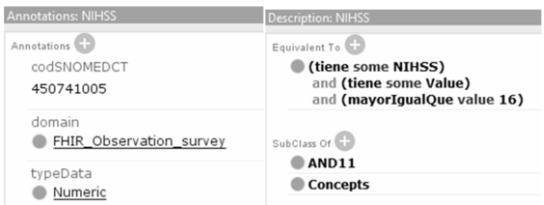


Figure 1 - Example of an annotated mapped concept and a concept-based premise

On the other hand, an internal logic was defined in the ontology, through the use of ObjectProperties, DataProperties, and the relationships between concepts (e.g., hierarchy, EquivalenTo, Class Axioms and others). In this way, it was possible to establish the premises defined in the clinical statements. For example, the NIHSS scale is greater than or equal to 16 (See Figure 1) and included in the tabulated document (through the use of annotations). In addition, an internal logic was redefined in the ontology between the different premises associated with the same clinical statement (See Figure 2).

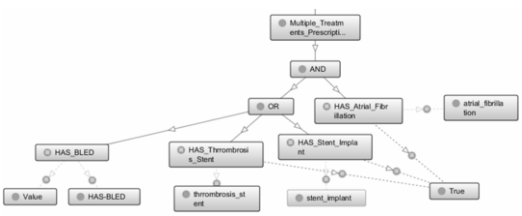


Figure 2 - Example of rule logic associations

Definition, Design and Implementation of the CDSS

Once the modeled information was available (mapped concepts, conditions, logic and recommendations), a PCDSS was implemented. For this purpose, the data of each patient was incorporated and the model designed to generate personalized prescription recommendations for each patient according to the sources analyzed was used (See Figure 3).

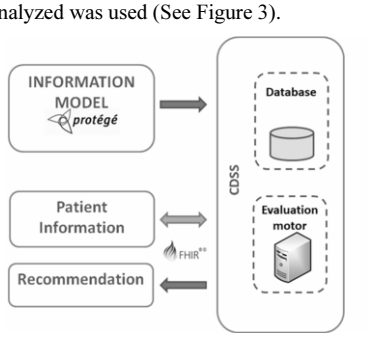


Figure 3 - System architecture

For the incorporation of data other than patient information, the system is able to automatically extract information from the previous model and incorporate it into the CDSS database. To do this, a Java project was created. The Owl API library was used to process the model created through the proposed methodology and was able to store the information so that it could be understood by the evaluation engine.

When the information arrives from a new patient, the tool created is able to (1) obtain the representative set of mapped concepts, (2) relating patient data to concepts through the use of reference terminology, (3) obtaining the premises defined in the information model, (4) evaluating each of the premises obtained with patient concepts and data, (5) evaluating the logic of the premises and clinical statements defined, and (6) giving a recommendation in real time.

Validation

First, clinical researchers validated the tool with retrospective data from real patients in the reference area of the HUVR. In addition, the clinical researchers conducted reviews of the recommendations throughout the tool development process.

In this way, it was possible verify that the PCDSS generated the most appropriate and personalized recommendations for each patient and that the data entered were correct in the clinical setting.

Results

A methodology was available to obtain clinical decision rules based on ontologies. The knowledge bases used were oriented to relevant medical decision-making scenarios in the context of integrated care.

The PCDSS was developed. The tool provides forms validated by clinical researchers that connected the questions in the form with SNOMED CT concept codes contained in the information model. The healthcare professional entered the patient information into the form, from which clinical recommendations are generated in the PCDSS.

The PCDSS was integrated into the workflow of healthcare professionals and an analysis was performed to determine whether the premises and logic of the clinical statements were met according to the data entered on the form. Besides, the PCDSS also generated personalized recommendations for the patient about prescription and/or deprescription of specific medications.

At the time this study was submitted, the PITeS-TIiSS project was in the experimental phase. The phases included clinical validation of the tool developed, training of clinical researchers and patient recruitment and follow-up.

Discussion

It is hoped that the tool will be very useful for support decision makingin clinical practice in both scenarios defined for treatment of CCP. The inclusion of clinical researchers from the HUVR reference area in this project, both in the definition of requirements and in the design of the technological infrastructure, has made it possible to design a tool adapted to the needs and preferences of healthcare professionals. The ontology played a key role in the design of the tool, as it served to model the information in natural, intuitive and understandable way after some adaptation processes. This was of great help to the developers, as it made it much easier for them to import the model into the CDSS. On the other hand, more time than expected has been invested in the modeling and implementation of the tool due to the desire to make it highly scalable from a functional point of view. In other words, the methodology allows the incorporation of recommendations on the prescription or deprescription of new drugs that in turn cross with previously modeled recommendations, with a minimum

impact on the previous model. A future identified line of work is to increase the scope of normalization of modeled concepts through SNOMED CT post-coordination.

Work is also underway to ensure that the tool is easy to use. In this way, the tool will allow the CDSS to include future scenarios for other drugs and, therefore, new recommendations, thus enriching the usefulness and knowledge base of the tool developed thanks to the functional scalability of the method.

Conclusions

In conclusion, thanks to the involvement of healthcare professionals, considerable progress has been made in a short time. Positive results have been obtained in various areas, as indicated below:

- Protocols have been established for the knowledge extraction phase, defining clear structures for the collection of information in the tabulated document. This has reduced modeling time, the number of meetings and the explanations necessary for healthcare professionals.
- The proposed methodology enables better use of the research team's time.
- Possible human errors have been reduced by including the models in the tool, thanks to the fact that a large part of this process is automated.
- According to the healthcare professionals, the use of the ontology has facilitated visualization and understanding of knowledge thanks to the relationships between concepts.

Once the validation has been carried out retrospectively, the pilot of the developed tool in this project will begin prospectively with real patients. This pilot will be carried out in the reference area of the HUVR in primary care and specialized care consultations (internal medicine)

Finally, once the pilot has been completed, the aim is to obtain data that will make it possible to check the reliability and effectiveness of the PCDSS through indicators. It is also intended to extract new knowledge from the decisions taken by healthcare professionals. The indicators used are as follows:

- Number and percentage of adverse events avoided in each group.
- Number and percentage of emergency visits in each group.
- Number and percentage of hospital admissions in each group.
- Number and percentage of avoidable referrals from primary care to hospital.
- Satisfaction of healthcare professionals.
- Quality of life related to the patient's health.
- Number of improvement proposals detected for the process of integrated care of CCP.

Acknowledgements

This research was funded in part by the Platform for Innovation in Medical Technologies and Health (ITEMAS Platform, codes PT13/0006/0036 and PT17/0005/0036) and by the PITeS-TIiSS project (code PI15/01213) of Carlos III Institute of Health, both co-funded by FEDER, European Regional Development Funds (European Union). Many thanks to other researchers for their useful contributions and suggestions. We are especially grateful to the researchers Francisco Núñez Benjumea, Jesús Moreno Conde, Samuel Salas Fernandez, Bosco Barón Franco, Lourdes Moreno Gabiño, and María José García Lozano, as well as to the other internists and biomedical informatics researchers of the Virgen del Rocío University Hospital.

References

[1] J. Casademont, E. Francia, and O. Torres, Age of patients admitted to Internal Medicine departments in Spain: a twenty years perspective,*Medicinaclínica* ***138***(7) (2012), 289-292.

[2] M. Ollero Baturone, M Bernabeu-Wittel, J. M. Espinosa Almendro, R. García Estepa, J. C. Morilla Herrera, B. Pascual de la Pisa & R. Sanz Amores. Atención a pacientes pluripatológicos: proceso asistencial integrado. (2018)

[3] Ministerio de Sanidad y Política Social. Unidad de Pacientes Pluripatológicos. Estándares y Recomendaciones. Madrid: Ministerio de Sanidad y Política Social. 2009.

[4] S. Dumbreck, A. Flynn, M. Nairn, M. Wilson, S. Treweek, S.W. Mercer, and B. Guthrie, Drug-disease and drug-drug interactions: systematic examination of recommendations in 12 UK national clinical guidelines, *BMJ* ***350*** (2015), h949.

[5] M. Peleg, Computer-interpretable clinical guidelines: a methodological review, *Journal of biomedical informatics* ***46***(4) (2013), 744-763.

[6]DW. Bates and AA. Gawande. Improving safety with information technology, *New England journal of medicine* ***348.25*** (2003), 2526-2534.

[7]P. Fraccaro, M. Arguello, J. Ainsworth, and I. Buchan, Adoption of clinical decision support in multimorbidity: a systematic review. *JMIR medical informatics* ***3***(1) (2015).

[8] F. Paganelli & D. Giuli. An ontology-based system for context-aware and configurable services to support home-based continuous care. *IEEE Transactions on Information Technology in Biomedicine* ***15***(2) (2011), 324-333.

[9] A. Galopin, J. Bouaud, S. Pereira & B. Séroussi. An Ontology-Based Clinical Decision Support System for the Management of Patients with Multiple Chronic Disorders. *In MedInfo* (2015) 275-279.

[10]F. Velickovski, L. Ceccaroni, J. Roca, F. Burgos, J. B. Galdiz, N. Marina, and M. Lluch-Ariet, Clinical Decision Support Systems (CDSS) for preventive management of COPD patients, *Journal of translational medicine*, ***12***(2), (2014), S9.

[11]Y. Böttiger, K. Laine, T. Korhonen, J. Lähdesmäki, T. Shemeikka, M. Julander, and M. L. Andersson, Development and pilot testing of PHARAO—a decision support system for pharmacological risk assessment in the elderly, *European journal of clinical pharmacology* ***74***(3), (2018), 365-371.

[12]Y. F. Zhang, L. Gou, T. S. Zhou, D. N. Lin, J. Zheng, Y. Li, and J. S. Li. An ontology-based approach to patient follow-up assessment for continuous and personalized

chronic disease management, *Journal of biomedical informatics* ***72***, (2017), 45-59.

[13]https://ec.europa.eu/eip/ageing/commitments-tracker/b3/pites-tiiss-complex-chronic-patient-personalized-decision-support_en [accessed on November 3, 2018]

[14]https://ec.europa.eu/eip/ageing/repository/andalusian-ehealth-strategy-system-diraya_en [accessed on November 3, 2018]

[15] http://www.ihe.net/ [accessed on November 3,2018]

[16]D. Kalra, and D. Lloyd, ISO 13606 electronic health record communication part 1: reference model, *ISO 13606* (2008).

[17]D. O'Mahony, D. O'Sullivan, S. Byrne, M. N. O'Connor, C Ryan, and P. Gallagher, STOPP/START criteria for potentially inappropriate prescribing in older people: version 2. *Age and ageing* ***44***(2) (2015), 213-218.

[18]A. Rodríguez-Perez, E. A. Lara, E.D. Silveira, E. P. Prados, J. N. Brophy, and C. P. Guerrero, DI-060 Translating into English a new tool to guide deprescribing: A cross cultural adaptation, (2016).

[19]M. Bernabeu-Wittel, M. Ollero-Baturone, and D. Nieto-Martín, Polypathological patients and prognostic scores. About the PROFUND index, *European journal of internal medicine* ***23***(4) (2012), e116.

[20]J. J. Gómez-Doblas, Implementación de guías clínicas, *Revista española de cardiología* ***59***(3) (2006), 29-35.

[21]P. Kirchhof, S. Benussi, D. Kotecha, A. Ahlsson, D. Atar, B. Casade, And G. Hindricks, Guía ESC 2016 sobre el diagnóstico y tratamiento de la fibrilación auricular, desarrollada en colaboración con la EACTS, *Revista Española de Cardiología* ***70***(1) (2017), 50-e1.

[22]A. L. Catapano, I. Graham, G. De Backer, O. Wiklund, M. J., Chapman, H. Drexel, and Ž. Reiner, 2016 ESC/EAS guidelines for the management of dyslipidaemias, *European heart journal* ***37***(39) (2016), 2999-3058.

[23]J. Walonoski, R. Scanlon, C. Dowling, M., Hyland, R. Ettema, and S. Posnack, Validation and Testing of Fast Healthcare Interoperability Resources Standards Compliance: Data Analysis, *JMIR Medical Informatics* ***6***(4) (2018), e10870.

[24]T. Sampalli, M. Shepherd, J. Duffy, and R. Fox, An evaluation of SNOMED CT® in the domain of complex chronic conditions, *International journal of integrated care****10***, (2010)

[25]T. Sampalli, M. Shepherd, J. Duffy, and R. Fox, An evaluation of SNOMED CT® in the domain of complex chronic conditions, *International journal of integrated care****10***, (2010)

Address for correspondence

Esther Román Villarán
ester.roman.exts@juntadeandalucia.es
Avda. Manuel Siurot S/N - 41013 Sevilla
Hospital Universitario Virgen del Rocío
Servicio Andaluz de Salud
Phone: +34 955 01 21 66 - (31 21 66)

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190326

Development, Implementation and Preliminary Results of an Electronic Reminder for HIV Screening Using a Service Oriented Architecture

Luciana Rubin, Natalia Pérez López, Alejandro Gaiera, Fernando Campos, Daniel Luna, Fernán Bernaldo González de Quirós

Department of Health Informatics, Hospital Italiano de Buenos Aires, CABA, Argentina

Abstract

30% of the Argentinian population and 58% of Plan de Salud HIBA patients are unaware of their HIV status. The Ministry of Health and US Preventive Service Task recommends physicians to assess HIV infection in persons aged 15 to 65. An HIV screening reminder integrated in an electronic health record (EHR) was created using FHIR to represent clinical information and CDS-Hooks to represent the exchange of information with a CDS service. The tool had a 1% intervention rate, and 67.4% acceptance rate. The number of HIV screening tests requested during the weeks after the CDSS implementation and in the same period in 2017 were obtained. 575 orders were requested in the 2017 period and 893 in the 2018. 89 (almost 10%) of these came from the electronic tool. The preliminary results indicate that this non disruptive, action oriented reminder can contribute to increased HIV screening orders.

Keywords:

Clinical Decision Support Systems, Electronic Health Records, HIV Infections

Introduction

In 2017 around 5,500 people became infected with HIV in Argentina, 6,500 were diagnosed with the virus (35% of them in advanced stages of infection), the perinatal transmission rate was 5% and 1,500 people died from causes related to the AIDS. With 122,000 people living in Argentina with HIV in 2017, 30% of whom were unaware of their situation. It is necessary to accelerate the pace of diagnosis to reach the goal of 90% diagnosis rate by 2020. This implies the need to increase efforts to actively offer HIV testing [1]. The Ministry of Health resolution 55-E/2017 [2] recommends reporting and offering HIV test. US Preventive Services Working Group (US Preventive Services Task Force) [3] also suggests physicians to assess HIV infection in adolescents and adults, aged 15 to 65.

In the Hospital Italiano de Buenos Aires (HIBA), only 42% of the Plan de Salud patients (HIBA's health maintenance organization) have a reactive or non reactive HIV result, leaving the remaining 58% with an undetermined HIV status.

Clinical decision support systems (CDSS) are one of the biggest benefits of electronic health records (EHR). Reminders are tools that can help physicians to order a diagnostic or therapeutic procedure adjusted to patient characteristics [4], and there is evidence that they can influence the physicians test-ordering behavior [5].

We could say that we are going nowadays through a fifth stage in the evolution of the CDSS, as a response to the problems presented in the four phases indicated by Wright [6]. This phase is based on the utilization of standards to share information between systems using service-oriented architectures implementations that adopts existing web-based technologies such as RESTful (REpresentational State Transfer) APIs [7].

Fast Healthcare Interoperability Resources (FHIR) [8] is the latest standard developed and promoted by the international organization Health Level Seven (HL7), permitting the exchange of information related to medical care. This includes clinical as well as administrative data, and contemplates a wide variety of settings: inpatient, ambulatory, emergency department, etc. FHIR helps us to achieve interoperability between systems, representing health data from heterogeneous sources using the same structure, this being the foundation for sharing information.

CDS-Hooks [9] is a standard that describes a "hook"-based pattern for invoking CDSS from within a clinician's EHR workflow. It triggers CDS calls from the EHR workflow returning as an answer information and suggestions for action.

The aim of this paper is to describe the development and implementation of an electronic reminder to physicians, integrated in an EHR, for HIV screening on the correspondent patients; using FHIR to represent clinical information and CDS-Hooks to represent the exchange of information with a CDS service.

The secondary aim is to describe the preliminary results of the tool's use without formal user training, and to describe the proportion of HIV screening requests in the weeks after the implementation and in the same period of the previous year.

Methods

Setting

HIBA is a non-profit healthcare academic center founded in 1853. It has a network of two hospitals with 750 beds (200 for intensive care), 41 operating rooms, 800 home care beds, 25 outpatient clinics and 150 associated private practices located in Buenos Aires city and its suburban area. Since 1998, HIBA has run an in-house-developed health information system, which includes clinical and administrative data. It has been recently certified by the HIMSS as level 7 in the Electronic Medical Record Adoption Model, being the first hospital in Argentina and the second in Latin America reaching this stage. The EHR is a fully-implemented web based, problem

oriented, patient centered record with customized functionalities depending on the level of care (outpatient, inpatient, emergency care and home care) and terminology web services [10][11]. Problems are referenced to SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms), the terminology standard used for health information contained in patient clinical data repository [12]

The hospital has a health insurance plan (Plan de Salud - PS) with 167,000 members. It is important to remark that these patients are enrolled with HIBA and get all their medical attention (laboratories tests, medical images, proceedings, etc) at the hospital.

The setting selected for the reminder implementation was the ambulatory.

Electronic Reminder

The reminder was created in a multidisciplinary team consisting of general physicians, health informaticians and engineers.

Its aim is to detect patients from 15 to 65 years without any HIV results, either positive or negative, HIV screening pending requests, structured HIV related diagnostics or HIV related medication stored in the EHR. If all these conditions are true for a patient with an appointment with some predefined specialty physician (for e.g., general practitioner, pediatrician, internal medicine, gynecologist) a non disruptive, action oriented reminder appears in the EHR header. The CDSS interface contains the following content (Figure 1):

- a sentence informing the physician about the patient's unknown HIV status (Figure 1: A)
- a "prescribe" button; action oriented option offering the screening test (fourth generation ELISA) (Figure 1: B)
- a "decline" button (Figure 1: C)
- a short explanation of the recommendation source with a hyperlink to the US Task Force website. (Figure 1: D)

The practitioner has three options: ignore the reminder (close it with the X button on the upper right side), accept the reminder and order the test from the same emerging window (from the prescribe button), or reject the reminder (from the decline button). In the last case scenario the practitioner is requested to choose between a set of structured and free text justification.

Figure 1 – HIV Screening electronic reminder interface. (A) patient's unknown HIV status, (B) "prescribe" test order button, (C) a "decline" button, (D) explanation of the recommendation and link to source.

The architecture of the tool focuses on the EHR that communicates with the FHIR server an the CDS service. When a physician enters in a patient's EHR ("patient-view" hook) and has an appointment with him, the EHR requests the FHIR server all the FHIR resources needed to call the CDS service (Figure 2). The EHR sends the request with the contextual data, like patient and practitioner identifiers. The FHIR server returns a bundle. This bundle has the FHIR resources packaged, such as "Patient", "Observation", etc. This FHIR server was implemented in HAPI FHIR.

When the EHR obtains the bundle, it calls the CDS service using CDS-Hooks standard. This call is made with the hook ("patient-view"), allowing the CDS service to execute only the rules related to that hook. The EHR also sends to the CDS service some contextual data needed for executing the rules. For example, the specialty that the physician is performing at that moment. The bundle and this contextual data are the inputs for the rules. The rule engine was implemented in Drools[13].

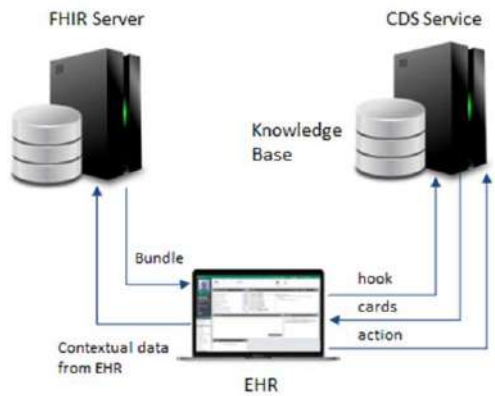


Figure 2 - CDS service architecture1

Once the rules have been executed and if they are true, the response is sent back to the EHR as information cards, suggestions for actions, etc. A log is also recorded for further analysis. In this case an action oriented suggestion card was used, suggesting the physicians to order the HIV screening test for the patient. When the EHR receives the response, it displays the recommendation so that the user can take an apprpriate action. Each card returned by the CDS service has a unique identifier that permits to monitor what happens with the provided suggestion (accept or reject it with their corresponding structured or free text justification) and is later used for analysis.

This architecture works, thanks to the use of HIBA's terminology server and the FHIR standard that enables the achievement of semantic and syntactic interoperability.

Data collection

The data was collected in the period of October 11 to November 21 of the year 2017 for pre-intervention and 2018 for post-intervention.

Two main analyses were performed.

[1] Images extracted from: https://pixabay.com/es/servidor-de-hardware-red-equipo-37258/ and https://es.wikipedia.org/wiki/Archivo:Database-152091_960_720.png

The first one was the extraction of the number of times that the reminder was triggered and appeared to the physicians, to how many patients, how many times it was intervened (defined as either click on the prescribe button or in the decline button), by how many physicians; as well as the justifications selected in case of declining the reccomendation.

The second analysis was to obtain the number of HIV screening tests requested during the following six weeks after the CDSS implementation and in the same period in 2017. The data considered was from the patients who didn't have any testing requests in the previous three months and didn't have any HIV results, in consistency with the rules implemented for the CDSS. Test requests received from the CDSS in 2018 were also determined. All of these analyses were made only for Plan de Salud patients.

Analysis

For the summary of the data, descriptive statistics were used. The categorical variables are reported as absolute and relative frequencies.

Results

Interface design, testing and deployment

The interface design was created with the User Experience team, through iterative mockups.

The tool was tested for a month in a testing environment, where rules were corrected as needed. Following which the final implementation was monitored with a random sample of patients showing 100% accuracy of HIV screening recommendations, as defined in the functional requirement.

On October 11, 2018, the services (FHIR server and CDS service) and the EHR were updated in the production environment. The physicians were informed about the new functionality through a notification placed in the EHR. As there was no formal training, it can't be guaranteed that the users have been able to internalize how to use the tool.

Preliminary results of reminder's use

Within six weeks of use since the implementation, following were the preliminary results.

The HIV test reminder was triggered 12,891 times to 470 physicians for 5689 patients.

An action was taken for 132 reminders (1% intervention rate), one for patient, by 58 physicians.

Of the 132 intervened reminders, 67.4% were accepted (HIV screening study ordered). The most frequent justification for rejecting the reminder was "The study will be ordered in the next consultation" and "The study was ordered by a different means" (Table 1).

Of all the medical specialties defined to receive the CDSS, the reminder was most accepted (measured by HIV screening tests ordered) by gynecology (34-38.2%), general practitioners (24-27.9%) , and internal medicine (11-12.4%).

Of the 2017 period, 575 ELISA HIV screening orders were requested. During 2018 period, the number increased to 893 ELISA HIV screening orders. 89 (almost 10%) of these were prescribed from the electronic tool.

Table 1 - HIV screening justification for CDSS rejections.

Justification	N:43
The study will be ordered in the next consultation	15 (34.9%)
The study was ordered by a different means	12 (27.9%)
Other (free text justification)	8 (18.6%)
The patient does not want to be tested	5 (11.6%)
The physician does not agree with the recommendation	3 (7%)

An intersting find in the 2017 period was that gynecology differed from the other specialties and only ordered 23 HIV screening tests. In the 2018 period after the reminder implementation, they ordered 61 HIV screening tests (2.6 times more), 34 of them (almost 56%) coming from the CDSS tool (Figure 3).

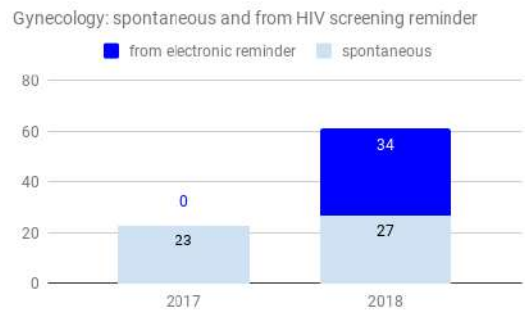


Figure 3 - 2017 vs. 2018 gynecology HIV screening orders

Discussion

This paper describes the development, implementation and preliminary results of an electronic reminder for HIV screening using a service oriented architecture to interact from the EHR with the FHIR server and the CDS service triggered by the physicians workflow inside the EHR.

The use of the same terminological standard in both EHR, the FHIR server and the CDS service, with adequate granularity, depth and breadth, was important for a successful implementation[14]. The incorporation of new technologies that ease the ability of outside digital tools to integrate into EHRs, as well as the utilization of standards for healthcare information exchange (such as FHIR), are said to be important enablers of a better health care model[15]. It is easy to envision that screening practices can be improved with education and electronic screening tools if they are ready to be implemented in health care facilities without the effort of development.

The preliminary results of this work show that the CDSS contributed to an increase of HIV screening test ordering in 2018. Clinical decision support systems have the potential to improve health care, especially when it is related with screening and diagnosing prevalent diseases. These results support the findings of three different studies which showed that an electronic reminder increased the HIV screening testing orders[16][17][18]. Also, similar informatics strategies

have been employed for HIV screening in cities with high HIV prevalence (New York and New Orleans), documenting significant improvements with EHR interventions[19] .

For these preliminary results it can be noted that the intervetion rate of the tool was low, but on the other hand, the majority of users who interacted with it ordered the HIV screening test, indicating that it was useful for them.

The low intervention rate could be related with a perceived stigma associated with an HIV diagnosis. The literature reveals that fear can be an important patient-driven factor in the avoidance of HIV screening[20]. This factor could influence both, the physician who has to suggest the screening recommendation and the patient who receives it. Nevertheless, this was contemplated when choosing only certains medical specialities to be recipients of the CDSS. Other reason to explain this low intervention rate could be that physicians may not be aware that this tool exists or know how to use the information within the CDSS [21], specially when non formal training was received. Lastly, the results are only based on six weeks of the tool's use. A year would be more appropriate time window to do further and more precise analysis, avoiding potential seasonal bias.

Another interesting thing was that the two most frequent reasons for rejecting the recommendation implied that the physician agreed with it but would perform it in the next patient consultation, or that had already ordered it by other means (for example, in a paper format). This indicates that the rejections don't imply a negative connotation towards the tool.

Analyzing the given periods in 2017 and 2018, the increase of HIV screening ordering was only partialy due to the CDSS. Educational programs of HIV screening awareness (not related with this CDSS) could have contributed to this finding. Several other factors may have contributed with the results.

The first one is the automatic provision of decision support. A systematic review identified it as a key factor for the success of CDSS based interventions [22]. Another factor identified is the integration of the CDSS to the already existing workflow of the patient visit, respecting the Five Corrects: "CDSS should be designed to provide the right information to the right person in the right format through the right channel at the right time."[23] and being a crucial part of a successful CDSS[24].

To guarantee certain rate of aceptability, we believe in an incremental and gradual rollout[25], incorporating one reminder at a time, instead of getting all the services available together. In this way we avoid overwhelming the user with new functionalities[26]. Also, as mentioned before, capacitation of users is fundamental to achieve success in CDSS implementations[24].

In addition, the tool was created by a multidisciplinary team which was aligned with the organizational motivation, another recommended best practice in the discipline[27].

Future Lines of Work

Future lines of work can be grouped in different themes:

Expand and improve the tool: Broadening and specializing the rules of the reminder, for example, not giving the recommendation for patients that are blood donors, as regular tests are performed on them in order to exclude infectious diseases. Also, it is required to exhaustively analyze the free text justifications to include other user cases.

Improve the analysis: With an appropriately large time window it will be easier to measure if there was an increase of patients with a known HIV status, and to determine if this was due to the use of the reminder. Another analysis improvement contemplated is to incorporate measures of patients (demographic information) and physicians (for example, how many years they have worked in the hospital and if they are residents, staff, etc). Systematic interviews will conducted for physicians who regularly reject or ignore the reminders in order to gather information about their behavior. Finally, a tendency analysis of the use of the tool is planned.

In order to raise awareness of the importance of HIV screening, it is absolutely necessary to improve the user training, with specific instructions for each medical specialty included within the scope of the reminder. Additionally of other consented reminders for preventive practices are the next step for this CDS program.

Conclusions

The preliminary results indicate that this non disruptive, action oriented reminder that appears in the EHR header, can contribute to the increased ordering for HIV screening.

However, significant opportunity for improvement remains. It is necessary to raise awareness about the importance of HIV screening so that the physicians decide not to refuse or ignore the reminder. It is also essential to make a careful analysis of the answers given by the physicians when they reject the reminder. It will be very helpful to analyze the free text that was introduced. With this information an evaluation will be made focused on whether the rules need some adjustment to include all the cases that may occur.

Acknowledgements

We thank Verónica Peuchot and Sonia Benitez for their valuable inputs and comments.

References

[1] E. Vih-sida, El VIH-sida en la Argentina, (2018) 4–5.
[2] M. de Salud, Resolución 55 - E/2017, (2017).
[3] T.U.S. Preventive, S. Task, and F. Task, Understanding Task Force Recommendations Screening for Human Immunodeficiency Virus (HIV), (2013) 1–4.
[4] A. Wright, D.F. Sittig, J.S. Ash, J. Feblowitz, S. Meltzer, C. Mcmullen, K. Guappone, J. Carpenter, J. Richardson, L. Simonaitis, R.S. Evans, W.P. Nichol, and B. Middleton, Development and evaluation of a comprehensive clinical decision support taxonomy : comparison of front-end tools in commercial and internally developed electronic health record systems, (2011). doi:10.1136/amiajnl-2011-000113.
[5] P.S. Roshanov, J.J. You, J. Dhaliwal, D. Koff, J.A. Mackay, L. Weise-Kelly, T. Navarro, N.L. Wilczynski, and R. Brian Haynes, Can computerized clinical decision support systems improve practitioners' diagnostic test

ordering behavior? A decision-maker-researcher partnership systematic review, *Implement. Sci.* **6** (2011) 88. doi:10.1186/1748-5908-6-88.
[6] A. Wright, and D.F. Sittig, A Four-Phase Model of the Evolution of Clinical Decision Support Architectures, *Int J Med Inf.* **77** (2008) 641–649. doi:10.1016/j.ijmedinf.2008.01.004.A.
[7] RESTful API, (n.d.). https://restfulapi.net/.
[8] FHIR, (n.d.). https://www.hl7.org/fhir/.
[9] CDS-Hooks, (n.d.). https://cds-hooks.org/.
[10] M. Franco, M.V. Giussi Bordoni, C. Otero, M.C. Landoni, S. Benitez, D. Borbolla, and D. Luna, Problem Oriented Medical Record: Characterizing the Use of the Problem List at Hospital Italiano de Buenos Aires, *Stud. Health Technol. Inform.* **216** (2015) 877. doi:10.3233/978-1-61499-564-7-877.
[11] V. Teixeira, L. Rubin, R. Rebrij, A. Tamborindeguy, R. Martineza, J.C. Bacigalupoa, and D. Luna, Implementation of an outsourced transnational service of clinical decision support system, *Stud. Health Technol. Inform.* **245** (2017) 1384. doi:10.3233/978-1-61499-830-3-1384.
[12] A.L. Osornio, D. Luna, M.L. Gambarte, A. Gomez, G. Reynoso, and F.G.B. de Quirós, Creation of a local interface terminology to SNOMED CT., *Stud. Health Technol. Inform.* **129** (2007) 765–9. http://www.ncbi.nlm.nih.gov/pubmed/17911820.
[13] M. Salatino, M. De Maio, and E. Aliverti, Mastering JBoss Drools 6 for Developers, Packt Publishing, 2016.
[14] K. Kawamoto, G. Del Fiol, D.F. Lobach, and R.A. Jenders, Standards for scalable clinical decision support: need, current and emerging standards, gaps, and proposal for progress., *Open Med. Inform. J.* **4** (2010) 235–44. doi:10.2174/1874431101004010235.
[15] R.M. Wachter, and M.D. Howell, Resolving the Productivity Paradox of Health Information Technology: A Time for Optimism., *JAMA*. **320** (2018) 25–26. doi:10.1001/jama.2018.5605.
[16] C. Kershaw, J.L. Taylor, G. Horowitz, D. Brockmeyer, H. Libman, G. Kriegel, and L. Ngo, Use of an electronic medical record reminder improves HIV screening, *BMC Health Serv. Res.* **18** (2018) 14. doi:10.1186/s12913-017-2824-9.
[17] M.B. Goetz, T. Hoang, C. Bowman, H. Knapp, B. Rossman, R. Smith, H. Anaya, T. Osborn, A.L. Gifford, and S.M. Asch, A system-wide intervention to improve HIV testing in the Veterans Health Administration, *J. Gen. Intern. Med.* **23** (2008) 1200–1207. doi:10.1007/s11606-008-0637-6.
[18] A.K. Avery, M. Del Toro, and A. Caron, Increases in HIV screening in primary care clinics through an electronic reminder: an interrupted time series, *BMJ Qual. &Amp; Saf.* **23** (2014) 250 LP-256. http://qualitysafety.bmj.com/content/23/3/250.abstract.
[19] X. Lin, P.M. Dietz, V. Rodriguez, D. Lester, P. Hernandez, L. Moreno-Walton, G. Johnson, M.M. Van Handel, J. Skarbinski, C.L. Mattson, D. Stratford, L. Belcher, B.M. Branson, and Centers for Disease Control and Prevention (CDC), Routine HIV screening in two health-care settings--New York City and New Orleans, 2011-2013., *MMWR. Morb. Mortal. Wkly. Rep.* **63** (2014) 537–41. doi:10.1016/j.annemergmed.2014.10.010.
[20] J.R. Marcelin, E.M. Tan, A. Marcelin, M. Scheitel, P. Ramu, R. Hankey, P. Keniya, M. Wingo, S.A. Rizza, F. North, and R. Chaudhry, Assessment and improvement of HIV screening rates in a Midwest primary care practice using an electronic clinical decision support system: A quality improvement study, *BMC Med. Inform. Decis. Mak.* **16** (2016) 1–11. doi:10.1186/s12911-016-0320-5.
[21] M. Lugtenberg, J.W. Weenink, T. Van Der Weijden, G.P. Westert, and R.B. Kool, Implementation of multiple-domain covering computerized decision support systems in primary care: A focus group study on perceived barriers, *BMC Med. Inform. Decis. Mak.* **15** (2015) 1–11. doi:10.1186/s12911-015-0205-z.
[22] K. Kawamoto, and D.F. Lobach, Clinical decision support provided within physician order entry systems: a systematic review of features effective for changing clinician behavior., *AMIA Annu. Symp. Proc.* (2003) 361–365. doi:D030002977 [pii].
[23] G.L. Alexander, Issues of trust and ethics in computerized clinical decision support systems., *Nurs. Adm. Q.* **30** (2006) 21–29.
[24] J.S. Ash, D.F. Sittig, K.P. Guappone, R.H. Dykstra, J. Richardson, A. Wright, J. Carpenter, C. McMullen, M. Shapiro, A. Bunce, and B. Middleton, Recommended practices for computerized clinical decision support and knowledge management in community settings: a qualitative study, *BMC Med. Inform. Decis. Mak.* **12** (2012) 6. doi:10.1186/1472-6947-12-6.
[25] C.B. Byrne, D.S. Sherry, L. Mercincavage, D. Johnston, E. Pan, and G. Schiff, Key Lessons In Clinical Decision Support Implementation, *Westat. Tech. Repport*. (2010).
[26] F. Lee, J.M. Teich, C.D. Spurr, and D.W. Bates, Implementation of physician order entry: user satisfaction and self-reported usage patterns., *J. Am. Med. Inform. Assoc.* **3** (1996) 42–55. doi:10.1136/jamia.1996.96342648.
[27] A. Wright, S. Phansalkar, M. Bloomrosen, R. Jenders, A. Bobb, J. Halamka, G. Kuperman, T. Payne, S. Teasdale, A.J. Vaida, and D.W. Bates, Best Practices in Clinical Decision Support, *Appl. Clin. Inform.* **01** (2010) 331–345. doi:10.4338/ACI-2010-05-RA-0031.

Address for correspondence

Luciana Rubin. Hospital Italiano de Buenos Aires. Juan D. Perón 4190 (C1181ACH) Buenos Aires, Argentina

Tel.(5411) 4959-0200 luciana.rubin@hospitalitaliano.org.ar

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190327

Effects of Computerized Guideline-Oriented Clinical Decision Support System on Antithrombotic Therapy in Patients with Atrial Fibrillation: A Systematic Review and Meta-Analysis

Ryota Sakurai[a], Kazuhiko Ohe[b]

[a] School of Medicine, International University of Health and Welfare, Chiba, Japan,
[b]Department of Biomedical Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

Abstract

A systematic review and meta-analysis was conducted to investigate the effects of computerized guideline-oriented clinical decision support system (CDSS) on antithrombotic therapy in patients with atrial fibrillation. PubMed, the Cochrane Library, and Web of Science were queried. Four studies were included in this meta-analysis. The proportion of appropriate antithrombotic therapy in accordance with clinical guidelines was significantly higher in the CDSS group than in the control group (risk ratio (RR): 1.03, 95% confidence interval (CI): 1.01 to 1.04, P = 0.004). Although the incidence of thromboembolic events was similar between the two groups (RR: 1.12, 95% CI: 0.88 to 1.42, P = 0.357), the incidence of major bleeding tended to be lower in the CDSS group compared with the control group (RR: 0.79, 95% CI: 0.61 to 1.01, P = 0.063). Computerized guideline-oriented CDSS may be effective for appropriate antithrombotic therapy as compared with control in patients with atrial fibrillation.

Keywords:

Computers, Decision Support Systems, Clinical, Atrial Fibrillation

Introduction

Atrial fibrillation is one of the most common cardiac rhythm disturbances and well-known as a strong risk factor of ischaemic stroke. Patients with atrial fibrillation are approximately five times more likely to have an ischaemic stroke compared to those without atrial fibrillation through all ages [1]. Accordingly, the prevention of ischaemic stroke in patients with atrial fibrillation is of great importance. Since the prevalence for atrial fibrillation has doubled in the last decade [2], more physicians have been involved in the treatment of atrial fibrillation, regardless of their specialties. As a result, especially in general or primary care, the need for clinical guidelines has been increasing. However, although antithrombotic therapy in accordance with clinical guidelines improves clinical outcomes as compared to undertreatment in high-risk patients for ischaemic stroke with atrial fibrillation [3], guideline-oriented antithrombotic therapy is underused [4-6]. Therefore, treatment adherence to clinical guidelines may be critical for care in patients with atrial fibrillation.

A computer-based clinical decision support system (CDSS) is thought to have the potential to improve clinical outcomes in patients with chronic diseases in the era of widespread electronic health record systems [7]. In this context, computerised CDSS, in conjunction with clinical guidelines, is expected to have a synergistic effect for management of atrial fibrillation, and the development of such CDSS will have a powerful impact on daily practice, especially for non-cardiology specialists. However, it has not been well characterized whether this type of CDSS is effective for atrial fibrillation, especially regarding appropriate prescription of antithrombotic agents according to clinical guidelines. To test the hypothesis that computerized CDSS implemented clinical guidelines improves appropriate antithrombotic therapy for atrial fibrillation in clinical practice, we conducted a systematic review and meta-analysis of clinical trials that compared computerized guideline-oriented CDSS with control.

Methods

Search Strategy and Eligibility Criteria

PubMed, the Cochrane Library, and Web of Science were queried for articles of any language from inception to November 2018. The search terms included “clinical,” “decision,” “support,” “system,” “atrial,” and “fibrillation.” For PubMed, the search details were ("decision support systems, clinical"[MeSH Terms] OR ("decision"[All Fields] AND "support"[All Fields] AND "systems"[All Fields] AND "clinical"[All Fields]) OR "clinical decision support systems"[All Fields] OR ("clinical"[All Fields] AND "decision"[All Fields] AND "support"[All Fields] AND "system"[All Fields]) OR "clinical decision support system"[All Fields]) AND ("atrial fibrillation"[MeSH Terms] OR ("atrial"[All Fields] AND "fibrillation"[All Fields]) OR "atrial fibrillation"[All Fields]). The same terms or relevant studies were also queried on the website of the U.S. National Institute of Health and relevant reviews. To increase internal validity, only randomized controlled trials were included in our study. The primary endpoint of our interest was the proportion of appropriate antithrombotic therapy in accordance with clinical guidelines. We also investigated the impact of CDSS on the incidence of systemic thromboembolic events, such as stroke, transient ischaemic attack or other thromboembolism, and major bleeding as clinical outcomes. If multiple follow-up reports existed in the same study, the outcomes during the longest follow-up period were analyzed. To investigate the usefulness of CDSS for physicians in daily practice, regardless of familiarity with care for atrial fibrillation, the inclusion criterion were the studies comparing the computerized guideline-oriented CDSS, defined as any that provide clinical support automatically generated by a computer according to clinical guidelines, with no CDSS. The exclusion criteria were the studies of CDSS not for physicians (nurses, patients, etc.) or not guideline-oriented CDSS.

Statistical Analysis

A random-effects model was performed to estimate the pooled risk ratio (RR). We chose RR, not odds ratio, since only randomized controlled trials were included and the proportion of appropriate antithrombotic therapy would be expected not so low as approximaled by odds ratio. The I^2 statistics and the Cochran's Q test were conducted to assess homogeneity among each study to confirm internal validity [8]. The possibility of publication bias was assessed visually at first by a funnel plot for asymmetry plotting of the standard error of log RR against the log RR. In addition, the Duval and Tweedie's trim and fill procedure would be conducted to estimate the possible impact of unpublished studies on the pooled estimate. A 2-sided P value of < 0.05 was considered to be statistically significant. In cases of the assessment of homogeneity, however, a 2-sided P value of < 0.10, instead of 0.05, was considered to be statistically significant [9]. If an I^2 statistic value was $> 50\%$, we also considered the results among each study as heterogeneous [10]. All analyses were performed using STATA 11.2 (Stata Corp., College Station, Texas, USA).

Results

Study Selection and Characteristics of Studies

This study was conducted according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses as much as possible [11]. Figure 1 shows a flow chart of study selection. After excluding articles on basis of title and abstract screening, we further excluded 2 studies due to no endpoints of our interest and another study due to not being guideline-oriented. As a result, 18,646 patients (10,313 patients assigned to the CDSS group and 8,333 to the control group) in 4 studies were included in this meta-analysis [12-15].

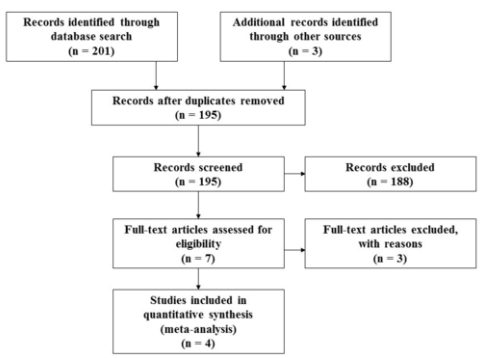


Figure 1– Flow Chart of Study Selection

The characteristics of studies are summarized in Table 1. All studies were performed as a cluster randomized controlled trial in outpatients. The follow-up period ranged from 8 months to 2.2 years. The targets of intervention were primary care physicians or general practitioners.

Pooled Estimates

Figure 2 shows the pooled estimate for the primary endpoint. Homogeneity was not rejected across individual studies by either the I^2 statistics or the Cochran's Q test ($I^2 = 0.0\%$ or P = 0.396).

Author (year)	Number of pts*	Follow up	Target	Clinical GL
Eckman, et al. (2016)	801:692	1 year	Primary care physician	2014 ACC/AHA/HRS GL for AF
Arts, et al. (2017)	522:259	8 months	GP	2013 Dutch GP GL for AF
Karlsson, et al. (2018)	7861:6156	1 year	Primary care physician	2012 ESC focused updated GLs for the management of AF
van Doorn, et al. (2018)	1129:1226	2.2 years	GP	2013 Dutch GP GL for AF

Table 1– Characteristics of Studies

ACC = American College of Cardiology; AF = atrial fibrillation; AHA = American Heart Association; ESC = European Society of Cardiology; GL = guideline; GP = general practitioner; HRS = Heart Rhythm Society; pts = patients.

* clinical decision support system:control.

The proportion of appropriate antithrombotic therapy was significantly higher in the CDSS group (from 55% to 85%) compared with the control group (from 50% to 84%, RR: 1.03, 95% confidence interval (CI): 1.01 to 1.04, P = 0.004).

Figures 3 and 4 show the pooled estimates for clinical outcomes. Homogeneity was also not rejected across individual studies by either the I^2 statistics or the Cochran's Q test ($I^2 = 38.2\%$ or P = 0.203 for thromboembolic events, and $I^2 = 0.0\%$ or P = 0.456 for major bleeding, respectively). Although the incidence of thromboembolic events was similar between the 2 groups (RR: 1.12, 95% CI: 0.88 to 1.42, P = 0.357), the incidence of major bleeding tended to be lower in the CDSS group compared with the control group (RR: 0.79, 95% CI: 0.61 to 1.01, P = 0.063).

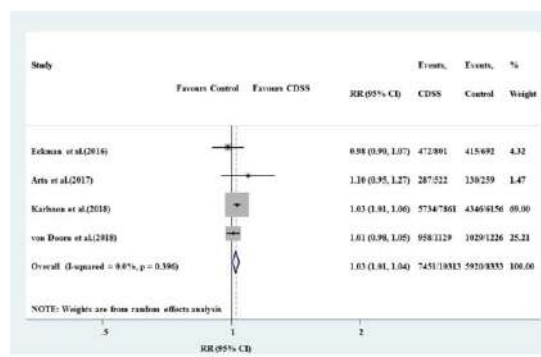


Figure 2– Forest Plot for the Proportion of Appropriate Antithrombotic Therapy

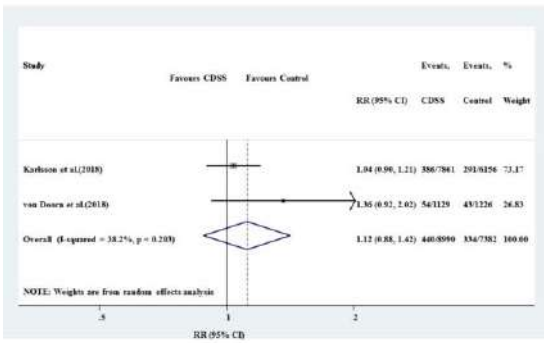


Figure 3– Forest Plot for the Incidence of Thromboembolic Events

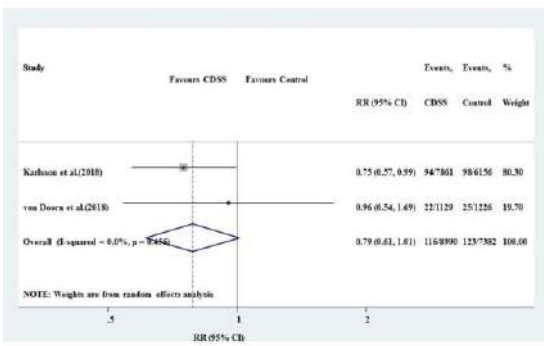


Figure 4– Forest Plot for the Incidence of Major Bleeding

Publication Bias

A funnel plot seemed asymmetric, especially for the incidence of thromboembolic events and major bleeding. Therefore, we conducted the Duval and Tweedie's trim and fill procedure, and the possible impact of an unpublished study favorable for control was suggested for the incidence of thromboembolic events and major bleeding, but not for the proportion of adherence to guidelines (Figures 5, 6, and 7).

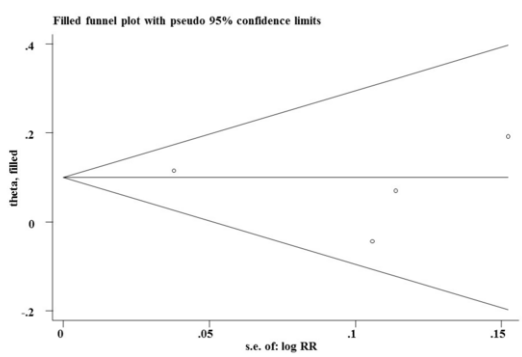


Figure 5– The Duval and Tweedie's Trim and Fill Procedure for the Proportion of Appropriate Antithrombotic Therapy

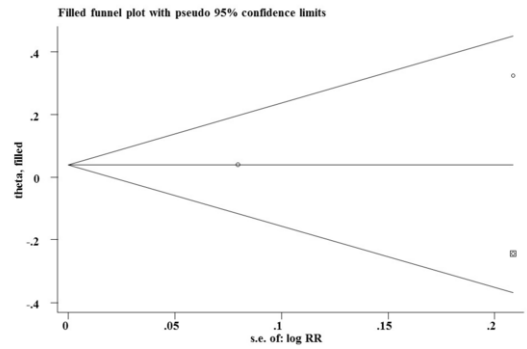


Figure 6– The Duval and Tweedie's Trim and Fill Procedure for the Incidence of Thromboembolic Events

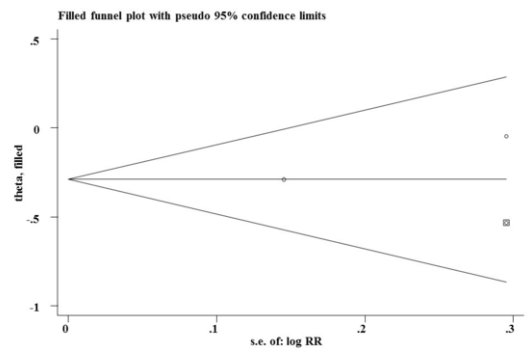


Figure 7– The Duval and Tweedie's Trim and Fill Procedure for the Incidence of Major Bleeding

Discussion

In this meta-analysis, computerized guideline-oriented CDSS demonstrated more favorable effects on the proportion of appropriate antithrombotic therapy as compared with control. Surprisingly, although the incidence of thromboembolic events was similar between the two groups, the incidence of major bleeding tended to be lower in the CDSS group than in the control group. Due to inclusion of only randomized controlled trials, heterogeneity among each study was not observed in any endpoints. To the best of our knowledge, this study reports favorable effects of computerized CDSS on guideline adherence in patients with atrial fibrillation by meta-analysis.

In real-world management of patients with atrial fibrillation, physicians must consider the risks of not only thromboembolic events, but also bleeding. There are several clinical guidelines incorporating some risk scores for stratifying patients, including the $CHADS_2$ or the CHA_2DS_2-VASc score for stroke risk, and the HAS-BLED, the RIETE, or the ATRIA score for bleeding risk. These clinical guidelines generally have numerous pages and are very complicated. In addition, patients with atrial fibrillation frequently have other cardiac diseases or comorbidities, such as coronary artery disease, valvular heart disease, hypertension, heart failure, thyroid dysfunction, and so on [2, 16]. In such complex clinical situations, especially in general or primary care practices, physicians may be extremely

troubled in making decisions by themselves. Accordingly, computerized guideline-oriented CDSS is helpful for non-cardiology specialists to deal with these situations, which may improve clinical outcomes in patients with atrial fibrillation. In fact, most studies included in this meta-analysis took bleeding risk into consideration as well as stroke risk [12-14], and this kind of CDSS may play an important role in decreasing the incidence of major bleeding, even with the similar incidence of thromboembolic events.

On the other hand, despite statistical significance, the range of the proportion of appropriate antithrombotic therapy varied from 55% to 85% in the CDSS group. From the viewpoint of physicians who use CDSS, the difficulties of handling computerized CDSS may be critical for adherence to clinical guidelines. Possible reasons for lower adherence to guidelines include a separate, nonintegrated CDSS apart from natural flow of patient care, requiring the click of a mouse to access information, too many tasks and too much information at once, and too many alert notifications contributing to alert fatigue [12-15]. To facilitate the use of CDSS and further increase adherence to guidelines, therefore, it is tremendously important for system developers to create and provide a fully integrated, simple, and user-friendly computerized guideline-oriented CDSS that is easily accessed by any physician within the limited time as part of clinical workflow, according to some standards or guidelines, e.g. ISO/IEC 25010 [17-19].

There may be several possible limitations in the present study. First, the number of studies included was relatively small in terms of a meta-analysis, partially due to possible publication bias. Therefore, the results, especially regarding the incidence of clinical outcomes, may not be conclusive. Second, the patient population may not be the same across individual studies. Therefore, we used a random-effects model instead of a fixed-effects model to calculate more conservative pooled estimate, although heterogeneity was suggested by neither the I^2 statistics nor the Cochran's Q test. Third, even an increased high internal validity, an external validity may be relatively low in a meta-analysis of randomized controlled trials with strict inclusion and exclusion criteria. Therefore, a real-world data such as registry may also be needed to validate our findings. Finally, due to the existence of several clinical guidelines, the quality of guidelines may not be the same. However, almost all guidelines are based on the $CHADS_2$ or the CHA_2DS_2-VASc score to recommend the treatment.

Conclusions

In this meta-analysis, computerized guideline-oriented CDSS is associated with more appropriate antithrombotic therapy in patients with atrial fibrillation as compared with control. To achieve further improvement of adherence to guidelines and clinical outcomes, it seems that the development of a fully integrated, simple, and user-friendly computerized CDSS in accordance with some standards or guidelines such as ISO/IEC 25010 would be more effective.

Acknowledgements

The authors thank Heidi N. Bonneau for her editorial review of this manuscript. The authors have nothing to disclose regarding this study.

References

[1] P.A. Wolf, R.D. Abbott, and W.B. Kannel, Atrial fibrillation as an independent risk factor for stroke: the Framingham Study, *Stroke* **22** (1991), 983-988.

[2] M. Zoni-Berisso, F. Lercari, T. Carazza, and S. Domenicucci, Epidemiology of atrial fibrillation: European perspective, *Clin Epidemiol* **6** (2014), 213-220.

[3] R. Nieuwlaat, S.B. Olsson, G.Y. Lip, A.J. Camm, G. Breithardt, A. Capucci, J.G. Meeder, M.H. Prins, S. Levy, and H.J. Crijns, Guideline-adherent antithrombotic treatment is associated with improved outcomes compared with undertreatment in high-risk patients with atrial fibrillation. The Euro Heart Survey on Atrial Fibrillation, *Am Heart J* **153** (2007), 1006-1012.

[4] I.M. Ogilvie, N. Newton, S.A. Welner, W. Cowell, and G.Y. Lip, Underuse of oral anticoagulants in atrial fibrillation: a systematic review, *Am J Med* **123** (2010), 638-645 e634.

[5] H. Gamra, J. Murin, C.E. Chiang, L. Naditch-Brule, S. Brette, and P.G. Steg, Use of antithrombotics in atrial fibrillation in Africa, Europe, Asia and South America: insights from the International RealiseAF Survey, *Arch Cardiovasc Dis* **107** (2014), 77-87.

[6] J.C. Hsu, T.M. Maddox, K.F. Kennedy, D.F. Katz, L.N. Marzec, S.A. Lubitz, A.K. Gehi, M.P. Turakhia, and G.M. Marcus, Oral Anticoagulant Therapy Prescription in Patients With Atrial Fibrillation Across the Spectrum of Stroke Risk: Insights From the NCDR PINNACLE Registry, *JAMA Cardiol* **1** (2016), 55-62.

[7] D.L. Hunt, R.B. Haynes, S.E. Hanna, and K. Smith, Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review, *JAMA* **280** (1998), 1339-1346.

[8] J.P. Higgins and S.G. Thompson, Quantifying heterogeneity in a meta-analysis, *Statistics in medicine* **21** (2002), 1539-1558.

[9] J.A. Sterne, D. Gavaghan, and M. Egger, Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature, *Journal of clinical epidemiology* **53** (2000), 1119-1129.

[10] J.P. Higgins, S.G. Thompson, J.J. Deeks, and D.G. Altman, Measuring inconsistency in meta-analyses, *BMJ* **327** (2003), 557-560.

[11] A. Liberati, D.G. Altman, J. Tetzlaff, C. Mulrow, P.C. Gotzsche, J.P. Ioannidis, M. Clarke, P.J. Devereaux, J. Kleijnen, and D. Moher, The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration, *BMJ* **339** (2009), b2700.

[12] M.H. Eckman, G.Y. Lip, R.E. Wise, B. Speer, M. Sullivan, N. Walker, B. Kissela, M.L. Flaherty, D. Kleindorfer, P. Baker, R. Ireton, D. Hoskins, B.M. Harnett, C. Aguilar, A.C. Leonard, L. Arduser, D. Steen, A. Costea, and J. Kues, Impact of an Atrial Fibrillation Decision Support Tool on thromboprophylaxis for atrial fibrillation, *Am Heart J* **176** (2016), 17-27.

[13] D.L. Arts, A. Abu-Hanna, S.K. Medlock, and H.C. van Weert, Effectiveness and usage of a decision support system to improve stroke prevention in general practice: A cluster randomized controlled trial, *PLoS One* **12** (2017), e0170974.

[14] L.O. Karlsson, S. Nilsson, M. Bang, L. Nilsson, E. Charitakis, and M. Janzon, A clinical decision support tool for improving adherence to guidelines on anticoagulant therapy in patients with atrial fibrillation at risk of stroke: A cluster-randomized trial in a Swedish primary care

setting (the CDS-AF study), *PLoS Med* **15** (2018), e1002528.
[15] S. van Doorn, F.H. Rutten, C.M. O'Flynn, R. Oudega, A.W. Hoes, K.G.M. Moons, and G.J. Geersing, Effectiveness of CHA2DS2-VASc based decision support on stroke prevention in atrial fibrillation: A cluster randomised trial in general practice, *Int J Cardiol* **273** (2018), 123-129.
[16] P.L. Hess, S. Kim, J.P. Piccini, L.A. Allen, J.E. Ansell, P. Chang, J.V. Freeman, B.J. Gersh, P.R. Kowey, K.W. Mahaffey, L. Thomas, E.D. Peterson, and G.C. Fonarow, Use of evidence-based cardiac prevention therapy among outpatients with atrial fibrillation, *Am J Med* **126** (2013), 625-632 e621.
[17] N.B. de Oliveira and H.H. Peres, Evaluation of a system of electronic documentation for the nursing process, *NI 2012 : 11th International Congress on Nursing Informatics, June 23-27, 2012, Montreal, Canada. International Congress in Nursing Informatics* **2012** (2012), 312.
[18] L.A. Virginio, Jr. and I.L. Ricarte, Identification of Patient Safety Risks Associated with Electronic Health Records: A Software Quality Perspective, *Studies in health technology and informatics* **216** (2015), 55-59.
[19] A. Idri, M. Bachiri, and J.L. Fernandez-Aleman, A Framework for Evaluating the Software Product Quality of Pregnancy Monitoring Mobile Personal Health Records, *Journal of medical systems* **40** (2016), 50.

Address for correspondence

Ryota Sakurai, MD, PhD, MPH. Prefer to be contacted by email.
School of Medicine, International University of Health and Welfare
4‐2 Kozunomori, Narita City, Chiba 286‐8686, Japan.
Email: yoshimune-tky@umin.ac.jp
Phone: +81-476-20-7701
Fax: +81-476-20-7702

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190328

Can openEHR Represent the Clinical Concepts of an Obstetric-Specific EHR — ObsCare Software?

Danielle Santos Alves[a], Priscila A. Maranhão[b], Ana Margarida Pereira[b,c], Gustavo M. Bacelar-Silva[b,d], Tiago Silva-Costa[d], Thomas William Beale[e], Ricardo J. Cruz-Correia[b,d]

[a] *Department of Nursing, Post-Graduation Program in Computer Science, Federal University of Pernambuco, Recife, Pernambuco, Brazil*
[b] *Center for Health and Technology and Services Research - CINTESIS, University of Porto, Portugal*
[c] *CUF - Porto Hospital and Institute*
[d] *VirtualCare*
[e] *Ars Semantica (UK); openEHR Foundation Management Board.*

Abstract

ObsCare is an obstetric-specific Electronic Health Record in use in nine Portuguese obstetric departments. Like other EHRs, it faces major challenges related to semantic interoperability and data quality. openEHR is proposed to address those needs. This study aimed to describe a summary representation of Obscare workflow and to validate whether archetypes in the openEHR Clinical Knowledge Manager repository can represent ObsCare clinical concepts. The study included the phases: a) ObsCare form selection; b) Description of the workflow care process; c) Detailed data extraction; and d) CKM models analysis. 379 variables were analyzed: 219 were fully represented in CKM repository; 99 were partially represented and needed archetype modification; and 61 were not represented and need new archetypes. To conclude, our study showed that the openEHR CKM repository requires further enhancements to be able to fully answer to the needs of an obstetric-specific EHR, the ObsCare software.

Keywords:

Electronic health records; Obstetrics, Reference Standards.

Introduction

An Electronic Health Record (EHR) contains the results of clinical and administrative encounters by providers (e.g: physician, nurse, nutritionist, and so on) with or for patients that occur during episodes of patient care [1]. The use of EHRs presents several advantages such as availability, data integrity, reduced costs, improved quality of care and health outcomes and enhances the communication between providers and users [1]. The correct recording of health information is essential since it will determine the conduct and decision-making of health professionals, as well as be used as a source for research and audits [1]. Specialist validation of models of EHR information can increase data comprehensiveness and overall quality. Obstetrics is an excellent opportunity to study the potential of EHR implementation. However, general EHRs are inadequate to record and exchange clinical information in the obstetric field, mainly because they do not integrate the content of this clinical area in a structured way and without specific obstetrics contents. A correctly structured EHR can be more easily used by obstetrics-related healthcare professionals and has clear advantages [1].

The ObsCare software was created in 2003 to meet the need for an obstetric-specific EHR and is currently in use in the obstetric departments of nine Portuguese Hospitals. It was produced for use as a clinical record for health professionals, allowing the prospective collection of structured data that can be used to facilitate and improve the analysis of the current situation of obstetric interventions in the Portuguese population (e.g. prevalence and reasons for cesarean section) [2,3].

The ObsCare-specific forms were developed by health information experts, based on World Health Organization (WHO) guidelines. These forms were iteratively updated over time, incorporating new knowledge and also the practical needs of obstetrics departments. ObsCare contains forms to record clinical and administrative data from appointments, hospital admissions, labor and abortion, surgical procedures, emergency department visits, and newborn data; it also includes a timeline representing the most relevant timepoints of each patient's care. In 2017, ObsCare was used by obstetricians and obstetric nurses to record data from 14,034 newborns [2].

However, ObsCare and other EHRs have major challenges related to semantic interoperability and data quality. In recent years, several projects exist to address issues of EHR modeling and implementation, focusing on the development of a future-proof EHR system [4,5]. One of the possible approaches is the use of openEHR, a set of open specifications for EHR architecture [5–7] which assumes a two-level framework, Reference Model (RM) and Archetype Model (AM) [8]. These two levels are applied to separate knowledge and information models [9,10]. While RM represents the semantics of storing and processing in the system, the AM is the keystone of openEHR. Archetypes are structured and complete representations of domain-level information concepts [11,12]; in others words, AM contains the knowledge enabling environment by defining domain level structure and constraints on the generic data structure residing in the RM [13] . Clinicians have an important role in the creation and improvement of these archetypes, which empowers them to directly influence EHR functioning and the quality of patient care [11,12]. Archetypes can be classified, regarding their topics, into two groups: 'general medicine' (e.g. drug administration, vital signs, etc) and specialist. Any speciality will routinely generate data of both categories, and therefore has requirements both for specific data points within the general archetypes, as well as completely specialist archetypes.

To date, there is a lack of research evidence on the possible use of openEHR archetypes to represent obstetric clinical EHR information and the available studies focus only on prenatal care data validation and the birth plan [14,15]. This study aimed to describe a summary representation of Obscare workflow and

to validate whether archetypes in the openEHR Clinical Knowledge Manager repository can represent ObsCare clinical concepts.

Methods

This study was developed based on the following steps:

a) **ObsCare form selection**

The authors chose the labor/abortion form from ObsCare, which includes detailed data on labor and abortion occurrences and procedures.

b) Description of the care process workflow

The authors identified the steps of the obstetric care process used to fill the selected ObsCare forms. Each step represents one ObsCare subform, built based on the available literature and integrating expert opinion over the 15 years of ObsCare clinical use. A clinical workflow related to the whole process was modelled, following the openEHR Task Planning Model [16], a recent specification of work plans in healthcare, and designed as a formal extension to the main openEHR EHR information models [16]. We used the Business Process Model and Notation (BPMN) for reasons of tooling availability to express the workflow model [17].

c) Detailed data extraction

The forms selected in step a) were reviewed by the authors to extract all the included variables. This review was performed by healthcare professionals, one of them a gynecology and obstetrics specialist. All the fillable fields available in the selected ObsCare form were included as variables. When the same variable was repeated in more than one ObsCare subform, only one was included in this analysis.

d) Clinical Knowledge Manager (CKM) analysis.

The identified variables and the underlying clinical concepts were searched for in the openEHR International Clinical Knowledge Manager (CKM) repository [18,19].

CKM is a system for collaborative development, management, and publishing of a wide range of clinical information artefacts. To perform the CKM analysis, the authors organized a variable list including all the ObsCare retrieved variables, and the search was conducted using the name of the clinical concept, keywords, and core-data items. When the existing CKM archetype included all the needed information, it was classified as "complete". We classified the existing archetypes as incomplete and considered reformulating them when they could represent the variables, but we considered that there was a need to improve the concept description and/or add additional items to the archetypes. The authors considered proposing new archetypes when the concepts were not available in the CKM repository.

We also analyzed the type, openEHR status, and specialty of the CKM archetypes that can be used to represent ObsCare variables.

Results

The ObsCare labor/abortion forms include eight sections: (1) abortion record, (2) labor data, (3) labor analgesia, (4) external cephalic version, (5) shoulder dystocia, (6) instrumental birth, (7) postpartum bleeding, and (8) complementary information. The use of the selected ObsCare form follows the workflow care process represented in Figure 1.

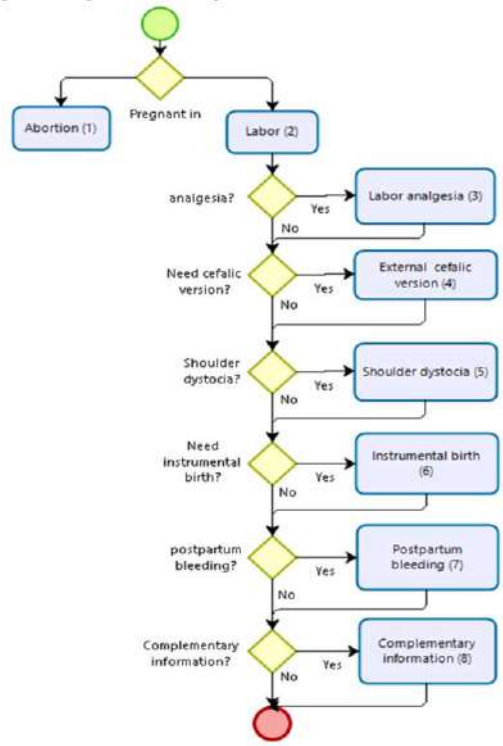


Figure 1 - ObsCare obstetric care process workflow represented in BPMN

We identified 379 variables in the selected ObsCare form. Fifty-eight percent (n=219) of the identified variables can be fully represented by the archetypes available in the CKM. Twenty-six percent (n=99) additional variables were partially represented by the existing archetypes with the need for some adjustment. Only 16% (n=61) were not represented by an existing archetype and a new archetype(s) will need to be created (Figure 2).

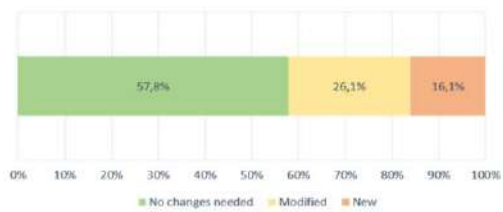


Figure 2 - Summary of openEHR repository status to represent the variables identified in the selected ObsCare form.

a) openEHR Archetypes: complete and requiring adjustment

318 ObsCare variables could be completely or partially represented by the 33 existing openEHR archetypes identified in table 1.

Table 1 - Archetypes used in the ObsCare form. The frequency of use and the need for modification is also provided.

	Total	Need modification	
openEHR archetypes	n	n	(%)
Procedure	68	42	(61.8)
Healthcare provider identifier	51	1	(2.0)
Individual healthcare provider	45	0	(0.0)
Pregnancy Summary	36	36	(100.0)
Problem/diagnosis	23	0	(0.0)
Medication summary	15	0	(0.0)
Medication order summary	12	1	8.3
Service	11	0	(0.0)
Examination/findings	9	5	55.6
Laboratory test result	8	0	(0.0)
Medication Management	4	0	(0.0)
Inspired Oxygen	4	2	50.0
Medication	3	0	(0.0)
Immunization Summary	3	0	(0.0)
Discharge Summary	3	3	100.0
Ventilator settings/findings	2	2	100.0
Transfusion	2	2	100.0
Fetal heart monitoring	2	0	(0.0)
Anatomical Location	2	0	(0.0)
Adverse reaction risk	2	0	(0.0)
Weight	1	0	(0.0)
Uterine Contractions	1	0	(0.0)
Symptom/sign	1	1	100.0
Specimen	1	0	(0.0)
Pregnancy/ Breast-Feeding Status	1	1	100.0
Palpation of fetus	1	0	(0.0)
Obstetric summary	1	1	100.0
Inspection_body_fluid	1	1	100.0
Height/Length	1	0	(0.0)
Head circumference	1	0	(0.0)
Fetal Heart Rate	1	1	100.0
Body surface area	1	0	(0.0)
Apgar score	1	0	(0.0)
Total	318	99	(31.1)

The archetype "Procedure" was the most frequently used to represent ObsCare variables (n=68), in relation with the high number of form variables that are used to describe labor-related interventions; nevertheless, it would require modification to adequately represent nearly 62% of those variables. "Healthcare provider identifier" and "Individual Healthcare provider" were the 2nd and 3rd most applied archetypes. These archetypes were used to represent data from the health professionals involved in obstetric care (e.g., the professionals who assisted the labor/birth or that were called for a complication) and would only need update to represent one variable (unknown healthcare provider assisting to labor).

Finally, "Pregnancy summary" and "Problem/diagnosis" archetypes are also frequently needed to represent ObsCare data. These archetypes correspond to records of obstetric information (e.g.: gestational age) and maternal or fetal diagnoses. However, "Pregnancy summary" could not fully represent all of the ObsCare variables.

Table 2 represents the type, speciality and openEHR status of the archetypes. Most of the archetypes used in the ObsCare form were 'Cluster' structures, closely followed by 'Evaluation' and 'Action' archetypes. Although most of the reviewed archetypes were generic, 14% of them were specific to obstetrics. Moreover, although half (50.8%) still have the status of 'Draft' - which makes editing suggestions possible - 44.5% are "Published" i.e. already formally validated by health professionals.

Table 2 - Archetypes type, speciality, openEHR status.

Type	*%*
Cluster	28.0
Action	26.4
Evaluation	25.5
Element	14.0
Observation	5.2
Admin	0.9
Speciality	*%*
General	85.9
Obstetric	14.1
openEHR status	*%*
Draft	50.8
Published	44.5
Team review	2.8
Initial/Predraft	1.6
Reassess	0.3
Total	**100.0%**

b) Variables not represented by existing openEHR archetypes

The authors found that around 16% of the variables from the selected ObsCare form were not represented in the CKM. These variables are related to several areas of obstetrics, such as placenta, bishop index, number of pulls during instrumental birth, anesthetic blockage space, effectiveness of anesthesia assessed by the doctor and the patient, passage to the operating room (hours) in the context of post-partum bleeding, type of abortion, and fetal position, among others.

Discussion

In this study, we described the status of the openEHR CKM to respond to the needs of a specific obstetric product — ObsCare. We identified 379 variables in the selected form, from which 26% would need the corresponding openEHR-archetypes to be modified and 16% would need new archetypes.

ObsCare has been used for the last 15 years, and the users (obstetricians, nurses, neonatologists, among others) have been validating and suggesting additional data, enriching the content of the form. ObsCare has eight forms (and their subforms) and a timeline, which presents the most relevant information from a particular patient, summarizing and making data visualization much easier. However, ObsCare developers acknowledge that this EHR needs improvements to become a future-proof tool, to strengthen data quality and precision, interoperability, and so on.

The openEHR platform, which supports interoperability and the capture of all clinical knowledge in a structured way separated from the software, provides the formalism and tools to directly link healthcare professionals and developers [5].

Another important point is related to obstetrics procedures. During the labor, birth and abortion care, healthcare professionals perform numerous actions related to diagnosis and treatment. For example, medication administration and blood collection. These diagnostic and treatment workups involve the cooperation of several important health professionals (e.g. nurses, technicians, obstetricians, anesthesiologists and so on) and that is reflected in the high number of variables present in ObsCare that are related to the description of procedures and the identification of the ones that perform them. Because of the complexity of labor and abortion care, a lot of archetype modifications were seen as necessary. In this step, the researchers experienced some challenges to establish the best way to use some particular archetypes.

A general criticism of the CKM and archetype-based modelling is that it is not always easy for healthcare specialists unfamiliar with the formalism or methodology to locate or determine the best concept to use within the 'general medicine' archetypes and specialist ones. Nevertheless, engaging specialists in modelling enables significant improvements in both the general archetypes (by addition or modification of specialist related attributes, e.g. in models like 'medication administration') and the specialist archetypes.

Accordingly, the archetypes relevant to the ObsCare forms include numerous 'general' archetypes, as well as others specific to the obstetric field.

Fortunately, many openEHR archetypes have the "draft" lifecycle status in CKM, meaning that change requests may be submitted in the current development phase.

The small amount of data required for the ObsCare forms not represented in the current openEHR archetypes, reinforces the need for proposals for new archetypes that represent the entire pregnancy cycle until six weeks postpartum, such as gynecological history, labor, birth, placenta, postpartum, newborn and so on [15]. Most of these variables would need obstetric-specific archetypes to be adequately represented. Such variables are indispensable for an obstetric EHR. These results were already found in another study and reinforce the need for more research in obstetric-specific modelling [15]. By these results, obstetric data could be organized and better used by other developers and health professionals.

On the other hand, the presence of variables with incomplete records and variables of doubtful meaning is a standard problem and is well-documented in EHRs [6,7]. ObsCare also exhibits these problems; we found that the selected ObsCare form would benefit from improvements on how the questions are formulated to ensure that all healthcare professionals correctly understand the content to be included in each form field. Consequently, we believe that the openEHR platform can be used to improve ObsCare.

The major contribution of this paper is to describe the relevant data, workflow, and the need for new and modified archetypes related to labor, birth, and abortion. Using agreed structured data expressed as archetypes would achieve semantic interoperability [8,10] between the EHR and other data users.

Governmental institutions, medical associations and other groups that produce clinical knowledge artefacts (e.g. clinical guidelines) could achieve significant benefits by incorporating domain-level information models in the form of archetypes, such as those described here for obstetrics, within their programs. The availability of computable artefacts would express obstetric information structures and needs in a way that can be directly used in the development of EHR and also for clinical decision support systems (CDSS). Using this structured data, the development of EHR could reduce the heterogeneity of similar applications and improve semantic interoperability [20]. openEHR presents the highest level of interoperability (level 4), where the data are electronically interpreted. In this case, the transferred messages are structured and contain standardized and coded data. This is the ideal state where systems exchange information using the same formats and vocabularies. In addition, openEHR improve EHR data quality, which is related to data accuracy, completeness, timeliness, among others [7,21]. ObsCare is a pioneer product in Portugal; neverthless, upgrading it to the openEHR platform is a challenging process and demands many steps, the first being to identify the ObsCare clinical concepts and match them with archetypes that are already available in CKM. Future steps will include proposing modifications to already existing archetypes and the creation of those that are not yet available.

Although the importance of user requirements capture prior to the construction of any computer system is much referenced in the literature, it was observed that there were few such studies available for IT systems in obstetrics [15,20,22].

In a study similar to the one reported here, that aimed to validate the minimum data set needed for a prenatal care EHR using archetypes, similar results were found, i.e. that significant CKM archetype reuse was possible along with a certain amount of modifications and proposed new models [15].

Overall, however, there is a relative paucity of international research studies related to formal modelling for health information systems or the use of the openEHR platform in specialist clinical areas, including in obstetrics in primary care [14,15,20]. Health information systems that use the openEHR platform offer great potential to improve clinical practice in obstetrics, where the same woman is seen many times by different physicians and other health professionals. The complexity of the diagnostic and therapeutic workups leads to a strong need for a higher direct involvement by healthcare professionals in the semantic modelling as well as workflow and usability design, contributing to the improvement, development and greater use of these systems [1,15,20,22]. The archetype formalism, methodology, and tools provide a significant step forward in enabling this involvement.

The current study has some limitations that must be considered. First, it is based on a single form from a wider obstetric-EHR; the labour/abortion form is one of the most commonly used ObsCare forms, nevertheless, it will be important to extend this approach to all other forms. Moreover, although ObsCare has been in clinical use, in Portugal, over the last 15 years, this study did not include an external validation of the extracted variables and they might not adequately represent other settings or countries' practices. Additionally, although this study allowed the identification of important clinical concepts present in ObsCare and matched them with already existing CKM archetypes, it is only an initial development and there is still need of further steps.

Conclusions

The present study showed that the openEHR CKM repository requires further enhancements to be able to fully answer to the needs of an obstetric-specific EHR, the ObsCare software. Our future perspective is modelling the missing obstetric archetypes and modify those that already exist. Following that, we will incorporate these standards into the future evolution of ObsCare.

Acknowledgments

The authors would like to acknowledge - Project "NORTE-01-0145-FEDER-000016" (NanoSTIMA) is financed by the North Portugal Regional Operational Program (NORTE2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

References

[1] G.M. Bacelar-Silva, P.M. Vieira-Marques, and R.J. Cruz-Correia, Identifying relevant data along selected medical care situation, *6th Iberian Conference on Information Systems and Technologies (CISTI 2011)*. (2011) 1–4. doi:10.1007/s12210-010-0111-3.

[2] VirtualCare, VC ObsCare – VirtualCare – Systems for Life, (2019). http://virtualcare.pt/index.php/portfolio/vc-obscare/.

[3] J. OLIVEIRA, S. TAVARES, N. MONTENEGRO, and C. RAMALHO, Influence of maternal age on caesarean delivery: experience in a tertiary hospital., *Acta Obstétrica e Ginecológica Portuguesa*. **12** (2018) 99–104. http://www.scielo.mec.pt/scielo.php?script=sci_abstract&pid=S1646-58302018000200003.

[4] L. Wang, L. Min, R. Wang, X. Lu, and H. Duan, Archetype relational mapping - a practical openEHR persistence solution, *BMC Medical Informatics and Decision Making*. **15** (2015) 88. doi:10.1186/s12911-015-0212-0.

[5] H. Leslie, openEHR - the world's record, (2015). https://www.researchgate.net/publication/277666797_openEHR_-_the_world's_record.

[6] K. Thiru, A. Hassey, and F. Sullivan, Systematic review of scope and quality of electronic patient record data in primary care, *BMJ*. **326** (2003) 1070–0. doi:10.1136/bmj.326.7398.1070.

[7] N.G. Weiskopf, and C. Weng, Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research, *Journal of the American Medical Informatics Association*. **20** (2013) 144–151. doi:10.1136/amiajnl-2011-000681.

[8] S. Garde, R. Chen, H. Leslie, T. Beale, I. Mcnicoll, and S. Heard, Archetype-based knowledge management for semantic interoperability of electronic health records, *Studies in Health Technology and Informatics*. **150** (2009) 1007–1011. doi:10.3233/978-1-60750-044-5-1007.

[9] J. Patrick, R. Ly, and D. Truran, Evaluation of a Persistent Store for openEHR, (2006) 7. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.488.1555&rep=rep1&type=pdf.

[10] L. Bird, a Goodchild, and Z.Z. Tun, Experiences with a two-level modelling approach to Electronic Health Records, *Journal of Research and Practice in Information Technology*. **35** (2003) 121–138.

[11] H. Leslie, Archetypes 101, (2006). https://www.researchgate.net/publication/228702894_Archetypes_101.

[12] T. Beale, Archetype Technology Overview, *OpenEHR Fundation*. (2018). http://www.openehr.org/releases/AM/latest/docs/Overview/Overview.html#_business_purpose_of_archetypes.

[13] T. Beale, and S. Heard, Archetype Definitions and Principles, 2007. https://specifications.openehr.org/releases/1.0.2/architecture/am/archetype_principles.pdf.

[14] J.M. Carrilho, Z.S.N. Reis, G.C. Osanan, and R.J.C. Correia, PROPOSIÇÃO DO PLANO DE PARTO INFORMATIZADO PARA APOIO A INTEROPERABILIDADE E HUMANIZAÇÃO, in: S.B. de I. em S. SBIS (Ed.), XV Congresso Brasileiro de Informática Em Saúde, Sociedade Brasileira de Informática em Saúde, Goiânia, Goias, 2016: pp. 713–720. www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis.

[15] D.S. Alves, V.C. Times, and M. de A. Novaes, Validation of Minimum Data of Archetyped Telehealth Clinical Report for Monitoring Prenatal Care, *Studies in Health Technology and Informatics*. **216** (2015) 64–68. doi:10.3233/978-1-61499-564-7-64.

[16] T. Beale, Task Planning Model Specification, *OpenEHR Fundation*. (2019). https://specifications.openehr.org/releases/PROC/latest/task_planning.html.

[17] F. Shamoon, H. Leitich, and J.S. De Bruin, Clinical Workflow Modeling in Obstetrics : Hepatitis B in Pregnancy, in: MEDINFO 2017: Precision Healthcare through Informatics, 2017 International Medical Informatics Association (IMIA) and IOS Press, China, 2017: p. 1336. doi:10.3233/978-1-61499-830-3-1336.

[18] OpenEHR, Clinical Knowledge Manager, *OpenEHR Fundation*. (2018). http://www.openehr.org/ckm/.

[19] S. El Heloua, T. Karvonena, G. Yamamotob, N. Kumec, S. Kobayashic, E. Kondod, S. Hiragib, K. Okamotob, H. Tamurab, and T. Kurodab, Generation of openEHR Test Datasets for Benchmarking, in: MEDINFO 2017: Precision Healthcare through Informatics, 2017 International Medical Informatics Association (IMIA) and IOS Press, China, 2017: p. 1266. doi:10.3233/978-1-61499-830-3-1266.

[20] G.M. Bacelar-Silva, R.F. Sousa-Santos, and R.J. Cruz-Correia, Creating openEHR content to different moments of care: Obstetrics emergency scenario, *Proceedings of CBMS 2013 - 26th IEEE International Symposium on Computer-Based Medical Systems*. (2013) 361–364. doi:10.1109/CBMS.2013.6627816.

[21] M.G. Kahn, M.A. Raebel, J.M. Glanz, K. Riedlinger, and J.F. Steiner, A Pragmatic Framework for Single-site and Multisite Data Quality Assessment in Electronic Health Record-based Clinical Research, *Medical Care*. **50** (2012) S21–S29. doi:10.1097/MLR.0b013e318257dd67.

[22] Z. Silveira, N. Reis, R. João Cruz Correia, and A. Da Costa Pereira, Sistemas eletrônicos de informação na assistência e pesquisa em saúde da mulher: para quando um maior envolvimento dos profissionais de saúde? Electronic information systems for women's health care, teaching and research: a plea for an higher involvement of health professionals?, *Rev Bras Ginecol Obste*. **33** (2011) 107–10. http://www.scielo.br/pdf/rbgo/v33n3/a01v33n3.pdf.

Address for correspondence

Danielle Santos Alves

Federal University of Pernambuco. Nursing Department. Av. Prof. Moraes Rego, 1235 - University City, Recife - PE – Brazil – Zip Code: 50670-901 Phone: (+55 81) 981087638. dsa3@cin.ufpe.br

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190329

An Obstetric Application Architecture for Information, Diagnosis and Control of Diabetes in High Risk Pregnancy

Danielle Santos Alves[a,c], Marcus Caio de Moura Ferreira Gomes[a,b], Magdala de Araújo Novaes[a,d]

[a] *Telehealth Center, Clinical Hospital, Federal University of Pernambuco, Recife, Pernambuco, Brazil*
[b] *Medicine Course, Federal University of Pernambuco, Recife, Pernambuco, Brazil*
[c] *Department of Nursing, Post-Graduation Program in Computer Science, Federal University of Pernambuco, Recife, Pernambuco, Brazil.*
[d] *Department of Internal Medicine, Federal University of Pernambuco, Recife, Pernambuco, Brazil*

Abstract

Hyperglycemia associated with pregnancy has been related to several unfavorable perinatal outcomes, as well as to the increase of incidence on future complications. Thus, the diagnosis of hyperglycemia in the pregnancy-puerperal context should be a global health concern. This article presents the development of a mobile application prototype which informs, imparts accurate diagnosis and provides tools for obstetric telemonitoring in women with pregnancy associated diabetes. After detailed analysis of the proposals for a new diagnostic strategy considering possible scarcity of resources, key elements were selected and inserted into the prototype, in order to cater for the most aspects and make it as thorough as possible. The application has been adapted to the Brazilian reality, however, having adjustments in compliance with international protocols, it has the potential to be used worldwide.

Keywords:

Mobile Applications; Diabetes, Gestational; Obstetrics

Introduction

It is known that hyperglycemia is one of the most common medical conditions during pregnancy. Besides that, the obstetric history of Gestational Diabetes Mellitus (GDM) is the main risk factor for the development of type 2 diabetes and metabolic syndrome in post-pregnancy women. Thus, it is feasible to assume that hyperglycemia during the pregnancy-puerperal cycle constitutes a relevant problem nowadays, considering not only the risks of worse perinatal outcomes (increased cesarean deliveries, traumas and distorts at delivery, macrosomia, respiratory distress and neonatal hypoglycemia, among other complications), but also the risk of developing future maternal diseases. This fact can also be extended to the children of these women, thus increasing the risk of the children developing, for example, obesity and type 2 diabetes in the future. Hence, it is evident that the diagnosis of GDM should be considered a global health concern [1-3].

Regarding the diagnosis, when hyperglycemia is detected during pregnancy it is classified as either Gestational Diabetes Mellitus (GDM) or diabetes mellitus in pregnancy (DIP). Women with slightly elevated blood glucose levels are classified as having GDM and women with substantially elevated blood glucose levels are classified as women with DIP. DIP may either have been pre-existing diabetes (type 1 or type 2) antedating pregnancy, or diabetes first diagnosed during pregnancy [3].

The predominance of hyperglycemia during pregnancy may vary depending on the diagnostic criteria used as well as the population studied. The International Diabetes Federation (IDF) estimates that approximately one in six live births (16.2%) in 2017 was related to some form of hyperglycemia in pregnancy. An estimated 86.4% of those cases were due to GDM, 6.2% due to diabetes detected prior to pregnancy, and 7.4% due to other types of diabetes (including type 1 and type 2 diabetes) first detected in pregnancy. Estimates regarding hyperglycemia in gestation in Brazil are inconsistent, but it is estimated that the prevalence of GDM in the Brazilian Health System is of approximately 18% [1,2].

GDM context in Brazil

Hyperglycemia (high blood glucose level) is currently inserted in the context of an actual obesity epidemic that has been observed in several countries. It is estimated that approximately 58% of the female diabetes mellitus (DM) cases in Brazil are attributed to obesity, whose causes are related, among other factors, to poor diet and nowadays' lifestyle [4].

In this context of increased prevalence of obesity, associated with the raising of the maternal age and also the lack of physical activity, especially in the last two decades, there was a progressive escalation in the number of women diagnosed with diabetes at childbearing age and during the pregnancy-puerperal cycle [2].

Considering the prevalence and the various short-term and long-term consequences of GDM for mothers and their children, a discussion forum on the topic was held in São Paulo, Brazil, on August 1, 2016. Participants in this forum were physicians specialized in assisting women with GDM: consultants from the Pan American Health Organization / World Health Organization (PAHO / WHO in Brazil) and technical advisors from the Ministry of Health, obstetricians of the Brazilian Federation of Gynecology and Obstetrics (FEBRASGO) and endocrinologists of the Brazilian Diabetes Society (SBD). The purpose of this gathering was to draw a proposal for the diagnosis of GDM in Brazil, taking into consideration that some countries follow protocols related to the diagnosis and management of hyperglycemia during pregnancy inserted in their respective contexts. The discussion took into account previous general agreements and studies, adding Brazilians particularities. Therefore, two strategies were proposed towards the diagnosis for GDM for the population, especially considering the fact that Brazil is a continental country, and that, consequently, it is subjected to scenarios of scarcity and fragility in the assistance in various places. The two strategies

are based primarily on the availability of the oral glucose tolerance test with 75g of glucose (75-g OGTT) [2].

These two strategies were presented and became part of a document – published by PAHO / WHO in Brazil, Ministry of Health, FEBRASGO and SBD in 2017 – called "Screening and diagnosis of Gestational Diabetes Mellitus in Brazil" (free translation), in which the main points of consensus on the forum were presented taking into account the distinctness in access to health services in Brazil [2]. However, in spite of the fact that these strategies were designed for the Brazilian context, where there is a lack of resources available in some areas, such strategies could be used in a worldwide perspective in places where access to the resources of laboratory tests in pregnancy is not so easy.

The use of mobile technology for gestational telemonitoring

Mobile applications in the context of an increasingly digitized world are inserted in the context of pregnancy, including, among other factors, access to information available as a source of support and advice and the use of several new devices – thus, they are able to change how the pregnancy and maternity are understood and practiced [5].

Diet control and physical activity in pregnancy, for instance, are essential issues in the treatment of hyperglycemia during pregnancy. And, even though health professionals are constantly trying to encourage healthy behaviors based on such issues, verbal information can be easily forgotten, and printed information may be lost as well. In this context, the use of applications in health interventions seems to be more appropriate for a better fidelity to the treatment, considering its various purposes and forms of communication rather than just texts (sounds, images, interactivity) [6].

A recent review involving telemonitoring effectiveness in obstetrics brought about the presence of several telemonitoring applications in this area of medicine. Some of them are related to cervical dilatation or preterm labor, GDM, maternal satisfaction, health care-related costs, birth weight and gestational age. Regarding the management of GDM, the telemedicine application observed in the studies focused mainly on mechanisms of transfer of glucose values from the patient to a provider, which reduced the need for frequent clinical visits and possible maternal, fetal or neonatal adverse outcomes [7].

Thus, in view of the high prevalence of hyperglycemia in the pregnancy-puerperal period and all the complications associated, as well as the fact that follow-ups through applications could generate better adherence to the treatment, we were highly motivated to the development of this work.

This article presents a mobile application prototype to inform, provide a correct diagnosis and supply tools for obstetric telemonitoring involving glycemic control in women with diabetes associated with pregnancy, taking into consideration the probability of a context of scarcity of laboratory resources.

Methods

This is an applied study based on another research (trial), carried out from January to October 2018. The proposed application for glycemic control in women with diabetes associated with pregnancy will be incorporated into an obstetric telemonitoring system for the detection and early intervention in the main gestational intercurrences, this system is being tested in the context of another parallel study by the authors.

The application is aimed at women in the pregnancy-puerperal cycle who have some alteration of glycemic levels and also for the health professionals involved with their care.

The study was conducted in three stages:

1. Detailed analysis of the strategies proposed by the document “Screening and diagnosis of Gestational Diabetes Mellitus in Brazil”;
2. Definition of key elements on what would be available in the application, aiming to provide users with the three pillars proposed by us in order to be as thorough as possible (I-D-F): (1) Information – which includes updated concepts about diabetes during pregnancy, with accessible language; (2) Diagnosis – which must be correct and timely, based on the context in which the patient is inserted and according to the established protocol (in this case, as defined in the document "Screening and diagnosis of Gestational Diabetes Mellitus in Brazil"); and (3) Follow-up – through the evolution of glycemic control, diet and patient weight;
3. For the development and presentation of the prototype, the Balsamiq® tool was used, which allows the construction of graphic interfaces.

Results

Analysis of the document “Screening and diagnosis of Gestational Diabetes Mellitus in Brazil”

The document in question was analyzed at length with the intention of providing a precise perception of the diagnostic process to be inserted in the prototype. The two strategies proposed in the document are basically based on whether the oral glucose tolerance test with 75 g glucose (75-g OGTT) is available for or not in the region where the pregnant / puerpera woman is.

In case of availability of such test, all pregnant women should perform the fasting blood glucose test (up to 20 weeks of gestational age) for diagnosis of GDM and Diabetes first diagnosed during pregnancy. All pregnant women with fasting blood glucose lower than 92 mg/dL should perform 75-g OGTT between 24 and 28 weeks. If the onset of prenatal care is delayed (after 20 weeks of gestational age), the OGTT should be performed as soon as possible. In this way it is estimated that 100% of cases of GDM should be detected [2].

In cases where just the fasting glucose is available, all pregnant women should perform fasting glucose at the beginning of prenatal care for the diagnosis of GDM and Diabetes first diagnosed during pregnancy. If the test results are below 92 mg/dL, before 24 weeks of gestational age, fasting glucose should be repeated for 24 to 28 weeks. It is estimated that in this way that 86% of cases of GDM are detected [2].

The diagnosis in the puerperal period also follows the strategies based on the availability of 75-g OGTT. Thus, in case of availability, the 75-g OGTT six weeks postpartum sets the gold standard for diagnosing diabetes after pregnancy (100% detection). If only fasting glucose is available, it is estimated that only 66% of the cases of changes in glucose metabolism, including DM [2], might be diagnosed.

Application of the key points in the development of the prototype

Taking as a starting point the recommendations of the document in question, a list of key elements which should be included in the prototype was made. This inclusion was done by using Balsamiq®, a medium-fidelity prototyping software, which

allows the creation of functional interfaces, giving the user an initial impression of the operation of the application.

The proposal to create a login for the access of to the application was raised due to the fact that the patients' information should be only accessed by them, and made it available to third parties, including health professionals, solely under their authorization.

As for the login, personal identification data is filled in, whether the user is a health professional, pregnant or puerpera, the gestational age or postpartum time, as well as a valid personal email and a password. The application user must agree to the terms of use of the application in order to obtain access (Figure 1).

Figure 1 – Screen in which there are fields for insertion of information that will compose the profile of the user

The application menu has been designed to be presented in a simple way, either on the home screen or based on a quick access (Figure 2), which enables the migration between topics without necessarily returning to the home screen. The topics were distributed as follows: My Profile; About Diabetes in Pregnancy; Diagnosis of Diabetes in Pregnancy; My glycemic control; Alerts; Settings; About; Logoff.

Regarding the Information pillar, we are concerned to inform patients, in clear and accessible language, of common questions (Figure 3) about diabetes, whether gestational or not. For this pillar, the "About Diabetes in Gestation" section has been targeted in the application, and the questions are: "What is Diabetes?", "What is Gestational Diabetes?", "Why Gestational Diabetes Happens?", "What are the risk factors for having Diabetes in pregnancy?", "How to diagnose Diabetes in pregnancy?" and "How to know if after pregnancy Diabetes will continue?".

Figure 2 – Proposed Start Menu and Quick Access Menu Screens

Figure 3 – Questions from the section "About Diabetes in Gestation"

As for the Diagnosis pillar, it will be correctly done through a series of simple questions, in which the patient or health professional should only answer "yes" or "no" (Figure 4) and choose the values of blood glucose and/or 75-g OGTT obtained during the diagnostic investigation. All the decision making and answers presented by the prototype follow the protocol previously established, be it in the context of pregnancy or puerperium.

Figure 4 – Question 1 that initiates the flow of questions for the correct diagnosis: "At the place of prenatal care is there 75-g OGTT available for the realization?" / Question requiring the value of the first fasting glycemia to set or not immediately the diagnosis: "What was the value of the first fasting glycemia obtained in this gestation?"

As for the Follow-up pillar, in view of the glycemic control, we devised a table synchronized with a chart - this table can be filled with the values of capillary blood glucose, and that will allow to follow the evolution (during a day and to the over several days) and the consequent need for glycemia correction (Figure 5).

In addition to the glycemic control, the Follow-up pillar also involves "Physical exercise, diet, and the weight control" section in which the patients receive information about the types of exercise and diet according to their contexts. Based on the calendar, the patient can enter the days on which she performed physical activity (as well as the nature of the activity and how long it lasted), and her daily diet, as well as the evolution of her weight throughout the days (Figure 5). This information would be saved and would enable a simple and practical analysis, later, by health professionals.

Figure 5 - Screens representing the graph of the section "My glycemic control - Glycemic curve" and "My glycemic control - Physical exercises, diet and weight control"

In this prototype, there is still the possibility of certain patients' profiles to be connected to monitoring by the responsible physician. In this way, the health professionals can have in their device the list of pregnant/puerperal women whom they accompany and can monitor how their blood glucose curve progress in real time. In this way these professionals could contact the user if necessary, advising her, for instance, to an early return to prenatal care.

The "Alerts" section has emerged as a proposal for the pregnant/puerperal women to control functionalities that will serve as true guidelines for their diet, physical exercises and medication (in case of using oral hypoglycemic agents and/or insulin) and for measuring capillary blood glucose. When activating these functionalities, they will be synchronized with the device, which will allow the user to receive, at the scheduled times, notifications that will act as real reminders to carry out those activities throughout the day. In addition, the user can activate features that begin a countdown to the beginning of the period of the laboratory tests of fasting glycemia and 75-g OGTT, when applicable.

Discussion

The recent proposal for the screening and diagnosis of GDM in Brazil has emerged taking into account different contexts of availability of resources (technical and financial), and proposals such as this should be encouraged, especially as they seek greater attention to prevailing conditions and with significant morbidity and mortality if not correctly diagnosed/managed.

In this context, consideration was also given to the fact that caring for hyperglycemia is not confined to pregnancy alone – because even if glucose tolerance normalizes rapidly postpartum in most women who have developed GDM, the risk of developing of type 2 DM or of glucose intolerance is significant. It is for this reason that the monitoring of women with GDM after childbirth is fundamental [2]. Our application proposal, therefore, also seeks to encompass the postpartum context, guaranteeing to inform, correctly diagnose and follow up these women throughout the pregnancy-puerperal cycle.

The control through the glycemic curve proposed by the application can be performed in the ward or at the ambulatory. In these contexts, it is important that the setting of the times and days when the capillary blood glucose dosage should be performed and the analysis of this curve should only be made by a trained physician, as well as possible corrections with drug therapies.

It is well known that mobile applications for self-management of diseases such as diabetes and general well-being (including diet and exercise monitoring) are widely available. However, users are faced with an enormous amount of new applications, which often do not use scientific evidence for their development and do not take into consideration the context in which the user is inserted in. In addition, users often do not seek or receive guidance from health care professionals about choosing the appropriate application [8]. Thus, the application in question appears to change this scenario, because it is a reliable tool, based on evidence and that will be used with the guidance provided by health professionals.

On top of that, the application presented here seeks to improve the relationship between women and health professionals, as well as their relationship to their health condition. Mobile applications of this nature allow patients to become more involved in self-management of their health. This type of follow-up increases confidence, adherence to treatment and participation by patients, giving them the knowledge to act and adopt behaviors to maintain and improve their health by taking actions, asking questions to health professionals and participating in the decision-making process related to their treatment [9]. Hence, it is well known that there is evidence of better health outcomes when patients are involved in self-management of their own diseases [8].

Literature surveys point to the growing number of applications available that are likely to be distributed between the two most popular mobile platforms, iOS (Apple Computer Inc.) and Android (Google Inc.) [10]. Thus, we expect the application to be available on both platforms in the future after its validation, aiming for free access and if possible, encouraged by public government policies.

Some specific potential impacts, such as the declining of morbidity and mortality due to diabetes in the gestation after the correct use of the application, might undoubtedly be observed if the information of this tool is used and followed in the right and early manner, along with the guidelines of health professionals. Nevertheless, they might be difficult to quantify in the future, having as a direct causal relation to the insertion of this application.

It is expected that, starting with the clinical trials with this application, it will be possible to measure the impacts of monitoring glycemic levels during pregnancy. Despite being based on gestational diabetes follow-up protocols advocated by the WHO, the Brazilian Society of Diabetes and the Brazilian Federation of Gynecology and Obstetrics, the use of computational protocols for the development of this prototype allows the parameterization of other protocols in their development. Thus, this application can be applied not only in Brazil but in other countries where the shortage of professionals and laboratory tests are a reality. In addition, from new and fast configurations, it is possible to use this application in other health contexts worldwide.

Conclusions

In the present work, a prototype was presented whose development was based on guidelines and scientific evidence for the monitoring of the glycemic changes during pregnancy, following the three pillars recommended by the authors: Information, Diagnosis and Follow-up (I-D-F), aimed at the early detection, timely treatment and follow-up in order to

reduce maternal and neonatal morbidity and mortality. It can also be used as self-management and a better adherence to treatment and follow-up by the pregnant / puerperal woman. The use of this application is expected to improve potential limitations on the scarcity of exams and professionals, which make it difficult to properly monitor these women and fetuses.

The use of mobile technologies and gestational monitoring applications has proven to be an important strategy to achieve better results in gestational follow-up. This application has great potential for development and application in future studies, assisting on the diagnostic process in the individual context of each patient, raising awareness on the general population about hyperglycemia in pregnancy and subsequent glycemic control in patients with this condition. Future application work in clinical practice will be carried out to evaluate the implementation of this application, analyzing users' opinions (patients and health professionals) and possible limitations or difficulties found in the use of this tool.

Acknowledgements

Thanks to Telehealth Center of Clinical Hospital of Pernambuco for making this research possible.

References

[1] IDF, IDF Diabetes Atlas, 8th ed., 2017. doi:http://dx.doi.org/10.1016/S0140-6736(16)31679-8.

[2] OPAS OP-A da S, BRASIL M da S, FEBRASGO FB de G e O, SBD SB de D. *Rastreamento e Diagnóstico de Diabetes Mellitus Gestacional No Brasil*. 1st ed. (OPAS, ed.). Brasília: OPAS; 2016. http://iris.paho.org/xmlui/bitstream/handle/123456789/34278/9788579671180-por.pdf?sequence=1.

[3] I.F. of G. and O. FIGO, The International Federation of Gynecology and Obstetrics (FIGO) Initiative on Gestational Diabetes Mellitus: A Pragmatic Guide for Diagnosis, Management, and Care, International Journal of Gynecology Obstetrics. 131 (2015) 1397–1405.

[4] L.S. Flor, M.R. Campos, A.F. de Oliveira, and J.M. de A. Schramm, Diabetes burden in Brazil: Fraction attributable to overweight, obesity, and excess weight, Revista de Saude Publica. 49 (2015). doi:10.1590/S0034-8910.2015049005571.

[5] S. Johnson, "Maternal Devices", Social Media and the Self-Management of Pregnancy, Mothering and Child Health, Societies. 4 (2014) 330–350. doi:10.3390/soc4020330.

[6] I. Borgen, L.M. Garnweidner-Holme, A.F. Jacobsen, K. Bjerkan, S. Fayyad, P. Joranger, A.M. Lilleengen, A. Mosdøl, J. Noll, M.C. Småstuen, L. Terragni, L.E. Torheim, and M. Lukasse, Smartphone application for women with gestational diabetes mellitus: A study protocol for a multicentre randomised controlled trial, BMJ Open. 7 (2017). doi:10.1136/bmjopen-2016-013117.

[7] D. Lanssens, T. Vandenberk, I.M. Thijs, L. Grieten, and W. Gyselaers, Effectiveness of Telemonitoring in Obstetrics: Scoping Review, Journal of Medical Internet Research. 19 (2017) e327.doi:10.2196/jmir.7266.

[8] S. Pais, D. Parry, K. Petrova, and J. Rowan, Acceptance of using an ecosystem of mobile apps for use in diabetes clinic for self-management of gestational diabetes mellitus, Studies in Health Technology and Informatics. 245 (2017) 188–192. doi:10.3233/978-1-61499-830-3-188.

[9] J. Sommer, M. Daus, M. Smith, and D. Luna, Mobile application for pregnant women: What do mothers say?, Studies in Health Technology and Informatics. 245 (2017) 221–224. doi:10.3233/978-1-61499-830-3-221.

[10] Arrais RF, Crotti PLR. Revisão: aplicativos para dispositivos móveis ("Apps") na automonitorização em pacientes diabéticos. *Journal of Health Informatics*. 2015;7(4). http://www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis/article/view/359.

Address for correspondence

Danielle Santos Alves

Federal University of Pernambuco. Nursing Department. Av. Prof. Moraes Rego, 1235 - University City, Recife - PE – Brazil – Zip Code: 50670-901 Phone: (+55 81) 981087638. dsa3@cin.ufpe.br

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190330

How to Assess Success of HIT Project Management: An Example of the Use of the Common Assessment Framework (CAF)

Kaija Saranto[a], Eija Kivekäs[a], Milla Rosenlund[a], Virpi Jylhä[a], Pia Liljamo[b], Sirpa Arvonen[c], Ulla-Mari Kinnunen[a]

[a] Department of Health and Social Management, University of Eastern Finland, Kuopio, Finland,
[b]Oulu University Hospital, Oulu, Finland
[c]Helsinki and Uusimaa University Hospital District, Helsinki, Finland

Abstract

The purpose of this study was to assess the project management of a national health information technology project focused on developing digital health care services. An e-questionnaire was sent at the end of the development project, and 93 experts (18%) from five university hospitals responded. The questionnaire was based on the Common Assessment Framework (CAF) to identify management success and also to get an overview of the project's performance. The questionnaire contained 11 background variables and 17 Likert scale items in five themes on leadership, strategy and planning, people, partnerships and resources, and processes, and one open-ended question. After analysis using descriptive statistical methods, the results showed that, overall, participants felt confident about management of the project. Criticism focused on the distribution of resources and lack of knowledge about the status of development activities in other hospitals. The CAF enablers criteria revealed subjects for further development.

Keywords:

Information technology; patient care management; systems analysis

Introduction

Deployment of digitalization is creating a variety of opportunities for the development of new ways to provide health care services. Digital services, meaning encounters with patients through virtual online appointments or electronic communication using various tools and technologies, is a growing means of providing health services [1]. One comprehensive platform for offering digital health services is patient portals, which nowadays have technical functionalities to support national interoperability at the levels of information technology (IT) infrastructure, applications, data, services, procedures, and policies, in addition to security [1–3]. Patient portals are comprehensive gateways providing health information, access to personal health records, the ability for patients to send secure messages to health professionals and manage appointments, decision support tools for self-care and diagnosis, and discussion forums [3–4]. While personal health records are managed by patients, patient portals are managed by health care organizations [4]. Results from many studies show that the change from face-to-face to online encounters through technology has already been accepted by both patients [5–6] and professionals [7–11], and effective treatment outcomes have been achieved [12].

Overall, the development of digital services is an international challenge [1–2] and, obviously, requires strategic goals at a national level. In Finland, the objective of the National eHealth and eSocial Strategy 2020 is to support the active role of citizens in promoting their own well-being by improving information management and implementing self-management and online services [13]. Strategic statements are crucial because they usually include a promise of allocation of resources.

Involvement of multidisciplinary team (physicians, nurses, and other experts in clinical areas as well as IT experts) is key for the development of digital health services [14–15]. The development process in regard to digital services encompasses a variety of actions for defining, procuring, planning, implementing, and providing new services; thus, it is prone to the risk of deterioration of processes and outcomes. According to the International Project Management Association (2006), project management is seen as an umbrella term encompassing planning, organization, monitoring, and control of all aspects of a project, along with motivations for all included to achieve the project objectives in a safe manner, on time, within the budget, and meeting performance criteria [16].

The implementation of health information technology (HIT) is executed as a project that often includes multiple subprojects. The success of HIT implementation has been widely evaluated, covering areas such as HIT adoption, acceptance of technology, and clinical quality, but there has been a lack of focus on or acknowledgment of the organizational context or human factors that could impact HIT implementation [17].

Due to the steadily growing number of projects in health service systems, the need for evaluation frameworks to assess project outcomes is obvious. The concept of project success is based on multiple assumptions regarding structures, processes, and outcomes as parts of a project [17–18]. As a result, many studies have emphasized splitting the concept of project success into two parts: project success, meaning the outcomes, and project *management* success, meaning structures and processes [17]. In particular, de Wit (1988) argued that product success should be assessed separately from the success of project management activities [19]. Radujković and Sjekavica (2017) analyzed various factors affecting project success and classified them into three categories: elements of project management competence, elements of organization, and elements of project management methodologies, methods, tools, and techniques [20]. Roberts (2018) used a framework that included technological, organizational, project-specific, and external factors to assess public health project success [21].

The European Public Administration Network launched the Common Assessment Framework (CAF) in May 2000 as the first European quality management instrument specifically tailored for and developed by the public sector itself. It is a common, basic, accessible, and easy-to-use model for all public

sector organizations across Europe and deals with all aspects of organizational excellence [22]. The CAF is designed to be used by public sector organizations, which distinguishes it from the European Foundation for Quality Management (EFQM) framework, which was founded within industry[23]. The CAF instrument is in the public domain, with the manual available online. Figure 1 presents the CAF framework with criteria [22].

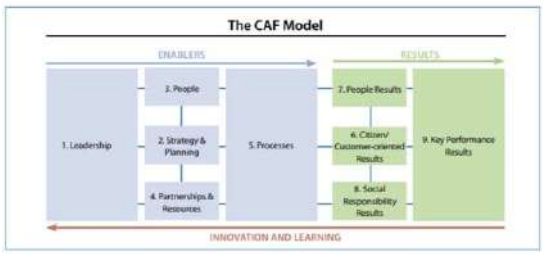


Figure 1. The CAF Framework [22]

The CAF framework is based on the premise that excellent results in organizational performance for citizens/customers and society in general are achieved through the leadership that drives people, partnerships, resources, processes, and strategy and planning. The nine criteria of the CAF represent the main factors that should be considered in evaluations looked at from two perspectives: Enablers and Results. Referring to de Wit's (1988) suggestion, the Enablers cover management success, and the Results cover project success [19]. There are five Enabler criteria (leadership, strategy and planning, people, partnerships and resources, and processes) and four Results criteria (citizen/customer-oriented results, people results, society results, and key performance results) [22].

In earlier studies, the CAF framework proved to be helpful for assessing the managerial practices of an organization based on the Enablers criteria and for the trends in results based on the Results criteria. The CAF self-assessment tools have proved to be helpful for finding areas in need of improvement [23–24].

The purpose of this study was to assess the project management success of a national HIT project focused on developing digital services in health care. There was particular interest in finding out how participants experienced the development project management from the perspective of the CAF criteria.

Methods

Study Context

This study was one part of a larger study evaluating a three-year national development project called Virtual Hospital 2.0. Funded by the Ministry of Social Affairs and Health, Virtual Hospital 2.0 was one of the main projects necessary to achieve the strategic goals of the national eHealth Strategy [13]. In particular, the project aimed to build Healthvillage.fi, a digital health service system comprising various digital portals that was developed through a collaboration among five university hospital districts and patient associations in Finland. Each university hospital was involved in the development work by either being responsible for creating a hub with certain care pathways or giving their expertise to other hospitals in their hubs. The first hub in Finland, informing and caring for mental health patients, was launched in 2009, and the second hub, for weight management, was launched in 2015. The success of the first two hubs was the guiding principle with the Virtual Hospital 2.0 project. To date, a total of 30 hubs and 86 pathways have been constructed in the project. Healthvillage.fi delivered three distinct types of services through hubs: 1) public health services available for all citizens, 2) digital care pathways for patients with a diagnosis and a care relationship, and 3) eLearning tools for healthcare professionals aimed at changing working practices.

The sociotechnical approach [15] guided the multidisciplinary teamwork in the development of portals and hub materials for patients with various diseases. The sociotechnical model helped to realize various dimensions of the development project: (1) computing hardware and software infrastructure; (2) clinical content; (3) human–computer interface; (4) people; (5) workflow and communication; (6) internal organizational policies, procedures, and culture; (7) external rules, regulations, and pressures; and (8) system measurement and monitoring [14–15]. Figure 2 illustrates the modified sociotechnical model adopted in the development of the work.

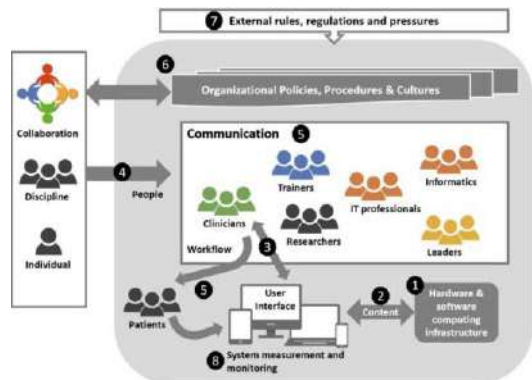


Figure 2. The Sociotechnical Model

In this study, patients were not informants because they have been involved in other studies [5].

Data Collection and Analysis

The data for three distinct studies were gathered with an e-questionnaire containing 11 background variables, 1 open-ended question, and 50 statements using a five-point Likert scale where 5 denoted strong agreement and 1 denoted strong disagreement. The participants (physicians, nurses, and other health care professionals, as well as IT designers [n=501]) were employed in the Virtual Hospital 2.0 project in the autumn of 2018. The concept of Enabler within CAF was applied to measure the management success of the project. The questionnaire contained five themes describing the CAF Enabler criteria: leadership, strategy and planning, people, partnerships and resources, and processes. The 17 items in total were operationalized based on the criteria and dimensions of the sociotechnical model used in the development project [14–15; 22]. The number and format of variables were decided by the evaluation team.

Research permission and ethical approval were received from project management. All questions (n=50) were validated by reference to earlier studies. Finally, the content validity of the questions was assessed by an external expert group, and minor changes were made based on their assessment. The link to the e-questionnaire was sent to each participant in the project (n=501) by their own organizational representatives. Two reminder messages were also sent by each organization.

The results were analyzed using descriptive statistical methods. The Likert scales were taken from the initial answer format of 1=totally disagree to 5=totally agree and modified by combining 1 and 2 to 1=weak, 3 to 2=undecided, and 4 and 5 to 3=strong agreement. The answers to open-ended question

were analyzed with content analysis. Narratives from answers to the open-ended question regarding the primary needs for development were used to confirm the opinions of the respondents.

This study was conducted in an ethical manner and in accord with best ethical research practice. Confidentiality and informed consent of the respondents were maintained so that none of the respondents could be recognized. The research data was stored electronically in secured servers at the University of Eastern Finland where only authorized persons in the research group had access . The data maintenance life cycle followed the principles of UEF, and the data will be destroyed when it is no longer needed.

Results

A total of 91 (18%) participants responded to the e-questionnaire. The majority of respondents were women (88%). Healthcare professions that were represented included nurses (48%), physicians (17%), or other clinicians (16%), and 19% represented various other types of expertise. Among the respondents, 80% had an academic degree. Half of the respondents to the survey represented two of the five university hospitals involved in the project. More than 60% of the respondents had worked in their organization for over 10 years, and 23% worked full-time, 21% part-time, and 56% worked on the project alongside the regular daily duties of their jobs. The participants represented development teams in a total of 14 different hubs. The three most frequently mentioned hubs were rehabilitation, childhood diseases, and heart diseases. Half of the respondents (n=41) replied to the open-ended question, out of which 30 focused on project management.

The project management success is described in the graphics of Figures 3 through 7, with the narrative in italics below the graphics being respondent comments used to support the results. In terms of leadership, the respondents were confident about the atmosphere, but there was some hesitation about the fluency of communication and managerial support (see Figure 3).

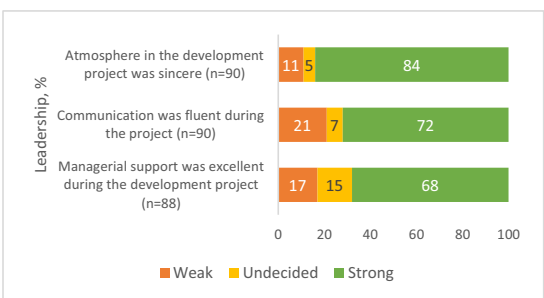


Figure 3. Leadership

The outcomes of the project should not stay inside the university hospitals.

The project management should focus on equal use of resources, both financial and competency-related, in the hospitals.

The respondents could not assess how the project was related to other digitalization projects in the country (Figure 4). They were also unaware of factors regarding the relationship with the development work in primary care.

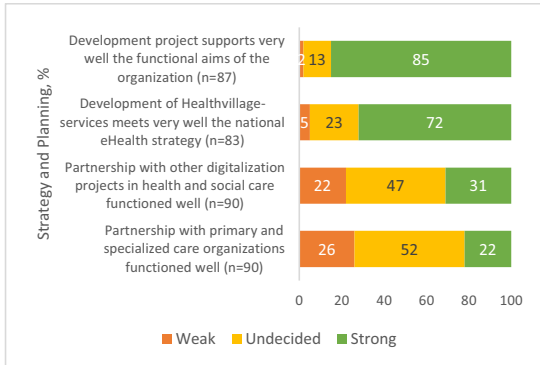


Figure 4. Strategy and Planning

The development work occurring in primary care and specialized care must be integrated.

The development work has produced a variety of material in digital format to support patients and professionals. Now is the time to really focus on dissemination of outcomes over the whole country.

The participants felt very positive about working on the project. Only a small minority had some criticisms (Figure 5).

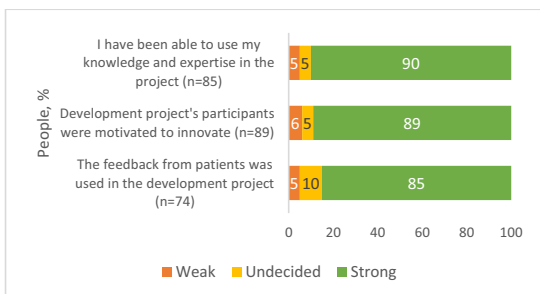


Figure 5. People

Working on this along with my daily duties has been sometimes overwhelming.

We as developers have worked hard on the project. Now I am worried about the rest of the clinicians, how to get them on board.

The development teams are not established based on the need of knowledge; pure enthusiasm is not enough.

Cooperation among various levels of the project seemed satisfactory. Some critical opinions were expressed in terms of working together with other participating university hospitals and the IT experts (see Figure 6).

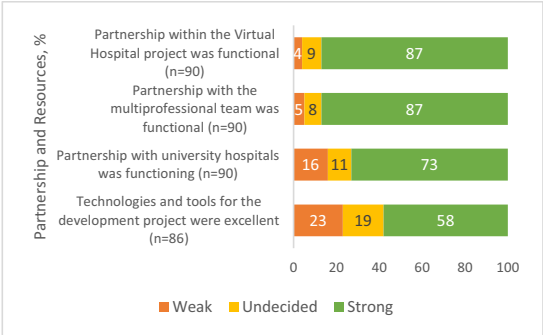


Figure 6. Partnership and Resources

In this project, resources have not been shared equally in terms of workload.

The model to create the hubs has now been tested and implemented many times. So, how to disseminate the knowledge further?

The opinions about the development process varied among the respondents. More than half were satisfied, with an almost equal distribution of those unsure about or against each item (see Figure 7).

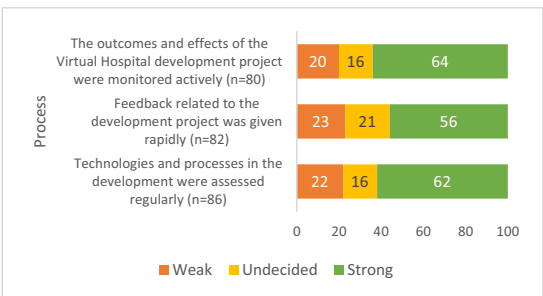


Figure 7. Processes

Too many important decisions have been made inside one university hospital. A national project involves shared and timely decision-making.

The support from IT departments has not functioned well; especially, we have been forced to wait for responses to our feedback.

Funding has not been followed during the project based on the knowledge and competencies in each university hospital due to a different situation in the development.

The use of the CAF's Enabler criteria in the assessment gave new insight for project management. In this case, the CAF was used for assessment at the end of the project, but as the digitalization of services continues, the procedures and policies assessed can be used to guide further development.

Discussion

The purpose of this study was to assess the project management success of a national HIT project focused on developing digital health care services. The specific interest was to assess the experiences of participants with regard to the project management. The data were gathered at the end of the development project, in August 2018. The data collection was organized through development managers so that the respondents had no contact with the researchers, and emails could be used for disseminating the link to the e-questionnaire. The low response rate (18%) may have been caused by the timing of data collection, given that late August is still a holiday season in Finland. Further, the link was sent by the development managers at the university hospitals, and the data collection occurred as the project was ending, so potential participants may have been receiving an excess of other emails. The link was opened a total of 229 times, but we do not fully know the reason why the questionnaire was left unanswered. In addition, because the national development project had also been a key focus among other researchers and stakeholders, some participation fatigue may have existed, particularly given that the project was ending in two months and an extensive amount of material on project outcomes was being collected at this time. What was learned by this could be applied in the future, particularly the involvement of researchers interested in national development projects should be planned at the beginning of a project, especially when pre- and post-project interventions are of interest [1].

While the project manager, team, organizational structure, culture, and atmosphere are still regarded as key components enabling a project's success, it is important to recognize all the factors influencing success [15, 20]. Although the CAF primarily focuses on the evaluation of management performance and the identification of organizational goals to make improvement possible, the ultimate goal is to contribute to good governance [22]. Roberts (2018) found in his study that leadership was the most important project success factor. He suggested that investing in the leadership and project management skills of project participants could improve the success of future projects [21]. In terms of leadership, this kind of multisite project can be challenging for supporting and guiding participants from different organizational cultures. Variations in the level of commitment to project aims may be difficult to monitor, and we found commitment especially challenging given that the project spanned approximately two-years. The strategies and planning we identified in this study referred mostly to national development initiatives in health care, as this project was one of the main development projects funded by the Ministry. Thus, we were surprised to discover that participants had limited knowledge about other national digitalization projects in health care. However, as the sociotechnical model highlights external rules and regulations should be taken into consideration for their influence on organizational policies [14–15].

Many studies have noted health care professionals' hesitation and fear about the implementation of new IT tools in daily practice [10–12]. In this study, half of the respondents were nurses, and they were enthusiastic about being involved in this development work. Participants expressed concerns regarding the rest of the clinicians and patients, and specifically how to ensure their voices were also represented. Further, opinions on partnership and cooperation were mainly very positive, which reflected on the multidisciplinary team's involvement as a part of the adopted sociotechnical model [14–15]. However, we found that while partnership between the university hospitals was satisfied, there were some criticisms around the technologies and tools. To address this, communication and collaboration between IT experts and end users should be of high importance, and considered in all IT projects [25].

Being a national development project, one aim was to create a model to develop digital services [4]. Thus, the processes during development were monitored and assessed. However, the participants wanted to know more regarding the progress, and criticized that the speed of feedback was unacceptable. Participants also stressed that their extensive work should be used for future implementation of digital care pathways.

Conclusions

Based on the project participants' opinions, we applied the CAF to reveal areas for development in project management. The five Enabler criteria of leadership, strategy and planning, people, partnerships and resources, and processes linked well with the sociotechnical model adopted in this development project. It can be argued based on the results that the length of the project is the basis of project management success, as it affects all criteria. However, more emphasis should be placed on communication as culture, leadership, resources, and competencies often vary between health care organizations.

Acknowledgments

The authors want to express gratitude to the Virtual Hospital 2.0 management for funding this external evaluation.

References

[1] World Health Organization, From Innovation to implementation - *eHealth in the WHO European Region*, (2016). http://www.euro.who.int/en/ehealth (accessed on March 23, 2019).[2] M. Rigby, A. Georgiou, H. Hyppönen, E. Ammenwerth, N. de Keizer, F. Magrabi, P. Scott, Patient portals as a means of information and communication technology support to PatientCentric care coordination – The missing evidence and the challenges of evaluation. A joint contribution of IMIA WG EVAL and EFMI WG EVAL, *IMIA Yearb Med Inform* (2015), 145-159.

[3] S. Kujala, I. Hörhammer, J. Kaipio, T. Heponiemi, Health professionals' expectations of a national patient portal for self-management, Int J Med Inform, **117** (2018), 82–87.

[4] Virtual Hospital 2.0. http://www.virtuaalisairaala2.fi/en/virtual-hospital-initiative (accessed on March 23, 2019).[5] K. E. Saranto, E. Kivekäs, P. Kuosmanen, and U-M. Kinnunen, Electronic health services in the patients' daily activities – willingness to use health village services, IOS Press, Amsterdam, *Studies in Health Technology and Informatics,* **247** (2018), 586–590.

[6] E. Ammenwerth, P. Schnell-Inderst, and Alexander Hoerbst, The impact of electronic patient portals on patient care: A systematic review of controlled trials, *J Med Internet Res,* **14** (2012), e162.

[7] E. Kivekäs, S. Mikkonen, E. Borycki, S. Ihantola, and K. Saranto, Physicians' Estimates of Electronic Prescribing's impact on patient safety and quality of care, *Applied Clinical Informatics,* Open **2** (2018), 1.

[8] C. S. Kruse, D. A. Argueta, L. Lopez, and A. Nair, Patient and Provider Attitudes Toward the Use of Patient Portals for the Management of Chronic Disease: A Systematic Review, *J Med Internet Res* **17** (2015), e40.

[9] M. Koivunen and K. Saranto, Nursing professionals' experiences of the facilitators and barriers to the use of telehealth applications: a systematic review of qualitative studies, *Scand J Caring Sci* **32** (2018), 24–44.

[10] U. Sharma and M. Clarke, Nurses' and community workers' experience of telehealth: A longitudinal case study, *Health Services Research,* **14** (2014), 164–170.

[11] I.M. Zadvinskis, J.G. Smith, and Y, Po-Yin, Nurses' experience with health information technology: Longitudinal qualitative study, *JMIR Med Inform,* **6** (2018), e38

[12] E. Basch, A.M. Deal, M. G. Kris, H. I. Scher, C. A. Hudis, et al., Symptom monitoring with patient-reported outcomes during routine cancer treatment: A Randomized controlled trial, *Journal of Clinical Oncology*, **34** (2015), 557–565.

[13] Ministry of Health and Welfare, Information to support well-being and service renewal, eHealth and eSocial Strategy **2020** (2015).

[14] D. Sittig and H. Singh, A new sociotechnical model for studying health information technology in complex adaptive healthcare systems, *Qual Safety Healthc* **19** (2010), 68–74.

[15] P-Y. Yen, A. S. McAlearney, C. J. Sieck, J.L. Hefner, and T. R. Huerta, Health information technology (HIT) adaptation: Refocusing on the journey to successful HIT implementation, *JMIR Med Inform* **19** (2017), e28.

[16] International Project Management Association, *ICB – IPMA Competence Baseline*, 3rd edition, Nijkerk, International Project Management Association, (2006).

[17] J. S. Ancker, L. M. Kern, E. Abramson, and R. Kaushal, The Triangle Model for evaluating the effect of health information technology on healthcare quality and safety, *JAMIA,* **19** (2012), 61–65.

[18] E. Howsawi, D. Eager, R. Bagia, and K. Niebecker, The four-level project success framework: application and assessment, *Organisational Project Management*, **1** (2014), 1–15.

[19] A. de Wit, Measurement of project success, *International Journal of Project Management,* **6** (1988), 164–170.

[20] M. Radujković and M. Sjekavica, Project Management Success Factors, *Procedia Engineering,* **196** (2017), 607 – 615.

[21] M. Roberts, Successful public health information system database integration projects: A qualitative study, *Online Journal of Public Health Informatics*, **10** (2018), e207.

[22] CAF Manual (2013). https://www.eipa.eu/wp-content/uploads/2017/05/CAF_Manual_2013.pdf (accessed on March 23, 2019).

[23] Moeller J. The EFQM excellence model. German experiences with the EFQM approach in health care, *Int J Qual Health Care,* **13** (2001), 45–49.

[24] Austrian Federal Chancellery, CAF works – better service for the citizens by using CAF, (2006). file:///C:/Users/saranto/Downloads/CAFworks_EN%20(1).pdf (accessed on March 23, 2019).

[25] R. Kukafka, S. Johnson, A. Linfante, J. Allegrante, Grounding a new information technology implementation framework in behavioral science: A systematic analysis of the literature on IT use, *J Biomed Inform,* **36** (2003), 218–227.

Address for correspondence

Kaija Saranto
email: kaija.saranto@uef.fi

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190331

Design and Evaluation of a Patient Monitoring Dashboard for Emergency Departments

Thomas Schmidt[a], Mikkel Brabrand[b], Annmarie Touborg Lassen[c], Uffe Kock Wiil[a]

[a]Health Informatics and Technology, University of Southern Denmark, Odense, Denmark,
[b] Department of Emergency Medicine, Hospital of South Western Jutland, Esbjerg, Denmark
[c] Department of Emergency Medicine, Odense University Hospital, Odense, Denmark

Abstract

Identification of risk patients in emergency departments is a complex interplay between clinicians, patients, working procedures, and information systems. Through a mixed methods approach, we have developed an novel patient monitoring dashboard that couples clinical data streams to convey the state and trajectory of patients admitted to the emergency d epartment. In this paper, we describe the design and implementation of the novel system by providing a description of how the project was conceived, divided into subsequent phases, and is currently being executed. Experience from this work highlights the importance of carefully assessing each on-site installation point, interdisciplinary partnerships through embedded presence, and continuous emphasis on the purpose and aim as a mean to greatly improve project momentum and buy-in.

Keywords:
Emergency service, hospital; Decision support systems, clinical; Monitoring, physiologic

Introduction

Identifying patients at risk of unforeseen deterioration in emergency departments (ED) remains an ongoing challenge despite numerous initiatives to formalize and systematize observations, tracking the state and trajectory of patients, and responding to changes to avoid complications.

In this paper, we provide a description of how this challenging clinical problem sparked a collaboration between engineers and clinicians resulting in the design, implementation, and evaluation of a novel patient monitoring dashboard. As we describe the prototype, process, and inter-disciplinary collaboration, we aim to provide a set of recommendations for similar attempts in assessing the impact of new health information systems. In doing so, we seek to extrapolate key findings that are decoupled from the initiating problem of detecting deterioration, and which instead point to more general challenges in systems design and implementation within secondary healthcare.

The inception of this project, originates from the work of Henriksen et al. who found that as many as one in three patients whose vital signs are within normality at admission, have an abnormal vital sign registration within the first 24 hours [1]. These patients have a higher risk of being transferred to the intensive care unit, experience heart or respiratory failure during admission, and a higher 7-day mortality. Identifying patients at risk in EDs spurred a research project collaboration between software engineers and clinical researchers to leverage their joint knowledge for proposing a way to alleviate the risk of unforeseen deterioration.

Related Work

A multitude of protocols and observation regimens have been introduced, evaluated, and often implemented in the last 20 years. Still, it is hard to come by evidence that undisputedly points to significant changes in patient outcomes due to the implementation of an Early Warning Score (or similar). The hospitals involved in this study relied on the DEPT triaging system for assessing the severity of patients [2]. DEPT deploys a mixture of objective (e.g., blood pressure) and subjective observations (e.g. pain), see Figure 1. The most severe observation determines the overall severity and priority of the patient.

Figure 1 - The DEPT Triage System

Part of the challenge relating to the utility of these systems, is not only due to the heterogeneity of patients, but also the variety in experience, background, and profession of attending clinicians [3]. Another major factor is the socio-technical complexities that come into play when mixing this variability with information technology.

Several sophisticated novelty detection systems have been proposed over the years. Clifton et al. published several studies on the application of Gaussian process regression for dealing with missing values and forecasting of trajectories [4], and also investigated extreme value theory as an mean to perform online learning of patient-specific models [5]. Clifton et al. are also one of the few groups that succeeded in advancing a system beyond the pilot evaluation stage, and actually tested their system in realistic settings [6]. Another substantial line of work originates from Edelson & Churpek who developed the CART score to predict cardiac arrest [7]. Perpendicular to the issue of modelling and prediction, we found a need to assess how to best visualize the system state.

Although there has been slow progress on systems designed for patients to monitor themselves, clinical dashboards and data aggregation have been subject to much work [8]. Generally, there seems to be less focus on taking a pragmatic stance to improve utilization and integration of existing streams of data, and integration of these into information systems in ways that improve the clinical perception of multiple patients simultaneously.

Methods

As the project has been ongoing since early 2013, we have split the methodology description into several stages. First, an initial stage from 2013-2015 focused on early scoping that took place at the ED of Odense University Hospital, and resulted in a prototype of a patient monitoring dashboard. The second stage from 2016-2017, was the system prototype maturation and integratation into additional information systems. The third stage ran from 2018-2019, and focused on a multidimensional effect evaluation.

From the very beginning of the project, it quickly became apparent that clinical epidemiologists and engineering researchers had substantially different approaches to designing and conducting research. As software engineers, we did not have any prior experience with the clinical, technical, and operational aspects of EDs. Consequently, we needed to design our study approach in a manner that embedded us within the context.

To achieve this, we designed the project using a mixed methods approach where we could explore several pathways to illuminate potential solutions to the problem of identifying patients at risk. This lead us down three distinct paths as illustrated in Figure 2, each path highlights a number of related activities.

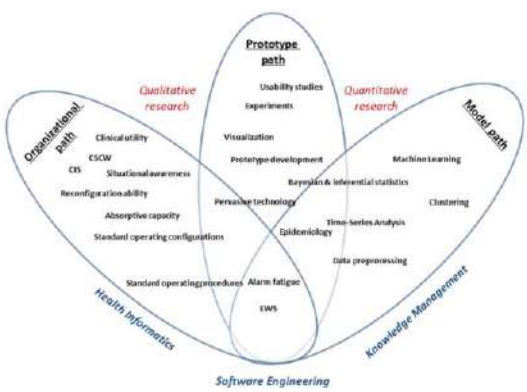


Figure 2 Methodological Project Paths

Firstly, as outsiders familiarizing ourselves with the context of our problem proved to be indispensable. To begin we mapped out an 'organizational path', which sought to clarify how clinicians dealt with the matter of patient processing, monitoring, and treatment. To do this, we deployed a number of tools, some were methodological by nature, and others were frameworks or research fields that helped us to understand the domain. The understanding of how clinicians worked with patients and the systems they utilizeed was then paired with data from actual patients. This data was the cornerstone of the 'modelling path', where we sought to understand patients and patient monitoring systems using a number of quantitative techniques, with the intention of deriving novel features and new ways of representing and understanding patients and monitoring systems. The 'organizational' and 'modelling' paths came together in the 'prototype' path where we attempted to validate assumptions from both of these paths. This was achieved via different prototype modalities, supported by usability studies.

Initial Stage – Early Scoping

Our first major effort was conducting a field study of the context. The first author of this paper followed physicians, nurses and nurse assistants on full work shifts at all times of the day and took note of how patients were attended to, how clinicians registered observations personally and collaboratively, and how information technology was used during the admission [9].

In parallel with this, we needed to establish a foundation for building better deterioration detection models with predictive analytic models. During the field study, we had observed several occasions where patients expressed individual normality that was not captured by the existing population-based threshold models. However, acquisition of vital signs from patients was not something the existing information communication (IT) configurations supported. Vital sign observations were registered in the department's electronic healthcare record according to protocol, but the registration frequency were low, and automated data extraction was not an option.

Coincidentally, a network diagram of the patient monitoring platform emerged while discussing other integration options with the local medical device engineers. This revealed that the ED had purchased a Philips IntelliVue™ solution that included a database server that automatically buffered all registrations from connected monitors. Thus enabling the HL7 Parameter Data Interface (PDI) allowed us to export vital sign readings at a configurable interval from the Philips database.

The Data Collection

Each bed at the department was equipped with a Philips IntelliVue MP30/50™ monitor. Each monitor had a Philips X2 Measurement ™ module, which could follow the patient during transfers. Each monitor had capacity to monitor heart and respiration rate through 3-lead electrocardiography, peripheral oxygen saturation (spO2) and pulse rate using pulse oximetry, and systolic/diastolic blood pressure (SBP/DBP) using non-invasive cuffs. To retrieve data we built a registration server that acquired registrations from all active monitors every minute using the HL7 PDI interface. The attending nurses were asked to register patients on the Philips monitors by name and social security number to enable later coupling with external health registries.

Prototype Design

The final phases of the initial stage focused on distilling the findings from all activities into a prototype system of a patient monitoring dashboard. Through incremental collaboration with clinicians, a functional prototype was implemented and evaluated in a pilot study involving 18 nurses and 50 patients [10]. The participating nurses were also asked to assess the system using the System Usability Scale [11], and results indicated that the overall design philosophy of the system made it easy to understand and use.

Second Stage - Refinement

After having successfully evaluated the prototype, we went on to acquire funding for further development of the system. This was achieved by funding from the Strategic Initiatives Program by the University of Southern Denmark through a 2M DKK grant. This gave us the opportunity to rewrite the entire prototype and explore new ways to deal with the

challenges of coupling registrations from the patient monitors with the department's clinical logistics system. The new system was dubbed the Patient Deterioration Warning System (PDWS). During this stage, the project grew in organizational complexity as an additional site was included, and the newly enforced General Data Protection Regulation had incentivized the IT departments across the region to impose stricter control with installation and evaluation of information technology for research purposes.

Third Stage – Effect Evaluation

Having rewritten the entire PDWS system, and tested deployment and operational stability of the system, the third stage of the project focused on conducting an effect evaluation. The intention of this larger trial was threefold:

1. To assess if deploying the PDWS into EDs would help clinicians identify patients who deteriorated unexpectedly.
2. To investigate if the design philosophy of the PDWS fitted well within clinical work.
3. To evaluate if utilizing the PDWS could help the departments in improving efficiency and quality.

Cluster Randomized Trial

The evaluation has been designed as a cluster randomized trial (CRT) consisting of three intervention and three control periods interleaved at each study site. Each period, lasting five weeks, has been separated by a one-week washout buffer. The study aims to include 10.500 patients, which will yield sufficiently material for evaluating the clinical hypothesis that the PDWS can reduce unexpected deterioration by 50%. The primary clinical outcome has been defined as patients who are transferred to the Intensive Care Unit, suffer from heart/respiratory failure, or die during admission. The evaluation stage will be assessed by journal review by clinical experts.

The CRT was initiated May 2018 and is scheduled to run until early 2019. During the period, project nurses and assistants will include patients at each site. Following the CRT period, data from all consenting patients will be transferred to a research repository and used for further research.

Study Sites

The first part of the study (field study, initial vital sign registration, and prototype pilot evaluation) was conducted at the Emergency Department of Odense University Hospital (OUH). OUH has an uptake population of 430,000 citizens and its ED has an capacity of 44 beds. The second part of the study expanded the participating sites to include also the ED of Hospital of South Western Jutland (HSWJ). HSWJ has an uptake population of 220.000 citizens and its ED operate 51 beds.

Although both EDs are situated in the Region of Southern Denmark, and are part of the same healthcare system, each ED organizes the patient admissions very differently. For example, at the ED of HSWJ, a patient is usually never transferred internally once admitted. Whereas in OUH, patients are seen and treated in a short term Acute Treatment Center for the first eight hours of admission, and then later transferred to the Center for Accellerated Patient flows for the remaining part of their admission.

Approvals

The project has been presented to the Scientific Ethics Committee of Southern Denmark, but does not need approval according to Danish legislation. The Danish Data Protection agency has accepted that we store data (J. nr. 17 14630). Research data are managed by OPEN (http://Open.rsyd.dk). Data from patients who provide informed consent will be used for the effect evaluation stage, and for future work in predictive analysis.

Results

As the project is still ongoing, and parts of it has already been described in other work, for the paper, we
seek to draw out some of the lesser known aspects – while still combining the entirety of the project to convey the key lines of results.

Data Acquisition

During the first part of the project, we acquired all registered vital signs over a two-year period as part of the initial prototype implementation. Although only a subset of these vital signs could be distinctly linked to specific admissions, the dataset in its entirety gave us a unique insight into the utilization of patient monitoring in the ED at Odense University Hospital. Table 1 lists the summary of all registrations. The number of registrations in relation to sensor type was evident, as we observed a decreasing number of registrations due to sensor nuisance.

Table 1- Vital Signs Registered During the Initial Stage

Vital Sign	#Registrations	Mean	Std.Dev
Heart rate	4,668,890	88 bpm	21 bpm
Respiration rate	4,491,545	20.4 rpm	5.7 rpm
Pulse rate	7,277,427	84.4 bpm	19.3 bpm
spO2	7,181,895	95%	3.8%
SBP	232,895	124 mmHg	26.4 mmHg
DBP	232,895	68.4 mmHg	17.9 mmHg

We evaluated the distributions of each vital sign type, and interestingly found distinct differences to what is considered normal vs. alarming thresholds according to the NEWS system [12]. Heart and pulse rate readings only exceeded the upper limit of 131 bpm in 3% of all cases, and just 0.2% of all readings were below 42 bpm. These numbers were much higher for respiration rate registrations where 16% surpass 25 rpm, and just 0.4% are lower than 8 rpm. 10% of spO2 registrations were lower than 91% oxygen saturation, and this was even without factoring in patients being administered additional oxygen. For systolic blood pressure, 7.5% of readings were lower than 90 mmHg which was quite substantial given the risk of low blood pressure [13].

Deriving Novel Metrics for Monitoring

Another line of work attempted to identify different models of normality for a range of clinical and patient specific factors [14]. Although we did not find support for the presence of such normality ranges, we instead identified clusters of patients who generally had vital signs within non-alarming thresholds, but with a distinctly higher mortality ratio. The main difference to similar clusters was a higher standard deviation. Thus, we identified a group of patients who seldom triggered any alarms, but still expressed high variability. This lead to the inclusion of the metric '*Relative severity*', which was implemented as an aggregated sum of shifts between states. Clinician could thus order patients by highest severity as prescribed by the attending physician, or by variability.

The degree of monitoring a patient was exposed to during admission was formalized in both departments by observation regimens. For example, a patient who was classified as

Figure 3 The Patient Deterioration Warning System - Client System

Orange was to be continuously monitored, and have vital signs registered in their electronic health record hourly. However, from our exploratory analysis of the vital sign dataset, we found that several factors, both patient and department specific, influenced the extent to which patients were monitored [15]. The findings of this work were utilized in the PDWS, as abnormal device utilization could be used as a marker for clinical concern.

The Patient Deterioration Warning System

The PDWS consists of a backend system implemented in Java™ using Spring v.4 with Hibernate for persistency and Envers for entity auditing, and a front-end implemented in AngularJs v.5. The PDWS interface is shown in Figure 3, and has been described in other work [10]. The system conveyed the state and trajectory of all admitted patients who at some point in time had been monitored using any of the department's vital sign monitors. All registered readings were scored using the DEPT system and aggregated into configurable periods of time with respect to the expected length of stay for a given ward. As the system automatically stored all vital signs from monitored patients every minute during their entire stay, and linked these with admission information, we were able to gradually establish a very detailed dataset of all patients admitted to the participating EDs.

The project as a whole made a giant leap forward when we figured out how to couple data from the monitors with data from the logistics system. The latter was essential for the daily operation of the department, and any changes in the admission flow of patients was instantaneously updated in the logistics system.

As portrayed by Figure 3, the PDWS was made available to clinicians using 24" all-in-one computers running Chrome OS. Nine of these were installed in offices at the ED of OUH, and six at the ED of HSWJ. Each computer was installed in the vicinity of the existing patient monitoring overview screen, and the clinical logistics system.

Discussion

As clinicians become increasingly reliant on computers to filter and present data, a sensible balance of work shared between information systems and clinicians must be sought. In this project, we sought to do so by simplifying a few of the streams of data that demanded the attention of clinicians, thereby aiming to mitigate the risk of information processing overload. The CRT is currently ongoing, and some preliminary observations are worth discussing.

Workflow Integration and Information Habits

Several aspects of the PDWS design philosophy, especially non-intrusiveness, and the assumption that clinicians will automatically utilize a novel system tailored to specifically address a problem most are familiar with in the settings, appeared to be challenged in the CRT. Self-evident as it may be, it became strikingly clear that installing yet another computer screen in the cramped work spaces of most clinicians did not offer much of a head start for evaluation of a new health information system. In most of the ED offices, the PDWS screen mentally disappeared in the clinicians information landscape. Futhermore, evaluation of new systems that aimed to enhance or replace existing systems, which were already highly embedded in current practice, was complicated. Old habits die hard, as was also evident by information acquisition. Essentially, the aim should be to make information acquisition easier with new system, rather than the old one. Deciding to obstruct access to systems already in use is seldom feasible, rather the alternative is to make the new noticeably easier to use. Several participants proposed embedding the PDWS components into existing IT solutions. This would however make it considerably more complex to conduct a targeted effect evaluation review.

Integration with Existing Information Systems

The real-time utilization of information from other hospital information systems, raised some concern amongst the organization. Legitimate concerns regarding unknown side effects were raised based on the new approach in utilizing existing systems. The solution was to conduct a PDWS code and performance audit on the supplier side to ensure that the PDWS did not strain existing resources or add risk.

Continuous Emphasis on Purpose and Aim

Two other crucial aspects we strived to address, which also relate to the issues above, were training in system use and awareness of intent and purpose. Prior to the CRT, the project and system was introduced to clinicians through newsletters, employee meeting presentations, and posters adjacent to each

screen. During the intervention periods we conducted several walkthroughs of the system. Still, we found that a major percentage of clinicians refrained from relating to the system. Project staff (nurses and nurse assistants), despite extra traning and expressed enthusiasm in the objective of the PDWS, were generally hesitant to act as system ambassadors in relation to coworkers. This points to the need for continous emphasis on the purpose of systems research in the settings, and awareness of the implications of employee turnover.

Future Directions

In several of the Danish Regions, work has been initiated to integrate data from medical devices faster and more seamlessly into decision making. Medical device data are to be enriched with data from other information sources (e.g., atttendending clinician, ordering speciality, patient info) to improve clinical, managerial, and operational services. The application of device utilization in the PDWS project, may serve as inspiration for new abnormality metrics. For example, raising flags when an abdominal ultrasound is ordered for a patient when it is not a part of the expected treatment plan.

Conclusion

Organizational, cultural, and technical aspects influence the detection of patients at risk of deterioration. Current patient monitoring solutions have shortcomings regarding incorporating the temporal and spatial organization of clinical work. A substantial ratio of automatically registered vital signs should be triggering more clinical concern than what is being observed. Coupling this with the fact that monitoring of patients in an ED is highly skewed, and that temporal trends are hard to derive from existing monitoring solutions, points to the need for better visualization tools.

Based on experiences from this work, we recommend that similar initiatives seek to: 1) *Carefully assess the settings of each installation point* to increase integration with existing workflows and information habits. 2) *Strive to establish interdisciplinary partnerships through embedded presence* to greatly improve the project momentum. Similarly, performance, load analysis, and openness regarding code reviews of integration points with external hospital information systems should be leveraged with stakeholder concerns. 3) *Continuously emphasize purpose and aim of a project and system,* as the typically high employee turnover in hospital departments, challenges long-term projects driven by user involvement. Consequently, a continuous effort is required to imprint the purpose of both research aim and the purpose of the evaluated information system.

Acknowledgements

This project was originally funded as a part of the Patient@home project, and subsquently by the Strategic Initiatives Program by University of Southern Denmark and Odense University Hospital. Additionally, this project would not have been possible without the support of the leadership and enthusiasm of both participating emergency departments.

References

[1] D.P. Henriksen, M. Brabrand, and A.T. Lassen, Prognosis and risk factors for deterioration in patients admitted to a medical emergency department, *PLoS One* **9** (2014), e94649.

[2] C. Skriver, M. Lauritzen, and J. Dahlin, Triage-Manual, (2009).

[3] C. Hands, E. Reid, P. Meredith, G.B. Smith, D.R. Prytherch, P.E. Schmidt, and P.I. Featherstone, Patterns in the recording of vital signs and early warning scores: compliance with a clinical escalation protocol, *BMJ Qua. Saf* **22** (2013), 719-726.

[4] L. Clifton, D.A. Clifton, M.A.F. Pimentel, P.J. Watkinson, and L. Tarassenko, Gaussian process regression in vital-sign early warning systems, *2012 Annu Int Conf IEEE Eng Med Biol Soc* (2012), 6161–6164.

[5] S. Hugueny, D.A. Clifton, and L. Tarassenko, Probabilistic Patient Monitoring with Multivariate, Multimodal Extreme Value Theory, in: H. Fred, Ana; Filipe, Joaquim; Gamboa (Ed.), Biomed Eng Syst Technol,, , Springer, 2011: pp. 199–211.

[6] D. a. Clifton, D. Wong, L. Clifton, S. Wilson, R. Way, R. Pullinger, and L. Tarassenko, A large-scale clinical validation of an integrated monitoring system in the emergency eDepartment, *IEEE J Biomed Heal Informatics* **17** (2013), 835–842.

[7] M.M. Churpek, T.C. Yuen, S.Y. Park, D.O. Meltzer, J.B. Hall, and D.P. Edelson, Derivation of a cardiac arrest prediction model using ward vital signs, *Crit Care Med* **40** (2012), 2102–2108.

[8] R. Kamaleswaran, and C. McGregor, A review of visual representations of physiologic data *JMIR Med Informatics* **4** (2016), e31.

[9] T. Schmidt, and U.K. Wiil, Identifying patients at risk of deterioration in the joint emergency department, *Cogn Technol Work* **17** (2015), 529–545.

[10] T. Schmidt, A. Lassen, and U.K. Wiil, A patient deterioration warning system for boosting situational awareness of monitored patients, in: 2016 IEEE Int Conf Healthc Informatics, (2016),129–138.

[11] J. Brooke, SUS-A quick and dirty usability scale, *Usability Eva. Ind* **189** (1996), 4--7.

[12] Royal College of Physicians, National early warning score (NEWS), Royal College of Physicians, London, 2012.

[13] A.K.B. Kristensen, J.G. Holler, J. Hallas, A. Lassen, and N.I. Shapiro, Is shock index a valid predictor of mortality in emergency department patients with hypertension, diabetes, high age, or receipt of β- or calcium channel blockers? *Ann Emerg Med* **67** (2016), 106-113.

[14] T. Schmidt, J. Hallam, A. Lassen, and U.K. Wiil, Clustering emergency eepartment patients – An assessment of froup normality, in: 37th Int. Conf. IEEE EMBS (2015), 6824–6829.

[15] T. Schmidt, C.N. Bech, M. Brabrand, U.K. Wiil, and A. Lassen, Factors related to monitoring during admission of acute patients, *J Clin Monit Comput* **31** (2016), 641-649.

Address for correspondence

Thomas Schmidt, tlf. +4524347434, email schmidt@mmmi.sdu.dk. Address: Syddansk Universitet, Maersk Mc-Kinney Moeller Instituttet, Campusvej 55, 5230, Odense M, Denmark

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190332

Development and Assessment of RecosDoc-MTeV to Improve the Quality of Direct Oral Anticoagulant Prescription for Venous Thromboembolic Disease

Brigitte Séroussi[a,b], Houda Ouarrirh[b], Ismaël Elalamy[c], Grigorios Gerotziafas[c], Isabelle Debrix[d], Jacques Bouaud[e,a]

[a] Sorbonne Université, Université Paris 13, Sorbonne Paris Cité, INSERM, UMR S_1142, LIMICS, Paris, France,
[b] AP-HP, Hôpital Tenon, Département de santé publique, Paris, France
[c] AP-HP, Hôpital Tenon, Service d'hématologie biologique, Paris, France
[d] AP-HP, Hôpital Tenon, Pharmacie, Paris, France
[e] AP-HP, DRCI, Paris, France

Abstract

Potentially inappropriate prescribing of direct oral anticoagulants is frequent with the most common errors being dosage, administration, and duration of therapy. We developed RecosDoc-MTeV, a documentary-based clinical decision support system (CDSS) for the management of direct oral anticoagulant prescription to prevent and treat venous thromboembolism. Simultaneously, the network of Parisian public hospitals (AP-HP, France) developed narrative clinical practice guidelines (CPGs) and a companion smartphone application to enhance medication and patient safety related to direct oral anticoagulant prescription. To assess the effectiveness of these CDS tools, we performed a retrospective review of 274 random patients hospitalized in 2017, which were either at risk of venous thromboembolism or actually treated for the disease. Consistency between the two CDS tools was measured at 96.7%. Administered treatments were compliant in 67.2% and 72.3% of the cases, with AP-HP CPGs and RecosDoc-MTeV, respectively. These results support that implementing CDSSs for the prescription of direct oral anticoagulants may ensure safe prescribing of high-risk medications.

Keywords:

Venous Thromboembolism, Clinical Decision Support System, Guideline Adherence, Clinical Practice Guidelines, Health Care Quality Assessment

Introduction

Anticoagulants are the mainstay of therapy for the acute and long-term prevention and treatment of numerous types of thromboembolic disorders. The prevention of thromboembolic stroke among patients with chronic atrial fibrilation is one of the primary indications for anticoagulation therapy. In addition, anticoagulants are indicated in, and increasingly prescribed for, the prevention and treatment of venous thromboembolic disease including deep vein thrombosis and pulmonary embolism. Until recently, anticoagulation therapy was dominated by parenteral anticoagulants and vitamin K agonists, e.g. warfarin. Since 2009, a new therapeutic alternative has appeared with direct oral anticoagulants. They have revolutionized the management of patients undergoing anticoagulant therapy due to their rapid onset of action, fixed dosage, and non-necessity of biological monitoring for their therapeutic effectiveness [1], becoming the first-line choice for treatment of venous thromboembolic disease and atrial fibrillation.

However, antithrombotic drugs and especially oral anticoagulants belong to the class of drugs causing the primary cause of serious adverse reactions and the primary cause of hospitalizations for adverse reactions [2]. It is estimated that more than 900,000 incidents of recurrent, fatal and non-fatal venous thromboembolic events occur in the United States annually [3]. In France, the annual incidence of venous thromboembolic disease is in the order of 50,000 to 100,000 cases responsible for 5,000 to 10,000 deaths. These situations could be avoided if clinical practice guidelines (CPGs) establishing the proper use of antithrombotic drugs were correctly implemented.

Numerous CPGs for the management of anticoagulation, like CPGs from the American College of Cardiology [4], have been published recently. However, the implementation of such narrative guidelines is complex: direct oral anticoagulants are contraindicated for patients with mechanical valve replacement, or with severe renal insufficiency. Clinical situations may sway patients and clinicians to favor one oral anticoagulant over another. Patients who prefer once-daily dosing will find both edoxaban and rivaroxaban to be more convenient than the twice-daily regimens for apixaban and dabigatran. Patients looking for single-drug treatment (especially outpatient treatment) will favor the use of apixaban or rivaroxaban, which do not require 5–10 days of pre-treatment with low molecular weight heparin as it is required for dabigatran and edoxaban. Besides, dosage, administration, and duration of therapy are quite different according to the situation (prevention vs. treatment), the patient, and the prescribed direct oral anticoagulants. As a consequence, the dissemination of narrative CPGs has had a limited effect in changing physician behavior and direct oral anticoagulant treatment remains underutilized in current clinical practice [5].

Clinical decision support systems (CDSSs) are receiving increased attention as tools to reduce costs and improve care. CDSSs embedding CPGs in their knowledge base have shown to be efficient tools to promote the adoption of CPGs by physicians [6], especially for the prescription of direct oral anticoagulants [7]. For instance, Karlsson et al. [8] demonstrated that a CDSS could increase guideline adherence with anticoagulant therapy in patients with atrial fibrillation.

This CDSS, integrated into the regular electronic health record (EHR), used medical record data to identify patients with a diagnosis of atrial fibrillation and at least one risk factor for stroke with no anticoagulant therapy and displayed a pop-up screen warning. In the same way, Borab et al. [9] evidenced that using CDSSs increased the proportion of surgical patients who were prescribed adequate prophylaxis for venous thromboembolic disease, correlated with a reduction in venous thromboembolic events. In most of the cases, when CDSSs are embedded into the EHR, decision support is implemented as the display of alerts triggered when discrepencies exist between physician prescription and guidelines. However, it has been reported that alert-based CDSSs may be counter-productive since healthcare professionals suffering from "alert fatigue" may ignore most of them [10].

We have developed a guideline-based CDSS called RecosDoc-MTeV to assist decision-making for anticoagulation therapy based on state-of-the-art knowledge combined with local CPGs as elaborated by Saint-Antoine and Tenon hospitals, two hospitals of the Parisian network of public hospitals, Assistance Publique – Hôpitaux de Paris (AP-HP, France). At the same time, because there were no national guidelines to support the managament of anticoagulation, the Assistance Publique – Hôpitaux de Paris published narrative CPGs to improve the quality of antithrombotic prescription within its 39 hospitals.

We conducted an analysis to assess the consistency of the propositions provided by RecosDoc-MTeV and AP-HP CPGs. Another objective was to evaluate the conformity to RecosDoc-MTeV and AP-HP CPGs of actual anticoagulation therapy prescriptions. The analysis has been performed on a sample of clinical cases randomly selected from patients hospitalized in 2017 in both Saint-Antoine and Tenon hospitals.

Materials and Methods

Description of RecosDoc-MTeV

RecosDoc-MTeV is a guideline-based CDSS applied to the management of anticoagulation therapy and the prescription of direct oral anticoagulants. The system relies on a knowledge base that models current state-of-the-art CPGs completed by the expertise of the hematologists of Tenon hospital (IE, GG) and the practice of the head of Tenon hospital pharmacy (ID).

The knowledge base of the CDSS is structured as a decision tree essentially made of two subtrees to represent the management of venous thromboembolic disease prevention and treatment (see figure 1). The prevention subtree explores various surgical situations, making the difference between orthopaedics surgery, cancer surgery, bariatric surgery, and other surgeries which have different risks of venous thromboembolic disease. In the same way, the different orthopaedics surgeries associated with different length of treatment are considered, e.g., hip replacement, knee replacement, hip fracture, and femoral neck fracture. For cancer surgery, a difference is made between the management of breast cancer, abdominopelvic cancers (either digestive, gynaecologic, or urologic cancers), and non-abdominopelvic cancers (either thoracic, otorhinolaryngologic, or brain cancers). In non-surgical situations, a pharmacological preventive treatment is recommended for hospitalized patients with high risk of venous thromboembolic disease. The treatment subtree is built in a similar way. It explores the special case of heparin-induced thrombocytopenia and some specific clinical conditions such as the co-occurrence of venous thromboembolic disease with an active cancer, a pregnancy, or any other clinical situations with acute organ failure.

At different points of the decision tree, it happens that evidence is missing to assess the risk of venous thromboembolic disease and guide the therapeutic strategy. In these cases, the risk is computed according to specific external scores provided by the literature and accessible on the Internet by external links (scores of Caprini [11], Padua [12], and Compass in the prevention subtree; scores of Wells and 4T in the treatment subtree).

In all cases, the recommended treatment varies with the value of creatinine clearance to assess the renal function, the value of the Body Mass Index (BMI) to assess overweight and obesity, and the existence of an antecedent of heparin-induced thrombocytopenia.

RecosDoc-MTeV has been developed according to the documentary paradigm of decision support [13] which allows for contextual interpretation of patient data and guidelines knowledge. Once the knowledge base is built, it can be used as an autonomous application and browsed by the physician user. At each depth level, a question is displayed in a closed-ended form to document a clinical criterion that may either concern patient information (renal failure, obesity, pregnancy, cancer), therapeutic history (prior heparin-based treatment) or one of the scores used to assess the risk of venous thromboembolic disease.

Starting from the root of the decision tree, the physician user navigates through the knowledge base while answering questions and thus instantiates the relevant patient criteria to establish the recommended treatment. Guideline-based patient-centered therapeutic recommendations are then provided when the navigation is completed, i.e. when a leaf of the decision tree is reached.

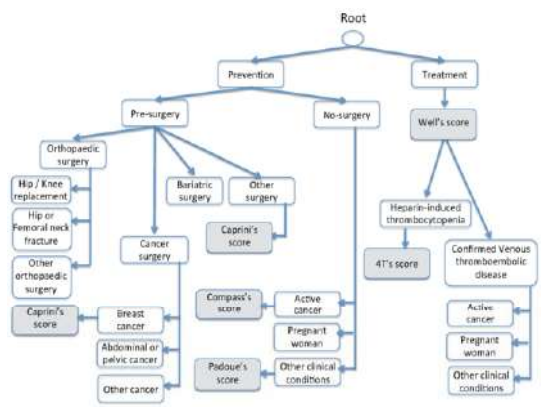


Figure 1– Excerpt of the top of the decision tree.

Description of AP-HP CPGs

AP-HP CPGs [14] have been developed in 2014 at the initiative of the Thrombosis working group of Assistance Publique – Hôpitaux de Paris to support the proper use of antithrombotics by residents and practitioners. AP-HP CPGs are described in a 55-page long document. The first part of the document is dedicated to the description of clinical conditions associated with

Figure 2 – Screenshots of RecosDoc-MTeV (left), and of the AP-HP app (middle) and narrative CPGs (right).

low, moderate, and high risk of thrombosis. Then, therapeutic protocols recommended for the different levels of risk are presented along with the recommended surveillance. In order to facilitate the availability of these recommendations and make healthcare professionals able to have at hand the essential information about anticoagulant drugs, AP-HP developed a smartphone application (app) named "Recos Thromboses" based on AP-HP CPGs.

Figure 2 displays the three types of interfaces: the structured patient-centered description of the patient profile as offered by RecosDoc-MTeV, the semi-structured organization of the information as provided by the AP-HP app, and tables as given by narrative AP-HP CPGs.

Selection of the population

In order to assess the consistency of the propositions provided by RecosDoc-MTeV and AP-HP CPGs, we randomly selected a sample of patients hospitalized for more than 48 hours in 2017, either in Saint-Antoine hospital or in Tenon hospital, stratified to cover the different clinical situations in the groups for prevention of, and treatment for, venous thromboembolic disease. We used ICD10 codes and CCAM codes (a French terminology to code medical acts) to identify relevant patients for the prevention group, e.g. patients undergoing orthopaedic surgery, cancer surgery, or bariatric surgery, and patients hospitalized for medical conditions at risk, such as acute renal failure, cardiac failure, respiratory failure, infection or sepsis, and pregnancy. We proceeded the same way for the treatment group to select patients diagnosed with venous thromboembolic disease according to ICD-10 codes (either deep vein thrombosis (I80, I81, or I82) or pulmonary embolism (I26)).

Quantitative analysis

We considered that the recommendations provided by AP-HP CPGs were the gold standard. We evaluated RecosDoc-MTeV by comparing the set of propositions issued by RecosDoc-MTeV, denoted $\{\text{RecosDoc-MTeV}\}_i$ to those generated by AP-HP CPGs, denoted $\{\text{CPG\_AP-HP}\}_i$ on each patient P_i of the randomized sample of clinical cases. We defined that AP-HP- and RecosDoc-MTeV-generated proposition sets were:

- *Identical* when $\{\text{RecosDoc-MTeV}\}_i = \{\text{CPG\_AP-HP}\}_i$
- *Consistent* when $\{\text{RecosDoc-MTeV}\}_i \neq \{\text{CPG\_AP-HP}\}_i$ and $\{\text{RecosDoc-MTeV}\}_i \cap \{\text{CPG\_AP-HP}\}_i \neq \varnothing$
- *Different* when $\{\text{RecosDoc-MTeV}\}_i \cap \{\text{CPG\_AP-HP}\}_i = \varnothing$

We assessed the quality of RecosDoc-MTeV by computing the frequency of identical and consistent propositions of both systems on the population sample. We also evaluated the clinical practices at Saint-Antoine and Tenon hospitals by computing the frequency the treatment actually received by a patient Pi was compliant with RecosDoc-MTeV (when included into $\{\text{RecosDoc-MTeV}\}_i$) and compliant with AP-HP CPGs (when included into $\{\text{CPG\_AP-HP}\}_i$). Computations have been made in the prevention and treatment groups.

Although hospitalized patients gave their consent for the re-use of their clinical data for research purposes at the time of admission, we asked for the authorization of the heads of the departments concerned with the patient cases we used in this study before proceeding with the collection and use of data.

Results

Study population

The target population for ICD-10 queries on Saint-Antoine and Tenon hospitals for the year 2017 was made of 6,881 patients. 355 were randomly selected, from which 81 were excluded (42 clinical cases were out of the scope, 10 medical records were unavailable, and 29 clinical cases had too many missing data) leading to a study population of 274 cases, 182 and 92 dispatched in the prevention and treatment groups, respectively. Patient characteristics are reported for the two groups in Table 1.

Comparison of RecosDoc-MTeV and AP-HP CPGs

For each clinical situation corresponding to each patient, the propositions of RecosDoc-MTeV were compared to the recommendations of AP-HP CPGs. The numbers for each configuration, either identical, consistent, or different are reported in Table 2.

Table 1– Main characteristics of the study population

		Prevention (n=182)		Treatment (n=92)	
		n	%	n	%
Sex					
	Female	118	64.8	46	50.0
	Male	64	35.2	46	50.0
Age group					
	< 40	30	16.5	7	7.5
	[40–75]	92	50.6	44	47.8
	> 75	60	33.0	41	44.6
Prevention group					
	Cancer surgery	59	32.4	—	—
	Bariatric surgery	7	3.8	—	—
	Orthopedic surgery	96	52.7	—	—
	Other surgeries	12	6.6	—	—
	Non-surgical management	8	4.4	—	—
Treatment group					
	Current cancer	—	—	35	38.0
	Other	—	—	57	62.0

Table 2– Consistency of RecosDoc-MTeV with respect to AP-HP CPGs

	Prevention	Treatment	Total
Identical	120 (65.9%)	50 (54.3%)	170 (62.0%)
Consistent	55 (30.2%)	40 (43.5%)	95 (34.7%)
Different	7 (3.8%)	2 (2.2%)	9 (3.3%)

On the whole, CDSS propositions differed from AP-HP recommendations in 3.3% of the cases (CI 95%: [1.5%–6.1%]). The nine clinical situations where the propositions of RecoDoc-MTeV were different of those provided by AP-HP CPGs are displayed in Table 3.

Table 3– Inconsistencies between RecosDoc-MTeV and AP-HP CPGs

Clinical situations	N	RecosDoc-MTeV	AP-HP CPGs
Hallux Valgus in obese patients	2	Pharmacological prevention	No prophylaxis
Hospitalized patients with a low risk of **venous thromboembolic disease**	4	Non-pharmacological prevention	No prophylaxis
V**enous thromboembolic disease** with cancer	2	Unfractionated Heparin	Unfractionated Heparin relayed by Vitamin K Agonists

Compliance of clinical practices with state of the art recommendations

For each patient, the administered treatment, both in the prevention and treatment groups, was compared to each of the propositions provided by RecosDoc-MTeV and AP-HP CPGs. Table 4 reports for each group the compliance level with each decision support resource.

The global compliance of administered treatments with AP-HP CPGs was measured at 67.2% (CI 95% : [61.2%-72.7%]). This compliance was significantly higher in the prevention group than in the treatment group (72.0% vs 57.6%, p = 0.024).

When administered treatments were compared to RecosDoc-MTeV propositions, the compliance level was 72.3% (CI 95%: [66.6%-77.5%]). No significant difference was observed between the prevention and treatment groups (73.6% vs 69.6%).

Table 4– Compliance of administered treatments with RecosDoc-MTeV and AP-HP CPGs

	N	Compliance / AP-HP CPGs		Compliance / RecosDoc-MTeV	
		n	%	n	%
Prevention	182	131	72.0	134	73.6
Treatment	92	53	57.6	64	69.6
Total	274	184	67.2	198	72.3

When compared, the two compliance measures are not significantly different. The contingency tables of the two compliance variables were built for each subgroup, prevention and treatment, to measure the agreement between the two decision support resources, RecosDoc-MTeV and AP-HP CPGs (Table 5).

Table 5– Contingency tables for the compliance variables with respect to RecosDoc-MTeV (RM) and to AP-HP CPGs (AP) in both prevention and treatment groups

	Prevention			Treatment		
	AP+	AP-	Total	AP+	AP-	Total
RM+	129	5	134	52	12	64
RM-	2	46	48	1	27	28
Total	131	51	182	53	39	92

For the prevention and treatment groups, the unweighted Kappa measures were 0.903 and 0.700, respectively, considered as “almost perfect agreement” and “substantial agreement”. However, in 26% of the study population, the administered treatment was neither compliant with RecosDoc-MTeV nor compliant with AP-HP CPGs. These non-compliant treatments represented 33% of the decisions in the treatment group with the use of drugs not mentioned in the resources. In the prevention group, the combined non-compliance reached 25% (46 cases), distributed in 15% of decisions where the therapeutic scheme was different than the one recommended, 7% of non-pharmacological prophylaxis, and 3% of decisions based on drugs not mentioned in the resources.

Discussion and conclusion

RecosDoc-MTeV is a guideline-based patient-centered decision support system applied to the prescription of direct oral anticoagulants for the management of venous thromboembolic disease. The system is developed according to the documentary paradigm of decision support [13] which allows the user to interactively navigate through the knowledge base to describe a given patient case and get the patient-specific recommendations. Compared to AP-HP CPGs considered as the gold standard in this study, we found on a randomized sample of 274 patient cases that 62.0% of the propositions were identical, 34.7% were similar and 3.3% were different. It should be noticed that most discrepancies were observed in the prevention group. The nine situations

where inconsistencies were observed concerned the management of the venous thromboembolic disease risk with propositions of pharmacological and non-pharmacological prevention by RecosDoc-MTeV in medium and low risk clinical situations respectively, whereas AP-HP CPGs recommended no prophylaxis at all. When comparing the treatments actually received by the patients to the propositions provided by both resources, we observed that clinical practices were compliant with AP-HP CPGs in 67.2% of the cases, and compliant with RecosDoc-MTeV propositions in 72.3% of the cases (no significant difference). It is interesting to notice that 46 therapeutic decisions in the prevention group and 27 therapeutic decisions of the treatment group were neither compliant with AP-HP CPGs nor compliant with RecosDoc-MTeV propositions. For these 73 decisions, RecosDoc-MTeV propositions and AP-HP CPGs were identical in 100% of the cases. This tends to evidence that in these specific situations, clinical practices are suboptimal and would be improved by the use of a CDSS such as RecosDoc-MTeV.

This study has various limitations. We have chosen AP-HP CPGs as the gold standard which could be questioned since AP-HP CPGs, published in 2014, are older than RecoDoc-MTeV, elaborated in 2016. In addition, the randomized sample incompletely covered the set of possible clinical cases since, for instance, we didn't find any medical record with a history of heparin-induced thrombocytopenia. Indeed there is no ICD-10 code for this disease, and all the ICD-10 codes that we used to target the disease didn't retrieve appropriate clinical cases. The study was also conducted on a small sample of clinical cases and on only two hospitals. Finally, compliance rates are probably overstated since we only took into account the class of anticoagulant drugs for the comparison of medical decisions with RecosDoc-MTeV and AP-HP CPG propositions, while dosage and duration of prescriptions were not considered (data were often missing).

We performed a retrospective review of 274 patients hospitalized in 2017 at Tenon and Saint-Antoine hospitals (Paris, France) who were candidates for direct oral anticoagulant treatment either to prevent or treat venous thromboembolic disease. We assessed the prescription of direct oral anticoagulants according to both RecosDoc-MTeV and AP-HP CPGs. Compliance with the CDSS was 72.3% when only taking into accont the drugs used. These results support that implementing CDSS for venous thromboembolic disease prevention and treatment in routine practice should improve the quality of care ensuring safe prescribing of high-risk medications. Further work needs to be carried on to assess in a prospective study whether the use of RecosDoc-MTeV upon order entry would increase the compliance rate of therapeutic decisions and thus improve the quality of care.

References

[1] A.T. Cohen, A.K. Gitt, R. Bauersachs, E.M. Fronk, P. Laeis, P. Mismetti, et al. The management of acute venous thromboembolism in clinical practice. *Thromb Haemost* **117**/7 (2017), 1326–1337.

[2] Agence nationale de sécurité du médicament et des produits de santé (ANSM). *Les anticoagulants en France en 2014: état des lieux, synthèse et surveillance*, ANSM, Paris, France, 2014.

[3] J.A. Heit. The epidemiology of venous thromboembolism in the community. *Arterioscler Thromb Vasc Biol* **28**/3 (2008), 370–372..

[4] American College of Cardiology (ACC). Anticoagulant Management. [www.acc.org/clinical-topics/anticoagulation-management. Last accessed March 26th, 2019]

[5] S. Chaturvedi S, A.G. Kelly, S. Prabhakaran, G. Saposnik, L. Lee, A. Malik, C. Boerman, G. Serlin, A.M. Mantero. Electronic Decision support for Improvement of Contemporary Therapy for Stroke Prevention. J Stroke Cerebrovasc Dis. 2019 Mar;28(3):569-573.

[6] S. Flottorp, A. Oxman, J. Krause, N. Musila, M. Wensing, M. Godycki-Cwirko, et al., A checklist for identifying determinants of practice: A systematic review and synthesis of frameworks and taxonomies of factors that prevent or enable improvements in healthcare professional practice. *Implement Sci* **8** (2013), 35.

[7] T. Ahuja, V. Raco, J. Papadopoulos, D. Green. Antithrombotic Stewardship: Assessing Use of Computerized Clinical Decision Support Tools to Enhance Safe Prescribing of Direct Oral Anticoagulants in Hospitalized Patients. J Patient Saf. 2018 Sep 25.

[8] L.O. Karlsson, S. Nilsson, M. Bång, L. Nilsson, E. Charitakis, M. Janzon. A clinical decision support tool for improving adherence to guidelines on anticoagulant therapy in patients with atrial fibrillation at risk of stroke: A cluster-randomized trial in a Swedish primary care setting (the CDS-AF study). *PLoS Med* **15**/3 (2018), e1002528.

[9] Z.M. Borab, M.A. Lanni, M.G. Tecce, C.J. Pannucci, J.P. Fischer. Use of Computerized Clinical Decision Support Systems to Prevent Venous Thromboembolism in Surgical Patients: A Systematic Review and Meta-analysis. *JAMA Surg* **152**/7 (2017), 638–645.

[10] R. Backman, S. Bayliss, D. Moore, I. Litchfield. Clinical reminder alert fatigue in healthcare: a systematic literature review protocol using qualitative evidence. *Syst Rev* **6**/1 (2017), 255.

[11] https://venousdisease.com/dvt-risk-assessment-online/ [Last accessed on March 2019]

[12] https://qxmd.com/calculate/calculator_388/padua-prediction-score-for-risk-of-vte [Last accessed on March 2019]

[13] B. Séroussi, J. Bouaud, E.C. Antoine. OncoDoc, a successful experiment of computer- supported guideline development and implementation in the treatment of breast cancer. *Artif Intell Med* **22**/1 (2001), 43–64.

[14] Comité des Médicaments de la COMEDIMS AP-HP. Bon usage des antithrombotiques. Recommandations AP-HP, Oct. 2014.

Address for correspondence

Brigitte Séroussi
Département de Santé Publique
4, rue de la Chine, 75020 Paris, France
brigitte.seroussi@aphp.fr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190333

Availability and Quality of Information Used by Nurses While Admitting Patients to a Rural Home Health Care Agency

Paulina S. Sockolow[a], Ellen J. Bass[a,b], Yushi Yang[b], Natasha B. Le[c], Sheryl Potashnik[a], Kathryn H Bowles[c,d]

[a]Drexel University, College of Nursing and Health Professions, Philadelphia PA USA; [b]Drexel University, College of Computing and Informatics; [c]University of Pennsylvania School of Nursing, Philadelphia, PA USA; [d]Visiting Nurse Service of New York, New York, New York, USA

Abstract

Home health care admission nurses need high quality patient information but that information is not uniformly available. Despite this challenge, these nurses must make four critical decisions at patient admission to construct the plan of care: (1) patient problems to address in the home health care episode; (2) patient medication management; (3) services in addition to skilled nursing; and (4) skilled nursing visit pattern. We observed 12 in-home admissions at a rural home health care agency and interviewed nurses before and after about these decisions. We analyzed content and quality of documents. To evaluate quality, for each decision we assessed concordance between documents. Interview responses provided context in the analysis. Across all admissions, documents and their contents were not uniformly present. Nurses rarely received visit pattern or medication management information. There was discordance in the number of patient problems among and between available documents and the plan of care. Electronic health record design recommendations include interoperability and structured, consistent, actionable information.

Keywords:

Home Health Care, Decision Making, Nursing Informatics, Documentation, Continuity of Care, Interoperability

Introduction

Home health care (HHC), skilled nursing care in the patient home, is a large and growing sector of the United States health care system. A smooth transition from hospital to HHC is critical to quality healthcare and to reducing hospital readmissions [1; 2]. However, HHC nurses often receive only half of the recommended information relevant to developing the plan of care (POC) [3]. The POC includes orders for: performing procedures and treatments; ongoing assessment; case management; and patient and caregiver education. Missing, incomplete or incorrect information is likely to lead to care delivery inefficiencies where care is missed or more care is delivered than is necessary [4], which in turn may effect patient outcomes and cost. Despite these challenges, HHC nurses must make four critical decisions when constructing the POC during the time-constrained admission:

1. *Problems* to be addressed in the plan of care;
2. *Medication* accuracy and medication self-management capability;
3. The *services* to be consulted (e.g., the disciplines involved in addition to skilled nursing such as physical therapy); and
4. *Visit pattern* (when and how often to visit the patient).

While each of these decisions could impact hospital readmission, nurses' information needs and how nurses use the information for decision making during the HHC admission are currently poorly understood.

Below we describe the information flow and processes related to the HHC admission from referring facility to admitting nurse documentation in the EHR for one rural HHC agency. As part of the reported intake process, the intake nurse at the HHC agency completes a standardized paper *intake form* containing information obtained from a telephone call with the referral facility such as a hospital, physician's office, or skilled nursing facility-(SNF) (a SNF provides skilled clinical services in a temporary residential setting). The referral facility faxes specific referral information from an ordering physician in a *referral document*. On the intake form, the intake nurse documents patient characteristics and clinical information such as orders, problems, activity level/weight bearing status, disciplines requested [5]. The hard copy intake form along with the referral documents are placed in the patient's paper chart.

After intake, the observed admission process continues when the admission nurse, with the paper patient chart, visits the patient at the patient's home. While in the patient's home, the nurse optionally requests the patient or caregiver for the discharge summary (also referred to as discharge instructions) or progress note (hereafter both are referred to as *transition documents*) which originated from the hospital, physician office or SNF. Such a transition document contains care instructions sent home with the patient for reference.

The admission nurse documents patient clinical information in the electronic health record (EHR). He/she records current health status in the mandated Outcome and Assessment Information Set (OASIS [6], a standardized data collection instrument). The EHR includes a free text POC with problems and patient health goals [5]. If not completed in the home, the nurse later completes the EHR admission documentation outside of the patient home [7].

This paper describes findings from a mixed methods observational study examining the availability and quality information that HHC admission nurses use to make the four critical decisions. The investigation detailed in this paper is part of a more extensive study that chronicles nurses' information and decision practices at the point of care. We intend to design health information technology (HIT) recommendations to enhance the activities involved in the HHC admission process and improve HIT standards for home health agency EHR systems.

Methods

The study was approved by the Drexel University Institutional Review Board. We reimbursed the agency for nurses' time.

Setting. The small, rural Pennsylvania HHC agency served a majority white, lower socio-economic population. Nurses used a commercial laptop-based point-of-care EHR from Allegheny Software Publisher.

Data Collection. Six nurses who admit patients consented to participate. Each nurse was observed admitting two patients for a total of 12 admissions. The sample size of 12 was expected to achieve saturation with respect to making decisions. Nurses were observed as they prepared to visit the patient (pre-home: Phase 1), in the patient home (in-home: Phase 2), and after (in-agency, when they completed the documentation: Phase 3). Two researchers (PS and EB) accompanied the nurse to observe each admission. A third observer sometimes accompanied. All three phases lasted 2 to 4 hours in total per admission.

At the end of each phase, researchers interviewed the nurses to determine after having reviewed the documents, what conclusions were made regarding each of the four critical decisions noted above. The interview was intended to assess the status of the nurse's ability to make these decisions given the information up until that phase. Interviews were audio-recorded and later transcribed. In the patient's home, researchers took field notes, audio-recorded, and photographed the EHR screens and transition documents (if accessed). The nurses completed the documentation in Phase 3 using a video recorded, think aloud protocol.[8] Researchers collected photocopies of patient paperwork and copies of relevant emails (input) and nurse generated documentation including the POC (output). During the data analyses, the research team met with the observed nurses and separately with intake nurses to clarify information and findings (member-checking) by asking the following questions:

1. To analyze medication accuracy and medication self management information (medication management): How is patient medication information communicated to the nurse?
2. With regard to the intake form, referral document, and transition documents, which do you consider:
   - the most important source of any patient information?
   - the most accurate source of information?
3. How do nurses prioritize these different sources of information?

Data Analysis.

Content of Documents and Interviews. Two authors independently searched the intake, referral, and transition documents for terms and phrases associated with the four decisions. They compared coding, came to concensus, and transcribed the terms and phrases onto a Microsoft® Excel spreadsheet for each admission. To identify nurse statements in the interview transcripts and audio recordings across the three different phases, two authors independently conducted top-down (a priori) thematic analysis sensitized to the four decisions. The researchers documented the quotations in the spreadsheet. Team members, including a HHC nursing expert (KB), reviewed the coding and made changes based on discussion. Consensus was reached.

Quality of Documents and Interviews. For each of the four decisions, we assessed, documented, and reviewed concordance of the availability and content of the information in each document and nurse interview. We calculated the proportion of matched information in each document and interview dyad if information was available. For *problems* and *services*, the comparison was to find missing or added patient problems or services. For example, at least one service listed on the intake form was not specified in the interview. For *medication self-management capability*, the content was analyzed using qualitative analysis and then compared to determine whether the assessment of medication management capability had changed across the phases. For *visit pattern*, content was compared to evaluate whether the number of visits in a comparable week had increased or decreased.

To provide more data in the quantitative analysis for decisions in which information was missing, interview responses were incorporated. Phase 1 interviews were included as input. Phase 2 interviews were grouped with output (electronic artifacts produced when documentation was completed-POC), as the interview took place following patient assessment. Phase 3 interviews were grouped with output.

Last, we compared the information that the nurse had before making a decision (input – intake, referral, transition document) and the decision the nurse documented (output – POC).

Concordance was analyzed based on the type of information. For decisions for which we compared presence or absence of information (i.e., visit pattern, medication management), we used 2 by 2 tables. For decisions for which we compared contents of lists (i.e., problems, services) we used 3 by 3 tables: We compared item presence\absence across sources.

Interview and member checking responses were qualitatively analyzed by two authors and compared. Quotations relevant to the documents and information available to the nurse in his/her decision-making were selected.

The mixed methods analysis intertwined the quantitative findings (counts of availability), which were primary, with the qualitative results. The latter provided context for interpretation of the quantitative findings. For example, discordance between phases for information availability was matched to relevant quotes.

Results

Across all admissions, the availability and contents of documents and interviews were variable.

Availability of Documents and Interviews. Across the 12 admissions, intake and referral documents were universally available and the POC was consistently produced. Seven transition documents were available before the visit (as part of the referral documents): Five of which were also available during the visit (requested in the home). Transition documents were available for 7 of the 12 observations. Eleven interviews were conducted. Not all nurses answered all interview questions, often because they did not yet have a decision. During interviews, one or more decisions were addressed in each phase in 11 observations.

Content of Documents and Interviews. Referral documents lacked uniformity across the various referral sources but generally contained diagnoses, medications, and the POC. The actual documents, the information contained within them, the location of the information, and the number of pages varied. Transition documents tended to be one page in length, and were not standardized across referral sources. Information could be illegible, variable, or missing.

Qualitative analysis of nurse interview responses reached saturation with 11 interviews (no new patterns being identified). In Phase 1, a prevailing theme among interviews was the lack of information available for making decisions. This theme applied to the decision about the *visit pattern* decision for which data were not available on any document. Nurses anticipated acquiring needed information from the

assessment and talking to the patient during the first visit: "*I have no idea how often I'm going to see the patient for until after I assess them.*" (Nurse 3).

Problems. Across the 12 admissions, all *intake* and *referral* documents contained patient problems. The 5 *transition documents* viewed in the home listed the problem(s). Two plans of care did not contain *problems*. Nurses replied to interview questions asking about the patient *problem* for 10 admissions in Phases 1 and 2, and for 7 admissions in Phase 3. Although *intake* and *referral* documents contained *problems*, nurses did not solely rely on these documents to identify problems to include in the POC. They also relied on the admission assessment at the first visit:

> *We didn't get any information yet and they just faxed that to me…so I only have like, 8 pages total on this patient and her history at all…We'll just find out when we get there.* (Nurse 6)

Medication list/management. Related to medication accuracy, *medication lists* were not uniformly present in the documents available to nurses; *medication management* information was scarce. A medication list was not present in any *intake* document. It was in 8 of 12 *referral* documents and in 4 of the 5 available *transition documents* viewed in the home. Nurses reported during member checking that they relied on the *transition document* because that medication list was more current. However only 5 patients had the transition documents available for the nurse. The POC did not contain a medication list. The list was documented in the medication reconciliation which appeared elsewhere in the EHR.

Regarding *medication management*, data appeared in 1 of 12 *referral* documents, 3 of 5 *transition documents*, and 3 of 12 POCs. An OASIS question (M2020) about patient medication management was universally answered. Nurse responses related to medication management were present for 4 interviews in Phase 1, 7 interviews in Phase 2, and 8 interviews in Phase 3. Nurses explained that they determined if a patient could manage medications if a patient demonstrated understanding the medications and when to take them, and had an organized pill box:

> *I think he certainly is fine. He had that list printed out, his pill boxes were all filled for the week already. He's definitely sufficient in monitoring that or managing that.* (Nurse 1)

Services. *Intake* forms universally contained *services* information whereas services information was contained in only 4 of 12 *referral* forms, 1 of 5 *transition documents*, and all POCs. Nurses replied to the *services* interview question for 11 of 12 admissions in Phase 1, 9 admissions in Phase 2, and 5 admissions in Phase 3. Nurses reported that skilled nursing was likely to be ordered for those referred from a SNF. Physical therapy was ordered if the patient needed to build strength. Occupational therapy was ordered for conditioning:

> "*OT can teach…measures to conserve energy as far as when they're getting dressed."* (Nurse 1).

Social work was ordered for reasons including advanced directives, long term planning, alcoholism, and the need for community resources to provide more help in the home. A home health aide was ordered to provide assistance with activities of daily living due to the patient's restrictions:

> *Normally if the only diagnoses marked on here are orthopedic- like … the hip fracture- then typically the main service going in would be therapy…Home health aide, that's kind of typical sometimes for that, just because of restrictions.* (Nurse 1)

Visit pattern. *Visit pattern* information was scarce on the documents nurses had access to in the patient home, yet it was universally documented in the POC. No *intake* forms contained visit pattern data. Visit information appeared on 2 of 12 *referral* documents. No *transition document* documented skilled nursing visit pattern, although one transition document did include visit pattern for other services. Nurses replied to the visit questions in 10 of 12 interviews in Phase 1, 9 interviews in Phase 2, and 7 interviews in Phase 3.

In their *visit pattern* decisions, nurses reported that they considered a number of factors. One consideration was insurance constraints, as Nurse 6 stated:

> "*Insurance will pay for us to go out every day, but if there's someone else in home able to do it, they won't pay for us to come out every day*."

A second consideration was laboratory draw orders, according to Nurse 1, "*It'll impact our next visit if they want us to draw the labs in the home*." Nurses also incorporated wound care concerns in their decision making.

> Nurse 3: "*I would say if she had a wound. Or a diabetic foot ulcer. Where we were coming out and we were treating that. Especially for the first couple weeks, I would want to see her three times a week, or even every day*."

Nurses also thought about patient fragility. Nurse 4 said, "*He looks like he's pretty fragile so I think we'll come out more to see him.*" Nurses also considered visit patterns of other services. For example, Nurse 3 said:

> *Physical therapy is coming out to see her. Occupational therapy is coming out to see her. We're almost seeing her every day, but different disciplines. So I like to adjust my time so I'm not running into physical therapy all the time… so we keep our eyes on you, evaluating how well you're progressing.*

Nurses rarely had visit pattern information in their documentation. After assessing the patient in the home, the nurse always documented the information in the POC.

During member checking, nurses stated that their prioritized list of documents in terms of accurate and trusted patient data was: 1) transition document, 2) intake, and 3) referral. Although the transition document was available less than half of the time, it was the nurses' preferred source for the medication list.

Quality of Documents and Interviews. Comparison of input and output for *problem* and *services* decisions involved documents and not interviews, as there were sufficient admission documents for analysis. However, *medication management* and *visit pattern* decisions were not available in documents: Inclusion of interview responses did not provide sufficient content for analysis due to few nurses having made these decisions when interviewed.

Problems. Comparison of the number of problems listed among and between *input* documents and the POC indicates complete discordance. Input documents contained lists of problems with some overlapping and some unmatched problems. For example, the *intake* forms consistently contained more problems as compared to the *referral* documents for the same admission. Two intake forms contained problems not on the related referral documents. Intake forms tended to have fewer problems as compared to the related transition documents. Referral documents tended to have more problems as compared to the associated transition documents. The presence of lists of patient problems that lack agreement does not provide a clear picture to the nurse, as illustrated by the following exchange.

> Interviewer: *"After you read through that [referral, intake documents], did you have any thoughts that there was more or less on the list of problems than what was noted there?"*
> Nurse 3: *"Not until I really get there. I have to actually ask*

the person and assess."

In addition to mismatches among input documents, for each observation there was a mismatch between each *input* document and the POC. *Referral documents* consistently contained more *problems* as compared to the related POC. Three POCs contained one or more problems not included on the related referral document. A similar pattern was present when intake forms were compared to POC, with one exception: For only two admissions the problems on the intake form matched the problems on the POC.

Medication management. Input document and Phase 1 interview data (input) were present for 5 admissions. Data output, constructed from POC and Phase 3 interviews was present for 9 admissions. Of the 6 admissions for which both input and output medication management data were present, half had input that matched output.

Services. Information was present on all intake forms. Data output regarding services from documents was present for all admissions. Of the 12 admissions with both input and output data, 3 did not have input that matched output. One admission had an additional service on the output and 2 admissions had one more service on the input as compared to the POC.

Visit Pattern. Due to scarcity of *visit pattern* data among input documents, interview data were included as input data. Input document and interview Phase 1 (before home) data were present on 6 admissions. Output data were universally present. Of the 6 admissions included in the analysis, 2 did not have input that matched output: These admissions had fewer visits on the input document.

In summary, data quality, when comparing input and output documents, was weak for patient *problem* and *services* decisions. *Medication management* and *visit pattern* decisions data quality was very weak due to frequent absence of data.

Discussion

This is the first study of the availability, content, and quality of information that a HHC nurse has at patient admission. The three input documents (i.e., intake, referral, and transition) available to nurses as they made the 4 clinical decisions had the same information source and referring facility/physician, and were intended for different audiences: clinicians, intake nurse, and patient (respectively). The availability, content, and quality of the documents were variable. Nurses' identification of the transition document as the most valued and trusted source for medication information makes it the top priority when recommending improvements in data transfer. Missed medications, wrong doses, or wrong medications can lead to adverse events and unfortunately are too common during transitions in care.[9]

The *intake* document had the patient *problems* listed, and tended to be the sole and consistent data source for *services*. During member checking, intake nurses stated that they controlled the telephone conversation with the referring facility to make sure the needed information was elicited. *Referral documents* tended to have a list of *problems* and a *medication list*. Agency management informed us they had strongly encouraged local SNFs and hospitals to send referral documents with the information needed during the admission.

The mismatch of *problems* among all documents was unexpected. The POC contains far fewer problems as compared to input documents. This mismatch and decrease in the number of problems require further investigation regarding: (1) whether problems are missed; (2) whether and why nurses choose to address certain problems over others; or (3) whether national or agency policy restricts the types and number of problems addressed during an episode. Considering the poor quality of documents, it is unclear how the nurse decides which input problems to include or not include on the POC.

Medication management input and output information was infrequently available which was problematic in light of nurses reporting the need for medication management information during admission [10]. Input documents (*intake, referral and transition documents*) rarely had *medication management* information available and infrequently had *visit pattern* information. Nurses relied on obtaining this information during the in-home patient assessment. Electronic transfer of this information could increase efficiency and decrease time spent collecting it. It also could alert nurses to patients likely to have issues with medications, a risk factor for readmission [11]. This unavailability calls for standardized measurement of cognition, physical function, and self-care ability to assist the next level of care in understanding what challenges may exist. Unfortunately, the Impact Act, which calls for a uniform comprehensive assessment across all post-acute care settings, does not include acute care settings, the most common referral source [12].

Services input and output information was available for almost all admissions. Availability of services input information was due to the agency's universal elicitation and documentation on intake forms. Despite this effort, quality of services data was not high. Considering that the source of intake information was the referral facility, nurses in HHC may assess information differently than staff in the hospital [1].

Visit pattern had a paucity of input data and universal inclusion in the POC which indicates the importance of the patient assessment in obtaining information for the visit pattern decision. The low quality of visit pattern information is a hurdle for a HHC agency which strives to apply frontloading (provision of 60% of planned skilled nursing visits during the initial 2 weeks of the home health episode [2]) of skilled nursing visits for high risk patients to reduce 30-day hospital readmissions. The rationale is to enable nurses to identify and intervene on issues early in the episode before a readmission is necessary [13].

The low availability and quality of information needed for all 4 important clinical decisions indicate an information deficit during the transition of care to HHC. As a result, nurses may rely on the patient or caregiver during the patient assessment for the missing information. However, the patient may not always be the most reliable historian and clinical information maybe incorrect or misunderstood by the patient [14]. From a human information processing perspective, adequate and accurate information is the basis for making subsequent appropriate clinical decisions [15], and providing safe patient care [16]. Further, nurses cannot determine which patients to visit early and often until after their first visit which may not occur until two or more days after hospital discharge.

To address these information deficits we recommend improving interoperability capability and EHR design. Interoperability is supported with the international data standard, the Continuity of Care Document (CCD) [17] which summarizes and categorizes patient transition document information useful for transitions. The CCD is intended to be shared by SNFs, hospitals, and other health facilities as patients transition between settings. While the CCD is structured to contain problems and the medication list [5], it does not contain medication management, services, or visit pattern information. Our analysis indicates that nurses rely on patient assessment in the home to support medication self-management and visit pattern decisions. If this reliance is due to the nurse's need to elicit this information in person,, including this information in

the CCD may not be an effective intervention. However, if this information seeking pattern is because this information has historically been unavailable and if the pattern is amenable to change, future research should address whether nurses would rely on input information if it were available. If so, the information should be considered for inclusion in the CCD. At a minimum, the referring clinician's recommendation regarding the timing of the first visit could be of great value. Screening tools indicating higher risk patients may help.[18]

Regarding *services* data, communication of functional status or previous service use from the referral source would provide valuable information about the need for continued services. However, this information is not contained in the CCD.

Anticipating that interoperability will replace the use of paper input documents, we provide EHR recommendations. First, the EHR should be redesigned to capture CCD data (i.e., problems, services, functional level) and, after nurse review and authorization, the data could populate the POC, eliminating nurse transcription of input data and reliance on transition documents. Further research is needed to ascertain whether, if additional information was included in the CCD, nurses would use this information and rely less on gathering information with the patient assessment.

A second recommendation acknowledges the value of the intake document. The intake nurse should be able to record problems and services needed. Upon nurse authorization, the selected information input by the intake nurse could populate the POC.

Research design is a strength of this study. The field study was designed to collect qualitative and quantitative data, to support our mixed methods analysis. Limitations of the study include the sample size (six nurses and 12 patient admissions) and the setting elements (we examined one rural home health agency that used one EHR). To improve the generalizability of findings, future work is needed that increases: (1) sample size; (2) heterogeneity of nurse and patient populations studied; (3) number of different EHRs examined, and (4) number and type of referring sources.

Conclusions

Home health care nurses at a rural agency who admitted patients relied on three paper documents for information to make four important clinical decisions included in the plan of care. Universally we found a mismatch in content related to the four decisions among input documents, between each input document and the POC, and between the set of input documents and the POC. None of the documents nor the POC contained structured, actionable, electronic data. EHR design recommendations include data interoperability from referring facilities, information incorporated electronically instead of on paper, and structured and actionable input and POC data. These recommendations address information deficiencies for 2 of the 4 important clinical decisions which underscores the importance of the home visit for information elicitation.

Acknowledgements

This research is supported by the Agency for Healthcare Research and Quality (AHRQ) Grant # R01 HS024537. The authors thank the home health care agency and the participants.

References

[1] L. Melby, B.J. Brattheim, and R. Helleso, Patients in transition--improving hospital-home care collaboration through electronic messaging: Providers' perspectives, *Journal of clinical nursing* **24** (2015), 3389-3399.

[2] J. Rogers, M. Perlic, and E.A. Madigan, The effect of frontloading visits on patient outcomes, *Home Healthcare Now* **25** (2007), 103-109.

[3] K.H. Bowles, J. Pham, M. O'Connor, and D.A. Horowitz, Information deficits in home care: A barrier to evidence-based disease management, *Home Health Care Management & Practice* **22** (2010), 278-285.

[4] D.G. Fairchild, J. Hogan, R. Smith, M. Portnow, and D.W. Bates, Survey of primary care physicians and home care clinicians, *JOURNAL OF GENERAL INTERNAL MEDICINE* **17** (2002), 253-261.

[5] P.S. Sockolow, C. Wojciechowicz, A. Holmberg, E.J. Bass, S. Potashnik, Y. Yang, and K.H. Bowles, Home care admission information: What nurses need and what nurses have. A mixed methods study., *Stud Health Technol Inform* **250** (2018), 164-168.

[6] Centers for Medicare and Medicaid Services, Outcome and Assessment Information Set, in, 2011.

[7] Y. Yang, E.J. Bass, and P. Sockolow, Distributed electronic documentation practices of homecare admission: Informing electronic health record design guidelines, in: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, SAGE Publications Sage CA: Los Angeles, CA, 2017, pp. 573-577.

[8] C. Lewis and C. Wharton, Cognitive walkthroughs, in: *Handbook of Human-Computer Interaction,* M. Helander, T.K. Landauer, and P. Prabhu, eds., Elsevier, Amsterdam, 1997, pp. 717-732.

[9] A. Brody, B. Gibson, D. Tresner-Kirsch, H.M. Kramer, I. Thraen, M.E. Coarr, and R. Ruppert, High prevalence of medication discrepancies between home health referrals and centers for medicare and medicaid services home health certification and plan of care and their potential to affect safety of vulnerable elderly adults, *Journal of the American Geriatrics Society* **64** (2016), 5.

[10] P.S. Sockolow, E. Bass, K.H. Bowles, A. Holmberg, Y. Yang, and S. Potashnik, Data visualization of home care admission nurses' decision-making., in: *Proceedings of the AMIA 2017 Annual Symposium; Precision Informatics for Health: The Right Informatics for the Right Person at the Right Time.,* N. Sarkar, ed., Washington DC, 2017.

[11] O'Connor M, Murtaugh CM, Shah S, Barrón-Vaya Y, Bowles KH, Peng TR, Zhu CW, and F. PH, Patient characteristics predicting readmission among individuals hospitalized for heart failure, *Medical Care Research and Review* **73** (2016), 3-40.

[12] CMS, IMPACT Act of 2014 & Cross Setting Measures, in, Centers for Medicare and Medicaid Services, Baltimore, MD, 2018.

[13] M. O'Connor, A. Hanlon, and K.H. Bowles, Impact of frontloading of skilled nursing visits on the incidence of 30-day hospital readmission, *Geriatric Nursing* **35** (2014), S37-S44.

[14] S. Meredith, P.H. Feldman, and D. Frey, Possible medication errors in home healthcare patients, *J Am Geriatr Soc* **49** (2001), 719-724.

[15] L.J. Greenfield, Doctors and nurses: A troubled partnership, *Annals of Surgery* **230** (1999), 279.

[16] E. Brown, Raue, P., Mlodzianowski, A., Meyers, B., Greenberg, R., Bruce, M., Transition to home care: Quality of mental health, pharmacy, and medical history information, *Int'l J. Psychiatry in Medicine* **36** (2006), 339-349.

[17] HL7/ASTM Implementation Guide for CDA® R2 -Continuity of Care Document (CCD) Release 1, in.

[18] Topaz M, Trifilio M, Maloney D, Bar-Bachar O, and B. KH, Improving patient prioritization during hospital-homecare transition: a pilot study of a clinical decision support tool, *Research in nursing & health* (2018).

Address for correspondence

Paulina Sockolow, pss44@drexel.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190334

Incongruence of Patient Problem Information Across Three Phases of Home Care Admission: There's a Problem with the Problem List

Paulina S. Sockolow[a], Natasha B. Le[b], Yushi Yang[c], Sheryl Potashnik[a], Ellen J. Bass[a,c], Kathryn H Bowles[b,d]

[a]*Drexel University, College of Nursing and Health Professions, Philadelphia PA, USA;*
[b]*University of Pennsylvania School of Nursing, Philadelphia, PA, USA;*
[c]*Drexel University, College of Computing and Informatics;*
[d]*Visiting Nurse Service of New York, New York, USA*

Abstract

In home health care, the patient problem list is an important component of the admission and care planning processes and determines the subsequent care received. We examined the information received from the referring facilities and its relationship with the final patient problem list generated at home health care admission. Researchers observed 12 admissions and collected available documents related to the admission and care planning process. Problems identified in documents provided to admission nurses (input documents) and in documents subsequently created by those nurses (output documents) were coded to form a standardized set of problem terms across the documents. Documents available, distribution of problems within the documents, and concordance between input and output documents were assessed. A varying number of the 17 unique problems found across the documents were distributed by document type. Patients were referred to home health care with more clinical problems than were documented in the output documents.

Keywords: Home Care Services, Decision Making, Nursing Informatics, Documentation

Introduction

An accurate, complete, and current list of patient problems is valuable for concisely communicating a patient's clinical status among numerous and diverse clinicians and across care settings [1]. Annually, 12 million patients in the United States are referred from hospitals, skilled nursing facilities (SNF), and physician offices to home health care (HHC) [2]. HHC patients are typically older adults with multiple chronic conditions and have on average nine problems (health related needs that may benefit from clinical intervention) [3]. These problems are often associated with multiple chronic medical conditions that affect patient function, self-care management, and hospitalization risk. As some problems are chronic and stable, not every problem is addressed during a HHC episode.

During the HHC admission, a nurse makes critical care planning decisions including identifying the problems to be addressed in the plan of care (POC) (i.e., orders for ongoing assessment, patient and caregiver education, case management, and performing procedures and treatments). Too often the HHC nurse conducts the admission visit with fragmented, incomplete, or inaccurate knowledge of the patient's clinical condition [4],[5]. Inaccurate or incomplete problem lists could lead to inappropriate, missed, or delayed care [1]. In addition, 20% of HHC patients in the United States are readmitted to the hospital within 30 days of discharge [6]. An adequate transfer of information leading to accurate care planning may assist in providing higher quality care and preventing early readmission for a large and growing population of older adults [7],[8].

This paper presents findings from a field study examining the information received from the referring facility and its relationship to the final patient problem list generated at HHC admission. This examination is part of a larger mixed methods study characterizing HHC admission and care planning practices at the point of care. The overall goal is to develop recommendations to improve the HHC admission process and to identify opportunities for technology standards that support transitions in care to HHC via electronic health record (EHR) systems.

Methods

We conducted observations of the admission process and analyzed paper and electronic documents to examine the relationship between HHC admission documents and the problems in the problem list. The Drexel University Institutional Review Board approved this study.

Setting. The research setting was a small, rural Pennsylvania HHC agency serving a majority white population with low socioeconomic status. Agency nurses used a laptop-based commercial EHR system from Allegheny Software Publisher, designed for use at the point of care. Nurses also received and reviewed paper documents from the referral facilities and generated additional paper and electronic documents. Six nurses volunteered for the study and provided consent for participation. Patients observed in the home provided consent. To facilitate data collection, the agency scheduled the nurses to return to the agency after visiting each patient in the home. The agency was reimbursed for the nurses' time.

Data Collection. The six nurses were each observed admitting two patients each (12 admissions total). Figure 1 highlights the admission process information flow from the referring facility to the admitting nurse documentation of problems in the EHR. Phase 1 includes the intake process where initial documents are prepared for the admission nurse. Typically, via a telephone call, the intake nurse collects information from the referring facility. The referral facility also faxes referral information. If the intake nurse determines that the patient is to be admitted, he or she documents patient information, including medical diagnoses, on a standardized paper intake form. The intake form and referral documents are added to the paper patient chart and the admission nurses thereby have access to them.

Phase 2 of the admission process occurs in the patient's home. Armed with the paper chart, the admission nurse visits the patient. If not part of the faxed referral, the nurse may ask the patient or caregiver for the Discharge Summary (also sometimes

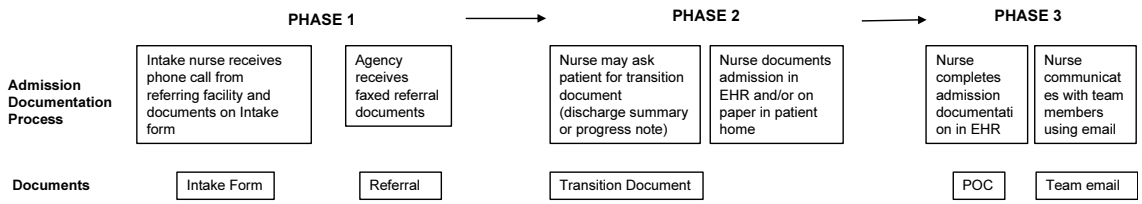


Figure 1. Phases of Home Care Admission Processes and Documents Created and/or Available.

referred to as discharge instructions) or Progress Note, called transition documents herein. With care instructions to reference, the documents are typically given to the patient at discharge from the referring facility.

In the home, the nurse documents in the EHR, on paper, or both [9]. Electronic documentation includes: (1) the patient's current health status recorded in the Centers for Medicare and Medicaid (CMS) mandated Outcome and Assessment Information Set (OASIS), [10] a standardized assessment instrument, and (2) a free text POC document with problems (nursing diagnoses) to resolve or support and related interventions, along with the patient's health goals.

Phase 3 is the completion of the admission and the creation of a team email. For this study and agency, this phase occurs at the agency and typically lasts approximately less than two hours. Not part of the EHR, the team email is a free text document summarizing the list of patient problems and the POC for the other nurses on the care team. Its purpose is to notify team members with timely, comprehensive information. This action avoids any delay in communication caused by inability to synchronize patient information on the laptop with the EHR database in a timely manner.

Research team members photocopied or photographed available documentation generated by the nurses including the POC, problem lists, and team emails.

The researchers, drawing on their HHC and nursing expertise, hypothesized that the intake document would paint the picture of the patient. They expected that additional problems would be added once the OASIS was completed. They also expected that the nurse would communicate all of the problems in the POC. The POC would then support problem prioritization.

Data Analysis. To standardize the vocabulary and reduce variability, the team used line by line coding to map each problem (medical and nursing) to terms in the Omaha System Problem Classification Scheme (Omaha System) [11]. This scheme is a standardized terminology often used in home and community-based settings. The scheme is recognized by the American Nurses Association as a recommended terminology. In 2007 it passed the Healthcare Information Technology Standards Panel (HITSP) Tier 2 selection criteria for Use Cases. It is integrated into the National Library of Medicine's Metathesaurus, Logical Observation Identifiers, Names, and Codes (LOINC®); and the Systematized Nomenclature of Medicine Clinical Terminology (SNOMED CT®) [12].

The Omaha System provides standardized signs, symptoms and problem labels for 42 problems organized in four domains (i.e., Environmental, Psychosocial, Physiological, Health Related Behaviors). Example problems include Cognition, Pain, Medication Regimen, Respiration, Circulation, Skin, Mental Health and Physical Function [13]. Omaha System expert, KB educated co-author NL about the Problem Classification Scheme definitions and problem signs and symptoms. Then the two co-authors conducted the coding together for two cases to establish coding rules. They then coded separately. To reach consensus they discussed terms on which they disagreed or were unclear. For example, the OASIS assessment "Able to bear weight and pivot during the transfer process, but unable to transfer independently" was coded in the Omaha System as the problem "Neuromusculoskeletal Function".

Each specific problem from each paper and electronic document was entered as a unique row into an Excel spreadsheet. Columns included one document per column, ordered by phases. Both nursing and medical problems were included in the analysis. Counts by document type for each problem were calculated and matches across the columns were identified.

Distribution of problems. After coding we identified the number of unique problems among the 12 admissions, regardless of which document contained the problem. We examined the overall distribution of problems among types of documents. We assessed which types of documents had more unique problems as compared to other documents. We calculated the median number of unique problems occurring on each type of document per admission.

We investigated the distribution of specific problems among the types of documents to discern patterns in appearance of problems in documents. We identified whether specific problems tended to occur more often in certain types of documents (e.g., pain appeared mostly on POCs). We also identified the set of problems that occurred most often in each type of document to see whether problems co-occurred among documents.

Concordance analysis. We assessed the concordance of problems documented (whether they matched across phases and documents) and discordance (whether they were missing from one or more phases or documents). We compared the content of the set of documents available to the nurse during the admission (input-referral, intake, transition document-discharge instruction/progress note) to the content of the electronic artifacts produced when documentation was completed (output-POC, team email). We determined whether either output document contained the complete set of problems that appeared across output documents. We conducted the same analysis for the three input documents. Then we compared the selected input document(s) and the selected output document(s) to determine concordance for problem sets between input and output. We identified problems that were matched between input and output documents.

For the quantitative analysis, we calculated the total number of problems for each observation and the median across all observations. We used the above comparison of input problems to output problems for each observation to calculate the number of problems that appeared in: (1) both input and output; (2) input and not output; and (3) output and not input. For these three categories, we then calculated the total number and median across all observations.

Case studies. We illustrate the analyses with two admission case studies. One case has all the documents. The second case lacked the transition document. We present the median number of unique problems on each document, the median number of problems on each dyad of input and output documents, and the total number of unique problems.

Results

The documents varied in content, length and detail. They contained a variety of problem terms to describe the patients' health needs.

Assessment of documents. Referrals, intake forms, POCs, and team emails were available for all observations. Seven transition documents (discharge instructions or progress note) were available before the visit (as part of the referral documents): Five of which were also available during the visit (requested in the home). The data on these documents were not standardized across referral sources; the information on a few documents was illegible or missing. Patient problems were found in all types of documents among the observations. However, two POCs contained no problems; instead they contained plans for assessment and health promotion interventions.

Identification of problems. Following standardization of problem terms to the Omaha System Problem Classification Scheme, we identified 17 unique problems across the documents and observations. Referral documents had the most problems per admission (median 7; 2-9) (Figure 2). Team emails had the next highest number of problems (4, 2-9). Intake and transition documents each had a median of 2 (ranges of 0-5 and 1-8 respectively). The POC had the least number of problems (1; 0-2). Overall, the median number of problems decreased from input to output.

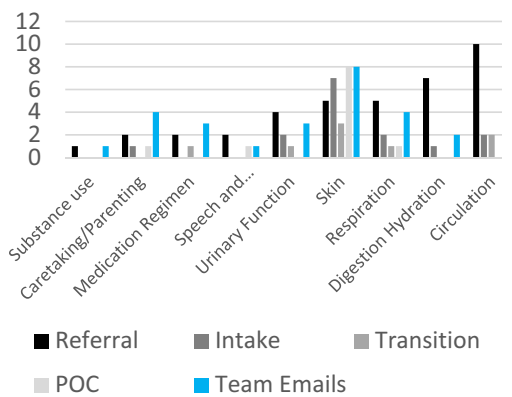


Figure 2. Occurrence of Actual Problems across Documents

Distribution of problems. Documents tended to differ as to which of the 17 problems occurred most often or at all (see Figure 2). Three problems occurred across all document types among the observations: "Skin" (occurred 28 times), "Respiration" (17 occurrences) and "Communicable/ infectious condition" (21 occurrences).

Concordance analysis. A complete set of problems was not found in any of the input or output documents. Accordingly, we used all available input documents, and both output documents in our concordance analysis of input documents as compared to output documents.

None of the 12 admissions had the same set of Omaha System problems on input documents as compared to output documents. All admissions had more problems on the input documents as compared to the output documents. For three admissions, the set of problems on the output documents was a subset of the problems on the input documents. The remaining nine admissions had one or more problems on the output documents that did not occur on the input documents, as well as having one or more problems missing on the output documents that were present on the input documents.

Although the set of problems did not exactly match between input and output documents, admissions did have individual problems that were matched between input and output documents. Here we report the medians (and ranges). In general, for each admission, a total of 8 (4-12) problems occurred on input and output documents. Among these problems, 3 (2-6) matched between input and output for each admission. Also, 3 (0-5) problems appeared on the input documents and not on the output documents. And, 1 (0-3) problem appeared on the output documents and not on the input documents.

Case studies. Analyses of two representative admissions are presented for further elucidation: an admission with all the documents, and an admission lacking the transition document. Admission 1 contained 8 problems on the referral, 4 problems on the intake form, 2 problems on the transition document, and 2 problems on the POC, and 9 problems on the team email. Among the input documents, 2 problems on the referral also occurred on the intake form and not on the transition document. Two other problems on the referral did not appear on the intake form but did appear on the transition document. The referral had 6 problems that did not appear on the intake form nor on the transition document. The intake form had 2 problems that occurred on neither the referral nor the transition document. Therefore, among the three input documents, 10 problems appeared. No single problem appeared on all three documents, and no single document contained all 10 problems. The 2 problems in the POC appeared in the team email. The 5 additional problems in the email did not appear in the POC. Of the 10 problems on the input documents, only 7 problems appeared in both the input and output documents. 3 problems appeared only in the input documents and not in the output documents. The dropped problems were Personal Care, Communicable/infectious condition, and Digestion-hydration.

A second example is admission 10 which lacked the transition document. This admission had 6 problems on the referral. Five of these problems also appeared in the intake form. No problems from the referral appeared on the POC; two problems did appear on the email. One problem on the referral did not appear on any other form. The POC problem, "Skin", did not occur on any other document. The team email had two problems, neither of which appeared on the POC. One of these problems, "Neuromusculoskeletal Function", did not appear on any other document. In summary, this admission had 8 problems; all but two of the problems appeared in the referral. Only 4 of the 8 problems on the input documents appeared on the output documents. The dropped problems were Urinary Function, Circulation, Nutrition, and Communicable/infectious condition.

Discussion

We investigated the content of problem data that the HHC nurse had at patient admission and the content of problem data that the nurse documented in the EHR and communicated outside of the EHR to the care team. This study focused on the source and congruence between documents available to the nurses when formulating the list of problems for the POC that they plan to address during the HHC episode. This study, in a rural home care agency, is the first of which we are aware, to examine this question. Findings are intended to inform point-of-care HHC EHR design, and to inform policy decisions related to interoperability along the transition in care from referral source to HHC.

Assessment of documents. The faxed referral documents contained the most problems compared to any other document.

The intake form, which was written by the home care agency nurse using information elicited from the referring facility, was universally available. The transition document was accessed in the patient home for less than half of the admissions. The team email, while universally present, was not part of the EHR. Instead, it was a communication work-around to address timeliness and information accessibility issues of the EHR. The POC, while universally present and part of the EHR, did not always contain problems and was in free text.

None of the documents were uniform or contained standardized information. Without standardization information can be lost or hard to find as our findings indicate.

Distribution of problems. The absence of any apparent pattern to the distribution of problems across documents provided no insight to nurse decision making. No document contained a complete set of patient problems. Thus, the initial referral document could not be relied on as the single source of problems. Others have reported that problem lists from hospitals and physician practices are likely to be inaccurate or out of date [1], [14], [15]. Further, these documents are often faxed from the referral source and contain multiple pages for a nurse to look through during the time constrained admission visit [16]. In addition, referral documents also contained both problems and interventions, especially for SNF patients, which puts further demands on the nurse's time to sort through dissimilar information. Similar to the other input documents in this study, referrals were not digitized or structured so as to be available for data management (e.g., copied to structured data fields for viewing or computing).

The intake form did not always contain the total set of problems appearing on the referral form. Intake forms matched the referrals in one admission and had fewer problems for the other admissions. The lack of completeness of the intake form and the occasional absence of the transition document suggest that neither could be relied on as a single source of patient problems.

Neither output document, team email nor POC, contained a complete record of patient problems. The team email, among all the documents, was second to the referral for the number of problems. The team email did not contain all problems documented in the POC. Nurses inconsistently documented some of the same problems in both email and POC. Redundant and conflicting documentation may require unnecessary expenditure of time and effort by the nurse, as well as introducing opportunity for transcription error and omission.

The POC, although part of the EHR, was not a complete source of problem information for the follow-up nurse. The POC unexpectedly contained fewer (or no) problems as compared to all other documents. Compared to team emails, the POC contained one quarter of the problems. For one quarter of the admissions, the set of problems on the POC did not match the set of problems on the team email. A potential reason for the mismatch may have been due to the fact that POCs tended to also contain interventions, evaluations, and health promotion activities, without a clear place designated where the nurse was required to document a finite problem list. Another potential cause for the mismatch is that the POC, being unstructured text, had problems which were not actionable: A nurse could not document against the problems when delivering care. As a consequence, the nurse was unable to determine if a patient problem was active or resolved. This data structure and flow do not support high quality patient care.

Our analysis of the distribution of problems among documents found that patients tended to have on average 8.5 problems at admission, which concurs with the finding of 9 problems in a prior study [3]. Findings also indicate that the problems skin, infection, and respiration are seen as important to communicate across all documents. This finding was clinically resonant with the nurse expert on the team (KB).

To the best of our knowledge, this is the first study to identify that patients are referred to HHC with more clinical problems (median 7) than are documented in the EHR plan of care (POC—median 1) or informally communicated to the team by email (median 4.5). None of these documents were in the EHR as structured data. This finding that patient referral documents contain more problems than are communicated to the care team indicates that admission nurses make decisions of which problems in the referral documents to include in the patient's care episode.

Concordance. The admitting nurse had access to the referral document, an intake form and infrequently, the patient's transition document. However, for no admission did the set of problems on the combination of these paper forms match the set of problems on the POC combined with the email.

Nurse decision-making related to inclusion and exclusion of problems between input documents and POC is unclear. We observed a decrease in number and mismatch of problems between input documents and output documents. The mismatch between input and output may be due in part to new information the nurse gathered during patient assessment in the patient home. Also, the EHR failed to provide structured language for describing patient problems and tracking their status over time.

EHR design recommendations. The decrease in number and mismatch of problems is an issue which could be addressed with EHR redesign and interoperability capability. Accordingly, we provide the following EHR design recommendations.

First, the EHR should capture and communicate problem data from the referral source. This recommendation relies on interoperability, the electronic movement of structured data among EHRs. The international data standard, the Continuity of Care Document (CCD), supports this capability [17]. The CCD is expected to be shared by health facilities as patients transition between settings. Relevant to this analysis, the CCD is structured to contain a list of patient problems [16]. Accordingly, we recommend that problems from the referral source be structured as per the CCD. Following communication of problems from the referral source to home care, the homecare EHR should map referral problems to a nursing problem terminology. Preferably the nursing terminology should be a standard terminology, such as the Omaha System [11]. The patient problem list should be viewable by the nurse. The problem should cascade through the EHR to the POC, avoiding nurse transcription of input problem data and the risk of losing information across documents.

Two recommendations are related to reducing redundant documentation. The software should enable the intake nurse to record problems viewable by the admitting nurse, and the problems should cascade through to the POC. In addition, the software should be designed to limit the number of times a problem is documented in the EHR, preferably to one. Our finding that the referral documents contained more problems as compared to the other documents, combined with our observation that referral documents tended to contain many pages [16] suggests the need to categorize referral problems to enable nurses to identify active problems. We suggest that problem modifiers such as resolved, actual, health promotion, or potential would be helpful to communicate the full array of patient issues [11]. Such descriptors would clarify the status of problems related to a hospitalization to retain the history of events and clarify which problems are resolved, are being actively worked on, or require preventive maintenance.

A fourth recommendation supports nurse workflow. The problems should be documented as structured data and be

actionable such that a nurse can document interventions and the status of the problem against the problems.

Two recommendations are related to eliminating the workaround of producing a team email to communicate the POC. We propose that the problem data should be readily viewable in a structured summary available to the care team, so as to eliminate the need to send the team email. Also, the EHR should be configured to support the timely update of patient information accessible by the care team.

A strength of this study is the research design. The field study was designed to collect and analyze quantitative data, to produce quantitative findings. Study limitations include a small sample size of six nurses and 12 patient admissions as well as the setting elements of one rural home health agency that used one EHR. Future work to further the generalizability of the findings would increase heterogeneity and sample size of nurse and patient populations studied, and the number of different EHR systems investigated.

The study was not designed to ask the nurse why selected problems were included in the POC and other problems were dropped. Future work to address this question would require near real time analysis of input and output documents and immediate review of this analysis with the nurse. In addition, future work could occur in acute care where a patient-centered assessment near discharge could identify patient problems from the patients' perspectives to share with the HHC admission team. Transmitting that list to HHC and studying the effect of having this information ahead of time are next steps.

Conclusion

This rural HHC agency relied on 3 paper input documents from the same referring source, each containing different sets of problems. The variation in problems may be due to the different intended audiences for the communication: clinicians at the referring facility, clinicians at the HHC agency, and patients. Input documents contained almost twice as many problems as compared to the output documents. However, no input document contained a complete list of problems. Following the home visit and patient assessment, new problems appeared in the output documents, and some problems on the input documents were omitted in the output documents without explanation or documentation of such. This observation underscores the importance of the home visit and further EHR functionality. Assessing the patient in the home often produced additional problems beyond those communicated from the referral source. The important output documents were the plan of care and a team email both of which were outside the EHR. This study illustrates serious issues related to EHR design and problem lists. The team email contained far more problems as compared to the plan of care. The usage of the team email highlights EHR design deficits related to the plan of care and team communication. A lack of timeliness of information; lack of structured, actionable data; and lack of interoperability provide opportunities to improve the design of this EHR.

Acknowledgments

AHRQ Grant: R01 HS024537, "Information Needs of Homecare Nurses During Admission and Care Planning." The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality.

The support and participation of the home health care agency management and nurses.

Carl Pankok, Jr. PhD CHFP for assistance with the analysis.

References

[1] A. Wright, J. Pang, J.C. Feblowitz, F.L. Maloney, A.R. Wilcox, K.S. McLoughlin, H. Ramelson, L. Schneider, and D.W. Bates, Improving completeness of electronic problem lists through clinical decision support: A randomized, controlled trial, Journal of the American Medical Informatics Association 19 (2012), 555-561.

[2] National Association for Home Care Hospice, Basic Statistics About Home Care, in, The National Association for Home Care & Hospice, 2008, pp. 1-21.

[3] M.D. Naylor, K.H. Bowles, and D. Brooten, Patient problems and advanced practice nurse interventions during transitional care, Public Health Nursing 17 (2000), 8.

[4] K.H. Bowles, J. Pham, M. O'Connor, and D.A. Horowitz, Information deficits in home care: A barrier to evidence-based disease management, Home Health Care Management & Practice 22 (2010), 278-285.

[5] S.E. Jewell, Discovery of the discharge process: A study of patient discharged from a care unit for elderly people, Journal of advanced nursing 18 (1991), 1288-1296.

[6] S.F. Jencks, M.V. Williams, and E.A. Coleman, Rehospitalizations among Patients in the Medicare Fee-for-Service Program, N Engl J Med 360 (2009), 1418-1428.

[7] D.M. Hartung, J. Hunt, J. Siemienczuk, H. Miller, and D.R. Touchette, Clinical implications of an accurate problem list on heart failure treatment, JOURNAL OF GENERAL INTERNAL MEDICINE 20 (2005), 143-147.

[8] The National Association for Home Care & Hospice, Basic statistics about homecare, in, The National Institution for Home Care and Hospice, Washington, DC, 2010, pp. 1-21.

[9] Y. Yang, E.J. Bass, and P. Sockolow, Distributed electronic documentation practices of homecare admission: Informing electronic health record design guidelines, in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, SAGE Publications Sage CA: Los Angeles, CA, 2017, pp. 573-577.

[10] Centers for Medicare and Medicaid Services, Outcome and Assessment Information Set, in, 2011.

[11] K.S. Martin, The Omaha System: A key to practice, documentation, and information management Health Connections Press, Omaha, NE, 2005.

[12] The Omaha System, The Omaha System: Solving the clinical data-information puzzle, in, 2018.

[13] K. Bowles, Omaha System in nursing research, Health Connections Press, Omaha, NE, 2005.

[14] D.R. Luna, M. Franco, C. Plaza, C. Otero, S. Wassermann, M.L. Gambarte, D. Giunta, and F.G.B. de Quirós, Accuracy of an electronic problem list from primary care providers and specialists, in: Medinfo, 2013, pp. 417-421.

[15] A. Singer, S. Yakubovich, A.L. Kroeker, B. Dufault, R. Duarte, and A. Katz, Data quality of electronic medical records in Manitoba: do problem lists accurately reflect chronic disease billing diagnoses?, Journal of the American Medical Informatics Association 23 (2016), 1107-1112.

[16] P.S. Sockolow, C. Wojciechowicz, A. Holmberg, E.J. Bass, S. Potashnik, Y. Yang, and K.H. Bowles, Home care admission information: What nurses need and what nurses have. A mixed methods study., Stud Health Technol Inform 250 (2018), 164-168.

[17] HL7/ASTM Implementation Guide for CDA® R2 - Continuity of Care Document (CCD) Release 1, in.

Address for correspondence

Paulina Sockolow, pss44@drexel.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190335

Evaluating the Scope of Clinical Electronic Messaging to Coordinate Care in a Breast Cancer Cohort

Bryan D Steitz[a], Mia A Levy[a,b]

[a] Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, USA,
[b] Department of Medicine, Division of Hematology and Oncology, Nashville, TN, USA

Abstract

Care coordination has received attention as an opportunity to improve healthcare delivery. Current work to quantify provider coordination has primarily relied on identifying shared patients, but neglects to understand communication patterns. We applied social network analysis to electronic health record (EHR) secure messaging data to compare networks of providers who share patients and networks of providers who communicate about patients. We studied 2175 stage I – III breast cancer patients who received outpatient treatment from 1758 providers at a large academic medical center in the southeastern United States. Patients in our cohort were involved in 94324 appointments and were the subject of 307144 message threads. We found that 9.9% of provider-provider pairs that shared patients were mutually involved in electronic communication about their patients. EHR data sources can be used to evaluate provider communication across a clinical enterprise, which can help identify opportunities to improve collaboration and reduce provider burnout.

Keywords:
Transitional Care, Workflow, Communication

Introduction

Care coordination is a multidimensional concept involving care integration and information sharing across providers and settings. Care coordination has been recognized as an approach to deliver high-quality care [1]. Inadequate care coordination is common, often caused by lack of communication between care providers [2]. Without effective care coordination, patients often experience poorer outcomes and higher costs [3,4].

Improving care coordination has been recognized around the world as an opportunity to improve healthcare delivery [1]. The United States [1], Australia [5], and many countries across Europe [6,7], have all implemented programs to improve coordination between care providers. One approach in the United Kingdom, Coordinate my Care, was implemented by the National Health Service to improve information sharing for palliative care patients [7]. Similarly, the SIKS project in Denmark aimed to integrate chronic disease care and improve care coordination [8]. With the aging global population and onset of complex chronic diseases, care coordination has become increasingly important as patients are required to receive care from multiple specialists across geographically distributed locations.

Health information technology and clinical information systems are central to care coordination. Electronic health record (EHR) systems offer important collaborative functionalities such as clinical messaging and shared patient charts [9], which are used by individuals in various clinical roles to support the shared goal of providing optimal patient care [10]. Similarly, technologies such as patient portals and personal health records allow patients to collaborate virtually with their care team and coordinate care without a clinic visit [11]. Other technologies such as health information exchanges can allow providers from multiple institutions access to patient data at the point of care [12].

Studying care coordination and communication between individuals is a critical way to gain insights for process improvement and optimization. One approach to evaluate care coordination and communication is through analysis of the extensive clinical data stored in the EHR [13]. The secondary use of routinely collected data from clinical environments can offer valuable insights into the collaboration patterns and routines of clinical personnel.

While the importance of care coordination is recognized, few studies have been conducted to quantify the scope of care coordination. Understanding provider communication patterns is a key first step to quantify care coordination. In our prior work, we found that medical oncologists treating breast cancer patients are connected to an average of 737 providers by a shared patient, and have as many as 114 new relationships per year [14]. In this study, we seek to advance the understanding of provider coordination by evaluating the scope of electronic communication between providers treating breast cancer patients at a single institution.

Methods

This study was conducted at the Vanderbilt-Ingram Cancer Center at the Vanderbilt University Medical Center (VUMC). VUMC is a large academic tertiary care center, located in middle Tennessee, with a large referral population from the Southeastern United Sates. This study was approved by the Vanderbilt University Institutional Review Board.

Study Population

We extracted data from the Vanderbilt University Tumor Registry on patients with stage I, II, or III breast cancer, who were diagnosed between January 1, 2011 and November 2, 2016. The Vanderbilt University Tumor Registry collects diagnosis and treatment data on all cancer patients who were either diagnosed or received part of their first course of treatment at VUMC [15]. Data from the tumor registry included a unique patient identifier, date of diagnosis, and staging information. Similarly, for each respective patient, we extracted all outpatient appointment data corresponding to an appointment after cancer diagnosis. Appointment data included a unique patient identifier, a unique provider identifier, and an appointment date. Patients who had at least one outpatient

appointment in the six-months following their diagnosis were included in the study. We also extracted secure clinical messaging logs from the EHR to understand the scope of communication between providers after the patient's diagnosis between January 1, 2011 and November 1, 2017. Messaging data included a unique provider identifier, message thread identifier, message date, and message length. In both the appointment and messaging data, we mapped each provider identifier to their unique national provider identifier (NPI) to determine specialty.

Network Representation

To understand physician relationships associated with the treatment of a patient in our cohort, we modeled the data as a social network. A social network consists of *nodes*, or entities between which a relationship occurs, and *edges*, a connecting tie between two nodes representing the existence of a relationship. Both nodes and edges can assume properties such as size or color to represent network features. To understand and compare relationships between providers, we create one network to represent potential coordination via outpatient appointments with a shared patient from our cohort, and another network to represent secure communications between providers regarding a patient in our cohort. We filtered the messaging network such that only provider-provider pairs who were connected by a shared patient were included. To visualize the network, we assign node and edge colors by network inclusion.

We created the outpatient appointment network using a method of temporal edge creation [14]. Temporal edge creation represents relationships that are likely to occur by treating a patient during overlapping time periods. To create the network, we start by extracting a list of all providers involved in the outpatient care of a patient in our cohort as the list of nodes. To create edges, we first order the list of all appointments by patient and appointment date. We iterate through the appointment list by patient, recording the first and last interaction between a provider and patient. For each patient, we record the list of provider pairs involved in the care of a single patient during an overlapping time period as our set of edges. We aggregate provider-provider pairs across patients by taking the sum of patients for whom both providers are involved in care over the same time period.

Similarly, to create the clinical communications network, we begin by extracting the list of providers involved in a clinical communication thread about a patient in our cohort as the node list. We create edges by taking the pairwise combination of providers involved in a single communication thread about a patient. Provider-provider pairs are summarized by a count of unique message threads in which both providers were involved.

Data Analysis

To analyze relationships between providers, we created two types of networks: a network of all providers involved in the care of a patient in our cohort and a network of providers involved directly in breast cancer treatment and their immediate connections. We defined medical oncologists, radiation oncologists, surgical oncologists, and plastic surgeons as providers involved directly in breast cancer treatment due to the frequency with which they are involved in the first course of breast cancer patients' treatment. We conducted all analyses using the igraph [16] package within R 3.3.1 [17].

To compare messaging between providers and patients, we calculated descriptive statistics. We assessed provider-level statistics by secure messaging involvement. To understand similarity between messaging and appointment networks, we compare node and edge sizes and analyze edge overlap. In the appointment network, we define edge size as the number of unique shared patients between provider-provider pairs; the node size is the number of unique patients treated by a provider. In the messaging network, edge size is defined as the number of shared message threads involving a provider-provider pair. We similarly quantify connectivity across the entire network to calculate *network density*. Network density is interpreted as the percentage of possible edges that are present in the network. For each individual provider, we calculated *degree*, or the number of direct relationships to a node.

Results

During our study period, there were 2267 patients diagnosed with Stage I – III breast cancer. 92 patients did not have an outpatient appointment within the six months following diagnosis, leaving 2175 patients in our study cohort. Table 1 summarizes the messaging data. Patients in our cohort had 94324 appointments with 1758 unique providers. There were a total of 625137 messages in 307144 messaging threads sent about patients in our cohort with an average follow-up time per patient of 32.9 months from diagnosis. 222168 (72.3%) message threads involved at least two individuals and 14484 (4.7%) involved at least two billing providers who were involved in at least one appointment (appointment providers). 1093 (62.2%) billing providers were involved in messaging about a patient in our cohort.

Table 1 – Message Thread Statistics

	Message threads	**Mean threads per patient**	**Mean threads per appointment**
Total	307144	142.3	3.7
Involve:			
One individual	88266 (28.7%)	41.1	0.45
At least two individuals	222168 (72.3%)	102.8	1.13
At least one appointment provider	112460 (36.6%)	52.1	0.57
At least two appointment providers	14484 (4.7%)	7.5	0.07
At least one cancer appointment provider	62856 (20.5%)	30.8	0.32
At least two cancer appointment providers	7371 (2.4%)	4.6	0.04
Patient	47496 (15.5%)	38.9	0.24
Patient and an appointment provider	16341 (5.3%)	14.9	0.83
Patient and a cancer appointment provider	6748 (2.2%)	8	0.03

Table 2 presents statistics for the full network of providers. There were 2610 provider-provider pairs connected by a shared patient and message thread, which accounted for 9.9% of the

total edges in the graph. There were 2610 provider-provider pairs involving 761 distinct providers who directly exchanged secure messages. Providers who sent messages communicated with an average of 6.9 other providers. Each provider-provider pair involved in messaging communicated an average of 5.7 unique message threads. 34.6% of message threads were sent within one day before or one day after an appointment with the respective provider. In 9.4% of provider appointments, providers communicated through a message thread within one day before or after the appointment.

Table 2 – Provider Network Statistics by Messaging Use

	Appointment and View Messages	Entire Graph
Number of Nodes	761	1758
Number of Edges	2610	26233
Average Shared Appointments	6.8	2.0
Average Shared Message Threads	5.7	3.9
Network Density	0.90%	1.70%
Mean Degree	6.9	29.8

Across the entire network, 25 providers were related directly to routine breast cancer treatment. These 25 providers were involved in 50.9% of all messaging threads between two appointment providers about our patient cohort. We present the messaging statistics for cancer specialists in Table 3. Surgical oncologists were involved in 6314 unique message threads, more than any other specialty. Likewise, each surgical oncologist, on average, was involved in 1064.2 unique message threads with 69.3 collaborators. Each medical oncologist, on average, communicated with 134.3 specialists, more than any other specialty. Across all four specialties, radiation oncologists had the fewest number of messaging threads (537) and number of provider-provider edges (91). 31.9% of all edges and 76.1% of all weighted edges involving a radiation oncologist were with another cancer specialist. Medical oncologists had 13.2% of edges with another cancer specialist, the fewest of any of the specialties involved directly in breast cancer treatment. However, when accounting for the volume of communications, 64.6% of the total edge weights involved a medical oncologist and another cancer provider. This 389.4% increase from percent of edges to percent of weighted edges was the largest of any specialty. Surgical oncologists had the highest percentage of weighted edges with other cancer providers (83.2). Medical oncologists had the largest overlap between shared patients and shared message threads with another provider (18.5), followed by plastic surgery (11.5), surgical oncology (11.0), and radiation oncology (7.3).

Discussion

In this work we analyzed communication patterns between providers during their treatment of breast cancer patients. Using social network analysis and graph statistics, we were able to quantify collaboration between providers by comparing physician messaging connectivity to patient sharing through outpatient appointments. Other studies have applied quantitative methods to study provider connectivity, but these studies have commonly relied on appointment data or claims data, which do not necessarily indicate clinical coordination or collaboration. In our previous work, we conducted a social network analysis of data from the Vanderbilt University tumor registry to detect differences in breast cancer provider connectivity by patient stage [18]. Another study by Hussain and colleagues used SEER-Medicare data and found that increased patient sharing led to improved mortality in Stage III colon cancer patients [19]. Similarly, Landon and colleagues used Medicare claims data to identify networks of physicians who likely have close relationships [20,21].

To our knowledge, this is one of the first studies to quantify provider collaboration using secure communications deployed throughout a clinical institution. Our quantitative approach is supported by the use of EHR messaging data. The EHR messaging functionality offers a secure and HIPAA compliant way for care team members to communicate about a patient's care. EHR functions such as secure messaging have been identified as a means for care coordination in previous work [22]. For providers within the same EHR system, care coordination is enabled by this form of asynchronous communication. Our results demonstrate the substantial volume of clinical communications between providers treating breast cancer patients to coordinate care. This form of communication is not available to providers who do not share the same EHR and practice at the same institution, which likely limits their ability to sufficiently coordinate care. Our results highlight the baseline need to implement electronic tools to enable care coordination between providers, particularly as EHR messaging functions are not availabile to providers practicing at separate institutions.

In our network, the 2175 patients in our cohort were the subject of over 222168 message threads between at least two individuals. 14484 of these threads involved at least two appointment providers. We found that each patient, on average, was the subject of 142.3 message threads. This staggaring number of communications represents substantial work required of the care team that is outside of an appointment. Our results indicated that 9.9% of the total provider-provider pairs were involved in secure communication in addition to sharing patients. However, medical oncologists had a much higher overlap of 18.5% among all connections and 48.4% among other cancer providers. A 2016 study by Shanafelt and colleagues found that physicians who reported high amounts of EHR use such as secure messaging were at higher risk of professional burnout from the amount of time spent on clerical tasks [23]. From our oncology provider network, we found that medical oncologists are involved in nearly 960 message threads, with one provider involved in 1942 threads over our study period. In the early stage breast cancer treatment, such as the population that we studied, the medical oncologist is tasked to coordinate care between specialists. We hypothesize that providers involved in large amounts of messaging could suffer from collaborative overload, which limits performance and productivity [24].

Our approach to analyze provider communication is not without limitations. Our study compares two common clinical tasks supported by the EHR: patient appointments and provider messaging. We do not account for other forms of direct communication, such as tumor boards, in-person discussions, pages, phone calls, or email messages. However, at VUMC, EHR-based secure communications are the preferred method of provider-provider communication and are heavily utilized within our institution. Our method also does not account for methods of indirect communication between providers, such as viewing clinical documentation or sending encounter notes between providers. We could also analyze other EHR artifacts, such as the transaction logs of orders and clinical notes, to understand providers who are communicating passively.

Table 3 – Appointment and Messaging Network Statistics for Physicians Involved Directly in Breast Cancer Treatment.

	Medical Oncology	Surgical Oncology	Radiation Oncology	Plastic Surgery
Number of Providers	8	9	3	5
Appointment				
Number of Appointments	21597	12870	5451	9525
Number of Provider-Provider Edges	4263	3648	1047	1412
Mean Specialty Node Size (range)	444.5 (140, 957)	327.9 (47, 575)	274.25 (25, 904)	216.6 (76, 553)
Mean Edge Weight (range)	8.1 (1, 686)	7.6 (1, 447)	8.8 (1, 622)	7.7 (1, 409)
Mean Degree per Node (range)	480.5 (310, 731)	401.1 (120, 628)	229.8 (32, 636)	264.4 (143, 550)
Edges with Cancer Provider (%)	285 (6.7)	246 (6.7)	131 (12.5)	158 (11.2)
Edges with Non-Cancer Provider (%)	3978 (93.3)	3402 (93.3)	916 (87.5)	1254 (88.8)
Edge Weight with Cancer Provider (%)	15742 (45.4)	13396 (48.0)	4808 (52.0)	5774 (53.4)
Edge Weight with Non-Cancer Provider (%)	18944 (54.6)	14500 (52.0)	4433 (48.0)	5048 (46.6)
Provider-Provider Messaging				
Number of Messaging Threads	5818	6314	573	805
Number of Provider-Provider Edges	1047	594	91	184
Mean Specialty Node Size (range)	959.5 (468, 1942)	1064.2 (243, 2668)	205.3 (14, 584)	180.6 (35, 386)
Mean Edge Weight (range)	6.5 (1, 419)	12.4 (1, 1748)	7.5 (1, 80)	5.0 (1, 54)
Mean Degree per Node (range)	134.3 (82, 217)	69.3 (37, 105)	31 (6, 78)	38.4 (22, 55)
Edges with Cancer Provider (%)	138 (13.2)	144 (24.2)	29 (31.9)	78 (42.4)
Edges with Non-Cancer Provider (%)	909 (86.8)	450 (75.8)	62 (68.1)	106 (57.6)
Edge Weight with Cancer Provider (%)	4426 (64.6)	6146 (83.2)	516 (76.1)	658 (71.3)
Edge Weight with Non-Cancer Provider (%)	2430 (35.4)	1242 (16.8)	162 (23.9)	265 (28.7)
Percent overlap among all edges	18.5	11.0	7.3	11.5
Percent overlap among edges with another cancer provider	48.4	54.9	21.4	48.7
Percent overlap among edges with another non-cancer provider	15.8	9.5	4.5	5.5

Understanding communication patterns between providers affords the opportunity to identify clinicians who work together closely, such that we can begin to identify clinic models for long-term chronic disease treatment and follow-up. We found that provider-provider pairs who were involved in messaging shared more patients than provider pairs who were not involved in messaging. These results suggest that providers who share multiple patients work together more closely than providers who share fewer patients, as hypothesized in other studies [19]. We also found that nearly 35% of message threads were within one day of an appointment. Appointment providers were involved in secure messaging within one day before or after nearly 10% of appointments. Similarly, there were an average of 7.9 messages sent per appointment. Work from Reid and colleagues found that implementing a co-located medical home clinic model provided more time to coordinate care between specialists and reduced provider burnout [25]. We hypothesize that providers who communicate closely may benefit from a co-located medical home clinic model by supporting inter-personal communication between specialists who collaborate frequently.

Our results indicate that only 1.4% (25) of the providers in our network were cancer specialists, suggesting that patients receive care during their cancer treatments from a breadth of providers. Unsuprisingly for our patient population, 50.9% of the messages between appointment providers involved these cancer providers. Results from our oncology specialist messaging network indicate that medical oncologists and surgical oncologists had the largest degree, or number of provider collaborations, with 134.3 and 69.3 collaborations per provider, respectively. This finding reflects the importance of these two specialties in coordinating early-stage breast cancer care. Surgical oncologists had a greater percentage of edge weights with other cancer providers and the largest overlap among edges with other cancer providers, compared to the medical oncologist. The medical oncologists had a greater number of provider-provider communications with non-cancer providers than any other cancer-related specialty. Similarly, medical oncologists had a 18.5% overlap between communications and shared patients with other providers. These results suggest that the medical oncologist is highly important in coordinating care between specialists, while the surgical oncologist is integral in coordinating care between cancer providers during cancer treatments.

Conclusions

Cancer treatment is complex, requiring multiple modalities delivered by many care providers. Initiatives have been implemented to improve coordination between care providers, but there does not yet exist a quantitative approach with which to evaluate care coordination. Current studies to understand coordination between providers have utilized payor claims and appointment data to suggest potential collaboration by identifying shared patients between providers. We employed a social network analysis approach to compare provider-provider relationships through shared patients and provider-provider communication about shared patients. We found that approximately 10% of relationships through shared patient were also involved in secure messaging. We also found that medical oncologists had the largest overlap between networks across all specialties, suggesting that medical oncologists are key to coordinating care across all providers associated with a patient. Applying social network analysis to EHR secure messaging data allows us to identify highly collaborative provider-provider pairs that can be used to further evaluate care coordination and drive improvements to healthcare delivery.

Acknowledgements

Bryan Steitz was supported by the 4T15LM007450 training grant from the United States National Library of Medicine.

References

[1] Institute of Medicine (US) Committee on Quality of Health Care in America, Crossing the Quality Chasm: A New Health System for the 21st Century, (2001), 1–360. doi:10.17226/10027.

[2] J.L. Ruth, J.M. Geskey, M.L. Shaffer, H.P. Bramley, I.M. Paul, Evaluating Communication Between Pediatric Primary Care Physicians and Hospitalists, *Clin Pediatr (Phila)*. 50 (2011), 923–928. doi:10.1177/0009922811407179.

[3]D. Peikes, A. Chen, J. Schore, R. Brown, Effects of Care Coordination on Hospitalization, Quality of Care, and Health Care Expenditures Among Medicare Beneficiaries, *JAMA*. 301 (2009), 603–16. doi:10.1001/jama.2009.126.

[4] C.E. Pollack, G.E. Weissman, K.W. Lemke, P.S. Hussey, J.P. Weiner, Patient Sharing Among Physicians and Costs of Care: A Network Analytic Approach to Care Coordination Using Claims Data, *J Gen Intern Med*. 28 (2012), 459–465. doi:10.1007/s11606-012-2104-7.

[5] C. McGUINESS, Development and initial validation of a measure of coordination of health care, *International Journal for Quality in Health Care*. **15** (2003), 309–318. doi:10.1093/intqhc/mzg043.

[6] World Health Organization Regional Office for Europe, Assessing chronic disease management in European health systems, (2015), 1–111.

[7] J. Riley, D. Madill, Coordinate My Care: A clinical approach underpinned by an electronic solution, *Progress in Palliative Care*. **21** (2013), 214–219. doi:10.1179/1743291X13Y.0000000060.

[8] M. Olejaz, A. Juul Nielsen, A. Rudkjøbing, H. Okkels Birk, A. Krasnik, C. Hernández-Quevedo, Denmark health system review, *Health Syst Transit*. **14** (2012), i–xxii– 1–192.

[9] M. Berg, G. Bowker, The Multiple Bodies of the Medical Record: Toward a Sociology of an Artifact, *Sociological Quarterly*. 38 (2016), 513–537. doi:10.1111/j.1533-8525.1997.tb00490.x.

[10] W. Pratt, M.C. Reddy, D.W. McDonald, P. Tarczy-Hornoch, J.H. Gennari, Incorporating ideas from computer-supported cooperative work, *Journal of Biomedical Informatics*. 37 (2004), 128–137. doi:10.1016/j.jbi.2004.04.001.

[11] C.L. Goldzweig, A.A. Towfigh, N.M. Paige, G. Orshansky, D.A. Haggstrom, J.M. Beroes, I. Miake-Lye, P.G. Shekelle, Systematic Review: Secure Messaging Between Providers and Patients, and Patients' Access to Their Own Medical Record: Evidence on Health Outcomes, Satisfaction, Efficiency and Attitudes, (2012), 1–63.

[12] J. Walker, E. Pan, D. Johnston, J. Adler-Milstein, D.W. Bates, B. Middleton, The Value Of Health Care Information Exchange And Interoperability, *Health Affairs*. 24 (2005), W5–10–W5–18. doi:10.1377/hlthaff.W5.10.

[13] Y. Chen, M.B. Patel, C.D. McNaughton, B.A. Malin, Interaction patterns of trauma providers are associated with length of stay, *J Am Med Assoc*. 25 (2018), 790–799. doi:10.1093/jamia/ocy009.

[14] B.D. Steitz, M.A. Levy, Temporal and Atemporal Provider Network Analysis in a Breast Cancer Cohort from an Academic Medical Center (USA), *Informatics*. 5 (2018). doi:10.3390/informatics5030034.

[15] Tennessee Cancer Registry [Internet]. [cited 2018 Nov]. Available from: https://www.tn.gov/health/health-program-areas/tcr.html (accessed October 28, 2018)

[16] G. Csardi, T. Nepusz, The igraph software package for complex network research, *InterJournal*. Complex Systems (2006), 1695. http://igraph.org (accessed November 3, 2018).

[17] R.C.T., R: A language and environment for statistical computing, (n.d.). https://www.R-project.org/ (accessed November 1, 2018).

[18] B.D. Steitz, M.A. Levy, A Social Network Analysis of Cancer Provider Collaboration, *AMIA Annu Symp Proc*. 2016 (2016), 1987–1996.

[19] T. Hussain, H.Y. Chang, C.M. Veenstra, C.E. Pollack, Collaboration Between Surgeons and Medical Oncologists and Outcomes for Patients With Stage III Colon Cancer, *Journal of Oncology Practice*. 11 (2015), e388–e397. doi:10.1200/JOP.2014.003293.

[20] B.E. Landon, J.-P. Onnela, N.L. Keating, M.L. Barnett, S. Paul, A.J. O'Malley, T. Keegan, N.A. Christakis, Using Administrative Data to Identify Naturally Occurring Networks of Physicians, *Medical Care*. 51 (2013), 715–721. doi:10.1097/MLR.0b013e3182977991.

[21] B.E. Landon, N.L. Keating, M.L. Barnett, J.-P. Onnela, S. Paul, A.J. O'Malley, T. Keegan, N.A. Christakis, Variation in Patient-Sharing Networks of Physicians Across the United States, *Jama*. 308 (2012), 1–24. doi:10.1001/jama.2012.7615.

[22] D.A. Chase, J.S. Ash, D.J. Cohen, J. Hall, G.M. Olson, D.A. Dorr, The EHR's roles in collaboration between providers: A qualitative study, *AMIA Annu Symp Proc*. 2014 (2014), 1718–1727.

[23] T.D. Shanafelt, L.N. Dyrbye, C. Sinsky, O. Hasan, D. Satele, J. Sloan, C.P. West, Relationship Between Clerical Burden and Characteristics of the Electronic Environment With Physician Burnout and Professional Satisfaction, *Mayo Clinic Proceedings*. **91** (2016), 836–848. doi:10.1016/j.mayocp.2016.05.007.

[24] R. Cross, R. Rebele, A. Grant, Collaborative Overload, *Harvard Business Review*. (2016). https://hbr.org/2016/01/collaborative-overload (accessed November 8, 2018).

[25] R.J. Reid, K. Coleman, E.A. Johnson, P.A. Fishman, C. Hsu, M.P. Soman, C.E. Trescott, M. Erikson, E.B. Larson, The Group Health Medical Home At Year Two: Cost Savings, Higher Patient Satisfaction, And Less Burnout For Providers, *Health Affairs*. **29** (2010), 835–843. doi:10.1377/hlthaff.2010.0158.

Address for correspondence

Bryan D Steitz
2525 West End Ave, Suite 1475.
Nashville, TN 37203, USA
Tel: (502) 320-6797
Email: bryan.steitz@vanderbilt.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190336

SERENE-IoT Project: How the Maturation Cycle Allows the Correct Development Process of Innovative Technologies in the Healthcare Domain

Samy Andrea Strola[a], Armand Castillejo[b], Alexandre Moreau-Gaudry[a]

[a] Université Grenoble Alpes, INSERM, CHU Grenoble Alpes, CIC 1406 (Innovative Technology), F-38000 Grenoble - France
[b] ST Microelectronics, 12 rue Jules Horowitz, B.P. 217, 38019 Cedex, Grenoble - France

Abstract

This article illustrates the maturation process of innovative technologies in the healthcare domain.

The role and the involvement of the stakeholders are explained, as well their interaction in the ecosystem. We focus on how the synergy between partners improves efficiency, boosts product development and accelerates market access.

The spiral of the innovation concept is introduced to illustrate the different stages ensuring the correct development of a medical technology. This iterative cycle drives the product maturation, according to medical needs and feedback, and ensures the correct implementation of the innovation in the clinical framework.

We finally illustrate how the innovation process is applied to successful drive the ongoing SERENE-IoT project: the spiral of innovation concept is specifically adapted to fit with the project requirements with the final ibjective to identify a panel of medico-economic indicators and to establish the medical service delivered and the opportunities in order to capture the full benefit of the proposed technology.

Keywords:
Delivery of Health Care, Medical Device

Introduction: The Innovation in Healthcare

The Clinical Investigation Center - Technological Innovation (CIC-IT) structures play a crucial and strategic role in the innovation process. The CIC-IT structures were established to accelerate the transfer of technological innovations in the healthcare domain to clinical practice: prognostic, diagnostic and therapeutic. One particular objective is to facilitate the translation from fundamental research to the industrial product [1]. To accomplish this mission, a CIC-IT must support the "Healthcare Innovation Technology" (HIT) as it matures [2, 3]. Thus, a CIC-IT will be able to support this innovation, either intermittently during one of the HIT maturation phases or in a more "exhaustive" way (i.e. accompanying the process of innovation maturation, from the genesis until the clinical evaluation).

A CIC-IT is able to participate in the evaluation of:

- The performance and safety of the proposed innovation,
- The balance between benefits and risks for the claimed added value,
- The usability prior to placing the HIT on the market,
- The expected medical service improvement.

However, in order to appreciate the potential contributions of a CIC-IT in the HIT maturation process, it is first necessary to describe this process and then identify the potential contributions of CIC-IT within it.

A HIT can be categorized in the following way according to [4]:

- "Breakthrough" innovation addresses a medical problem by "disrupting" the current state of the art, introducing ruptures in the area of usages and involving strong modifications regarding the indication or the use of a Medical Device (MD);
- "Integration" innovation integrates new technological/scientific concepts to upgrade a MD, in response to a clinical need;
- "Incremental" innovation is characterized by a significant improvement of a MD, without disrupting the conditions of use in the clinical practice.

It also happens that a technological innovation, which was not created specifically for a medical need, can be used in the medical environment and can have strong benefits in the area of the healthcare.

Regardless of the category, a schematic maturation process for innovation in healthcare technologies has to be clear in order to be helpful for its use in different contexts. The actors and the technical, scientific and medical aspects involved in this process, as well the regulatory, economical and industrial frames will be discussed in the next sections of this article.

Stakeholders in the Healthcare Ecosystem

Because the healthcare ecosystem has a particularly high number of stakeholders linked with complex interactions that vary depending on the countries, culture and local regulation, bringing innovation into this ecosystem is complex. Innovation in healthcare can impact all the stakeholders at various levels in the very complex ecosystem of health [5, 6]. Ideally all the stakeholders should take into account the many factors involved from the genesis phase and all along the maturation process [7]. We propose a generic view of the healthcare value chain (summarized in Table 1) distinguishing five key player roles.

The Patient

The patient (or user associations, or target population) is the first player concerned by innovation in healthcare. The innovation should bring a higher quality of health service and should improve the benefit/risk ratio of the diagnostic or therapeutic act involved. The patient is obviously central to the maturation process, both in the upstream phases of innovation

(identification of a specific need) and in the downstream phases (during clinical evaluations in order to evaluate the performance and the security of the MD and the delivered medical service). As an example, patient associations can contribute by identifying needs uncovered by routine practice, by providing expertise on the reality of needs and on the constraints of use.

The Healthcare Professional

Healthcare professionals are the professional stakeholders in primary contact with the patients. They are responsible for good patient care. All medical and paramedical specialties are concerned. All the actors in this category can contribute to the design and maturation of the innovation in their domain of competence, ideally from the earlier phases. Healthcare professionals are at the root of meeting medical needs: they provide the first answers to drive the future steps of maturation. The early involvement of medical experts along the medical maturation process will help to determine the relevance and the early adoption of the proposed innovation for its future implementation in the clinical framework.

Other Healthcare Actors

Other healthcare actors can be also identified:

- Public Health Authorities are responsible for the evaluation of the healthcare benefit, the definition of the policies for reimbursement of the acts, the definition of the management and cost coverage of health products;
- Other competent authorities are in charge of monitoring the safety of MD;
- The local authorities and the healthcare insurance organizations.

These stakeholder groups lead to the identification of priorities for the health of the population at the national level and can be involved in the regulations, laws and authorization for clinical investigations. In fact, the expertise of the High Authority of Health may lead to the reimbursement of the proposed innovation and then rapidly accelerate its deployment.

The Researcher (or Innovative Technology Provider)

The researcher, working in a public institution or in a private company, designs, develops, validates and implements the technological-scientific innovative solution. Ideally, this solution responds optimally to a previously identified medical need. In the case of pure technological-scientific "breakthroughs", the proposed innovative solution can give rise to "clustered" innovations, made from and inspired by the first idea.

The Industrial Manufacturer

At the end of the successful maturation process, the industrial manufacturers will intend to commercialize the product resulting from the innovation. More generally, this stakeholder group represents all the economic actors that will contribute to the relevance of the innovation in the market.

The industrial manufacturer is responsible for the quality of the innovation, from the prototype to the certified market product. This actor is the only one in charge to provide and distribute the MD in compliance with the quality assurance policy.

The success of these actors is the result of early, close and well-coordinated collaboration.

Table 1 – Stakeholder Analysis for Innovation in the Healthcare Ecosystem.

Stakeholder	Role	Benefits	Costs
Patient	Clinical need, Purchaser	Improvement of life quality	None
Healthcare professional	Clinical feedback, Purchaser	Improvement of practice	User learning
Researcher	Innovator	Intellectual property, dissemination	Development
Manufacturer	Product, Market access	Turnover	Regulatory, Quality

The Spiral of the Innovation

The cycle of innovation represents the stages of the development of an innovative MD and identifies the collaboration and interactions among the stakeholders [8]. Different actions have to be anticipated from the idea (whose intellectual and industrial property has to be protected) until it achieves market penetration (which has to be planned, ideally from the beginning) and beyond. These actions will accelerate the availability of innovation responding to the unmet clinical need and can be described as a "spiral of innovation" cycle, as summarized in Table 2 [9, 10, 11, 12].

Table 2 – Stages of the MD Innovation Cycle.

Stage	Key phases
Concept and Design	Medical challenge, Needs, Methodology, Benefits/Risks
Tests on pre-product	Users feedback, Usability, Maturation, Market study, Intellectual property analysis
Clinical prototype	Compatibility with clinical environment, Risk analysis
Product and Marketing	Clinical trials, European Conformity (CE) mark, Regulatory frame, Usability, Market access and product follow-up, Indicators

Rigorous methodologies, specific to clinical trials of this domain, will be used to obtain the objective evidence, adapted to the degree of maturation of the innovation.

Thus, the development of an innovative MD, compatible with first clinical uses, will enable the achievement of first evaluations in the current clinical practice. First evaluations may be undertaken within the legislative framework whatever the class of the MD, CE marked or not. This first biomedical research will determine the feasibility of the use of the innovative MD for the patients.

A real process of maturation of the innovation will follow thanks to the validation within specific biomedical research, in order to obtain the evidence of the maturation degree of the innovation. In the case of an unmarked CE product, this research may contribute to demonstrating a favorable risk/benefit ratio in view of the CE marking application. Specific biomedical research will provide additional evidence that will impact future clinical studies in order to demonstrate the benefit. The implementation of biomedical research at a larger scale will generate the objective evidence (or not) of the

benefit for a target population, from which the public health benefit of the innovation can be estimated.

SERENE-IoT (Secured & EneRgy EfficieNt healthcarE solutions using IoT technologies): A Successful Example of Innovation Process

Context of the project

Since the mid-2000s, connected devices have impressively increased their place in the market of Internet of Things (IoT) thanks to the widespread use of smartphones and numerous dedicated consumer applications. In particular, a new branch appeared in healthcare: the Internet of Medical Things (IoMT). Yole Development estimates that there are more than 45M IoMT devices today and that there will be more than 235M in the market in 2020 [13].

This digital revolution is rapidly changing healthcare systems towards a new concept focused on smart health. The convergence of healthcare and high-tech mass-market ecosystems is coming to a point where things and people are increasingly connected, as with the IoT, and where healthcare is partly divided between hospital, private structure and home. This opens the gates to new technological challenges in order to ensure solutions in the domains of energy efficiency, data collection/transmission/storage security, connectivity and architecture and device management and configuration. In particular, these solutions have to be applied facing new European regulations (2016/679 – 2017/745 – 2017/746) which drive the development of MD (software included) and data privacy.

The objective is to move from the existent smart watch model to a new smart health system. Benefits are multiple: the patient experience and follow-up are improved because of remote health personalized data monitoring that is able to detect abnormalities faster to alert healthcare professionals, device maintenance needs are optimized via remote configurations, healthcare professionals can remotely monitor patient health and treatment compliance and device performance far from their offices, patient families can help in the follow-up of patient wellness and healthcare systems reduce human errors and falsifications.

The long term goal is to provide a global Big Smart Data platform to aggregate and correlate large volumes of data from disparate MD with health outcomes to deliver advanced insight on disease prevention and prediction.

Objectives of the Project

SERENE-IoT will consider the beneficial impacts for the patient with an increased quality of life and for the health system with an expected cost saving, while also increasing access to care for the population. SERENE-IoT addresses the needs of patients remotely followed by professional caregivers by developing advanced smart e-health IoT devices and architecture in Europe.

Following the definition proposed in the Yole report [13], SERENE-IoT has the ambition to cover 4 market segments: implantable devices, self-monitored devices, professionally oriented devices and assistance devices for people lacking autonomy.

The core values of the project are:

- High quality of healthcare services;
- High level of trust (security, safety, privacy, robustness);
- Efficient execution of required operations and tasks;
- Interoperable and compatible information technology systems;
- Reduced costs compared to current traditional care.

In this context, SERENE-IoT focuses on 3 medical challenge domains:

- Domain 1: Remote Healthcare by moving care services from hospital to home;
- Domain 2: Early detection of Methicillin-Resistant bacteria;
- Domain 3: Fall Prevention.

In line with the medical innovation cycle until the step “Clinical Prototype”, the goal is to develop 3 clinical prototypes of new MD supporting security, safety and privacy and their complete validation inside a secured end-to-end IoT system platform. With this, the project will contribute to the standardization of a healthcare data structure to ensure MD and system interoperability.

Therefore, SERENE-IoT develops the following connected clinical prototypes (see Figure 1):

- A low-power medical IoT module validated with two MD (class IIx) providing homebased remote healthcare service;
- An early detection system (first low-power mobile detector) of Methicillin-Resistant Staphylococcus aureus bacteria;
- A fully wireless insole for fall detection and gait monitoring.

Figure 1 – Clinical Prototypes of the SERENE-IoT Project.

In the 3.5 year timeframe of the project, the goal is to clinically validate these devices (with mono and multicentric studies) in the operative environment under the coordination and supervision of the Grenoble-Alpes University Hospital (CHUGA) and through the expertise and the “Medical Innovation Cycle” concept developed by the CIC-IT of Grenoble.

The three demonstrators will be used to provide the necessary validation of advanced concepts needed by European industry for the development and manufacture of products and services in remote medical-care. Importantly, SERENE-IoT will contribute to the evaluation of a secure, end-to-end, IoT system platform in “real-life” scenarios (including the use of the proposed healthcare data structure), while demonstrating the resulting benefits. Certification and industrialization phases will follow the SERENE-IoT project.

Partnership (Consortium)

The SERENE-IoT consortium (over three countries: France, Germany, Spain) aims to provide a balanced and holistic view leveraging different types of partners (large firms, SMEs and academics) covering the overall Health Innovation Technology ecosystem (from IDM/OEM to end-user service provider).

Looking at Figure 2 below it is evident that none of the involved partners can cover the whole value chain and be successful on its own. Only a close collaboration of the different stakeholders will enable technological innovations, and then products, that support a real and qualified medical service validated by healthcare professionals to be achieved at the end. This is the key reason to assemble the SERENE-IoT partners in a single project sharing a common goal. SERENE-IoT should be considered as a business "seed" for the demonstrators, who will establish all the conditions to contin-ue later on though medical innovation up to the introduction of a product on the market in maximizing its chance of success.

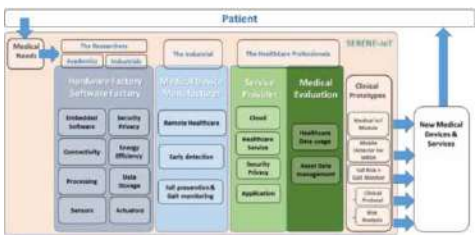


Figure 2 – Only a Close Collaboration Among the Various Stakeholders can Successfully Propose Technological Innovations.

Role of STMicroelectronics

STMicroelectronics is a global leader in semiconductors with the expertise and resources to propose disruptive microelectronics-based platform solutions to make a positive contribution to lives of people. As a large firm able to provide solutions to MD manufacturers, STMicroelectronics is in the strong position to assume the European coordination of SERENE-IoT across the healthcare value chain. From this vantage point, STMicroelectronics can push disruptive technologies across the value chain to consolidate learnings to the next evolution of its portfolio.

Project Flowchart and Short Description of the Tasks

Figure 3 below shows the different model steps, starting from a medical problem expressed by health experts (H), industry (I) or research (R) entities always around the patient (P) and resulting in real needs, through placing the "Product" on the market.

Based on our experience, the time scale for the complete cycle can be around 10 years while the time from the "medical problem + ideas" to "Clinical Prototype" is 3 to 4 years. Note that the blue/green area domain in Figure 3 includes the model steps ("Product concept", "Technical Prototype" and "Clinical Prototype") devoted to developing the technology needed to solve the identified medical problem. The remaining steps ("Pre-product" and "Product") are devoted to clinical intensive evaluation/test and maturation to bring "Clinical Prototype" to a real and certified product in accordance with the protocol defined, having its certification and having followed the required industrialization phases. These two last steps are the unique responsibility of the industrial manufacturers developing the product and, because of the difficulty of performing this in a collaborative way, is not part of the SERENE-IoT project.

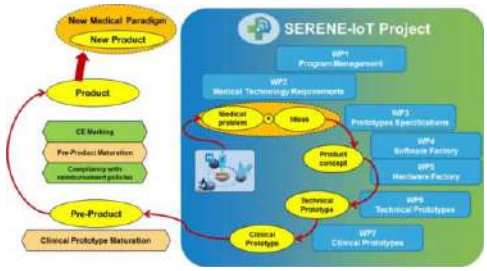


*Figure 3 – The **S**piral of **I**nnovation (resuming the main stages of innovation diffusion) **S**pecifically **A**dapted for the SERENE-IoT **P**roject.*

Role of CIC-IT Grenoble

The role of the CIC-IT Grenoble in SERENE-IoT is to support and drive the project within the innovation cycle, with the aim to clinically validate the proposed devices at the CHUGA.

The CIC-IT Grenoble has been involved in the early stages of the innovation cycle in order to share expertise in the clinical domain with the consortium, to accelerate the complete understanding of the medical need and to facilitate the definition of the functional and operational specifications of the proposed HIT via iterative feedbacks with the clinicians. This collaboration should contribute to shortening the HIT maturation process, from idea to demonstration of the benefit for the target population.

The missions of the CIC-IT Grenoble in the SERENE-IoT project are:

- To define user cases identifying the scenario for future implementation of the clinical prototypes in the hospital framework, in order to anticipate the possible drawbacks and to target the opportunities;
- To work on the usability of the clinical prototypes, either in support of the parallel technological developments (for example, to define the software / hardware specifications as requested by the clinician) and to collect information on the device usage for the further optimization based on user feedback (i.e. patients, clinicians, medical and technical services);
- To establish the clinical studies (i.e. writing of clinical protocols) to evaluate and validate the implementation of the clinical prototypes in a real environment.

The outline of these missions for the CIC-IT Grenoble will be:

- To demonstrate the feasibility of acquiring the relevant information through the IoMT devices developed in the SERENE-IoT project;
- To understand the possibilities for future standardization of the proposed end-to-end chain;
- To study the problem of deploying and integrating such architecture in a complex infrastructure;
- To ensure the safety and the integrity of the collection, transfer and access of the patient data along the IoMT end-to-end chain in compliance with the current regulations, in particular with the General Data Protection Regulation (GDPR);

- To estimate the medical service delivered with the objective to optimize the care of patients, the work of caregivers and to identify medico-economic indicators;
- To use the "big data" generated in the study in the development of diagnostic, prognostic and therapeutic predictive models.

Conclusions

In conclusion, innovation in the healthcare domain has an important multidimensional facet that involves the synergy of stakeholders at technical and scientific, juridical and regulatory, industrial and economic levels.

The innovation, from the idea to the final product, involves different phases that can be resumed schematically in an iterative process. This innovation cycle has multiple consequences in terms of relationships between the consortium partners. With respect to a dynamic and efficient behavior, a proficient work emerges that ensures a beneficial agreement that gives sufficient incentives to market actors in order to constantly improve products and services in the healthcare system.

In this paper, we demonstrated that the "the spiral of the innovation" concept can drive effectively a HIT. In particular, the maturation cycle of the SERENE-IoT project currently benefits from the contribution of the CIC-IT of Grenoble. First, the proposed MD will be developed faster than the current state-of-art: the objective is to finalize the first clinical validation for the proposed prototypes in 3.5 years, the time of the project, compared to the 8-10 years, as reported in other projects. Second, the active involvement of the CIC-IT (via its support concerning regulation knowledge, its contribution in terms of clinical use cases, needs and feedback and the work on usability) allows the partners to improve the technological developments. Finally, the clinical studies will evaluate and validate the implementation of the clinical prototypes in the clinical frameworks.

In the long run, this methodology will decrease the costs of innovation and increase the potential synergies of partners in healthcare domains. We consider that this represents a collective challenge and an opportunity that all players in healthcare must embrace in order to achieve a faster access to the latest innovative technologies providing a better and cost-effective benefit for the population.

Acknowledgements

SERENE-IoT project is a project labelled within the framework of PENTA, the EUREKA cluster for Application and Technology Research in Europe on NanoElectronics.

The authors gratefully acknowledge Dr. Cecilia Hughes for her careful proofreading of the manuscript.

References

[1] J. Rousseau. *Éditorial. IRBM*, Volume 7, numéro 4, page 149 (2006).

[2] L.A. Marsch and D.H. Gustafson. The role of technology in health care innovation: a commentary. *J. Dual. Diagn.*, 9(1) : 101-103 (2013).

[3] S. Poonen, Health Care Innovation: Harnessing New Technologies to Benefit Patients, (2018). https://www.forbes.com/sites/forbestechcouncil/2018/04/02/health-care-innovation-harnessing-new-technologies-to-benefit-patients/#318bf39c5a88 (accessed July 12, 2018)

[4] O. Ciani, P. Armeni, P.R. Boscolo , M. Cavazza, C. Jommi, R. Tarricone De innovation: the concept of innovation for medical technologies and its implications for healthcare policy-making. *Health Policy and Technology*; Vol. 5, Issue 1, 47-64 (2016).

[5] Institute of Medicine (US) Committee on the Public Health Effectiveness of the FDA 510(k) Clearance Process; Wizemann T, editor. Balancing Patient Safety and Innovation: Workshop Report. Washington (DC): *National Academies Press (US)*; 2010. 4, The Medical Device Industry Innovation Ecosystem. Available from:https://www.ncbi.nlm.nih.gov/books/NBK209786/

[6] D. M. Beswick; A. Kaushik; D. Beinart; S. McGarry; M. Khoon Yew; B. F. Kennedy; P. L. Santa Maria; Biomedical device innovation methodology: applications in biophotonics; *J. of Biomedical Optics*, 23(2), 021102 (2017).

[7] Shah SG, Robinson I, AlShawi S. Developing medical device technologies from users' perspectives: a theoretical framework for involving users in the development process. *Int JTechnol Assess Health Care*; 25(4): 514–21 (2009).

[8] Barriers to innovation in the field of medical devices. In: Medical Devices: Managing the Mismatch (An outcome on the Priority Medical Devices project); *Background Paper 6*, World Health Organization (2010).

[9] A. Moreau-Gaudry, L. Pazart, Development of innovative technologies in health: The CREPS cycle, Concept – Research – Experiment – Product – Safe Care ; *IRBM* 31, 12-21 (2010).

[10] B. D. Smith, R. Tarricone, V. Vella. *Journal of Medical Marketing*. Vol. 13, Issue 1, 37-43 (2013).

[11] P.Yocket et al., Biodesign: The Process of Innovating Medical Technologies, *Cambridge University Press*, Cambridge (2015).

[12] A. D.McCarthy et al., Unmet needs: relevance to medical technology innovation?; *J. Med. Eng. Technol.* 39 (7), 382–387 (2015).

[13] Connected medical devices market and business model 2017; *Yole development Market & Technology report* (2017).

Addresses for correspondence

Strola, Samy Andrea
Université Grenoble Alpes, INSERM, CHU Grenoble Alpes,
CIC 1406 (Innovative Technology),
F-38000 Grenoble - France
E-mail: sstrola@chu-grenoble.fr

Castillejo, Armand
ST Microelectronics
12 rue Jules Horowitz
B.P. 217, 38019 Cedex, Grenoble - France
E-mail: armand.castillejo@st.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190337

Modeling the Personas of Primary Care Communication Modality Usage: Experiences from the R-Health Direct Primary Care Model

Si Sun[a], Pei-Yun Sabrina HsueH[a], Sasha Ballen[b], Marion Ball[a,c]

[a]IBM T.J. Watson Research Center, Yorktown Heights, NY, USA,
[b]R-Health Inc, Elkins Park, PA, USA,
[c]Johns Hopkins University, Baltimore, MD, USA

Abstract

Recent advancements in mediated communication technologies enable unfettered communication between patients and their providers. We explore patients' utilization of different direct and mediated communication modalities in a Direct Primary Care (DPC) clinic and the patient characteristics that predict their communication behavior patterns. Based on this knowledge, we developed 2 patient personas that explicate the nuances of patients who tend to prefer visiting the clinic in person versus those who use mediated modalities more often. We hope this study may inform future work in understanding and supporting patient-provider communication in a new technical environment. The results suggest that patients and their health team alike may be incentivized to voluntarily adopt and utilize multi-modality communication in a DPC setting.

Keywords:

Patient-Specific Computational Modeling, Patient Engagement, Primary Health Care.

Introduction

Patients living with chronic conditions ideally are expected to take on more responsibilities of their own care due to the limited resources (e.g., physician time) in traditional clinical visits [1]. Evidence has shown, however, that the long-term and nuanced nature of chronic conditions makes it more effective for physicians to involve in patients' self-management actively [2; 3]. The direct primary care (DPC) model offers to address this conflict by allowing patients to have unfettered access, either in person or through mediated communication channels, to their physicians for everyday health decision-making tasks. It is hoped that the collaborative relationship between patients and their physicians facilitates the much-needed frequent shared decision-making and gives patients the tools and confidence to manage their health with the active support of their physicians.

The advancement in medicated communication technologies for health, such as mHealth and patient portals, enables patients' multimodality access to their physicians and medical records. Those technologies are especially relevant for patients with chronic conditions due to their need for long-term shared decision-making. A recent review investigated 47 tools for mediated communication between chronic patients and their health care team [4]. The authors noted great diversity in functionalities such as direct text-based communication between patients and care teams, the reporting of patient-generated data, and the delivery of behavior-related interventions [4]. Another recent study suggests that patient portal use is associated with improved glycemic control in patients with diabetes mellitus [5]. It is therefore important to explore factors associated with the usage of mediated patient-provider communication technologies and identify potential opportunities to promote engagement.

With unprecedented freedom of choice for patient-provider communication, we are able to observe patients' behavior patterns across multiple communication modalities and learn their preferences of different modality for various situations. The extraction of patient characteristics for communication patterns will further allow us to better reach different patient segments in health care settings, such as care management [6], public health [7], and community wellness [8]. For example, in the care management setting, understanding the behavioral characteristics of patients may help care managers evaluate intervention effectiveness [6]. The increased patient understanding will further enable targeted communication for coordination of care among and between patient and care team, which has been identified as one of the key success factors of care delivery models for high-need patients [7]. It is also expected to help design health communication campaigns for population health management. For example, the Center for Disease Control in the U.S. provides Health Communication Works toolkit [8], which leverages a previously developed risk communication model to identify patient persons and to offer an evidence-based approach to audience-centered messaging.

Prior studies examining patient factors associated with mediated communication technology adoption found that diabetes patients who use mediated communication channels tend to be younger, male, with higher morbidity, and use more medications [9]. There is a paucity of studies on patients' communication patterns beyond technology adoption and the usage of different functionalities, such as the number of messages in each conversation and how frequently they send compared to receive messages. Less is known about patient characteristics other than demographics as well as access to and acceptance of technologies, such as their health trajectory over time and medication types.

In this paper, we answer the key research question: *What are the patient personas for communication modalities use in a direct primary care setting?* Communication modalities examined in this paper include office visits, secure messaging, video chats, EHR portals, and phone calls. Patient personas are collections of patient characteristics that describe a typical patient and their communication modality use behaviors. A simple example is a patient who goes to the clinic in person once a year and talks to her doctor through secure messaging infrequently may be a relatively younger, healthy person who lives further away from the clinic. We develop such personas for the purpose of clarifying the starting point for technology, intervention, and policy design.

Methods

Research Site

The research site of this study is a R-Health clinic based in New Jersey. The R-Health Direct Primary Care (DPC) model is a value-based model of care focused on delivering personal, convenient, and effective primary care to its members for a regular monthly fee. Providers are free from the typical incentives of volume-based fee-for-service primary care, and thus able to engage patients across multiple communication modalities. Members can use secure messaging, video chats, electronic health record (EHR) portals, traditional phone calls and office visits to communicate with their health care team. Since they are not required to ensure that interactions happen under a billable event, providers are free to engage with the patients in a manner that is most effective. Therefore, patient-provider communication behaviors are mostly organically developed through long-term negotiation between patients and their healthcare team. By the end of October 2018, a total of 2427 patients have visited the R-Health clinic and 2487 patients have communicated with their health care team through at least one mediated communication modalities.

Data Set

The data analyzed in this study is a comprehensive data set across all communication modalities available at R-Health. For in-person visits, the data set includes information such as date and time of visit, reason for visit (e.g., new patient visit, follow-up visit, emergency visit), status of visit (i.e., canceled or not). For mediated communication, the data set specifies the communication medium, messages nested in conversations for text-based communication, and the type of conversation (e.g., support and standard). The data set also includes demographics, lab tests, vitals, medication forms and intake routes, medication intolerance statuses, medication fulfillment routes, allergy statuses, and Hierarchical Condition Categories (HCC) [10].

To assure meaningful data analysis, we processed the data set at the patient level by re-organizing multiple records per patient (e.g., records on mediated communication and vitals) into frequencies (for cases of categorical entries) or distribution (for cases of continuous entries). To avoid producing results due to recording inconsistency, we also constructed binary variables for those that reflect frequencies (e.g., whether they have allergies). All variables are then standardized to have a mean of 0 and a standard deviation (SD) of 2. This process improves the interpretability of results from clustering and regression by removing the effects of scales.

Missing data are handled in two different ways. For variables that describe behavior counts (e.g., total number of conversations a patient is involved in), where missing data represent no behavior, missing data are replaced with 0s. For variables that describe properties (e.g., age) and where 0s stand for a meaningful state (e.g., interquartile range (IQR), median absolute deviation (MAD)), missing data are replaced with the variable median. Missing data are handled after the standardization of the dataset to avoid influences on the standardization process.

The final dataset includes all 678 patients in the cohort and 224 variables, including 103 features (i.e., independent variables), and 121 responses (i.e., dependent variables). Due to the large numbers of features and responses analyzed in this paper, traditional statistical methods (e.g., multivariate linear regression) are not able to capture the full picture depicted by the rich data set, and we adopted more sophisticated machine learning methods in an attempt to paint a more complete picture.

Data Analysis

We conducted data analysis in three steps. First, we explored the variable space with cluster analysis by partitioning cases around medoids (PAM) [11]. PAM is a more robust version of the k-means approach because it "minimizes a sum of dissimilarities instead of a sum of squared Euclidian distances"[11]. Then, we excluded the variables that yield the same medoids, which are less differentiable of personas. Excluding those variables enables the building of more concise machine learning models and subsequently more distinct personas. After that, we conducted cluster analysis again and performed multi-task regression with clustered structure (MTR_CMTL) [12] on the remaining 75% training and 25% testing data. We also performed the same regression on meaningful patient segments to provide in-depth insights into the nuances of different patient groups.

Machine Learning Framework

The goal of this study is to extract patient personas associated with various communication behavioral patterns in terms of modalities (e.g., office visit, secure messaging) and behavioral outcomes (e.g., frequency, intensity). In addition, given a set of patient markers that include social determinants (e.g., distance to care), clinical characteristics (e.g., allergies, medication change) and health status based on vital signs (e.g., body mass index (BMI), blood pressure), the task of patient persona extraction is cast as classifying the communication behavioral patterns by simultaneously modeling all markers and their effects.

Since some of these communication patterns might share a similar structure of patient markers, we adopt a multi-task learning framework, in which a multitude of patient communication behavioral patterns are modeled together to improve the accuracy of the classification from modeling each of them alone. Our dataset contains patient markers associated with communication patterns. The change of markers over 6 and 12 months are coded as features to be incorporated into modeling. The classification of each communication pattern using all patient markers and marker changes is thus considered as a learning task.

The objective of the multi-task learning framework is thus to learn the shared structure among the multiple learning tasks while minimizing the classification errors of each learning task. The formulation in Equation (1) combines both the objective value of empirical risk $\ell(\mathrm{W})$ and $\Omega(\mathrm{W})$ from a convex form of k-means clustering, which places penalty to enforce a clustering of the weight vectors w_i's towards the mean of the cluster.

$$min_{\mathrm{W}\in\mathrm{R}^{\mathrm{d}\times\mathrm{m}}}\ \ell(\mathrm{W}) + \lambda\Omega(\mathrm{W}) \qquad (1)$$

This formulation enables us to balance the need for finding the shared structures among the multiple learning tasks and for identifying the regression models that fit to the data empirically. The set of m learning tasks learn linear regression models over $\mathrm{X} = \mathbb{R}^d$ from pairs of input features (i.e., patient makers) and each output variable (i.e., continuous response indicating the occurrences of communication pattern), $(x_i, y_i), \mathrm{i} = 1, \ldots, \mathrm{n}$, where $x_i \in \mathrm{X}$ and $y_i \in \mathrm{Y}$. X and Y are the sets of input features and continuous response output variables respectively, W is the set of weight vectors, and m is the number of tasks.

We then experiment with how modeling multiple communication patterns together could improve prediction accuracy. The models learned this way can be further evaluated

with the error measures, e.g., root mean squared error (RMSE), mean average percentage error (MAPE). We also qualitatively assess the weights over the multiple learning tasks to better understand the reasonableness, the simplicity, and the usefulness of the model.

Results

Cohort

The cohort involved in this study are patients between 19 and 65 years of age (mean = 49.1, SD = 10.6) and with diagnosed obesity, hypertension, or high cholesterol. The cohort includes 421 (62.1%) females and 257 (37.9%) males. The diagnoses are reflected in the ICD 10 codes available in the EHR, E66* (including .01, .09, .3, .8, .9), E78* (including .00, .01, .1, .2, .3, .5, .6, .9), R03, I10, I11.9, I12.9, I15.8, I15.9, I16.0. According to those criteria, 678 patients are included in this study.

Modeling

After the first round of cluster analysis with the PAM method, variables that produce the same medoids between groups are deleted. This process points to 63 (out of 131) variables for further analysis. Interestingly, the reasons for communication (e.g., follow-up visits), properties of treatments (e.g., the form of the medication), demographics (e.g., age, sex), and access to technologies (e.g., email, phone) are not among the variables that differentiate the groups. However, almost all variables that describe patient-provider communication patterns and patients' health status are preserved due to their ability to differentiate the groups.

We then performed cluster analysis on the remaining variables and two distinct clusters emerged, describing two groups of patients with vastly different communication behavior patterns. The first group relies heavily on mediated communication (i.e., the mediated group), such as secure messaging, video chats, and phone calls, and the second group tends to visit the clinic for face-to-face interactions more often (i.e., the direct group). Figure 1 shows a 2D visualization of the two groups. Multi-task learning models further identify factors associated with the two different behavior patterns and enable the development of patient personas. However, the multi-task learning models fitted for the data in this study have a high error, as indicated by the root mean squared error (RMSE = 0.5977) and mean absolute percentage error (MAPE = 1.8148), considering the standardized data set used for analysis has a standard deviation of 2. The rest of the results section describes results from the two types of models.

Patient-Provider Communication Behavior Patterns

Overall, patients in the mediated group made fewer appointments. When considering the possibility of recent enrollments, those patients also have fewer average appointments since their first visit to the clinic, whereas patients in the direct group make more appointments and have more average appointments per year.

Patients in the mediated group tend to receive more messages from the clinic compared to patients in the direct group and tend to send more messages than receiving them. When examining distinct conversations that are comprised of multiple messages, we found that patients in the mediated group tend to be involved in more conversations as well. The two groups do not differ in terms of the average number of messages in each conversation. Interestingly, the distribution of the messages, but not the conversations, differs between the two groups. Patients in the mediated group have a consistent number of messages in each conversation, marked by their low IQR and MAD values, while patients in the direct group tend to exhibit a larger variance in terms of the number of messages in each conversation.

When patients in the mediated group communicate with their providers, either in person or through mediated channels, those communication behaviors tend to occur after 9:00 am and their direct group counterparts tend to do so before 9:00 am, indicating a high likelihood of late night and early morning

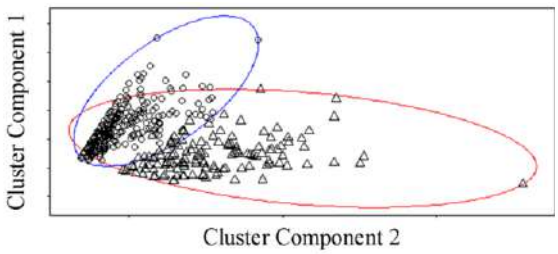


Figure 1 – A 2D visualization of 2 patient groups defined by multi-dimensional communication behavior patterns. The 2 components explain 70.28%[1] of the variation between the 2 groups.

emergencies. In addition, patients in the mediated group tend to be transferred to other clinicians in the clinic's network more often and those in the direct group tend to be referred out. The latter is common when the patients' health condition requires specialists, and they tend to stay with the DPC clinic for ongoing care.

Predictive Features of Communication Behavior Patterns

The multi-task learning framework (more specifically, MTR-CMTL) revealed that patients' health status and its change are the best at predicting patients' communication behavior patterns. Table 1 shows the top 7 features that contributed the most to each of the seven learning tasks of communication behavioral pattern classification.

Overall, patients whose home address is closer to the clinic tend to communicate with the clinic more often across all modalities. Patients with more lab test records, medication dose changes and refills tend to visit the clinic more frequently and communicate with their health care team more frequently through emails and phone calls, but less often through secure messaging.

However, this observation seems to be in contrast with those from the change in patient vitals. For example, patients whose weight, BMI, and body temperature increased over the last 6 months and 12 months tend to be in contact with the clinic less often, either in person or through mediated communication channels. This observation may be a result of noises added from combining the two patient groups. Considering this possibility, we segmented the data from the two groups and performed multi-task regression on each group separately. The distinctive predictors now clearly have different effects on the two groups.

Predictive Features of the Two Emerging Personas

In the rest of this section, we report the results from the cluster analysis, enhanced by the two sets of multi-task regression. Table 2 shows the summary statistics and top predictive features of the two groups (i.e., the mediated and the direct group).

According to the vitals taken at patients' latest visit to the clinic, patients in the mediated group tend to have lower average body

Table 1 – Top 7 features contributing the most to the 7 MTR-CMTL learning tasks, color-coded by the size and direction of coefficients. All the features are reported as after standardization. Med_dose_change: medication dosage change; n_tests: number of lab tests; med_refill: the count of the person having had medication refills done as shown in EHR; 6month_temp_change and 12month_temp_change: body temperature change in 6 months and in 12 months; home_clinic_dist: distance between the home zip code and the clinic; 12moth_hr_change: heart rate change in 12 months.

	In-Person Behaviors		Mediated Behaviors				
Features	**1. n_appt**	**2. ave_appt**	**3. n_ex_msg**	**4. n_in_msg**	**5. heavy_convo**	**6. IQR_msg**	**7. MAD_msg**
med_dose_change	0.170	0.133	-0.002	-0.001	-0.023	0.047	0.044
n_tests	0.170	0.146	-0.004	0.002	-0.015	0.034	0.027
med_refill	0.139	0.116	0.018	0.016	-0.001	0.030	0.032
6month_temp_change	-0.087	-0.058	-0.039	-0.029	-0.014	-0.046	-0.045
12month_temp_change	-0.047	-0.023	-0.024	-0.018	-0.010	-0.036	-0.039
home_clinic_dist	-0.021	-0.020	-0.037	-0.038	-0.034	-0.002	0.001
12month_hr_change	0.022	0.013	0.023	0.020	0.006	0.026	0.028

Table 2 – Top 7 features contributing the most to the 7 MTR-CMTL learning tasks segmented by two user groups, color coded by the size and direction of their coefficients. The numbers in the table heading correspond to the task names in table 1 (i.e., 1. n_appt, 2. ave_appt, etc.). The grey cells with white numbers are coefficients of top features that do not overlap between 2 groups. Additional features to be explained here. 6m_temp: body temperature in 6 months. 12m_o2: blood oxygen level at 12 months. 12m_bp_s_change: systolic blood pressure changes in 12 months.

	Direct Group (RMSE = 0.850, MAPE = 1.304)							Mediated Group (RMSE = 1.065, MAPE = 1.401)						
Features	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
med_dose_change	.165	.120	-.020	-.019	-.015	.041	.047	.123	.082	.122	.112	-.090	.098	.065
n_tests	.176	.146	-.010	-.012	-.014	.022	.019	.076	.090	.009	.053	-.003	.020	.031
med_refill	.139	.114	-.004	.000	-.007	.019	.022	.093	.086	.225	.183	.026	.088	.083
6m_temp_change	-.093	-.068	-.009	-.003	-.012	-.033	-.031	-.070	-.041	-.184	-.176	-.079	-.066	-.089
12m_temp_change	-.038	-.016	-.025	-.022	-.033	-.028	-.032	-.055	-.033	-.048	-.025	.056	-.025	-.040
6m_o2_change	.001	-.005	-.020	-.022	-.029	-.028	-.031	-.022	-.014	.088	.094	.232	.058	.019
6m_temp	.001	.014	.019	.020	.022	.021	.034	.013	.017	.024	-.016	.006	-.013	-.010
12m_o2	-.004	-.003	-.017	-.016	-.019	-.019	-.026	-.019	-.008	.186	.136	.207	.014	-.014
home_clinic_dist	-.031	-.028	-.002	-.003	-.001	.005	.007	-.012	-.013	-.268	-.270	-.243	-.050	-.041
12m_bp_s_change	.019	.015	.010	.009	.011	.010	.006	.012	.005	.166	.184	.135	.033	.014

weight and higher blood oxygen percentage. However, a higher weight is associated with more direct and mediated communication. Patients in the mediated group are more likely to have reported an allergic reaction at least once and experienced more types of allergies compared to the direct group. Having reported an allergy at least once is associated with the increase in all communication behavioral patterns in the mediated group. Meanwhile, patients who have reported more different types of allergic conditions are more likely to use mediated communication channels less often. For the direct group, patients who have more allergic reaction reports are more likely to be associated with higher usage of all communication channels.

Compared to the latest vitals and reports of allergies, features that have better predictive powers for communication behavior patterns are health status 12 months before patients' latest clinical visit and their health status change over 12 months and 6 months. Patients in the mediated group tend to have lower body weight as well as higher diastolic and lower systolic blood pressure 6 and 12 months prior. Also, patients in the mediated group tend to experience decreased systolic and diastolic blood pressure and decreased heart rate over 6 and 12 months, while they are more likely to experience a slight elevation in body temperature during the same period.

The direct group, in comparison, tend to experience increased blood pressure, weight and BMI over 6 and 12 months. For the mediated group, the decreases in blood pressure, weight, and BMI over 6 months and 12 months respectively are associated with a higher number of clinical visits. While the improvement in health status over 6 months for patients in the mediated group is associated with lower usage of mediated communication channels, the improvement over 12 months is associated with higher usage of mediated communication channels. For the direct group, increased body weight and body temperature are associated with decreased in-person clinical visits.

Patients in the mediated group tend to change their medication less frequently but have more instances of refills, which can be achieved through mediated channels. Those in the direct group, on the other hand, tend to have more lab tests that have to be performed in person. Surprisingly, patients in the mediated group with home addresses further away from the clinic tend to contact the clinic less often both in-person and through mediated modalities. This observation is the same as the direct group.

Those observations seem to suggest that patients in the mediated group tend to have better general health status compared to those in the direct group, as suggested by the proxy outcome measures relevant to their health condition.

Discussion

Cluster analysis and multi-task regression revealed two distinct patient personas with nuances in terms of factors that predict communication modality usage. Figure 2 illustrates the two personas, highlighting their major behavioral differences and the variables that predict the major differences. The personas offer guidance on the modalities care teams may prioritize to reach out to patients with different characteristics in the DPC setting.

The results suggest that patients who frequently and more steadily use mediated communication modalities tend to be in better health and have improved health over time, *vice versa* for patients who more frequently visit the clinic in person. Due to the retrospective nature of the analysis, it is unclear whether more frequent mediated communication improved health or patients with better health have fewer needs for in-person clinical visits. Analysis over longer periods of time comparing patients' health upon transferal to 6 months and 12 months after

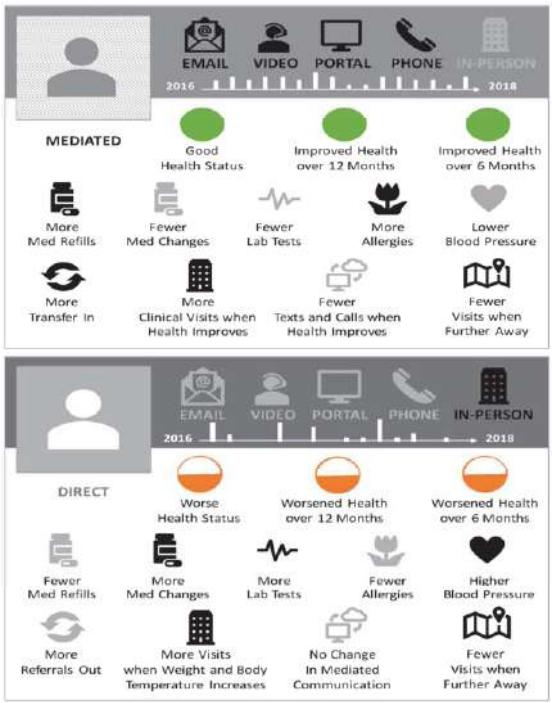


Figure 2 – Two personas for the mediated group and the direct group highlighting the differences in their choice of communication modalities and factors associated with those choices. (left) Mediated group; (right) Direct group.

transferal may produce a clearer picture with causal implications.

Mediated communication modalities are widely used among R-Health patients and have provided viable and valued alternatives to in-person visits, especially for patients who live further away from the clinic and need short and prompt consultations for emergencies and quick questions outside of normal business hours. This unfettered access to physicians tends to create a larger workload for the health care team in a traditional care delivery model [3], but in the flat-fee DPC care model physicians are incentivized to use whichever modality their patients prefer to assure frictionless interactions wherever and whenever necessary to deliver appropriate and effective primary care.

The freedom of choice between multiple communication modalities also enhances patient engagement and satisfaction. This is supported by the patient satisfaction survey at R-Health in August 2018, where many patients expressed their appreciation of multi-modality communication with their health care team in the form of free text comments: "I love the personal care and attention. I get texts etc. from dr. I love it.", "… the service is very convenient. I'm able to set appointments, see my test results, and consult with my doctor on my terms.", "Can consult with doctor over phone and over portal.". The personas, therefore, are a depiction of patients' free choices of communication modality as suited to their preferences and situations.

The personas emerged from the data used in this study do not completely agree with the literature. For example, prior studies found that diabetes patients who use mediated communication channels have worse health [9]. Another study on patients at a facility of Department of Veterans Affairs argued there is no significant association between health status and patient portal usage [14]. Our findings suggest that patients who prefer mediated communication tend to be in better health status. Also, we found that health status and health status change, more than the other factors examined in this study (e.g., demographics, reasons for visits), seem to have stronger associations with communication modality choice. These findings also add to the debate on other issues, such as whether demographics are associated with patients' choices of communication modality ([9] suggests a connection, [14] and our analysis did not identify such connection).

The results of this paper should be interpreted in the context of the reliability of the models we built. While the clustering model performs well at distinguishing between the two groups (i.e., explaining 70.28% of the variation of the data points in the two groups), the regression models are affected by high noise (i.e., RMSE = 0.5977). As a result, although the analysis points to two distinct groups of patients with different choices of communication modalities, confidence in the predictive features are low. Nevertheless, the results reported in this paper is one of the first on patient-provider communication modality choices and predictive features in the DPC setting to the best of the authors' knowledge. The research team is compiling a new and larger data set with more variables and cases in order to build more robust models and reduce noise ratio. More recent results based on the new data set will be presented at MEDINFO 2019.

Conclusions

Patients who are members at R-Health Clinic use multiple communication modalities to connect with their health care providers in the DPC model. Our analysis suggests that there are two distinct groups in modality choices, including a direct group and a mediated group. Multi-modality communication seems to thrive in the DPC model, and mediated communication is highly utilized by both patients and their health care team. The unfettered access to care is expected to foster better patient-provider relationship. The increased level of patient understanding would further lend support to many downstream applications such as designing targeted care management and communication strategies and personalized care plans, as suggested by the U.S. National Academy of Medicine report [7]. More research is needed to explore the characteristics of patients who prefer different modalities and use this knowledge to better engage patients in primary care settings.

References

[1] T. Ostbye, Is There Time for Management of Patients With Chronic Diseases in Primary Care?, *The Annals of Family Medicine* **3** (2005), 209–214.

[2] J. Dwarswaard, E.J.M. Bakker, A. van Staa, and H.R. Boeije, Self-management support from the perspective of patients with a chronic condition, *Health Expectations* **19** (2016), 194–208.

[3] S.H. Kaplan, S. Greenfield, and J.E. Ware, Jr., Assessing the Effects of Physician-Patient Interactions on the Outcomes of Chronic Disease, *Medical Care* **27** (1989), S110–S127.

[4] T. Voruganti, E. Grunfeld, T. Makuwaza, and J.L. Bender, Web-Based Tools for Text-Based Patient-Provider Communication in Chronic Conditions: Scoping Review, *J. Med. Internet Res* **19** (2017), e366.

[5] S.S. Coughlin, L.B. Williams, and C. Hatzigeorgiou, A systematic review of studies of web portals for patients with diabetes mellitus, *mHealth* **3** (2017), 23–23.

[6] P.-Y.S. Hsueh, S. Das, C. Maduri, and K. Kelley, Learning to Personalize from Practice: A Real World Evidence Approach of Care Plan Personalization based on Differential Patient Behavioral Responses in Care Management Records, in: AMIA '18, 2018.
[7] P.V. Long, editor, Effective Care for High-need Patients: Opportunities for Improving Outcomes, Value, and Health, National Academy Of Medicine, 2017.
[8] G.E. Cole, P.A. Keller, J. Reynolds, M. Schaur, and D. Krause, CDC MessageWorks: Designing and Validating a Social Marketing Tool to Craft and Defend Effective Messages, *Social Marketing Quarterly* **22** (2016), 3–18.
[9] R. Sun, M.T. Korytkowski, S.M. Sereika, M.I. Saul, D. Li, and L.E. Burke, Patient Portal Use in Diabetes Management: Literature Review, *JMIR Diabetes* **3,** (2018)
[10] Hierarchical Condition Category (HCC) model, (2018). www.elationhealth.com/hcc/.
[11] M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, K. Hornik, M. Studer, P. Roudier, J. Gonzalez, and K. Kozlowski, "Finding Groups in Data": Cluster Analysis Extended. Version 2.0.7-1.
[12] H. Cao, J. Zhou, and E. Schwarz, RMTL: An R library for multi-task learning, *Bioinformatics* (2018).
[13] G. Pison, A. Struyf, and P.J. Rousseeuw, Displaying a clustering with CLUSPLOT, *Computational Statistics & Data Analysis* **30** (1999), 381–392.
[14] S.S. Woods, C.W. Forsberg, E.C. Schwartz, K.M. Nazi, J.H. Hibbard, T.K. Houston, and M. Gerrity, The Association of Patient Factors, Digital Access, and Online Behavior on Sustained Patient Portal Use, *J. Med. Internet Res* **19** (2017), e345.

Address for Correspondence

Si Sun, si.sun1@ibm.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190338

SIGICAM: A New Software to Improve the Patient Care Supported by a Constraint-Based Model

Carla Taramasco[a], Rodrigo Olivares[a]

[a] *Universidad de Valparaíso, Valparaíso, Chile*

Abstract

Health facilities are care centers that receive patients with different requirements. The management of patients falls to the clinical staff trained for this activity. However, given the demands of the population, the task of managing beds is sometimes too complicated when carried out manually. In this work, we propose the design and implementation of a technological platform that provides an improved optimization approach. It manages the patient-bed allocation efficiently, by considering hospital resources given the number of units and patient diagnosis. This tool was deployed in hospitals of the Atacama regional health service in Chile, boosting the work of the clinical staff of the health facility.

Keywords:
Beds, Health Facilities, Health Resources

Introduction

Recently, health facilities are approaching digitalization as a process of improvement to boost their services. They highlight this approach as a critical factor for the management of health resources and an increase in the efficiency of the clinical services. Informatics services such as Electronic Medical Record [1], traceability in clinical documents [2], medical examinations [3], among others, have taken health to the next level in order to enhance the quality of patient care.

When we focus our efforts on improving patient care, we inevitably entrust ourselves with the task of offering the best available resource [4]. Here, we must define what tasks should performed by the clinical staff and also, we allocate resources to the patient. This decision is not trivial. For instance, what is the ideal bed for a patient? When answering this question, it is necessary to know patient requirements and both features and characteristics of available beds. For instance, patients may need complex clinical tests, specialized care, mandatory isolation, and more. Also, it is essential to be able to classify beds by aspects like age, gender, critical levels, and others.

In the current literature, the patient bed assignment problem has been studied as a particular case of the scheduling problem [5,6]. It corresponds to a typical and recurrent task in hospitals or medical centers. It consists of finding an assignment of available beds for patients with critical medical assistance [7]. Studies have resolved this problem by using a structured mathematical model [8, 9]. Other works have used an efficient approximate algorithm to solve a linear programming model that uses soft and hard constraints but does not include critical medical information [10].

In this work, we propose software supported by a new mathematical approach that improves the patient care process just in time when a bed is required. Our software works from emergency control to the hospitalization of the patient. In this process, the mathematical model uses features of patients for finding the best available beds. It generates a list of potenial beds, and the user decides which bed will be best for the patient.

The software has been deployed in the health services of the Atacama, a region of Chile. During its use, we have detected that the patient care process has improved due to users (nurses) having total visibility of available beds and the assignment process is more efficient.

The remainder of this paper is organized as follows, first we detail the problem statement. Second, we present the proposed software. Finally, the conclusions of this work are detailed.

Problem statement

People demand more and better services, this is a fact. In many situations, these demands trigger a collapse in public health services, especially in hospitals and medical centers [11]. People arrive at hospitals, those who require admission need to be assigned a bed, and at this moment, they become in patients of the health service. According to their needs, each patient presents variables such as age, risk, gender, if they require isolation (or not), among others [12]. Furthermore, beds are a scarce resource that should be optimized, but this is not always the case. In Chile, this process is directed by a national policy. Therefore, we must respect it. Ministry of Health defines the protocol for receiving patient requirements for bed availability. The detail of this protocol is shown in Figure 1.

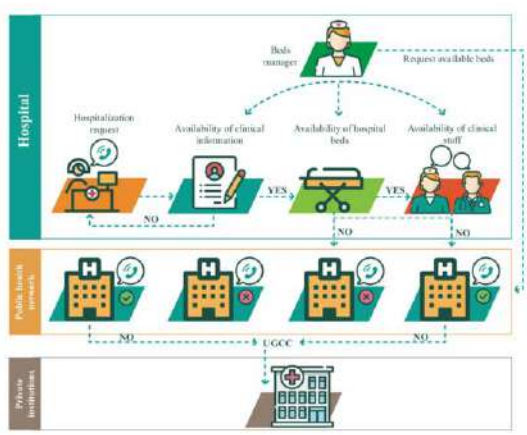


Figure 1– Request protocol for bed availability.

There beds are not always available, and those that are vacated must be treated correctly to satisfy the needs of patients. In

order to accomplish that, it is necessary to know the properties of each bed to ensure efficient assignment. The properties of beds are the type (basic, critical), classification (neonatal, pediatrics, adult, older people, etc.), gender, type of insulation, among others. All these properties must be considered when assigning a patient to the best available bed. However, what would happen if the only available bed is an adult type and a teen needs it? Current optimization models and mixed linear methods determine that it is not possible to assign it [13].

To address this problem, we developed new software, boosted by a new binary model bed assignment problem, that minimizes the inter-movement of patients, and also considers constraints. There is exsisting work that takes a similar approach [14].

Search process

When we designed the software, we defines a search process based on two-stages, as shown in Figure 2, and Figure 3.

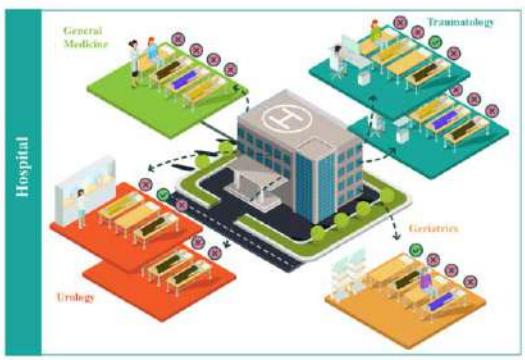


Figure 2– Internal Search Process (hospital).

Figure 3– External Search Process (public health network).

The first stage allows us to find the best bed into the hospital that belongs to the regional health service. If it is not possible to find this match, the next stage begins. We take the best available beds and we perform the assignment patient-bed searching in all public assistance network. We designed this process in order to initially reduce the search space, and to use potential solutions that exhibit good performance from the beginning.

For the software to work correctly, we need to control the following aspects:

- The configuration of hospitals and units (set of rooms) is known and this set does not change during the decision period.
- Needs of patients will not change during the decision period. For example, if a patient needs isolation, their critical level is maintained during the phase of bed search.
- Beds remain unavailable while in use by patients, until discharge.

Decision variables

To initiate the model, we describe the binary decision variables: bed (b). This variable is used to determinate if a bed is available or it is assigned to a patient: 1 if the bed b of a specific type and it belongs to a particular room is assigned to a patient, 0 otherwise. In our case, all beds belong to a specified room, and they are defined by a type.

Regional health service have several hospitals, and each hospital is characterized according to its complexity. If a hospital has intensive care units (ICU) or it has critical patient units (CPU), we can say that this hospital is defined as complex. Not every hospital is designated as complex because of limited resources/equipment. Complex hospitals and normal hospitals, both have basic beds available to assign. However, only complex hospitals have *critical* beds.

Additionally, we use $f_r = 1$ and $m_r = 1$ if at least one woman and one man, respectively, is in room r, 0 otherwise. This value is computed after to search for an availability bed. Finally, we define $y_{pl} = 1$ if patient p is consistent with age policy l, 0 otherwise.

Constraints

Next, we present the following constraints that we should cover but could be violated. It is important to mention that hospitals, units, and rooms, have non-variable capacity during the resolution process.

- c_1: **Assignment**. Each patient is assigned to only one bed, and only one bed is assigned to a patient.
- c_2: **Risk and dependence**. It indicates the severity of the patient and it is classified into 3 categories:
  - *Maximum risk* is assumed for the patient. A critical bed must be assigned to the patient. In case that there is not one available critical bed, a normal bed will be equipped as a critical bed.
  - *Medium risk* is assumed for the patient. A critical bed will be found but if there is not, a normal bed will be assigned.
  - *Minimum risk* is assumed for the patient. A normal bed will be assigned to the patient.

  This constraint is extremely relevant, however, due to the shortage of critical beds, there is a high possibility that even being a critical patient the assigned bed will not be the critical type. Therefore, this constraint can be violated. Complexity hospitals can transform a normal bed into a critical bed.
- c_3: **Isolation**. In the health systems, contagious diseases develop that in some cases can be more serious than the disease originated abroad. This constraint may not be satisfied because hospitals follow internal protocols to modify a room using

special equipment. This process allows the patient to be "isolated" in a room with other persons.

- c_4: **Unit policies**. Each hospital offers a set of units that it groups a set of rooms. A patient-derived to a unit should be assigned to one bed of a room of this unit. However, if it is not possible finding one bed in a specified room, by default, he/she will be assigned to one bed into a general room.
- c_5: **Gender policies**. If a patient declares belonging to the male gender or the female gender, the assigned bed must belong to a room with this characteristic. This constraint may not be satisfied because it is possible to modify a room as in the isolation restriction.
- c_6: **Age polities**. This constraint considers the choice of the hospitalization area. Since the beds are heterogeneous according to the age range, it is possible to determine 4 types: neonatology, pediatrics, adult, elderly. If there is no available beds with the desired age feature, an adult bed will be assigned to the patient. A preliminary study conducted to there are more adult beds than any other type. Therefore, we can conclude that this constraint may not be covered.
- c_7: **Distance**. If a patient requires one bed and no beds are available at the current hosptial, it is necessary to transfer the patient to the nearest hospital. Nevertheless, the nearest hospital may not be the best alternative according to patient requirements. In this case, it is necessary for finding the best bed in another hospital that may not be the closest. This case can be violated.

Resolution

To give an efficient response to the model, we perform this model by using an approximate algorithm. This technique is known as bat algorithm.

Bat algorithm was proposed Xin-She Xang [15, 16] and it is inspired by eco-localization behavior (biosonar) that belong microbats. Rules that led this algorithm are:

1. All bats use echolocation to perceive the distance they have to objects or obstacles. They can differentiate between food, prey, or obstacles in their way.
2. Bats randomly fly, with velocity v_i position x_i. When they are looking for a prey they adjust their frequency f_i by changing their wavelength and volume A_o. Bats can adjust their frequency and adjust the pulse emission radius$r \in [0,1]$.
3. The volume varies from a positive value A_0 to a minimum value A_{min}.

The bat algorithm proposes the change of position using three equations:

1. Calculate the frequency of the bat.

$$f_i = f_{min} + (f_{max} - f_{min})\beta \quad (1)$$

2. Calculate the speed of the bat considering the frequency as input data.

$$v_i^d(t+1) = v_i^d(t) + (x_i^d(t) - \hat{x}^d)f_i \quad (2)$$

3. Calculate the new position of the bat.

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \quad (3)$$

Updating position can alter the binary domain of the problem. However, the algorithm proposes an adaptation to the domain by using the sigmoid function and a probabilistic selection. Algorithm 1 describes the procedure to find a solution.

Algorithm 1: Bat algorithm

Input: Parameters of the problem
Output: List of available bed.

for all bat x_i, $(\forall_i = \{1, \dots, n\})$ **do**
  for all dimension d, $(\forall_d = \{1, \dots, m\})$ **do**
    $x_i^d \leftarrow Random\{0,1\}$
  end for
  $A_i \leftarrow Random[0,1]$
  $r_i \leftarrow Random[0,1]$
  $fit_i \leftarrow cost\_function(x_i)$
end
$globalfit \leftarrow +\infty$
while $t < T$ **hacer**
  for all bat b_i, $(\forall_i = \{1, \dots, n\})$ **do**
    $rand \leftarrow Random[0,1]$
    if $rand < A_i$ **the**
      $A_i \leftarrow \alpha A_i$
      $r_i \leftarrow r_i^{t=0}(1 - e^{-\gamma t})$
    end
  end
  $\{minfit, minindex\} \leftarrow min(fit)$
  if $minfit < globalfit$ **then**
    $globalfit \leftarrow minfit$
    $\hat{x}^d \leftarrow x^d_{minindex}$
  end
  for all bat b_i, $(\forall_i = \{1, \dots, n\})$ **do**
    $\beta \leftarrow Random[0,1]$
    $rand \leftarrow Random[0,1]$
    if $rand < r_i$ **then**
      For all dimension d, $(\forall_d = \{1, \dots, m\})$ **do**
        $x_i^d \leftarrow \varepsilon \bar{A}$
        $rand \leftarrow Random[0,1]$
        if $rand < \frac{1}{1+e^{-x}}$ **then**
          $x_i^d \leftarrow 1$
        else
          $x_i^d \leftarrow 0$
        end
      end
    fin si
    $rand \leftarrow Random[0,1]$
    if $rand < A_i$ **y si** $fit_i < globfit$ **then**
      $f_i \leftarrow f_{min} + (f_{max} - f_{min})\,\beta$
      for all dimension d, $(\forall_d = \{1, \dots, m\})$ **do**
        $v_i^d \leftarrow v_i^d + (x_i^d - \hat{x}^d)f_i$
        $x_i^d \leftarrow x_i^d + v_i^d$
        $rand \leftarrow Random[0,1]$
        if $rand < \frac{1}{1+e^{-x}}$ **then**
          $x_i^d \leftarrow 1$
        else
          $x_i^d \leftarrow 0$
        end
      end
    end
  end
  $t \leftarrow t + 1$
end
retornar $\hat{x}$

To find an optimal solution, it is usually necessary to combine the exploitation and exploration processes. In this context, the

variability of the solutions is given by the adjustment of the volume A and the radius r, as shown in equations (4) and (5).

$$A_i\,(t+1) = \alpha A_i\,(t) \quad (4)$$

$$r_i\,(t+1) = r_i(t=0)[1-e^{-\gamma t}] \quad (5)$$

Software

Finally, the implemented software has successfully been deployed in Government Regional Health Services. The user can access in http://www.sigicam.cl. Primary functions are described as follows:

- Each user logins to the application by using a specific account. After login, the software alerts the user if patients have been waiting more 12 hours for an available bed. At this moment, the users can see a "waiting list" (highlighted in red).

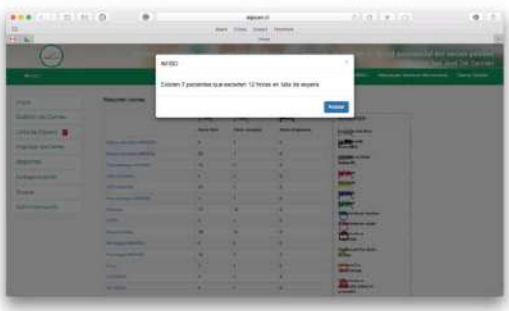


Figure 4– After login.

- From the "waiting list", the user can assign a specific bed using a list of beds deployed in the "floor map".

Figure 5– Assign manually a bed.

- When a patient is registered in the platform, the users can decide if it is the application itself who assign the best available bed or not. The system shows a list of best beds according to features or conditions of the patient. Here, the optimization model works to help users quickly find a bed. This list is rated.

The software provides additional functions, such as internal movements, external movements (government hospitals or private health facilities), daily monitoring and update of risk-dependence, statistical information, reports, among others.

Finally. we can say that our software supports the nurse's work, which in turn directly impacts the quality of patient care. Furthermore, the platform highlights bed occupation rates by the hospitals and shows how efficiently the beds are being used. This information includes the stay times of each patient and, above all, the waiting times that each patient suffers when there are no available beds.

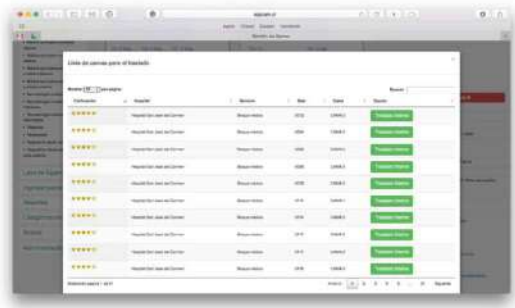


Figure 6–Semi-automatic assignment of beds.

Conclusion

In this work, we present new software to improve the patient care supported by a constraint-based model. The software can find the best available bed for patients and it alerts when one of them waits for more than 12 hours. For that, we design a constraint-based model using hard and soft constraints according to the needs of patients. Using those constraints, we implement the bat algorithm to optimize bed selection.

SIGICAM is a web-based implementation that coverages the resolution of the model on-demand for expert users. This software includes the comunication between the optimization approach and the functional features of the software.

As future work, we plan to develop a self-adaptive approach to the bat algorithm in order to improve the quality of the reached solutions. This self-adaptive approach will be based on the principle of the autonomous search, following the work conducted by Soto et. al. [17].

Acknowledgements

Authors acknowledge FONDEF IDeA ID16I10449, STIC-AMSUD 17STIC-03, MEC 80170097, and CORFO National Center on Health Information Systems (CENS).

References

[1] S. R. Raman, L. H. Curtis, R. Temple, et al. Leveraging electronic health records for clinical research. American Heart Journal, 202:13–19, Aug 2018.

[2] H. F. Hedian, J. A. Greene, and T. M. Niessen. The electronic health record and the clinical examination. Medical Clinics of North America, 102(3):475–483, May 2018.

[3] P. Tracol. Materials vigilance and traceability. Orthopaedics & Traumatology: Surgery & Research, 102(1): S95–S103, Feb 2016.

[4] A. Kassem, M. Hamad, C. El Moucary, E. Nawfal, and A. Aoun. MedBed: Smart medical bed. In 2017 Fourth International Conference on Advances in Biomedical Engineering (ICABME). IEEE, Oct 2017.

[5] J. Marynissen and E. Demeulemeester. Literature review on multi-appointment scheduling problems in hospitals. European Journal of Operational Research, Mar 2018.

[6] B. Cardoen, E. Demeulemeester, and J. Belin. Operating room planning and scheduling: A literature review. European Journal of Operational Research, 201(3):921–932, Mar 2010.

[7] V. L. Smith-Daniels, S. B. Schweikhart, and D. E. Smith-Daniels. Capacity management in health care services: Review and future research directions. Decision Sciences, 19(4):889–919, Dec 1988.

[8] S. Brailsford and J. Vissers. OR in healthcare: A european perspective. European Journal of Operational Research, 212(2):223–234, Jul 2011.

[9] I. B. Vermeulen, S. M. Bohte, S. G. Elkhuizen, H. Lameris, P. J. M. Bakker, and H. La Poutré. Adaptive resource allocation for efficient patient scheduling. Artificial Intelligence in Medicine, 46(1):67–80, May 2009.

[10] R. Guido, M. Carmela Groccia, and D. Conforti. An efficient matheuristic for offline patient-to-bed assignment problems. European Journal of Operational Research, 268(2):486–503, Jul 2018.

[11] The Lancet. The collapse of the Venezuelan health system. The Lancet, 391(10128):1331, Apr 2018.

[12] P. B. Ginsburg and D. M. Koretz. Bed availability and hospital utilization: estimates of the roemer effect. Health Care Financing Review, 5(1):87, 1983.

[13] A. Mazier, X. Xie, and M. Sarazin. Real-time patients assignment: A method for improving emergency department flow. In 2010 IEEE Workshop on Health Care Management (WHCM), pages 1–6. IEEE, Feb 2010.

[14] A. M. Turhan and B. Bilgen. Mixed integer programming based heuristics for the patient admission scheduling problem. Computers & Operations Research, 80:38–49, Apr 2017.

[15] X.-S. Yang. A new metaheuristic bat-inspired algorithm. In Proceedings on Nature Inspired Cooperative Strategies for Optimization (NICSO), volume 284 of Studies in Computational Intelligence, pages 65–74. Springer, 2010.

[16] X.-S. Yang. Bat algorithm: literature review and applications. IJBIC, 5(3):141–149, 2013.

[17] R. Soto, B. Crawford, R. Olivares, C. Galleguillos, C. Castro, F. Johnson, F. Paredes, and E. Norero. Using autonomous search for solving constraint satisfaction problems via new modern approaches. Swarm and Evolutionary Computation. Elsevier BV. 30:64-77, 2016.

Address for correspondence

Carla Taramasco and Rodrigo Olivares are professors at Universidad de Valparaíso, Chile.

Emails: carla.taramasco@uv.cl, rodrigo.olivares@uv.cl.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190339

Performance Evaluation of Clinical Decision Support Systems (CDSS): Developing a Business Intelligence (BI) Dashboard

Vania Teixeira[a], Analía Mori[a], Andres Usera[a], Juan Carlos Bacigalupo[a], Daniel Luna[b]

[a] Department of Biomedical Informatics, Universal Soluciones Tecnológicas, Montevideo, Uruguay,
[b] Department Biomedical Informatics, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina.

Abstract

This document describes the development of a Business Intelligence (BI) dashboard for tracking the drug-drug interaction (DDI) alerts implemented as Clinical Decision Support Systems (CDSS) in Electronic Health Records (EHR). CDSS are known for their potential to reduce medical error. The use of requirements in the development of BI dashboards is crucial to obtain successful software. In this work, the requirements were analysed using a score methodology, considering the relevance of the indicators and visualization methods. CDSS effectiveness and acceptance have been questioned, so it is fundamental to monitor their behaviour and performance. The dashboard was designed in order to satisfy the needed indicators. Using BI as a tool for monitoring the CDSS performance made it possible to operationalize the EHR content repository, maximizing the understanding in relation to the override and, by inference, to optimize the CDSS system by opening new lines of work.

Keywords:

Decision Support Systems, Clinical; Drug Interactions, Software

Introduction

Medical errors are a serious problem for the health system. The stages defined in the medication cycle are prescription, transcription, dispensation, administration, and monitoring. It is known that at least 50% of errors are generated during prescription. This can be due to lack of information about the medication used or the patient's background [1,2].

Clinical Decision Support Systems (CDSS) are recognized as a significant contribution to structured electronic prescription. They provide information about a specific need and responses that are similar to human reasoning. While there are no clear short-term benefits, previous work has shown that warning systems could reduce medical errors by 81% [3]. CDSS are able to identify up to 89% of medication-related errors and prevent 23% of them [4].

Despite providing warnings about potential harm and the initial promising outcomes, users often do not adhere to CDSS messages, overriding 49% to 96% of the time [5–7]. Reasons for nonaderance are diverse and reflect the complexities of clinical practice, where usually there is more than one correct decision when managing a specific case. Moreover, design and implementation could lead to ineffective alerts. Redundant alerts and those with lack of scientific evidence or usability are usually overridden and cause alert fatigue [8].

It is widely recognized that CDSS are based on rigorous scientific evidence, so that the advice they provide is equal or superior to the average in a health care system. However, contrary to other analytic decision software that is part of a medical device (for example, automatic infusion pumps), CDSS have no regulatory standards, and like any other software, they have flaws [9].

In this context, it is necessary to deploy a monitoring system during the implementation and ongoing use of CDSS. Not only would such monitoring allow design validation, detection usability problems, and inconsistencies in the knowledge base and rules, but it also grants a cycle of continuous improvement in order to optimize the tool.

In the last decade, there has been significant growth in the literature on the use of Business intelligence (BI) in the healthcare field. Electronic Health Records (EHR) contain massive clinical datasets. BI emerged as a technological tool with the potential to collect, manipulate, and analyse the dataset in the EHR repository in order to improve the evidence-based decisions and quality practices [10–12].

As was explained in our previously published report, CDSS of drug-drug interaction (DDI) alerts have been implemented in a Uruguayan healthcare network [13].

The purpose of this paper is to describe the development of a CDSS-DDI dashboard for a federated Health Information System (HIS) in Uruguay using Business Intelligence (BI) tools.

Methods

Definition of Key Progress Indicators

Previous to the availability of BI tools, CDSS alerts were manually monitored by the medical informatics staff. Manual monitoring was based on the team objectives, the literature and their knowledge of CDSS use. Different Key Progress Indicators (KPI) of that manual monitoring using Microsoft Excel where then the basis for the initial BI dashboard.

The objective of the KPI is to monitor the user's alert fatigue and alert adequacy, as well as design and usability aspects.

DDI alerts provided by the Buenos Aires Italian Hospital (HIBA) web service have been used in our system for more than two years and have allowed us to successfully collect valuable data. However, due to the amount of information generated, conventional tools are inefficient for data analysis and reporting.

In 2018 with already defined indicators, we incorporated a BI tool, Tableau, for the CDSS dashboard construction.

CDSS Dashboard Requirements

The CDSS dashboard requirements were established in agreement with the BI developer. The requirements were designed regarding priority levels from 0 (not a priority) to 5 (very high priority).

System Architecture

Our Healthcare Data Warehouse (DW) was built using Pentaho Data Integration (PDI) to extract the relevant data from the EHR and Enterprise Resource Planning (ERP). Both use DB2 as a transactional database. For the DW a combination of Postgres and MySQL was used.

The DataMart for CDSS alerts is a subset of this Data Warehouse. Each service, databases or Extract-Transform-Load (ETL) tools, runs as a Docker Container in a cluster. This allows optimizing resource allocation during the ETL process. For the consumption of this DataMart, Tableau was chosen, which is based on VizSQL, a proprietary language that integrates SQL data consumption with graphical visualization grammar. This allows postponing the actual specification of the final use of a specific element. The advantages of Tableau Hyper were used in the memory data engine to speed up analytical query processing. A Drug ID could be used as a dimension or measure filter (distinct count) by the end user in a very intuitive way. No SQL, no MDX, no extra burden and no need to wait for a new logical cube modification.

The validation of the cube and its data were carried out in iterative cycles with the medical informatics team and the BI developer, based on the comparison of previous results with some of the basic KPI obtained by queries and processed in Microsoft Excel.

As an initial stage, it was defined that the dashboard should not be integrated with the Electronic Medical Record (EMR) application or in other medical management tools but that it would be accessed from Tableau by the medical informatics team.

Results

Definition of Indicators

Some Key Progress Indicators (KPI) were previously defined and manually measured twice a year to monitor the clinician's behaviour and CDSS performance. The established KPI were:

- Number of CDSS-DDI alerts per 1000 prescriptions
- Number of CDSS-DDI per 1000 patients
- Number of CDSS-DDI per 1000 medical consults
- % of overridden CDSS-DDI alerts

Definition of Requirements

Considering the KPI and other possible data to be retrieved such as medical specialities, alert incidence per DDI pair and override justifications, a list of requirements for the BI dashboard was developed. The requirements with priority level over 3 where covered in the dashboard. Table 1 shows a description of the requirements with their priority assessment.

Table 1– BI dashboard requirements

Priority (0 – 5)	Requirement	Description
5	EHR as a source of information	Exporting data from EHR of three healthcare centres in Uruguay.
4	Automatic update	Real-time update.
5	Remote access	Provide access to specified dashboards.
4	Ad-Hoc Calculations	Related to the predefined indicators and metric.
5	Information retrieval	Data reporting based on dates.
4	KPI visualization	Real-time display of KPI evolution.
4	Advanced Chart Types	Ad-Hoc report with the selection of multiple dimensions and metrics: number of alerts, override justifications, DDI pairs, override, medical specialities.
5	Dataset exploration	Explore one CDSS case to audit the EMR.
5	Data security	Sensitive patient data should remain confidential.

Data Validation

Data architecture issues such as semantics, integrity, and security were analysed. A semantic standard was established to define, for example, what was understood by "Consult," what type of health professionals could be responsible for the consult, if it included a prescription or not, and if there was an alert or not. Regarding data integrity, all the components were evaluated, leading to the detection of some initially incomplete logs. For example, when the clinician decided to override the alert, the active drug information was not recorded in the data warehouse. For data security, data items that should be visible or editable according to the user´s profile and license duration were assessed.

Design and Development of the Dashboard

The requirements were embedded in the software as three modules:

1. KPI
2. Overall CDSS performance
3. Detailed CDSS performance

Module 1 was deployed as line charts. Module 2 (Figure 1) was developed in order to monitor the rate of prescriptions with DDI alerts, the rate of overridden or accepted alerts and the quantity and type of justifications for the override. Module 3 (Figure 2) was designed as an advanced chart to combine and visualize multiple dimensions and their relation (quantity of alerts, override percentage, medical specialities and DDI pairs involved).

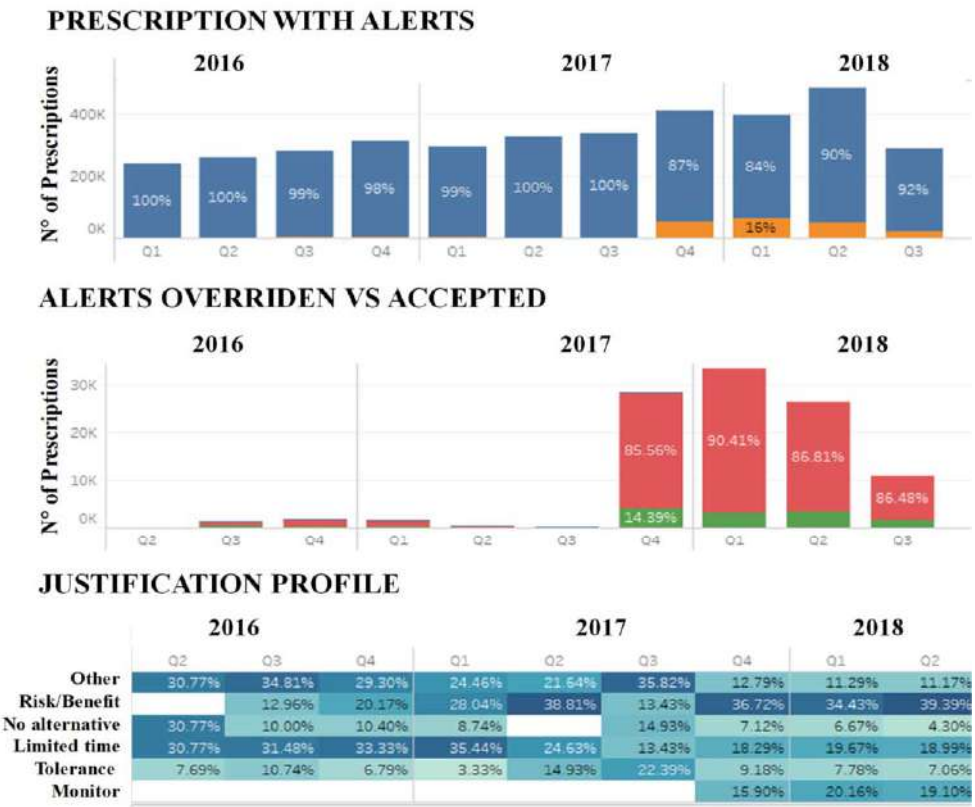


Figure 1- Module 2: The figure shows prescriptions with (orange) and without (blue) CDSS alerts, overridden (in red) vs accepted (in green) alerts and the justification profile for clinician alert override (higher percentages in darker tones).

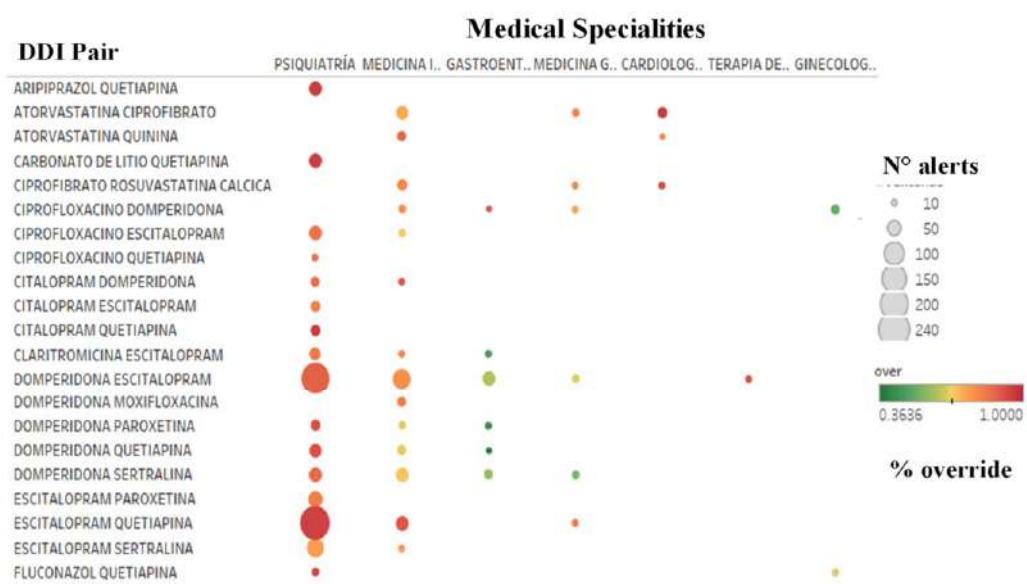


Figure 2- Module 3: medical specialities are in the x-axis and DDI pairs in the y-axis.The quantity of alerts is represented in the size of the spheres. The override percentage is represented in a colour range from green (lower override %) to red (higher override %

Discussion

CDSS can reduce errors and prevent the harm caused by them by alerting the clinicians. However, their effectiveness and acceptance have been questioned. The recommendation to improve these issues respecting the five rights is: "CDSS should be designed to provide the right information to the right person in the right format through the right channel at the right time" [14,15]. In order to achieve the five rights, it is fundamental to monitor CDSS behaviour and performance.

The EHR is a repository where the clinical data is stored, some structured and others unstructured. Data mining technologies are necessary to extract quality data and inference rules from the information stored in order to provide CDSS real-time monitoring. The benefit in employing BI is that it works with unstructured data while other tool doe not [16].

According to the literature, the use of BI technology with EHR allowed operationalizing the data warehouse of the EHR to improve the quality and safety delivered by the healthcare system [11,12,17,18]. This improvement is due to the potential of BI to support evidence-based practice and decision-making process [12].

To the best of our knowledge, there is no study regarding the use of BI for tracking CDSS. However, BI seems like a valid alternative to efficiently and effectively monitor CDSS. To test the integration of these technologies on a first pilot scale, a BI dashboard was developed for the previously implemented DDI alert system [13].

The BI dashboard for CDSS-assisted DDI alerts was developed as customized software. It was designed in order to satisfy the needed KPI. Before the dashboard, KPI were measured once or twice a month through a manual process that included Microsoft Excel spreadsheet calculations dependent on data requests to the engineer. The KPI were defined to track the prevalence of CDSS alerts and overrides. In order to have the number of alerts independent of the number of prescriptions, persons, and consults, CDSS prevalence was settled in relation to these attributes. The acceptance of the CDSS alerts was determined by the override percentage.

The use of requirements in the developing of BI dashboards has been reported as a crucial issue to obtain successful software [19,20]. In this work, requirements were analysed using a score methodology [20], considering the relevance of the indicators and visualization methods. The requirement analysis was outlined with a mixed approach that considered the goal of the tool, the characteristics of the user, and the nature of the information. The chart types were selected to visualize several variables in the same chart in order to obtain a process behaviour overview. At the same time, it enables the evaluation of data below or above the specified threshold.

BI is only useful if the information provided is built on quality-assured data. Otherwise, logic and inference rules can be flawed [12,21]. Hence, data validation is a critical issue even though it is not always considered by software developers. During data validation for this project, incomplete logs in the data warehouse were detected, caused by the lack of knowledge of their relevance by the EHR developers. Semantics and security were also reviewed. When CDSS-DDI are implemented in multiple healthcare centres, it important to recognize that one data field can have several meanings, especially if the EHR has several types of users. Unless the data field has the same meaning, it is impossible to integrate or communicate efficiently across the organization(s). Regarding data security, it is relevant to consider that the information managed is related to the patients and, in this way, it is sensitive information. Previous reports have described techniques to adopt secure barriers for EHR [22]. For this dashboard, we adopted techniques such as access control and data encryption.

In the current work, a modular approach was used to develop the DDI dashboard. Since it was important to have rapid visualization of the system behaviour, three modules were designed. The modules varried based on their level of specificity: the first is aimed for large-scale CDSS monitoring with KPI, the second provides in-depth analysis, and the last module mixes several variables in one chart to simultaneously convey large-scale and granular information. The indicators previously tracked manually, new indicators, and measurements that were previously unable to be processed manually were included in the development of the dashboard.

The increased autonomy and flexibility of users to get information and the efficiency and quality of the reports are only some of the benefits of using the developed dashboard. For example, the dashboard enables detection of trends in alerts overridden by the clinicians so the usefulness of these alerts can be analysed.

On the other hand, several challenges have been described in the literature for the appropriate integration of BI with the EHR. These were mainly non-technical factors such as organizational issues and lack of governance [23,24]. In this work, the most noticeable challenges were those related to data architecture.

Limitations

The current work has faced limitations related to data architecture. As mentioned above, initially, some crucial information for the dashboard was incomplete in the data warehouse. A second issue was to join data that, historically, were in separate data silos.

Future Lines

Considering the benefits obtained with this initial dashboard, the scope will be extended to other CDSS alerts and implemented in the EHR. Furthermore, based on the findings of this work, the data integrity of the EHR repository is being thoroughly reviewed. Based on last year's monitoring, a work line was initiated to optimize some drug pair rules in order to reduce the rate of overrides. Finally, the application of BI to other issues related to healthcare decision-making processes has been triggered.

Conclusion

The development of a BI dashboard for tracking CDSS-assisted DDI alerts in an EHR was described throughout this document. Alongside the definition of indicators and requirements, data architecture was crucial, the latter being the most challenging issue during development.

Using BI as a tool for CDSS performance monitoring made it possible to operationalize the EHR content repository, to maximize understanding of overrides, and to optimize the CDSS.

Acknowledgements

We would like to thank Jose Abadie and Rodrigo Martinez for their assistance with the data warehouse in the EHR and Álvaro Tamborindeguy for supporting the development of the Business Intelligence area at Universal Soluciones Tecnológicas. We

would also like to thank all the members of the Biomedical Informatics Department of Hospital Italiano de Buenos Aires.

References

[1] L.L. Leape, D.W. Bates, D.J. Cullen, J. Cooper, H.J. Demonaco, T. Gallivan, R. Hallisey, J. Ives, N. Laird, and G. Laffel, Systems Analysis of Adverse Drug Events. ADE Prevention Study Group., *JAMA* **274** (1995) 35–43.

[2] L. Daniel, H. David, G. Schpilberg, Mónica Hernandez, E. Soriano, M. Martinez, A. Gomez, G. Cifarelli, and F. Bernaldo de Quiros, Validación de la base de conocimiento de un sistema notificador de interacciones farmacológicas, *V Simp. Informática En Salud **31*** (2002).

[3] D.W. Bates, J.M. Teich, J. Lee, D. Seger, G.J. Kuperman, N. Ma'Luf, D. Boyle, and L. Leape, The Impact of Computerized Physician Order Entry on Medication Error Prevention, *J Am Med Inform Assoc* **6** (1999) 313–321.

[4] D.W. Bates, A.C. O'neil, D. Boyle, J. Teich, G.M. Chertow, A.L. Komaroff, and T.A. Brennan, Potential Identifiability and Preventability of Adverse Events Using Information Systems, *J Am Med Inform Assoc* **1** (1994) 404–411.

[5] H. Van Der Sijs, J. Aarts, a Vulto, and M. Berg, Overriding of Drug Safety Alerts in Computerized Physician Order Entry, *J Am Med Inform Assoc* (2006) 138–148.

[6] A.B. McCoy, E.J. Thomas, M. Krousel-Wood, and D.F. Sittig, Clinical Decision Support Alert Appropriateness: A Review and Proposal for Improvement, *Ochsner J* **14** (2014) 195–202.

[7] K.C. Nanji, D.L. Seger, S.P. Slight, M.G. Amato, P.E. Beeler, Q.L. Her, O. Dalleur, T. Eguale, A. Wong, E.R. Silvers, M. Swerdloff, S.T. Hussain, N. Maniam, J.M. Fiskio, P.C. Dykes, and D.W. Bates, Medication-Related Clinical Decision Support Alert Overrides in Inpatients, *J Am Med Inform Assoc* **25** (2017) 476–481.

[8] J.S. Ancker, L.M. Kern, E. Abramson, and R. Kaushal, The Triangle Model for Evaluating the Effect of Health Information Technology on Healthcare Quality and Safety, *J Am Med Inform Assoc* 19 (2012) 61–5.

[9] E.H. Shortliffe, and M.J. Sepúlveda, Clinical Decision Support in the Era of Artificial Intelligence, *JAMA*. **320** (2018) 2199–2200.

[10] M. Reimers, Leveraging Business Intelligence to Make Better Decisions: Part I, *J Med Pract Manage* **29** (2014) 327–30.

[11] W. Zheng, Y.-C.J. Wu, and L. Chen, Business Intelligence for Patient-Centeredness: A Systematic Review, *Telemat Inform* **35** (2018) 665–676.

[12] W. Bonney, Applicability of Business Intelligence in Electronic Health Record, *Procedia - Soc. Behav Sci* **73** (2013) 257–262.

[13] V. Teixeira, L. Rubin, R. Rebrij, A. Tamborindeguy, R. Martinez, J.C. Bacigalupo, and D. Luna, Implementation of an Outsourced Transnational Service of Clinical Decision Support System, *Stud Health Techno. Inform* **245** (2017) 1384.

[14] G.L. Alexander, Issues of Trust and Ethics in Computerized Clinical Decision Support Systems, *Nurs Adm Q* **30** (2006) 21–29.

[15] C.B. Byrne, D.S. Sherry, L. Mercincavage, D. Johnston, E. Pan, and G. Schiff, Advancing Clinical Decision Support - Key Lessons in Clinical Decision Support Implementation, 2010.

[16] J. Lamont, Business Intelligence: The Text Analysis Strategy, *KM World* **15** (2006) 8–10.

[17] P. Brooks, O. El-Gayar, and S. Sarnikar, A Framework for Developing a Domain Specific Business Intelligence Maturity Model: Application to Healthcare, *Int J Inf Manage* **35** (2015) 337–345.

[18] R. Gaardboe, T. Nyvang, and N. Sandalgaard, Business Intelligence Success Applied to Healthcare Information Systems, *Procedia Comput Sci* **121** (2017) 483–490.

[19] N.H.Z. Abai, J.H. Yahaya, and A. Deraman, User Requirement Analysis in Data Warehouse Design: A Review, *Procedia Technol* **11** (2013) 801–806.

[20] P. Rezaei-hachesu, T. Samad-Soltani, S. Yaghoubi, M. GhaziSaeedi, K. Mirnia, H. Masoumi-Asl, and R. Safdari, The Design and Evaluation of an Antimicrobial Resistance Surveillance System for Neonatal Intensive Care Units in Iran, *Int J Med Inform* **115** (2018) 24–34.

[21] J. Jordan, and C. Ellen, Business Need, Data, and Business Intelligence, *J Digit Asset Manag* **5** (2009) 10–20.

[22] C.S. Kruse, B. Smith, H. Vanderlinden, and A. Nealand, Security Techniques for the Electronic Health Records, *J Med Syst* **41** (2017) 127.

[23] J. Glaser, and J. Stone, Effective Use of Business Intelligence, *Healthc Financ Manage* **62** (2008) 68–72.

[24] W. Yeoh, and A. Koronios, Critical Success Factors for Business Intelligence Systems, *J Comput Inf Syst* **50** (2010) 23–32.

Address for correspondence

vania.teixeira@universal.com.uy

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190340

The Value of Teledermoscopy to the Expertise of General Practitioners Diagnosing Skin Disorders Based on ICD-10 Coding

Esmée Tensen[ab], Femke van Sinderen[ab], Leonard Witkamp[ab], Monique W.M. Jaspers[a], Linda W.P. Peute[a]

[a]Amsterdam UMC, University of Amsterdam, Department of Medical Informatics, Amsterdam Public Health Research Institute, Meibergdreef 9, Amsterdam, Netherlands
[b]Ksyos Health Management Research, Amstelveen, Netherlands

Abstract

Early recognition of skin cancer is vital to enhance patient outcomes. Teledermoscopy (TDsc), a telemedicine service, supports general practitioners (GPs) in gaining fast access to dermatologists' feedback to detect skin cancer. This study aimed to assess if GPs gain expertise in diagnosing skin disorders after continued use of TDsc, based on diagnosis classification by the International Statistical Classification of Diseases and Related Health Problems (ICD-10). A retrospective study was conducted on TDsc consultations sent by GPs to teledermatologists in the Netherlands (July 2015 - June 2018). GP sensitivity and confirmed cases in diagnosing skin disorders slightly increased over time. However, the total positive predictive value showed a decrease. In three years, 43 melanomas were diagnosed by the TD for which the GP did not provide a (correct) pre-diagnose. Though GPs appear to improve their expertise in skin disorder detection after continued TDsc use, TDsc remains imperative to early melanoma detection.

Keywords:
Skin cancer, telemedicine, ICD-10

Introduction

Melanoma, a malignant tumor of the skin, evolves fast and is currently recognized as the deadliest type of skin cancer [1,2]. Early recognition and treatment is essential in improving these patient outcomes. When diagnosed in the early phase, melanoma is almost always curable. General Practitioners (GPs) have a vital role in detecting and diagnosing melanoma early, as often the first care contact patients visit. However, literature shows that GPs expertise in detecting melanoma is insufficient [3-5]. Previous studies showed low agreement between GPs and dermatologists in diagnosing suspicious skin lesions [4,5]. In addition to the inadequacy in melanoma detection, long waiting times before consultation of a dermatologist also limits early melanoma detection. A telehealth service that supports GPs in diagnosing melanoma could be a solution to optimize early detection and access to specialist skin care services.

Teledermoscopy consultation (TDsc) is a growing online service for melanoma detection. In TDsc, dermoscopic images (10-30x magnification) are sent via a secured internet connection to a (tele)dermatologist, who examines these suspicious skin lesions online. The (tele)dermatologist then provides a patient's caregiver with an accurate diagnosis and advice on the need for referral based on the assessment of the dermoscopic images [6]. This service would thus provide the GP with direct feedback on the correctness of their pre-diagnosis of a patient's skin disorder. Research has shown that GPs expertise in melanoma detection increases after training [3] and that fewer melanomas were missed by experienced and trained consultants [6]. It therefore remains a question, whether a telehealth service will also enhance GP expertise in melanoma detection after continual system use.

Ksyos is a healthcare organisation in the Netherlands which provides TDsc consultation between the GP and teledermatologist [7]. While performing TDsc the GP takes an overview, detailed and dermoscopic picture of the suspicious lesion and sends it together with some patient characteristics and a pre-diagnosis in the Ksyos digital health record system to a teledermatologist (TD). The TD assesses the consultation and provides a diagnosis. Recording this diagnosis is mandatory for the TD but optional for the GP. Figure 1 gives an overview of this TDsc consultation process. Data from the Ksyos digital health record system shows that GPs indicate that they learned from the TDsc consultations and the total number of teleconsultations per GP is indeed decreasing over time. However, the question is if frequently diagnosing TDsc consultations and exposure to this service increases the expertise of GPs in diagnosing skin lesions after continual use.

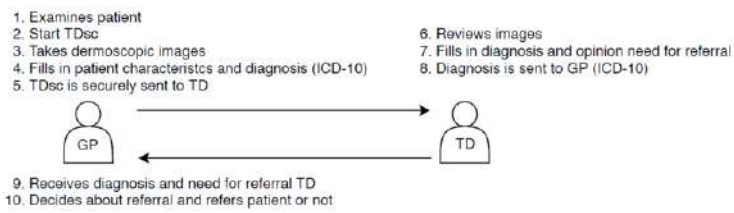


Figure 1 – Overview of the teledermoscopy consultation process between general practitioner (GP) and teledermatologist (TD)

Table 1 – 2x2 table to calculate sensitivity and positive predictive value

		Diagnosis TD (golden standard)	
		Diagnosis **confirmed** with ICD-10 code of TD	Diagnosis **not confirmed** with ICD-10 code of TD
Diagnosis GP	Diagnosis **present** according to ICD-10 pre-diagnosis of GP	**a**	**b**
	Diagnosis **absent** according to ICD-10 pre-diagnosis of GP	**c**	**d**

a = True positives, Diagnosis present according to ICD-10 pre-diagnosis of GP and confirmed by TD
b = False positives, Diagnosis present according to ICD-10 pre-diagnosis of GP and **not confirmed** by TD
c = False negatives, Diagnosis absent according to ICD-10 pre-diagnosis of GP and confirmed by TD
d = True negatives, Diagnosis absent according to ICD-10 pre-diagnosis of GP and **not confirmed** by TD

Positive predictive value (PPV) of GP = a / (a+b)
Sensitivity of GP = a / (a+c)

In the Netherlands, diagnoses given by the GP and TD within a TDsc consultation are classified to the corresponding ICD-10 code. The International Statistical Classification of Diseases and Related Health Problems (ICD) is an international diagnostic classification standard for reporting health conditions and diseases released by the World Health Organization [8]. This ICD-10 code can be used to analyse, compare and monitor diagnoses worldwide. Since July 2015, classifying diagnoses in compliance to this ICD-10 code has been mandatory in specialized care for reimbursement according to the Dutch Healthcare Authority (Dutch: Nederlandse Zorgautoriteit, NZa). And since this date, the 10th revision of this standard (ICD-10) has also been used in the Ksyos digital health record system. However, the effect of TDsc on GPs expertise in diagnosing skin disorders, based on the ICD-10 codes in the Netherlands has not been systematically investigated.

The aim of this paper is therefore to assess if, since the introduction of the ICD-10 codes in July 2015, GPs are gaining expertise in diagnosing (specific) skin problems conform ICD-10 after one, two and three years of TDsc use. In our analysis we specifically focused on the development in the number, ICD-10 type and correctness of GPs in pre-diagnosing skin disorders compared to the TD diagnoses. In doing so, we aimed to address the added value of continued use of TDsc for GP melanoma detection.

Methods

Setting and study population

We conducted a retrospective study in the Netherlands on the value of TDsc to GPs expertise in diagnosing skin disorders in three years' time. All TDsc consultations sent by affiliated GPs to TD between July 2015 and June 2018 and completed before the 18th of October 2018 were extracted from the Ksyos digital health record system.

A cohort of *experienced* GPs was selected for inclusion in the data analysis. GPs were classified as *experienced* if they performed at least five TDsc consultations for each of the included years, respectively (July 2015 - June 2016, July 2016 - June 2017, July 2017 - June 2018), and did not perform any consultation before January 2015. Unexperienced GPs were excluded in this study. The pre-diagnosis of the GP was compared with the diagnosis of the TD on the group level of the ICD-10 codes. Examples of group levels included are among others: C43-C44 *Melanoma and other malignant neoplasms of skin*, D00-D09 In situ neoplasms, D10-D36 *Benign neoplasms*, L20-L30 *Dermatitis and eczema*. The ICD-10 diagnosis of the TD was considered as the golden standard for the diagnosis confirmation.

Statistical analyses

The total percentage of cases a GP correctly diagnosed is calculated as the number of confirmed diagnosed GP cases by TD divided by the number of obtained TDsc cases minus the number of patients for which the TD did not provide a diagnose.

To assess the surplus value of continuing use of teledermoscopy the GP sensitivity and GP positive predictive value (PPV) in diagnosing skin disorders were calculated overall (total) and on the ICD-10 diagnosis group level for three subsequent years. The 95% confidence intervals (CI) of the PPV was calculated according to the Wilson score interval method without a correction for continuity [9].

The 2x2 table in table 1 shows the formulas used to calculate the PPV for each diagnosis category separately. The PPV was calculated as the proportion of consultations where the ICD-10 code was present according to the pre-diagnosis of the GP and confirmed by the TD divided by the total number of diagnosis scored within this group ICD-10 code by the GP. The sensitivity was calculated as the number of cases where the ICD-10 pre-diagnosis of the GP was confirmed by the TD divided by the total number of diagnosis scored within this group based on the ICD-10 codes of the TD.

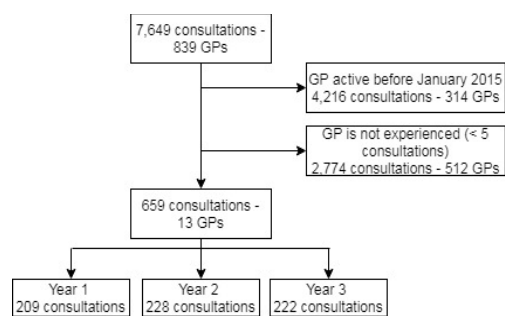


Figure 2 – Number of consultations

Results

In total 7,649 completed TDsc consultation requests by 839 GPs were extracted from the Ksyos digital health record system (figure 2). Overall, 314 GPs were active before January 2015 and for this reason the corresponding 4,216 consultations in the study period were excluded. In addition, 512 GPs did not perform five or more consultations for each study year respectively and were excluded. After exclusion, thirteen experienced GPs and 659 consultations were analysed (minimum of 20, maximum of 157 consultations per GP). The total number of consultations performed by the included GPs for each year were 209, 228, 222 respectively and were assessed among 32 TDs.

Table 2 – Number of (in)correct diagnosed GP cases and positive predictive value over three years' time

	Obtained TDsc cases	Not diagnosed by TD (%)	Not diagnosed by GP* (%)	Diagnosed by GP* (%)	Confirmed diagnosed GP cases by TD (%)	PPV [95% CI]	Incorrect pre-diagnosed by GP (%)
Year 1 July 2015- June 2016	209	26 (12.4)	170 (92.9)	13 (7.1)	11 (6.0)	0.85 [0.58-0.96]	2 (15.4)
Year 2 July 2016- June 2017	228	21 (9.2)	180 (87.0)	27 (13.0)	19 (9.2)	0.70 [0.52-0.84]	8 (29.6)
Year 3 July 2017- June 2018	222	33 (14.9)	155 (82.0)	34 (18.0)	25 (13.2)	0.74 [0.57-0.85]	9 (26.5)

Notes: PPV is calculated as the number of confirmed diagnosed GP cases by TD divided by the number of total obtained TDsc cases minus number not diagnosed by TD and GP.

* Total number = obtained TDsc cases minus the cases not diagnosed by TD

CI= confidence interval; **PPV**= positive predictive value

Sensitivity and PPV of GP in TDsc

No diagnosis was provided by the TD in 26 (12.4%), 21 (9.2%) and 33 (14.9%) of the cases. These cases were excluded from the statistical analyses. As shown in table 2, the percentage of confirmed GP diagnosed cases is increasing over years from 6.0 percent in year 1 to 13.2 percent in year 3. However, the GP's PPV is slightly decreasing over time from 0.85 [95% CI 0.58-0.96] to 0.74 [95% CI 0.57-0.85]. As presented in figure 3 the GP PPV though is increasing over time for diagnosis category L20-L30 *Dermatitis and eczema*. The number of incorrect pre-diagnosed cases by the GP is marginally increasing from 15.4 percent in the first year to 26.5 percent in the third year.

The sensitivity for all the diagnoses categories together increased from 0.07 in year 1 to 0.10 in year 2 and 0.14 in year 3. The sensitivity for the diagnosis categories C43-C44 *Melanoma and other malignant neoplasms of skin*, D10-D36 *Benign neoplasms*, L20-L30 *Dermatitis and eczema* (figure 4) also increased. The sensitivity for *melanoma* (C43-C44) improved from 0 out of 12 cases confirmed (0.0%) in year 1, 3 out of 22 cases (13.6%) in year 2, and 4 out of 16 cases confirmed (25.0%) in year 3 (total 7 out of 50 cases diagnosed as melanoma). The sensitivity for *Benign neoplasms* (D10-D36) improved from 5 out of 65 cases confirmed (7.7%) in the first year to 13 out of 79 cases in the third year (16.5%).

Added value of the teledermoscopy

Thirty-four diagnoses were scored by the TD to fourteen ICD-10 categories not chosen by the GPs (table 3). In these cases the GP did not fill in a diagnosis. Five consultations were scored by the TD as *no assessment possible* and for these cases no diagnosis was filled in by the GP as well.

The number of diagnoses inserted by the GP increased over time from 13 (7.1%) to 27 (13.0%) and 34 (18.0%) in year 1, 2 and 3 respectively. Conversely, this indicates to the added value of TDsc, which therefore starts out as relatively high as 92.9 percent of diagnoses are provided by the TD and were not filled in by GPs starting with TDsc. This number then slightly decreases after GP's continuing use of TDsc to 82.0 percent of the patients diagnosed by TD in year 3.

In three years' time, TDs diagnosed the majority of the patients as benign neoplasm (year 1 35.5%, year 2 43.0%, year 3 41.8%) or other disorders of the skin and subcutaneous tissue (year 1 37.7%, year 2 23.2%, year 3 33.3%).

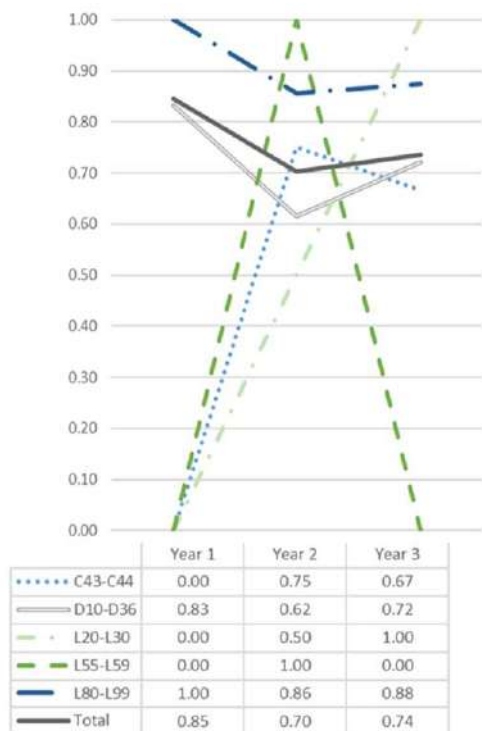
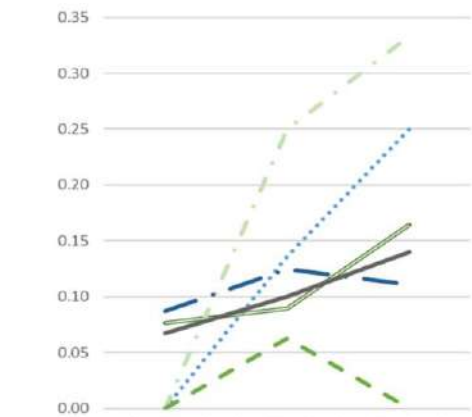


Figure 3 – PPV per diagnose category

In the first year, patients in TDsc were not diagnosed by the GP with the ICD-10 code *Melanoma and other malignant neoplasms of skin* (C43-C44). In the second year GPs diagnosed two consultations as benign neoplasms, while the TD diagnosed these patient cases as *melanoma or other malignant neoplasms of the skin* (C43-C44). In the third year one consultation was diagnosed as *radiation-related disorder* (L55-L59) by the GP, that was diagnosed as melanoma by the TD.

Overall, 12 (6.6%), 19 (9.2%) and 12 melanoma cases (6.3%) were diagnosed by the TD in three years' time respectively and not or not correctly pre-diagnosed by the GP. This suggests that in 7.4 percent of the total TDsc consultations performed in three years' time melanoma was detected, which were not and/or not-correctly pre-diagnosed by the GP. Also, 1 out of 3 (35.8%) of TDsc consultations were diagnosed as benign by the TD and not and/or not-correctly pre-diagnosed as such by the GP.



	Year 1	Year 2	Year 3
C43-C44	0.00	0.14	0.25
D10-D36	0.08	0.09	0.16
L20-L30	0.00	0.25	0.33
L55-L59	0.00	0.06	0.00
L80-L99	0.09	0.13	0.11
Total	0.07	0.10	0.14

Figure 4 – Sensitivity per diagnose category

Discussion

We conducted a retrospective study in the Netherlands to assess the potential value of TDsc to GPs expertise in diagnosing skin disorders over three years' time. Overall, of the total number of cases included in the study, the GPs provided more pre-diagnoses for each subsequent year from 7 percent in year 1 to 18 percent in year 3. Our study shows that the total percentage of correct pre-diagnosed cases of GPs is low, but during continual use of TDsc it slightly increased over time from 6 percent to 13 percent.

However, the number of incorrect pre-diagnosed cases also increased over time from 15.4 to 26.5 percent. This corresponds to a decrease in the overall PPV for all diagnosis categories. A possible explanation might be that after continual TDsc use, GPs might select the more difficult cases for TDsc and handle the less complex cases themselves. GP PPV specifically increased over time for diagnosis category L20-L30 *Dermatitis and eczema*. This may indicate that GPs become more attune to correctly diagnosing patients with dermatitis and eczema over time.

In general, GP sensitivity over all diagnosis categories showed a slight increase from 0.07 in year 1 to 0.14 in year 3. More importantly, GPs appear to become more sensitive in accurately pre-diagnosing skin disorders in the categories: C43-C44 *Melanoma and other malignant neoplasms of skin* and D10-D36 *Benign neoplasms*. As TDsc is especially important in early diagnosis of melanoma, these results are promising. When GP clinical expertise in recognizing melanoma improves, the sensitivity in melanoma detection increases. This finding is supported by a Cochrane review on accuracy of dermoscopy that shows that this sensitivity increases with more clinical expertise [6].

What we know from a previous study is that 95.1% of the GPs learned from the TDsc [10]. TDsc provides the GPs with direct feedback on the correctness of their pre-diagnosis by the confirmation of the TD. However, the system does not provide any active feedback on the GP performance. Literature on audit and feedback mechanisms shows that feedback leads to minor improvements in professional practice, but the effect is influenced by the way which the feedback is delivered [11]. New studies on analysing and advancing the effect of the feedback mechanism incorporated in TDsc consultations might lead to a higher GP learning curve.

Due to the rising number of correct pre-diagnosis of GPs, the percentage of TDsc in which the TD provides the GP with ICD-10 diagnoses decreases over time from 93 percent in the first year to 82 percent in the third year. This decrease might imply advancement in GP skin disorder diagnosing expertise. However, of the total TDsc consultations included in this study, in the subsequent three years 6.6%, 9.2% and 6.3% were diagnosed as melanoma by the TD for which the GP did not provide a (correct) pre-diagnose. In addition, two consultations which were pre-diagnosed by the GP as a benign neoplasm and a radiation-related disorder, were classified by the TD as *melanoma or other malignant neoplasm of the skin* (C43-C44). The high number of incorrectly diagnosed cases by GPs, indicates that the added value of TDsc after three years is still very high. However, these cases were not histopathological proven. Also, fourteen ICD-10 diagnosis categories were given by the TD which were not pre-diagnosed by GPs at all. This could indicate that GPs are unfamiliar with these diagnoses.

The results of this study reveal a potential learning effect of TDsc on GP skin disorder diagnose expertise. An increase in the number of pre-diagnosed consultations by the GP, an increase in diagnosis sensitivity and a modest increase in positive predictive value after three years for specific diagnose categories of TDsc usage were seen. However, overall, GPs pre-diagnose expertise of skin disorders appeared low in this study. Especially *Benign neoplasms* (D10-D36) and *other disorders of the skin and subcutaneous tissue* (L80-L99) appear difficult to diagnose by the GPs since a pre-diagnosis for these disorders was often lacking.

In this study GP were not obliged to fill in a pre-diagnosis, but the number of provided pre-diagnoses appears comparable to normal practice. In the study of Rijsingen et al. they assessed the quality of referral letters of GPs to the dermatologist of patients with skin tumours [4]. Their study showed that GPs do not always provide a diagnosis for suspicious lesions in referral letters to the dermatologist. A diagnosis was missing in 18.3% of the cases. In addition, only two out of eight melanoma were correctly pre-diagnosed in the GP referral letters. In our study, GPs detected 14 percent of all melanoma diagnosed by TDsc. The positive predictive value of GPs in melanoma detection for both studies was equal, 0.67.

Table 3 – Diagnosis categories chosen by Teledermatologist (TD) and not by General Practitioner (GP)

Category	Description
B00-B09	Viral infections characterized by skin and mucous membrane lesions
B35-B49	Mycoses
C81-C96	Malignant neoplasms, stated or presumed to be primary, of lymphoid, haematopoietic and related tissue
D65-D69	Coagulation defects, purpura and other haemorrhagic conditions
D70-D77	Other diseases of blood and blood-forming organs
H60-H62	Diseases of external ear
I80-I89	Diseases of veins, lymphatic vessels and lymph nodes, not elsewhere classified
L00-L08	Infections of the skin and subcutaneous tissue
L40-L45	Papulosquamous disorders
L60-L75	Disorders of skin appendages
M30-M36	Systemic connective tissue disorders
O94-O99	Other obstetric conditions, not elsewhere classified
Q80-Q89	Other congenital malformations
T08-T14	Injuries to unspecified part of trunk, limb or body region
	No abnormalities

One of the limitations of this study is that in Dutch GP practice diagnoses are registered according to the International Classification of Primary Care (ICPC) and recorded according to SOAP notes (Subjective, Objective, Assessment, Plan). The ICD-10 classification of diseases was thus a new classification method where GPs were unfamiliar with, when it was implemented in 2015. GPs becoming more familiar with ICD-10 coding might therefore have contributed to the increase in the number of (correct) ICD-10 pre-diagnoses registered by the GPs in our study. Hence, in this study we only included GPs who started with TDsc after the ICD-10 was implemented in the Ksyos system, to research how their pre-diagnosing patterns changed according to the ICD-10 coding in three years' time. Also, GPs might use TDsc in general practice to fasten the face-to-face consultation with the patient and not solely for support in diagnosing. If a GP correctly pre-diagnosed the patient, but did not fill-in this diagnosis in the system, this would affect the GP PPV in diagnosing skin disorders. During the time of our study it was not mandatory for the GP to fill in the ICD-10 code in the system and this would thus not be seen in our data. Though this study has several limitations, the strength of the study is that we had a large database available of TDsc users and were able to include only GPs who started with TDsc when the ICD-10 coding system was first implemented in the Ksyos system and had continued and frequent use of the system in the past three years. We could therefore accurately address their progress in expertise in pre-diagnosing skin disorders based on ICD-10 coding.

Conclusions

TDsc supports GPs in assessing suspicious lesions of patients and their need for referral to a dermatologist. Continual use of TDsc over the years appears to slightly enhance GP sensitivity in diagnosing skin disorders based on ICD-10 coding. However, GPs PPV for the main ICD-10 codes showed a decrease over the years. Though GPs become more perceptive in recognizing benign neoplasms (D10-D36) and melanoma (C43-C44), TDsc detected a high number of melanoma not correctly pre-diagnosed or otherwise not detected by GPs in this cohort. Hence, TDsc has the potential to enhance GP expertise in skin disorder diagnosing, but remains essential in early melanoma detection even after GP continued and frequent TDsc use.

References

[1] Dutch public health and health care https://www.volksgezondheidenzorg.info/onderwerp/huidkanker/cijfers-context/sterfte-en-overleving#!node-sterfte-huidkanker (last accessed 25th of November 2018)

[2] National Cancer Institute Cancer Stat Facts: Melanoma of the Skin https://seer.cancer.gov/statfacts/html/melan.html (last accessed 25th of November 2018)

[3] P. Carli, V. de Giorgi, E. Crocetti, L. Caldini, C. Ressel, and B. Giannotti, Diagnostic and referral accuracy of family doctors in melanoma screening: effect of a short formal training, *Eur J Cancer Prev* **14** (2005), 51–55.

[4] M.C.J. van Rijsingen, S.C.A. Hanssen, J.M.M. Groenewoud, G. van der Wilt, and M.P. Gerritsen, Referrals by general practitioners for suspicious skin lesions: the urgency of training, *Acta Derm Venereol* **94** (2014), 138–141.

[5] A.F. Duarte, A. da Costa-Pereira, V. Del-Marmol, and O. Correia, Are General Physicians Prepared for Struggling Skin Cancer?-Cross-Sectional Study, *J Cancer Educ* **33** (2016), 321-324.

[6] J. Dinnes, J.J. Deeks, N. Chuchu, L. Ferrante di Ruffano, R.N. Matin, D.R. Thomson, et al., Dermoscopy, with and without visual inspection, for diagnosing melanoma in adults, *Cochrane Database Syst Rev* **12** (2018).

[7] Ksyos healthcare organization https://www.ksyos.nl/en/ (last accessed 23rd of November 2018)

[8] The International Statistical Classification of Diseases and Related Health Problems (ICD) http://www.who.int/classifications/icd/en/ (last accessed 23rd of November 2018)

[9] E.B. Wilson, Probable Inference, the Law of Succession, and Statistical Inference, *Journal of the American Statistical Association* **22** (1927), 209-212.

[10] E. Tensen, J.P. van der Heijden, M.W.M. Jaspers, and L. Witkamp, Teledermatologie, *Bijblijven* **33** (2017), 101-113.

[11] N. Ivers, G Jamtvedt, S. Flottorp, J.M. Young, J. Odgaard-Jensen, S.D. French SD, et al., Audit and feedback: effects on professional practice and healthcare outcomes, *Cochrane Database Syst Rev* **6** (2012).

Address for correspondence
Esmée Tensen, e.tensen@amc.uva.nl, +31 20 5666204

Conflicts of interests E.T., F.v.S are employed (part-time) by Ksyos, and L.W. is the director of Ksyos. The remaining authors state no conflicts of interest.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190341

Decentralized Privacy-Preserving Platform for Clinical Data Sharing and Analysis

Sui Jun Tong[a], Yong Yang[a], Wen Sun[a], Eryu Xia[a], Shao Chun Li[a]

[a] *IBM Research - China, Beijing, China*

Abstract

Collection and management of clinical data for administration and analysis is a time-consuming and complex task, especially when multiple data providers been involved. Even if people are willing to take on the burden for it, there is still no mature solution to protect data privacy for distributed data providers. Distributed ledger is an emerging technology that supports decentralized data sharing and management. Based on this, we present a platform which enables distributed and truthful data collection and serves privacy-preserving needs in clinical data management. Our system, built on Hyperledger Fabric, used smart contract to execute data aggregation and provide basic analysis methods. The system used ledger and world status to record data access history and other metadata. This decentralized platform enables data providers to proactively share and protect their data, Thus can simplify clinical data collection procedure and promote efficient collaboration between providers.

Keywords:

Data Sharing, Clinical Data Analysis, Decentralized Data Management

Introduction

Data analysis of electronic medical records has shown its great power in risk prediction, new treatment or cure detecting, and precision medicine [1]. To validate results and reduce bias, collecting data from multiple sources is a desired approach in such clinical research. However, since medical data are mostly isolated and stored with different healthcare providers, data collection and sharing for analysis is usually a time-consuming and complex task.

Even if people are willing to take on the burden for such tasks, traditional centralized data collection still has some weaknesses. First, data leaks may occur during the data sharing and consumption process; technically, data providers have no idea on who accessed the data. Second, the data authenticity and quality cannot be guaranteed. It is difficult to detect in a timely manner if the dataset has been modified or faked. Finally, the centralized approach sometimes prevents data providers from a fair collaboration, and the contributions of different providers are not traceable. Hence, a more transparent and trustful way for clinical data collection, sharing, processing, and analysis is highly desirable.

The emerging distributed ledger technology may offer a solution for the above issues. Derived from Bitcoin [2], distributed ledger technology, such as blockchain, is now widespread in both financial and non-financial areas. Popular blockchain frameworks, like Hyperledger Fabric [3,4] and Ethereum [5,6], allow people to build customized blockchain networks with self-defined smart contracts. The possibility of using blockchain for healthcare data management has recently gained more attention in both industry and academia [8, 9, 10, 11]. However, how to leverage blockchain to build a decentralized, privacy-preserving platform for clinical data sharing and analysis still remains to be explored.

In our work, we present a prototype platform which enables distributed and truthful data collection and serves privacy-preserving needs in clinical data management. We build our platform based on Hyperledger Fabric [4], which allows data providers to join the network by loading their data to a local peer. To guarantee data security and privacy, source data will not be exposed to a public network. Data consumers can only use the predefined smart contract to get access to the data, and all the data transmissions in our system will be encrypted. The ledger stores the metadata, access APIs, and access control policy of all datasets. All data access will be recorded to the ledger, which can be easily traced and verified. In addition, since the contributions of the distributed datasets can be defined and calculated, reward mechanisms can also be devised to encourage data providers to collaborate with each other. The decentralized nature of the proposed platform makes it easy to be deployed and scaled up in practice.

Methods

Compared to other medical data management systems [1,7,8], the proposed platform provides a secure, trustful, and convenient way for clinical data sharing and analysis. The framework of the proposed platform can be found in *Figure 1*.

As shown in Figure 1, the distributed data are integrated into the proposed platform in the data storage layer. The data providers, such as hospitals, keep their data on local databases and provide secure connections for local peers. The database servers are not part of the blockchain network, and thus cannot be accessed by other peers from other origins.

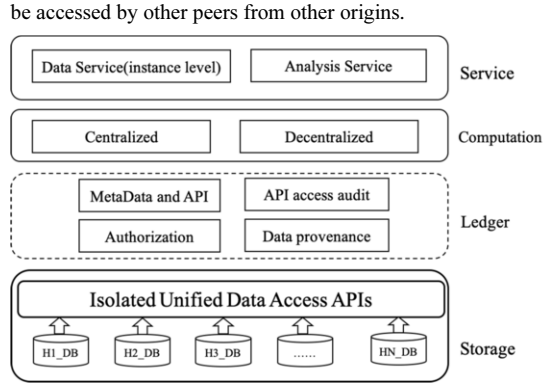


Figure 1– Framework of clinical data sharing platform

Data stored in the local databases will follow a common schema, so we can use unified data access APIs to query the data. Data providers are free to join or leave the network, and their data can no longer be accessed after they leave the platform.

The ledger layer can be generally regarded as the permissioned blockchain services, which is the main component of our platform. In our platform, ledger services are customized for different purposes. First, we use the ledger to store the metadata and APIs. After loading their dataset to the local databases, data providers need to invoke a transaction to record the metadata to the ledger, which contain the data schema and basic information of the loaded data. Each kind of data schema is associated with a list of APIs, which are predefined and announced with the data schema. Metadata and APIs are represented as the World States in our HyperLedger Fabric [3,4] implementation, and stored in CouchDB [4]. The ledger also record the authorization policy for the dataset. Access control can be managed by channel, by organization, or even by a certain peer. When a user invokes a query in our platform, the smart contract will verify the authorization of the user from the ledger. When transactions are successfully executed, the transaction information, such as transaction invoker, accessed datasets, and results will be recorded to the ledger, which can be used for auditing and contribution calculations. Thanks to the irreversibility and undeniability of blockchain, all the data recorded in the ledger are trustful. The final results can also be associated with a list of ledger transactions, which can be easily used to verify the data provenance and the analysis process.

Based on the distributed ledger layer, we offer two kinds of computing engines by different smart contracts. The centralized computing engine requires the driver to send sub-queries to different data provider peers, aggregate the query results, and return final results to the requester. It is similar to the MapReduce [12] computing framework. The decentralized computing engine is used to manage data providers' own information, executes typical blockchain transactions happened inside the local peer, and directly records the results to the distributed ledger.

We also provide two kinds of data services targeted to two types of demands. Instance level data services support ad hoc queries over row-level data records, and can be dispatched and executed in a local dataset. The analysis services contain fine-grained statistical APIs and predefined modeling pipelines which are more complicated and may involve multiple datasets. With the common data schema, a list of APIs, for both instance level data services and analysis services, are published to all the providers. The data provider does not have to accept all the APIs; they can choose the APIs that they would like to support during the dataset register, and disable some APIs if they have any concerns.

Results

Our prototype is easily implemented and receptive to adaptations and extensions in practice. Here we demonstrated an implementation on HyperLedger Fabric, upon which we simulated a research pipeline using a real-world dataset.

Implementation

The key components of our implementation include a permissioned blockchain network and functional smart contracts. First, we built a Fabric network to grant all the healthcare stakeholders the ability to join the platform. Then we write Chaincode as Smart Contracts to manage the data access APIs and providing data services on our demands.

Network Setup

Generally speaking, a clinical data collection and sharing platform has three kinds of stakeholders:

- **Data provider.** This can be a hospital or other healthcare providers who has the medical records and is willing to share the data with authorized users for research purposes. The data provider stores the source data and takes the responsibility for the data privacy and security.
- **Data consumer.** This can be a researcher who needs real-world clinical records for research, such as an insurance company or a pharmaceutical company for a business need, or a healthcare administrator for public health surveillance. A hospital can also be a data consumer. A data consumer uses the data and needs to pay for data services.
- **Miner.** This is a trustful third party who does not keep any data and only provides computing services, A miner can be a cloud service provider or other powerful server nodes.

Accordingly, our blockchain network consists of three kinds of nodes:

1. Data peer nodes, which store ledger data and maintain a connection to their local databases;
2. Miner nodes, which store ledger data, have the access to query or write the ledger data, but cannot access the source data; and
3. Client nodes, which provide data services, receive data query requests, and invoke transactions with miner nodes.

Each node has a unique identity, which cannot be faked by others. For security purposes, a network can have many channels, and each channel can share the same ledger. It is possible to isolate data access records by ledger. To build a Fabric network, we need to first generate the channel configuration, then create encryption materials and certification for the origin of each peer.

Smart Contracts

Smart contracts are the business logic of a blockchain network. In our platform, metadata and the access APIs are regarded as the data provider's assets. Smart contracts need to specify the data schema design, define what kind of assets the data provider should providing, and release a list of associated APIs. A smart contract also contains a detailed scoring or charge mechanism. As mentioned above, there are many different kinds of smart contracts serving different functions (also known as chaincode in Hyperledger Fabric framework):

- **Data chaincode.** This is installed and run on the data provider peer and has direct access to the source data. Each data provider installs their unique data chaincode, which holds the connection for its local dataset.
- **Service chaincode.** This is installed and run on the miner peer and acts as the central service. It receives queries from clients, sends transaction proposals to data chaincode, aggregates data, and returns final results to the client. Service chaincode uses a "chaincode invoke chaincode" way to execute a client's query and only return final results to the requester. Intermediate data will not be shown to the client side. Many miners could be equipped with the same chaincode.
- **Miner.** This is installed and run on every peer. It controls the dataset register and manages assets and

access records for audit purposes. The whole channel using the same management chaincode.

Transaction Flow

At the beginning, each data provider should load its datasets to the platform. Before registration, all the data providers should reach an agreement on the data schema, which can be defined based on standard practices or by some authoritative organizations. The common data schema would then be released to the platform.

Figure 2 shows a typical dataset registration transaction flow, which includes the following steps:

1. Data schema holder (could be any user) connects to network through Client/SDK.
2. Client/SDK proposes a request to management chaincode to create new schema.
3. Management chaincode creates a new table and updates it to the ledger.
4. Data provider formats their local datasets accordingly upon notification of the schema.
5. Data provider proposes a transaction request through Client/SDK to register an new dataset.
6. Data chaincode creates connection for local databases and updates metadata of the local datasets to the ledger.

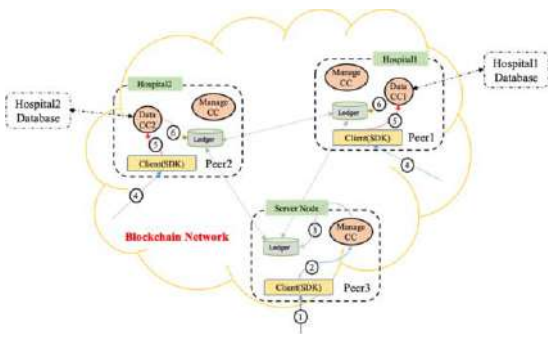


Figure 2– Transaction flow of dataset register

After loading the dataset, some predefined query can be executed on our platform. A common query workflow is shown as in *Figure 3*, which includes the following steps:

1. Data requester connects to network through Client SDK.
2. Client SDK proposes a request to service chaincode.
3. Service chaincode obtains all the available data chaincodes and proposes child query transaction to them.
4. Data chaincode answers the transaction proposal and returns query results to the service chaincode.
5. Service chaincode aggregates all the intermediate data, and returns final results to the requester.

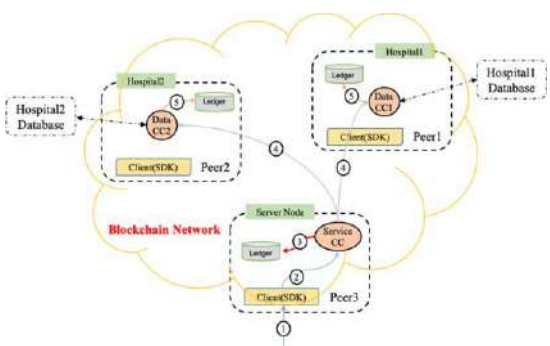


Figure 3– Transaction flow of query

Simulation

To demonstrate the use of our prototype, we simulate a use case using a real dataset. The medical record data was independently collected from three hospitals (below referred to as data centers) and organized based on the same data schema.

In this scenario, we create a channel with four organizations, include three data provider organizations and a data service organization. Each data provider has one peer node in the network. When the data provider detects a new data schema in the ledger, it first loads its own data to the local database (relation database such as MySql), and then registers the dataset with a secure connection to that database server via its peer. This peer will run a data chaincode after dataset registration, which is used to interact with the local database. Besides, we have another peer that provides data query and analysis services. That peer installs a server chaincode to call data chaincode and aggregates the results.

Table 1– Execution time for sample queries

Query	Parameter	Result	Avg. Execution Time (s)
getSum	Null	26974	3
count	mace=1	955	3
getRowdata	age<60	{rowdata}	12
getAvg	Age	67.3	3
getColumn	mace	{column data}	10
getLrModel	x_xolumns, y_xolumns	{model weight}	300
evaluate	model	auc:0.862	5

Table 1 shows the execution time for some sample queries. Here three kind of basic queries as well as LR model training and evaluation is used for testing. *GetSum* is use to compute the sum of all the dataset. *Count* is used to count the total records that satisfy the given constraint. *GetRowdata* is used to query instance level data and returns all the row data that satisfy the given constraint. *GetAvg* returns the mean value of a given column. *GetLrModel* performs the training process of a logistic regression model on the given columns. *Evaluate* applies a given model on the whole dataset, and returns the AUC score as the model evaluation result.

As we can see from the table, basic statistical queries could be performed with relatively low latency. But the performance was worse when the data processing logic became more complicated, such as complex instance level data services or modelling tasks. Considering the source data are scattered in multiple distributed data sources, a couple of minutes for a model training task should still be acceptable for typical use cases.

Conclusions

In our work, we present a prototype platform which enables distributed and truthful data collection and serves privacy-preserving needs in clinical data management. Through this, a researcher can easily collect the data they want, and the data provider will get fair payment and also keep their source data safe. Once we have enough datasets and performed further research on the platform, we can also use our platform as an audit system for clinical research reports.

References

[1] The Office of the Nat. Coordinator for Health Information Technology, "Report on health information blocking," U.S. Department of HHS, Tech. Rep., 2015.

[2] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system" White Paper, 2008.

[3] Elli Androulaki, Artem Barger, Vita Bortnikov, Christian Cachin, Konstantinos Christidis, Angelo De Caro, David Enyeart, Christopher Ferris, Gennady Laventman, Yacov Manevich, Srinivasan Muralidharan, Chet Murthy, Binh Nguyen, Manish Sethi, Gari Singh, Keith Smith, Alessandro Sorniotti, Chrysoula Stathakopoulou, Marko Vukoli´c, Sharon Weed Cocco, Jason Yellick, "Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains," arXiv:1801.10228 [cs.DC], 2018

[4] Hyperledger. https://www.hyperledger.org/projects/fabric

[5] Ethereum. https://www.ethereum.org

[6] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," Ethereum Project Yellow Paper, 2014.

[7] James Cunningham, John Ainsworth, "Enabling Patient Control of Personal Electronic Health Records through Distributed Ledger Technology," International Medical Informatics Association (IMIA) and IOS Press. 2017

[8] Asaph Azaria, Ariel Ekblaw, Thiago Vieira and Andrew Lippman, "MedRec: Using Blockchain for Medical Data Access and Permission Management," 2nd International Conference on Open and Big Data, 2016

[9] GEM. https://gem.co/health/

[10] Yue X,Wang H, Jin D, Li M, JiangW, "Healthcare data gateways: Found healthcare intelligence on blockchain with novel privacy risk control," *Journal of Medical Systems*, **40**(10):218, 2016.

[11] Beninger P, Ibara MA, "Pharmacovigilance and biomedical informatics: a model for future development," Clinical Therapeutics, 2016.

[12] Jeffrey Dean, Sanjay Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," in Proc. 6th USENIX Symp. on Operating Syst. Design and Impl., 2004, 137–150.

Address for correspondence

Shao Chun Li, IBM Research – China, 28, Zhongguancun Software Park, Haidian District, Beijing, China. Email: lishaoc@cn.ibm.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190342

User-Centered Design of the C3-Cloud Platform for Elderly with Multiple Diseases – Functional Requirements and Application Testing

Lamine Traore[a], Ariane Assele-Kama[a], Sarah N. Lim Choi Keung[b], Liran Karni[c], Gunnar O. Klein[c], Mikael Lilja[d], Isabella Scandurra[c], Dolores Verdoy[e], Mustafa Yuksel[f], Theodoros N. Arvanitis[b], Rosy Tsopra[a, g], Marie-Christine Jaulent[a]

[a] Inserm, Sorbonne Université, Univ Paris 13, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances pour la e-Santé, LIMICS, F-75011 Paris, France
[b] Institute of Digital Healthcare, WMG, University of Warwick, Coventry, UK
[c] Örebro University School of Business, Informatics, Örebro, Sweden
[d] Department of Public Health and Clinical Medicine, Unit of Research, Education, and Development Östersund Hospital, Umeå University, Umeå, Sweden
[e] Asociacion Centro De Excelencia Internacional En Investigacion Sobre Cronicidad – Kronikgune, Spain
[f] SRDC Software Research Development & Consultancy Corp, Ankara, Turkey
[g] AP-HP, Assistance Publique des Hôpitaux de Paris, Paris, France

Abstract

The number of patients with multimorbidity has been steadily increasing in the modern aging societies. The European C3-Cloud project provides a multidisciplinary and patient-centered "Collaborative Care and Cure-system" for the management of elderly with multimorbidity, enabling continuous coordination of care activities between multidisciplinary care teams (MDTs), patients and informal caregivers (ICG). In this study various components of the infrastructure were tested to fulfill the functional requirements and the entire system was subjected to an early application testing involving different groups of end-users. MDTs from participating European regions were involved in requirement elicitation and test formulation, resulting in 57 questions, distributed via an internet platform to 48 test participants (22 MDTs, 26 patients) from three pilot sites. The results indicate a high level of satisfaction with all components. Early testing also provided feedback for technical improvement of the entire system, and the paper points out useful evaluation methods.

Keywords:

Multimorbidity; Evaluation Studies; User-computer Interface.

Introduction

The World Health Organization estimates that 63% of all annual deaths (~36 million people) is attributable to non-communicable, chronic diseases [1] and the number of people with multiple comorbidities has increased considerably for some years, mainly due to the ageing of the population [2]. Elderly patients with multimorbidity (i.e., having at least two chronic diseases) [3] are at higher risk of safety incidents [2]. These could include incidents such as delayed diagnosis, no recommended treatment, drug side effects, drug interactions, over/under dosage of drugs, complications, infections, etc [2,4]. Increased risks in patient safety, in this context, may be explained by many reasons. Firstly, patients with multimorbidity are often polymedicated, with a potential decrease in treatment adherence and a possible increase of drug side effects. Secondly, patients may receive contradictory advice or treatments, due to the application of different guidelines that are designed to manage only single disease pathways. Thirdly, patients with multimorbidity are often cared for by several health and social care (HSC) professionals, who are not always coordinating and communicating throughout the patients' journey. For example, there is often a lack of communication between general practitioners (GPs) and secondary care specialist centres. Finally, these patients are often more vulnerable than others, due to their multiple diseases and their advanced age, which makes the care process even more complex.

C3-Cloud[1] is a European Commission supported Horizon 2020 research and innovation project, which aims at improving the provision of integrated care to patients with multimorbidity via enhanced ICT solutions. The research aims at resolving guidelines' conflicts (by reconciliation of varying recommendations from individual disease guidelines), supporting clinical decision making through clinical decision support services, and facilitating communication among multidisciplinary care team (MDT) members through an interoperable platform, which integrates patients' health records from existing Electronic Health Record (EHR) systems [5]. The project mainly focuses on elderly patients (65+) with diabetes, heart failure, renal failure and depression in different comorbidity combinations. Three European pilot sites are involved in the study: Osakidetza (Basque Country, Spain), RJH (Region Jämtland Härjedalen, Sweden) and SWFT (South Warwickshire NHS Foundation Trust, UK).

The C3-Cloud system consists of a variety of components:

- Coordinated Care and Cure Delivery Platform (C3DP),
- Patient Empowerment Platform (PEP),
- Clinical Decision Support Modules (CDSM),

[1] http://C3-Cloud.eu, Federated Collaborative Care Cure Cloud architecture for the needs of multi-morbidity and managing poly-pharmacy.

- Interoperability Middleware, which includes modules of technical and semantic interoperability, as well as privacy and security.

All these components constitute the solution that will be used for the technological trial of the C3-Cloud application. Following a user-centered development (UCD) approach [6], the solution has to be evaluated iteratively during its life cycle; during development and implementation with a restricted number of participants, during deployment with a larger number of users as well as during the routine phase for prospective cost-benefit analyses and real impacts.

The objective of this study is three-fold: the report of the user-centered functionality testing, the conclusions to further improve the C3-Cloud system and, for the community, to present useful methods among the UCD evaluation methods often used in health informatics.

Methods

This evaluation consisted of a number of questionnaires to collect user feedback regarding system functionality. The questionnaires were created based on a Delphi approach [7]. In the context of the project, a total of 51 pilot application requirements (PARs) and 72 use cases were defined covering the scope of all high level C3-Cloud components to depict expected functionality. Three different types of users interact with the system: multidisciplinary team members (MDTs) also known as health and social care professionals; patients; and informal care givers (ICGs), and are accordingly studied. This is in line with previous informatics research, e.g. OLD@HOME[8–10] and more recent research regarding patient access to health information [11–13]

This qualitative inductive study directly separated the demands or requirements with the use cases needed. For example, "as a patient, I need to access drug interaction information" is a PAR for the pilot sites and "Enabling patients to access self-management material" is a use case of the PEP component. A full list of user scenario descriptions and PARs is presented in deliverable D8.1 [14].

Evaluation Procedure

Following the Delphi approach [7], the evaluation framework was developed:

Brainstorming: Formulation and Evaluation of an Initial List of Relevant Questions

Based on an analysis of 51 PARs and 72 use cases, the first step was to formulate a list of questions, starting by a simple mapping (1 PAR to 1 use case) to identify possible links between the PARs. Secondly, the questions were grouped by profiles identified during the PARs' analysis process, in order to define one questionnaire per user profile. The workflow was based on a C3-Cloud key scenario linked to a use case and defined by application testing criteria.

Refining and Prioritization: Internal Review Based on a Cognitive Walkthrough by Experts

The initial questions were reviewed by the three pilot sites as well as by the technical partners of the C3-Cloud project. The results of the cognitive walkthrough [11] by five IT and clinical reviewers, allowed us to validate, modify, delete or add questions based on the updated PAR list and covered system functionality. Based on the review from the aforementioned experts, although addressing different professionals and individuals, the questionnaires could be appropriated to, two user profiles: 1) MDTs and 2) Patients & ICGs, as grouped respondents of the questions.

Think Aloud

During the test sessions, the think aloud method [11] was encouraged and the pilot site managers, who moderated the sessions, noted all comments of the participants. Feedback from the different pilot sites could be complementary. If feedback was raised more than once, it was reported only once. Examples, issued from the feedback, and how they were handled by technical partners to improve the C3-Cloud components, are reported in the Result section.

Evaluation Set-Up

We implemented an online application that allowed participants to answer questionnaires. The application site is available at https://c3cloud.irsan.eu. In the questionnaire, participants responded with [Yes], [No] and [NA] (for functionality, which was Not Available). When the response was [No], both MDTs and Patients/ICGs had the opportunity to specify and explain the problem faced by writing free text comments.

Participants

Overall, 26 elderly patients (> 65 years) and 22 MDTs from the three pilot sites: Osakidetza, RJH, and SWFT; participated in the testing. At the time of testing, only an English-language demonstrator and materials were available, and local sites considered this when recruiting test participants.

Test Sessions

The participants received login credentials for the online demonstrators of the C3DP (for MDTs) and the PEP (for Patients/ICGs) as well as training materials including a walkthrough that guided them through certain activities on the demonstrators. For the Osakidetza and RJH pilot sites, a language facilitator moderated each session, and was available for translation of the material and any other question raised by the participants. Think aloud notes taken during the test sessions generated a summary report.

Results

Overall 57 questions were formulated; 33 for MDTs and 24 for Patients & ICGs. Below, detailed results of the application testing for MDTs and Patients/ICGs, respectively, are reported, as well as examples of the questions in the questionnaires.

Evaluation by MDTs

Questions were categorized by the following main topics: Care Plan; Decision Support Module; Patient Data; Communication; and Notifications. The MDT responses [Yes], [No] or [NA] are reported, in percentages, in Table 1.

Table 1– Summary of MDTs average response rates to C3-Cloud application testing

MDTs' Response rates			
C3DP Categories	[Yes]	[No]	[NA]
Care plan	94 %	4%	2%
Decision Support Module	72%	0%	28%
Patient Data	75%	6%	19%
Communication	79%	3%	18%
Notification	89%	5%	6%

Questions related to Care Plan received the highest amount of [Yes] responses, 94%. The following questions received positive responses by all (N=22) participants:

- *"Are you able to create a new specialized care plan for the patient?"*
- *"Are you able to define new or update the planned intervention?"*
- *"Are you able to define new or update self-care activities (like exercise recommendations)?"*

The question "Are you able to update an existing care plan?" scored: [Yes] 82%, [No] 9% and [NA] 9%. The [No] responses were complemented with MDT comments, revealing that some update functionalities were missing. For example:

- *"I can update some elements of the plan as goals and activities and training material, but I cannot update the health conditions of the patient."*

From the free text feedback, further details about "health conditions," in a new care plan creation, were considered, together with improvement proposals from MDTs and responses from the technical partners. For example:

- MDT proposal: *"When creating a new care plan 'Addressed Conditions' may need rephrasing. It is also unclear how it is decided what conditions the list here suggests. The list can be very long if many conditions apply or are possible."*
- Technical partner improvement feedback: *"Addressed conditions will be removed, the SNOMED-CT codes of the 4 main diseases will be added."* Regarding the Decision Support Module, an average of 72% of the participants provided [Yes] answers to questions such as:
- "Does the Clinical Decision Support Module give you advice about treatment options such as i) new safety, treatment or lifestyle? ii) starting/stopping of medication, based on the most recent context of the patient including changes in recent remote monitoring results?"

There were zero [No] answers and [NA] responses were rated at 28% on average, meaning that all accessible functionalities were approved by the participants.

For Patient Data, average responses were [Yes] 75%, [No] 6% and 19% for [NA]. There were highly rated [Yes] responses >90% for questions such as:

- "Are you able to access patient data after the Care Plan Manager approved your membership to the Multidisciplinary Care Team?"
- "Are you able to review the patient's Health Records?"

A lower percentage of [Yes] responses were received, for example, regarding questions like:

- "Are you able to access the readings, that have been uploaded by patients manually or via remote monitoring systems such as wireless medical sensor devices?" with [Yes] 55% respective [NA] 36%.
- "Are you able to follow-up patients' activities, such as complications, side effects via questionnaires?" 68% responded [Yes] and 18% [No] which were completed with MDT comments like "Not recorded" and "Did not find them..."
- "Are you able to access information completed by the patient such as files uploaded via the PEP?" 64% responded [Yes] and 32% responded [NA].

Communication questions were related to message exchange between the MDT members or invitations to another specialist to join the care team. On average, 79% responded [Yes], 3% [No] and 18% responded [NA]. Questions regarded, e.g.:

- "Messaging - Are you able to send messages to other members of the MDT via asynchronous messaging?"
- "Invitation - Are you able to invite another specialist to join the patient's Care Team?"

Finally, responses related to **Notification** functionalities received an average of 89% [Yes], with questions such as

- "Are you able to notify the existence of the updated care plan to Care Team Members and to the patient?"
- All participants (100%) answered positively on
- "Can you see the upcoming activities in your calendar and in the Activities section of your dashboard?"

Evaluation by Patients and ICGs

For the Patient Empowerment Platform (PEP), Patient and ICG responses were categorized in the following main topics: Care Plan; Patient Empowerment; Patient Data; Notifications and Communication. The responses [Yes], [No] or [NA] are reported, in percentages, in Table 2.

Table 2 – Summary of Patients & ICGs average response rates to C3-Cloud application testing

Patients and ICGs Response rates			
PEP Categories	[Yes]	[No]	[NA]
Care plan	87%	4%	10%
Patient Empowerment	68%	8%	24%
Patient Data	48%	4%	48%
Communication and Notification	32%	5%	63%

Questions related to the Care Plan received average responses of 87% [Yes], 4% [No] and 10% [NA]. Examples of questions:

- "As a Patient or Informal Care Giver, are you able to access the care plan?" 96% of participants responded positively, and 4% responded [NA]
- "Do you receive enough advice and support about how to follow the care plan?"

[No] responses were provided, with comments such as:

- Not clear how to work through the system.
- Not obvious what needs to be done, when and how. Some guidance notes would be helpful.
- More time is needed.
- Needs to be more simple with clear single click pathways trough each of the components for the patient. A lot of the technical material in each patient activity is not needed by the patient and therefore confusing.

Patient Empowerment average responses were 68% of [Yes], 8% of [No] and 24% of [NA]. Example of questions:

- "Do you think that the information given could help you to improve your health and wellbeing?"
- 85% responded [Yes], 15% responded [NA].
- "Are you able to learn about treatment options through the PEP?" 61% [Yes], 8% [No], 31 %[NA].

- "Are you able to learn about drug benefits through the PEP?" 58% [Yes], 11% [No], 31 %[NA].
- [No] responses were completed, with patients' comments as: - *Not recorded, - Depends on which materials are presented, - I have not stopped/answered.*

Patient Data gathered average response rates of 48% for [Yes], 4% for [No] and 48% responded [NA]. Some examples

- "Are you able to access the readings uploaded to the system from remote monitoring systems (e.g., wireless medical sensor devices) from the PEP?"
- 42% responded [Yes], 8% responded [No], 42% lacked the functionality with [NA] responses.
- "Are you able to upload documents, such as a picture, to your PEP?" 54% responded [Yes] and 46% [NA].
- [No] responses were completed, with patients' comments such as: *"Not available", "I have not tried it*".

Communication and Notification average responses were 32% [Yes], 5% [No] and 63% [NA]. Example of questions:

- "Messaging - Are you able to contact MDT members via messaging from the C3-Cloud Platform?" 58% [Yes], 0% [No], 42% [NA]
- "Video calls - Are you able to join a video conferencing session with MDT members?" 8% [Yes], 92% [NA]
- "Notifications - Are you able to schedule an appointment with your Primary Care Provider?" 16% [Yes], 15% "No and 69% [NA]

Proposed Actions Based on the Evaluation

Based on the technical partners' feedback regarding the overall responses from participants, below is a summary of issues to be handled to improve the first version of the C3-Cloud system. These issues are currently being followed up in the project in relevant work tasks.

- Bugs – errors related to the expected functionality should be fixed.
- Training needs to be improved – related to the uncertainty of users regarding the functionality of the system, next steps, scope.
- Local configuration – local customizations need to be implemented during deployment.
- Evaluation questionnaires – More time to complete the tests and a way to report issues repeatedly..
- Language – issues related to native language usage on the platform, both for local configuration and content.
- Feature improvement – issues related to aspects such as unclear labels, layout, etc. Some comments for features could be out of scope but noted for future recommendations.
  - Incomplete/unclear specification – insufficient information to implement the improvement, for example: Care plan content, and concept issues – specific issues relating to care plan clarifications are required.
  - New feature – a new feature requested.
  - Visual guideline – information missing regarding visual guidelines or available accessibility settings.
  - Test data unrealistic – value ranges incorrect: can be improved with realistic test samples as provided by pilot sites.
  - Scope clarification – the scope of the project needs to be clarified [to whom?/technicians/developers/?] in order to implement improvements.

Discussion

Our findings indicate that with the help of a user-centered design methodology [6] it was possible to define the functional ICT requirements for the C3-Cloud project and further refine them through an early application testing by end-users. A combination of different techniques that complement one another should preferably be used as their collective application will be more powerful than applied in isolation [11]. Therefore we used different evaluation methods.

The first approach used, inspired by the Delphi method, permitted "to obtain the most reliable opinion consensus from the group of experts by subjecting them to a series of questionnaires in-depth and interspersed with controlled opinion feedback"[7] . The development of the questionnaires for application testing was therefore preceeded by and based upon the creation of pilot application requirements which were matched with the use case scenarios, and later reviewed by both IT and clinical experts. However, we are aware that a second review round by the experts may have improved the formulation of the questions.

During the test sessions, we observed some limitations regarding the contextual coupling and synchrony between the test enviroments that were based on PARs and case scenarios and the perceived relevance of some items of the questionnaires. This was highlighted by some of the responses from the participants such as: "not finding the option"; "not knowing"; "not been informed about it"; "not clear how to work through the system"; "not obvious what needs to be done"; "more time needed". Although we encouraged the think aloud method during the application testing, in some cases feedback was too vague to be interpreted. For example, such feedback included statements such as "Depends on which materials are presented" and "I have not stopped/answered". In such cases, participants need to be more specific about the problem they face, and the observers need to make sure they fully understand what the user means at that specific moment.

The analysis of feedback also highligted that it would have been advantageous to apply a shorther questionnaire that addressed some site-specific smaller discrepancies of the C3-Cloud platform and the language-related barriers. Most notably, the prerequisite of a good command of English caused both limitations regarding the recruitment of participants and an accurate understanding of nuances in the non-English speaking countries. Additionally, some participants felt that the questionnaire was not sufficiently detailed to allow them to express all their concerns with the C3-Cloud platform. Specifically, some suggested that it would have been more efficient to answer questions directly in the evaluation walkthrough document, which in turn would have enhanced their understanding of the entire test module in advance. Further, two participants did not feel that they wanted to use the patient empowerment platform in its current format at all.

Notably, as the majority of the participants were elderly, early testing in the development is crucial, as one could expect that such users would experience difficulties with the comprehension and adaptation of the novel C3-Cloud platform, but it could be hard to know in advance where in the system the unintended effects arise. This finding is in line with current research on evaluation methods in health IT, used for early detection an addressing of the unintended consequences of IT usage [12]. However, the overall response of the participants was that the system has great potential to simplify and enhance their engagement in and understanding of the care process. This was reflected by the high overall rate of positive [Yes] responses.

This study also demonstrated that with a multi-faceted user-centered design methodology it was possible to perform an early evaluation of a complex ICT infrastructure, involving different groups of end-users from three different pilot sites (in Spain, Sweden, and the United Kingdom), which in turn further consolidated the European-wide collaboration within the C3-Cloud project. This constructive evaluation was performed at an early stage of the development to achieve a fast improvement of the C3-cloud system. The results summarized in the section "Proposed Actions Based on the Evaluation" were communicated as user feedback to the techical partners, who have reconfigured and updated the C3-Cloud components accordingly. Incorporating the end-user's requests for change and modification has been completed before the pilot application deployment. The C3-Cloud platform, which has been developed by the results of this evaluation, will be subjected to further and more rigorous testing based on real life experience, with actual MDT members and patients/ICGs in the three pilot sites during the planned pilot study starting in April 2019.

Conclusions

This study demonstrates how an integrated application can be tested against the requirements, as elicited through an extensive European collaboration, to improve care for the elderly with multiple diseases. Mainly, application testing was performed without any adverse incident. The online platform worked well throughout the application testing sessions. The aim of this evaluation was not only to appraise the system's functionality, but also to investigate how to improve the C3-Cloud application and its implementation further. The results obtained reflect insights from MDTs, patients and informal care givers for both user-facing components: C3DP and PEP. Overall, this application testing is an early evaluation exercise in order to adapt the system, where needed, and to get the first users' feedback for further development. Integrating the questions into the evaluation walkthrough document, so that participants can answer the questions as they are testing the relevant sections, would make it simpler for the participants.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 689181. Project website: http://c3-cloud.eu/
Heartfelt thanks to the C3-Cloud research team and the contributing authors of Deliverable 9.3: Malte von Tottleben (Empirica), Veli Stroetmann (Empirica), Eric Sadou (INSERM), Eugénia Lamas (INSERM), Laurent Toubiana (INSERM), Pontus Lindman (MEDIXINE), Marie Sherman (RJH), Lei Zhao (WARWICK), Gokce B. Laleci Erturk-men (SRDC), Antonio de Blas (OSAKIDETZA), Nicolás González (OSAKIDETZA), Esteban de Manuel (KG), Marie Beach (SWFT), Göran Ekestubbe (CAMBIO).

References

[1] World Health Organization, Global status report on noncommunicable diseases 2010, World Health Organization, Geneva, Switzerland, 2011.
[2] A. Nobili, S. Garattini, and P.M. Mannucci, Multiple diseases and polypharmacy in the elderly: challenges for the internist of the third millennium, *J. Comorbidity*. 1 (2011) 28–44.
[3] World Health Organization, Multimorbidity, World Health Organization, Geneva, 2016. http://www.who.int/iris/handle/10665/252275 [accessed October 11, 2018].
[4] S.M. McPhail, Multimorbidity in chronic disease: impact on health care resources and costs, *Risk Manag. Healthc. Policy*. (2016). doi:10.2147/RMHP.S97248.
[5] G.B. Laleci Erturkmen, M. Yuksel, B. Sarigul, M. Lilja, R. Chen, and T.N. Arvanitis, Personalised Care Plan Management Utilizing Guideline-Driven Clinical Decision Support Systems, *Stud. Health Technol. Inform.* 247 (2018) 750–754.
[6] D.A. Norman, and S.W. Draper, eds., User centered system design: new perspectives on human-computer interaction, Erlbaum, Hillsdale, N.J., 1986.
[7] N. Dalkey and O. Helmer, An Experimental Application of the DELPHI Method to the Use of Experts, *Manag. Sci.* 9 (1963) 458–467.
[8] I. Scandurra, M. Hägglund, and S. Koch, Specific demands for developing ICT systems for shared home care : a user centred approach, *J. Inf. Technol. Healthc.* 3 (2005) 279–285.
[9] M. Hägglund, I. Scandurra, D. Moström, and S. Koch, Integration architecture of a mobile virtual health record for shared home care, *Stud. Health Technol. Inform.* 116 (2005) 340–345.
[10] I. Huvila, H. Enwald, K. Eriksson-Backa, N. Hirvonen, H. Nguyen, and I. Scandurra, Anticipating ageing: Older adults reading their medical records, *Inf. Process. Amp Manag.* 54 (2018) 394–407.
[11] M.W.M. Jaspers, A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence, *Int. J. Med. Inf.* 78 (2009) 340–353.
[12] F. Magrabi, E. Ammenwerth, H. Hyppönen, N. de Keizer, P. Nykänen, M. Rigby, P. Scott, J. Talmon, and A. Georgiou, Improving Evaluation to Address the Unintended Consequences of Health Information Technology:. a Position Paper from the Working Group on Technology Assessment & Quality Development, *Yearb. Med. Inform.* (2016) 61–69.
[13] C. Grünloh, G. Myreteg, Å. Cajander, and H. Rexhepi, "Why Do They Need to Check Me?" Patient Participation Through eHealth and the Doctor-Patient Relationship: Qualitative Study, *J. Med. Internet Res.* 20 (2018) e11.
[14] C. C3-Cloud, Project Deliverable 8.1 Use Cases and Requirements Specifications of the Pilot Application, Aug-2016. Available from: http://c3-cloud.eu/sites/default/files/pdf/D8.1_V1.0.pdf. [accessed on November 23, 2018].

Address for Correspondence

Lamine TRAORE laminet@gmail.com Address : Campus des Cordeliers, INSERM U 1142 – LIMICS, 15 rue de l'école de médecine, Esc D - 2ème étage, 75006 Paris France

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190343

Automatic Sleep Stages Classification Combining Semantic Representation and Dynamic Expert System

Adrien Ugon[a,b], Carole Philippe[c], Amina Kotti[b,c], Marie-Amélie Dalloz[b,c], Andrea Pinna[b]

[a] *ESIEE-Paris, Cité Descartes BP99, Noisy-le-Grand, 93162, France,*
[b] *Sorbonne Université, CNRS, Laboratoire d'Informatique de Paris 6, Paris, F-75005, France,*
[c] *Assistance Publique-Hôpitaux de Paris (AP-HP), Groupe Hospitalier Pitié-Salpêtrière Charles Foix, Service de Pathologies du Sommeil (Département R3S),Paris, France*

Abstract

Interest in sleep has been growing in the last decades, considering its benefits for well-being, but also to diagnose sleep troubles. The gold standard to monitor sleep consists of recording the course of many physiological parameters during a whole night. The human interpretation of resulting curves is time consuming. We propose an automatic knowledge-based decision system to support sleep staging. This system handles temporal data, such as events, to combine and aggregate atomic data, so as to obtain high-abstraction-levels contextual decisions. The proposed system relies on a semantic reprentation of observations, and on contextual knowledge base obtained by formalizing clinical practice guidelines. Evaluated on a dataset composed of 131 full night polysomnographies, results are encouraging, but point out that further knowledge need to be integrated.

Keywords:

Sleep Stages, Expert Systems, Semantics

Introduction

In the last decades, sleep has been considered more and more seriously either for wellness reasons, or for diagnosing sleep troubles. Its impact on quality of life and on health is henceforth well known. To diagnose a sleep trouble, the gold standard sleep exam is the polysomnography, which consists in recording, during sleep, the course of a set of physiological signals and then observe the brain activity, recorded by electroencephalography (EEG), the eye movements, recorded using electrooculography (EOG), and the muscle tone recorded using electromyography (EMG).

Sleep staging is a fundamental preliminary step to the diagnosis of sleep troubles. During this task, a sleep expert visually browses the polysomnographic curves in 30-second epochs, to assign one of the five sleep stages defined, since 2007, by the American Academy of Sleep Medicine in their international guidelines [1]. A sleep stage is assigned by considering different criteria observed on EEG, EOG, and EMG curves, and by considering the dynamics of sleep. Guidelines for visual scoring of sleep staging define, firstly, for each sleep stage, a set of criteria that need to be met for a sleep stage to start to be scored ("transition rules"). Secondly, another set of criteria is defined for a sleep stage to continue to be scored ("continuity rules"). Five sleep stages are defined: W (Wakefulness), N1 (Non-REM 1), N2 (Non-REM 2), N3 (Non-REM3) and R (REM). N1 and N2 correspond to light sleep; N3 corresponds to deep sleep; REM stands for Rapid Eye Movements. Even if software dedicated to sleep scoring include automatic scoring functionalities, they are not used in routine practice, since sleep physicians are not satisfied by their results [2].

Common approaches found in litterature for automatic sleep stages are based on machine learning techniques. Features, extracted from the acquired signals, are used to feed a classifier, that will be able, after a learning step, to make a decision on new samples. Many of these approaches use a single channel – generally a single EEG channel – to make the decision. Using an open dataset, they might not be compliant with the current guidelines, but with older ones, defined by Retschaffen & Kales in 1968 [3]. Machine learning approaches mainly ignore the domain of knowledge; moreover, sleep dynamics are unsufficiently integrated into the decision, since epochs are considered independantly from each other (through the segmentation of the analysis and decision).

Alternatively, semantic approaches are seen as satisfactory solutions to have a formal description of knowledge on a domain without ambiguity. Based on a formal modelisation of concepts and relationships of the field, they allow reasoning. This formalization results in semantic networks, conceptual graphs or, more often, in ontologies. They might be seen as a description of a state of the universe at one time, focusing on one interest domain, listing concepts and specifying relationships between them. However, the dynamic aspects that make the universe change from one state to another remain an issue to be modelised. From one time to another time, the world changes; thus concepts and relationships between them change also. The description of the universe at one time is different to the description of the universe at another time.

Expert systems are decision support systems composed of an observations base and a knowledge base. The observations base contains facts and events observed in the current state of the universe. Knowledge base contains inference rules allowing the system to make sense or assemble existing observations and facts and generate new events or new facts that will then be added to the observations base. Expert systems are knowledge-based systems and can be designed without any data. In the medical field, we benefit from existing knowledge that needs to be formalized into compliant inference rules.

In this paper, we present a dynamic expert system, using a formal representation of obervations using conceptual graphs and formalizing rules governing sleep-stages transitions and sleep events occurrences. The next section will focus on methods. Then, results will be presented and be followed by a discussion. Finally, we will give our conclusions in the last section.

Methods

An expert system with four modules

Our expert system is composed of four different modules: (1) the *events vector*, a time vector containing all events observed during sleep; (2) the *sleep stages vector*, a time vector containig all sleep stages assigned to each epoch during sleep; (3) the *events fusion knowledge base*, a knowledge base containing fusion strategies to fuse events (4) *five contextual sleep stage assignment knowledge bases*, five knowledge bases containing inference rules to apply in five different sleep contexts.

All these modules can be considered as an expert system. The observation base is composed of the *events vector* and *sleep stages vector*, composed each of facts of, respectively, low and high abstraction level. The knowledge base is composed of the *events fusion knowledge base*, and of the *five contextual sleep stages assignment knowledge bases*.

Events vector

All events observed during sleep are gathered in a time vector, where each event is identified by an identifier and a label. The vector captures the semantic type of each event, given by the identifier, but also its start and end. If necessary, other useful information can be added to the event; for instance, the lowest value of the saturation can be specified and attached to a desaturation. Each event is then formalized as a conceptual graph, giving its label, its start, its end and all other information that might be useful for the final decision.

Each event is represented by a conceptual graph. Conceptual graphs are a formalism that was introduced by John F. Sowa in 1984 [4]. Semantic concepts are linked by labelled relationships. Fusion algorithms have been defined to combine several conceptual graphs that share concepts.

For example, Figure 1 shows an event vector with all events occurring during the epoch. Representations of event 1 and event 2 are given on Figure 2.

Figure 1 – Events vector

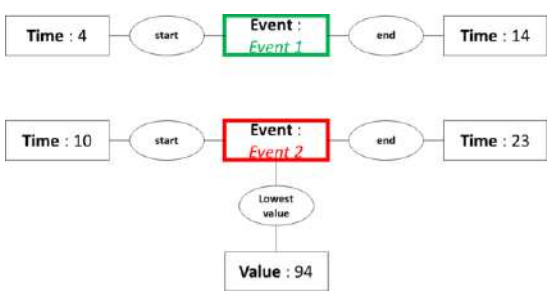


Figure 2- Conceptual graphs of Event 1 and Event 2

Sleep Stages vector

All sleep stages assigned by the expert system are stored in a sleep stages vector. Sleep stages are considered as high-abstraction-level information, obtained by combining and fusionning atomic information observed on other channels. This vector is initialized by specifying that, just before the start of the exam, i.e. the last epoch before the start of the exam, the patient was awake. Thus, we initialize the sleep stages vector with the W sleep stage assigned to the last epoch preceding the start of the exam.

An example sleep stages vector is given in Figure 3. Each cell represents a 30-s epoch. The vector starts with seven epochs of N2, followed by 15 epochs of N3, followed by eight epochs of R, followed by 2 epochs of W and ends with 2 epochs of N1.

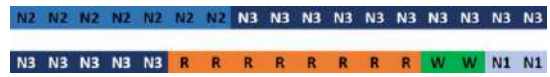


Figure 3 – Sleep stages vector

Events fusion knowledge base

Considered separately, events observed on each acquired channel may not be sufficient to asisgn a sleep stage. However, they may be different observations of the same physiological event. To be combined, it is hence necessary to fuse conceptual graphs to get a new conceptual graph representing a higher-abstraction-level event. Initially, they are separated; it is necessary to formalize the causal relationships that link them. Abstraction rules, combining all linked events to generate the physiological fact, can be applied to add the physiological fact to the observation base (the events vector).

Each rule of the events fusion knowledge base is formalized as a conceptual graphs fusion strategy. Fusion strategies follow the principles of the maximal join operator. Criteria required to fuse two conceptual graphs are defined. The resulting conceptual graph is entirely defined, on the basis of the concepts – and their values – of the fused conceptual graphs.

Figure 4 illustrates an example of fusion rules using conceptual graphs. Event 1 and event 2 are fused into Event 12, starting at starting date of Event 1, and ending at ending date of Event 2. Criteria to fuse these two events could be defined on start dates, end dates, or on the lowest value of event 2.

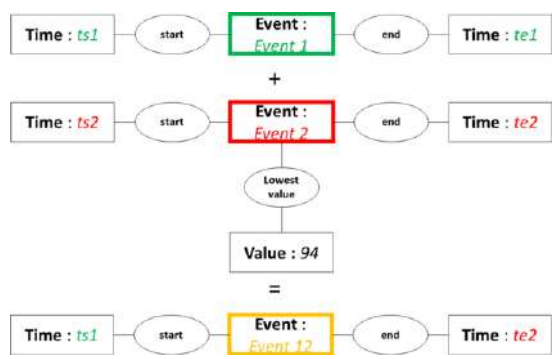


Figure 4 – Example of events fusion

Five contextual sleep-stage assignment knowledge bases

Depending on the current sleep stage in the previous epoch beeing considered as a sleep context, rules to assign a sleep stage to the current epoch are different, when applying the guidelines of the AASM. As a consequence, five different knowledge bases were defined, each containing the rules to apply in a given sleep stage. Depending on the current stage, each knowledge base contains rules to start a new sleep stage and rules to apply to continue scoring the current sleep stage. For each knowledge base, sleep stages are ordered. The inference step evaluates whether rules of a given sleep stage are met or not, testing all sleep stages successfully until one given sleep stage can be assigned. Ordering sleep stages allows to formalize priority rules, defined in the guidelines.

Rules to assign a sleep stage are based on binary criteria. Each criterion is met whether a high-abstraction-level event is met, or not.

Figure 5 illustrates an example of the five contextual knowledge bases. Depending on the sleep stage assigned to the previous epoch, a knowledge base is selected. Then, criteria are evaluated until one is met. It allows to assign a sleep stage to the current epoch.

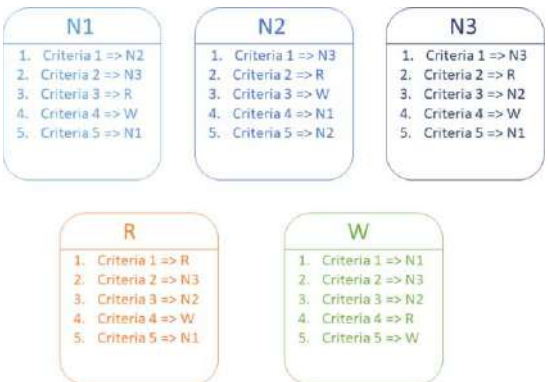


Figure 5 – Contextual Decision knowledge bases

Decision algorithm

Previously of the decision process, it is required that following steps have been done : (1) split the recording into 30-second epochs, (2) extration and recognition of low-abstraction-level atomic events, (3) representation of all extracted events by conceptual graphs, consistantly to a defined terminology.

Figures 6 shows a flowchart of decision process. It is composed of four steps.

1 - Move to the next epoch

A decision is made for each epoch of the recording, starting from the first to the last, moving from one epoch forward at each step, once a decision has been made and a sleep stage has been assigned to the epoch. At each step of the decision process, a "reading head" is moving to the next epoch.

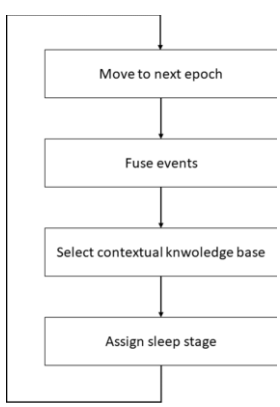


Figure 6– Flowchart of decision process

2 – Fuse events

We assume that events extracted previously to the decision process are low-abstraction-level events. To be able to make a decision, we need high-abstraction-level events. These events are obtained by events fusion using fusion strategies [5,6]. Fusion strategies used to fuse events are defined and formalized into the *events fusion knowledge* base. Representing events by conceptual graphs allows to fuse delayed events observed in different channels, which is necessary in physiological processes. Using events fusion, it is possible to abstract events from a given abstraction level to a higher abstraction level; it is also possible to integrate literal data to get a semantic event; it is finally possible to aggregate, or combine, events. This step is based on an expert system.

3 – Select contextual knowledge base

Depending on the sleep stage assigned to the previous epoch, we know the sleep context of the subject at the start of the current epoch. Regarding this context, the right knowledge base can be chosen, in order to select the right rules to apply to make the decision and assign the sleep stage.

4 – Assign a sleep stage

Having identified which knowledge base to use, and regarding events of the low- and high- abstraction levels oberved and inferred on the events vector, the system is able to assign a sleep stage to the current epoch. To achieve this task, ordered criteria are evaluated. Each criterion is defined as a logical combination of high-abstraction-level events observed. Rules are evaluated following a predefined order until a criterion is met. This allows to assign a sleep stage. Criteria are ordered according to priority rules.

Example

According to the guidelines, the N2 sleep stage should be assigned to an epoch, if a K-Complex, not associated to an arousal, is the first half of the current epoch.

Firstly, K-complexes and arousals need to be scored (manually or automatically). Then, K-complexes and arousals are fused, to classify K-Complexes into 2 categories: K-complexes associated to an arousal, and K-Complexes not associated to an arousal. In the next step, given an epoch, the right time period is observed to check whether a "K-complex non associated to an arrousal" was observed. For those epochs where this criterion is met, the N2 sleep stage is assigned. Otherwise, the criteria of other sleep stages are tested.

Results

Description of the dataset

The dataset used for evaluating our method was already used in [7]; Gathering 131 full night polysomnographies, it is composed of 148,407 epochs (1,237 hours). The first subset is composed of 101 polysomnographic recordings of patients suspected of suffering from Sleep Apnea Syndrome; the second subset is composed of polysomnographic recordings of ten control subjects having spent three consecutive nights at the hospital. Polysomnographic recordings were performed in the sleep pathologies unit of La Pitié-Salpêtriére Hospital (AP-HP, Paris, France) using the Grael HD- PSG TM device which is produced by Compumedics Limited® from Australia. All recordings have been fully scored by a sleep expert in accordance with AASM guidelines [1] using Profusion Sleep TM Software from Compumedics®. All 131 recordings were visually scored by two experienced sleep experts, in accordance with the clinical practice guidelines defined by the AASM, without using the automatic pre-analysis functionality. This scoring is used as gold-standard for our work.

Each recording is composed of more than 30 channels for the full-night polysomnography. In this study, we only need some

channels to be compliant with the clinical practice guidelines defined by the AASM: the EEG, the EOG and the submental EMG. Six EEG channels were recorded, following the international 10–20 system with a sampling rate of 256Hz: O1, C3, C4, Fp1, A1, A2. 2 EOG were recorded with a sampling rate of 256Hz: Left EOG and Right EOG; One channel of the submental EMG was recorded with a sampling rate of 256Hz. As recommended in the clinical practice guidelines defined by the AASM, EEG and EOG channels were filtered using a 0.3–35 Hz bandpass filter; EMG submental channel was filtered using a 10–100 Hz bandpass filter.

This study was approved by The Committee for the Protection of Human Research Participants, Paris VI (Comité de Protection des Personnes Ile-de-France VI, Paris, France).

Control subjects were received by a physician to be informed in detail about the purpose and the procedure of the research study. Informed consent was provided and signed by all subjects. A sleep physician performed a physical examination and a questionnaire to validate that they had no suspicion of a sleep disorder.

All subjects signed an informed consent form. All individuals included in our database followed instructions to refrain from alcohol and caffeine ingestion and to avoid engaging in prolonged and/or strenuous exercise before sleeping. All polysomnographies were recorded in a quiet, darkened room.

Measures

To evaluate the performance of our automatic sleep slating tool, we used different measures.

Agreement rate is a measure used to assess the results of the automatic analysis globall, i.e. including all sleep stages. It is defined as the ratio of epochs scored with the same sleep stage by the expert and the automatic analysis for each recording.

$$agreement\ rate = \frac{Nb\ epochs\ with\ same\ sleep\ stage}{Total\ number\ of\ epochs}$$

Because sleep stages have an unbalanced number of epochs, Cohen's kappa was also used. This measure takes into account that some epochs are scored identically by the automatic analysis and by the expert by chance

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

Where p_0 is the relative observed agreement among raters (identical to agreement rate); and p_e is the hypothetical probability of chance agreement.

Other measures are defined to assess results obtained for each sleep stage. These measures are defined for binary classifications. All of them are defined considering the following confusion matrix

Table 1 – Confusion Matrix

		Automatic analysis	
		C	Not C
Expert analysis	C	TP	FN
	Not C	FP	TN

TP refers to True Positive elements. Both expert and automatic analysis have classified these events in the class of interest C.

TN refers to True Negative elements. Both expert and automatic analysis classified these events as not belonging to the class of interest C.

FP refers to False Positive elements. Classified as belonging to the class of interest C by the automatic analysis, they were classified as not belonging by the expert.

FN refers to False Negative elements. These events were classified as not belonging to the class of interest by the automatic analysis, whereas they were classified as belonging by experts.

Recall and precision are widely used in the field of machine learning. Recall is the ratio of elements belonging to the class of interest, that were correctly identified by the automatic analysis. Precision is the ratio of elements belonging effectively to the class of interest among all elements that were considered as belonging to the class of interest by the automatic analysis. Recall and precision are defined by following formulas:

$$\begin{cases} recall = \dfrac{TP}{TP + FN} \\ precision = \dfrac{TP}{TP + FP} \end{cases}$$

Sensitivity and specificity are used in the medical field to evaluate a diagnosis test. Sensitivity allows to know the performance of the test to identify sick individuals (=positive elements). It is equal to the predefined recall. Specificity allows to know the performance of the test to identify healthy individuals (=negative elements). Sensitivity and specificity are defined by following formulas:

$$\begin{cases} sensitivity = \dfrac{TP}{TP + FN} \\ specificy = \dfrac{TN}{TN + FP} \end{cases}$$

To balance recall and precision, the F Mesure was proposed, defined by the following formula:

$$F\ Mesure = 2 \times \frac{recall \times precision}{recall + precision}$$

To give more weight to recall than to precision, the F2-Measure was proposed as an extension of the F-Measure, with the following formula:

$$F_2 - Measure = \frac{(2^2 + 1) \times recall \times precision}{2^2 \times precision + recall}$$

Results

On the dataset of 101 polysomnographic recordings of patients suspected to suffer from sleep apnea syndrome, the obtained average agreement rate was 51.5; the average kappa was 0.36. All measures for all sleep stages are detailed in Table 2. Considering the control subjects, we obtained an average agreement rate of 56.55 and an average κ of 0.43 (see Table 3).

Table 2 – Measures obtained on Sleep Apnea Syndrome patients dataset

Measure	W	N1	N2	N3	R	MVT
Recall / Sensitivity	54.3	30.3	58.6	44.4	54.4	0.0
Precision	75.3	21.1	54.0	51.6	53.4	0.0
Specificity	92.64	86.26	73.86	91.4	91.6	99.5
F Measure	58.8	22.3	53.4	42.9	50.3	0.0
F2 Measure	54.5	25.3	55.8	42.5	51.6	0.0

Table 3 – Measures obtained on control subjects dataset

Measure	W	N1	N2	N3	R	MVT
Recall / Sensitivity	54.9	32.7	63.2	54.4	61.8	0.0
Precision	72.8	24.1	59.9	50.4	69.5	0.0
Specificity	92.7	86.4	74.9	86.1	91.6	96.7
F Measure	60.4	25.1	60.5	49.0	63.7	0
F2 Measure	56.6	28.1	61.9	51.4	62.2	0

As we can see, there are some differences between results obtained on the patients dataset and on the subjects dataset. This can be explained by the quality of the signals to process. Because of sleep troubles, signals from patients are very noisy, which make signal processing methods less efficient. Moreover, sleep troubles generated a highly disturbed sleep with many transitions; and as discussed by Thomas Penzel [8], transitions between sleep stages present the biggest source of differences between human sleep scorers.

Discussion

Our results may appear to be less performant than other results published in recent literature. All these works are mainly based on machine learning algorithms, and ignore, during the evaluation step, the intra- and inter-raters concordance. It was assessed by Danker-Hopfe et al. in 2009 [9]. Inter-raters rates are given by sleep stage in Table 4.

Table 4 – Inter-rater agreement rate (kappa)

W	R	N1	N2	z	N3
0.8608	0.9054	0.4608	0.7188		0.7285

As we see, most of results given in recent published works claim to have a higher concordance that what another human sleep expert would obtain. It means that their results are obtained by overfitting, which is not suitable.

On the contrary, our approach follows rules of practice guidelines. Our results show that these rules still need to be improved and expanded. Knowledge base might be enriched with transitions rules or experience knowledge.

The performance of identifying atomic events remains also to be evaluated.

Weaknesses of our approach include a relaince on the need to formalizing a complete knowledge base of rules to be applied to score sleep stages; medical knowledge is often incomplete, and knowledge acquired from experience is hard to formalize.

Strengths of this approach include its robustness, its flexibility, and its upgradeability. Its ability to use easy time-based rules helps contextualize decisions. Semantic reasonning is fully understandable and customizable by experts. Furthermore, in absence of formalized knowledge it can be hybridized with machine learning approaches.

Conclusion

In this paper, we present an approach representing information by semantic concepts. Temporal aspects are taken into account by reasonning on a dedicated framework, inspired by the Turing machine. Sleep stages are assigned by using different expert systems, one for each sleep context considered.

Acknowledgement

This work was performed within the Institut Universitaire d'Ingénierie en Santé of Sorbonne Université and the Labex SMART supported by French state funds managed by the ANR within the Investissements d'Avenir programme under reference ANR-11-IDEX-0 0 04-02.

References

[1] B.R. Brooks, R. Gamaldo, C. Harding, R. Lloyd, R. Marcus, and C. Vaughn, The AASM Manual for the Scoring of Sleep and Associated Events, American Academy of Sleep Medicine, Darien. Illinois, 2017.

[2] P. Escourrou, N. Meslier, B. Raffestin, R. Clavel, J. Gomes, E. Hazouard, J. Paquereau, I. Simon, and E. Orvoen Frija, Quelle approche clinique et quelle procédure diagnostique pour le SAHOS ?, *Revue des Maladies Respiratoires*. **27** (2010) S115–S123. doi:10.1016/S0761-8425(10)70017-6.

[3] A. Rechtschaffen, and A. Kales, A Manual of Standardized Terminology, Techniques, and Scoring System for Sleep Stages of Human Subjects, US Government Printing Office, Washington Public Health Service, Washington DC, 1968.

[4] J.-F. Sowa, Conceptual structures: Information processing in mind and machine, United States, 1982.

[5] A. Ugon, J.-G. Ganascia, C. Philippe, H. Amiel, and P. Lévy, How to Use Symbolic Fusion to Support the Sleep Apnea Syndrome Diagnosis, in: M. Peleg, N. Lavrač, and C. Combi (Eds.), Artificial Intelligence in Medicine, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011: pp. 45–54. doi:10.1007/978-3-642-22218-4_6.

[6] A. Ugon, J.-G. Ganascia, C. Philippe, H. Amiel, and P. Lévy, Multi-layers Symbolic Fusion with conceptual graphs to support the Sleep Apnea Syndrome Diagnosis, in: Proceedings of Fusion 2013 Conference, Istanbul, Turkey, 2013.

[7] A. Ugon, A. Kotti, B. Séroussi, K. Sedki, J. Bouaud, J.-G. Ganascia, P. Garda, C. Philippe, and A. Pinna, Knowledge-based decision system for automatic sleep staging using symbolic fusion in a turing machine-like decision process formalizing the sleep medicine guidelines, *Expert Systems with Applications*. **114** (2018) 414–427. doi:10.1016/j.eswa.2018.07.023.

[8] T. Penzel, X. Zhang, and I. Fietze, Inter-scorer Reliability between Sleep Centers Can Teach Us What to Improve in the Scoring Rules, *JCSM*. **9** (n.d.) 89–91. doi:10.5664/jcsm.2352.

[9] H. Danker-Hopfe, P. Anderer, J. Zeitlhofer, M. Boeck, H. Dorn, G. Gruber, E. Heller, E. Loretz, D. Moser, S. Parapatics, B. Saletu, A. Schmidt, and G. Dorffner, Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard, *Journal of Sleep Research*. **18** (2009) 74–84. doi:10.1111/j.1365-2869.2008.00700.x.

Address for correspondence

Dr Adrien UGON

ESIEE-Paris

Département Santé, Energie, Environnement

2 boulevard Blaise Pascal

BP 99

93162 Noisy-le-Grand Cedex

France

Tel : +33 (0)1.45.92.65.36

Mail: adrien.ugon@esiee.fr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190344

Development and Preliminary Evaluation of a Visual Annotation Tool to Rapidly Collect Expert-Annotated Weight Errors in Pediatric Growth Charts

P.J. Van Camp[a,c], C. Monifa Mahdi[c], Lei Liu[a,c], Yizhao Ni[b,c], S. Andrew Spooner[b,c], Danny T.Y. Wu[a,b]

[a] Department of Biomedical Informatics and [b] Department of Pediatrics, University of Cincinnati, Cincinnati, OH, USA,
[c] Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Abstract

Patient weights can be entered incorrectly into electronic health record (EHR) systems. These weight errors can cause significant patient harm especially in pediatrics where weight-based dosing is pervasively used. Determining weight errors through manual chart reviews is impractical in busy clinics, and current EHR alerts are rudimentary. To address these issues, we seek to develop an advanced algorithm to detect weight errors using supervised machine learning techniques. The critical first step is to collect labelled weight errors for algorithm training. In this paper, we designed and preliminarily evaluated a visual annotation tool using Agile software development to achieve the goal of supporting the rapid collection of expert-annotated weight errors. The design was based on the fact that weight errors are infrequent and medical experts can easily spot potential errors. The results show positive user feedback and prepared us for the formal user-centered evaluation as the next step.

Keywords:

Electronic Health Records, Data Curation, Patient Safety

Introduction

A child's weight is a vital parameter in pediatrics, with its importance ranging from general development to precise medication dosing. The latter is particularly important since many drug doses are weight-based. According to previous studies, about 20% of pediatric medication errors resulted from "improper dose/quantity," which is a much higher rate than that in the adult setting [1,2]. Consequently, weight errors can lead to under- or over-dosing and further cause patient harm [3]. Pediatric weights are recorded incorrectly in various ways: typing errors, conversion errors (e.g. scale in pounds, system in kg), and weight estimations instead of actual measurements. Capturing weight errors with high sensitivity and specificity, however, can be very difficult. Manual chart reviews are often required and they are time-consuming and impractical to comply with in busy clinical routines. Moreover, capturing weight errors is especially difficult in neonates and growing children in which age and gender only provide a rough estimate of the expected weight range. A myriad of medical conditions, both acute and chronic, can also drastically change weight patterns. Given that it is in these more complex cases that drugs are often prescribed, detecting incorrect weights and administering proper weight-based dosing are critical.

There has been previous research in developing informatics solutions to detect weight errors. One approach is based on z-scores, a score that indicates how many standard deviations an element is from the mean. Error detection methods using this approach are often rule-based and tend to use empirical cut-offs to determine weight errors based on age and gender. One well-known example is the program provided by the Centers for Disease Control and Prevention (CDC) of the United States [4]. These rule-based systems are efficient and good at capturing outliers, but often have low performance in subtler and/or more complex charts.

Our literature review shows a lack of automated algorithms to detect weight errors with high performance – both retrospectively and in real time. The closest we can find is the method developed by Dymont *et al*., which is based on exponentially weighted moving averages of standard deviations with rule-based decision [5]. This method was developed for the purpose of cleaning research data and identifying implausible values in growth data. While this method took personal weight history into account and can be optimized in individual datasets, it was not derived from training on all weight patterns in a dataset and therefore may fail to detect weight errors in more complicated patterns. Another limitation of this method is that it requires data after the point of interest to perform analysis (*i.e*., in a retrospective manner) and may not be very useful in real-time detection of weight errors as a clinical decision support tool.

To address this gap, we aim to develop an advanced algorithm with supervised machine learning techniques to detect pediatric weight errors, which can retrospectively detect weight errors and be deployed as a real-time clinical decision support tool. To achieve this goal, a critical first step is to collect a large body of expert-annotated weight errors. However, there are three challenges in collecting such an annotated dataset: 1) the low prevalence of weight errors (less than 5%) [6], 2) the tremendous amount of time required to manually review patient charts to identify true errors [7,8], and 3) various weight patterns with limited information. To overcome these challenges, we developed a web-based annotation tool utilizing visual analytics principles and human-computer interaction techniques to efficiently and effectively annotate pediatric weight charts. With this tool, weight data are converted into a set of growth charts and presented in an interactive user interface to support rapid collection of expert annotations. In this paper, we report the design and preliminary evaluation of the visual annotation tool.

Methods

Clinical Setting and Weight Data

The raw weight data were collected from the electronic health record (EHR) system of the Cincinnati Children's Hospital Medical Center (CCHMC) between January 2010 and June 2018. CCHMC is an urban, quaternary children's health system

in Cincinnati, Ohio in the United Sates with more than 600 inpatient beds and over 33,000 admissions as well as 1.2 million patient encounters annually. Each weight record consisted of an identifier, gender, and weight value (in kg) plus timestamp of each measurement. The patient identifiers were anonymized, and the timestamps were converted into the age in days when the weight was taken (with decimals to incorporate time of day). The data was stored in a file-based (SQLite) database to support downstream data manipulation and analysis. The weight data were summarized statistically to understand the distribution of key characteristics, including age, gender, and weight variability.

Pediatric Growth Chart

The display of weight data in our visual annotation tool was informed by the clinical growth charts developed by the World Health Organization (WHO) and the Centers for Disease Control and Prevention (CDC) in the United States [4]. This visualization of child's weights has been routinely used in practice both in paper-based charting and EHRs. The growth charts come with age-matched percentiles based on the distribution of the general population. In our design, the percentiles were adjusted based on gender and age groups. Specifically, each patient has two sub-charts: one from birth to age two (0 - 24 months) and the other from age two to eighteen. The reason of this split was due to practical considerations where following up on patient weights in the infant period (0 – 24m) often has much higher variability. Plotting weight points in the infant period with other points in childhood on the same growth chart would be impractical and may introduce challenges in our annotation tasks. Figure 1 shows an example of a growth chart for boys between 2 and 20 years old as reproduced from the CDC paper [1].

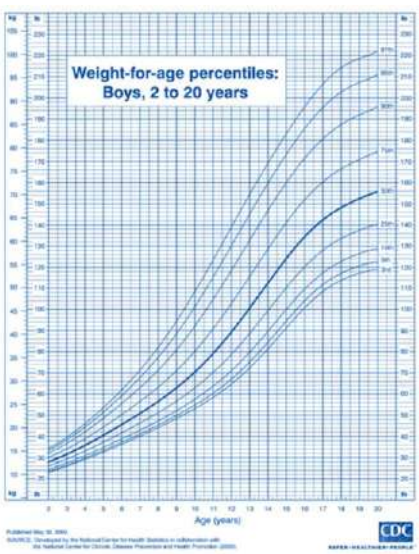


Figure 1 – CDC Growth Chart Example [4]

System Design and Evaluation

Our visual annotation tool was developed following the principles of the Manifesto for Agile Software Development [9]. The development focused on end users (medical experts) and their interactions with the software. The research team provided a working prototype over time to collect user feedback and respond to the changes. Specifically, the visual annotation tool was designed and evaluated in three phases: 1) Design of core functionality and user interface, 2) Iterative refinement through group discussion, 3) Experimental workflow and pilot testing. The detail of each phase is described in the following sections.

Phase 1 - Design of Core Functionality and User Interface

In the first phase, a focused group of clinical informaticians (PVC, SAS, and DTW) worked closely together to brainstorm the core functionality and user interface of the visual annotation tool., This group searched the literature and reviewed known data visualizations to inform the design of the tool with the goal of supporting effective and efficient collection of expert-annotated weight errors. In this phase, only the facial validity of the tool was evaluated. That is, the system was presented to and reviewed by the three group members to check if it had the potential to achieve the design goal.

Phase 2 – Iterative Refinement through Group Discussion

In the second phase, the visual annotation tool was refined in an iterative process by a larger group called "decision support analytics workgroup" (DSAW) at CCHMC. This group of clinical informaticians and medical experts met weekly to collaborate on projects and regularly discussed the weight error detection project and reviewed our visual annotation tool. The iterative refinement through group discussion lasted for several weeks, and an informal usability testing was conducted at the end of this phase when the visual annotation tool was functional enough. In this information usability testing, user feedback was collected and categorized, and the changes were made to generate the first stable version of the tool.

Phase 3 – Experimental Workflow and Pilot Testing

With the stable version of the tool, the research team developed a protocol to conduct a formal user study to examine the relationship between different combinations of tool settings and the efficiency and effectiveness of annotations on the weight data. The visual annotation tool was enhanced in two ways to support this user-centered evaluation and data analysis. First, it was able to record every click made by the user and produce usage logs. Second, it had an administration interface to manage users, assign datasets, and control tool settings.

To prepare for our formal user study, the protocol for the experiment was tested to refine the workflow. A small group of medical experts and clinical informaticians was recruited. Each participant spent an hour with the research team. In each session, a short introduction of the weight error detection project was given, followed by a brief tutorial to the visual annotation tool. Then, the participants were assigned two datasets, each containing 225 patient charts. The patient charts were randomly selected from the weight data (the charts with 4+ data points) and verified by the research team. After finishing the annotation tasks, the participants were asked to provide feedback on the interface design and workflow through a structured survey with open-ended questions. Specifically, the usability of the tool was measured by the Systems Usability Scale (SUS), a 10-question survey with a modified Likert scale from one to five with five being the highest [10]. After filling out the SUS survey, a short, structured interview was conducted to collect detailed feedback that was not captured by the SUS.

In terms of data analysis, the composite SUS scores were calculated based on the guideline, which range from 0 to 100. A system with a SUS composite score of 68 or above is considered as having "above average" usability. In addition, the time of task completion was calculated based on the usage logs. This experiment protocol was reviewed and approved by the University of Cincinnati Institutional Reviewed Board (#2017-2075).

Software Implementation

The application was developed using Shiny in R-studio. Shiny is a framework that supports the quick development of interactive web applications straight from R. This deployment has benefits in easy access with no extra client software installation. While Shiny has several benefits and can save much programming time, it has limitations like any other framework and libraries. Some inline JavaScript and jQuery were added to provide additional, optimized user experience. The visual annotation tool was hosted on an internal Shiny server behind the firewall and could only be accessed through a web browser by users at the University of Cincinnati (UC) and CCHMC in order to protect the weight data with the highest standard.

All chart annotations and usage logs were stored in a separate SQLite database. This allowed the original weight data to be left untouched and only a minimal data to be stored. The annotated patient charts and the process of annotation can be recreated by combining the data from both sources. As mentioned in Phase 3, an administrative interface was created as a separate Shiny application to support the workflow. The design and functionality of this administrative interface are out of the scope of the present paper.

Results

Weight Data

Table 1 shows the descriptive statistics of the weight data. Over 600,000 patient and their weight data points were extracted from CCHMC's EHR system. Maximum number of weights per patient and weight value taken were 2312 times and 632.2 kilogram, respectively. 47% of the patient (N=283,971) had at least four weight points, accounting for 88.6% of the individual weights (N=4,193,008). Random charts selected out of this subset were used to explore the graphical representation of the weight data and support the tool evaluation.

Table 1 – Statistical Summary of Weight Data

Characteristics		All Patients	Patients with 4+ data points
Number of Records		4,733,317	4,193,008
Number of Patients		604,147	283,971
Number of Females		291777	136404
Num. of Weights per Patient	Min	1	4
	Median	3	8
	Average	7.8	14.8
	Max	2312	2312
Weight Value (kg)	Min	0.0	0.0
	Median	23.2	23
	Average	33.1	32.9
	Max	632.2	632.2

Phase 1 - Design of Core Functionality and User Interface

Three core functions of the visual annotation tools are described in the following sections, including 1) individual patient chart display and annotation, 2) multiple charts in a grid, and 3) annotation flow.

Individual Patient Chart Display and Annotation

The display of individual patient charts was designed based on the CDC examples. Weight data were converted into growth charts with the *x*-axis being the age in days and the *y*-axis being the weight values. Digital display of weight charts had several benefits compared with their paper-based counterparts. First, the display was designed to automatically crop the *x*-axis to reflect the proper age range of the weights recorded. Second, the population percentiles were automatically adjusted based on age and gender. Next, the display always kept the upper (97th) and lower (3rd) percentiles in a chart, allowing the trend of the weights being interpreted correctly so that changes in weights would not be over- or under-estimated. Finally, users can zoom in a certain area of a chart and zoom out if needed. Figure 2 shows 4 patient charts with likely complex underlying pathology resulting in atypical weight patterns.

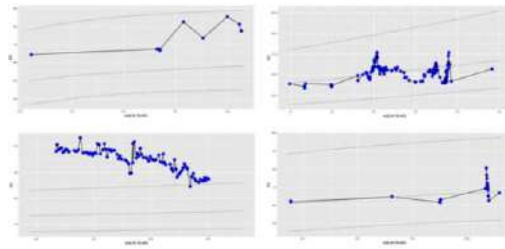


Figure 2 – Patient Charts with Complex Weight Patterns

Another core functionality of the tool was to label incorrect weight points (a.k.a. annotation), which was implemented using color coding. Users can click each weight point to indicate its error type. All the weight points were colored in blue by default. The first and the second click of a point turned the color into red and orange, respectively, and a third click brought it back to blue. The color indicated different error type: Blue:) a point is not an error (default setting), Red) a point is definitely an error with high clinical importance and likely causes patient harm (e.g. medication errors), and Yellow) a point may be an error with medium or low clinical importance. This color coding allowed users to focus on potential weight errors and quickly annotate patient charts.

In the cases where the weights in a patient chart were too confusing or too complicated to annotate, two additional checkboxes were provided to indicate such issues for the research team's information.

Multiple Charts in a Grid

Annotating individual patient charts one-by-one can be very time-consuming. As such, the idea of displaying multiple charts at once in a grid was introduced. Due to the fact that the majority of patient charts did not contain a weight error, it is expected that human eyes can scan through the charts very quickly and identify weight errors based on the trend. That is, patient charts containing errors or with complex weight trends would stand out visually and then could be viewed and annotated in the individual chart display. The grid size was designed to be adjustable (i.e. 3x3 or 5x5) through the administration interface. An example of the grid display is presented in Figure 3.

Annotation Flow

The individual patient charts can be annotated in two ways. One way was to annotate a chart immediately after clicking it in the grid. The other was to first go through the whole dataset in the grid views and select all charts of interest, and then annotate the selected charts one-by-one as a second step. The former is called "one-step annotation" (1SA), which was implemented by creating a pop-up of the individual patient chart when it is clicked in the grid (as shown in Figure 3). The latter is called "annotation" (2SA), where clicking a chart in the grid would only highlight it without a pop-up. After browsing thought all grids (whole dataset), the selected charts were displayed one-by-one as individual patient charts for annotation.

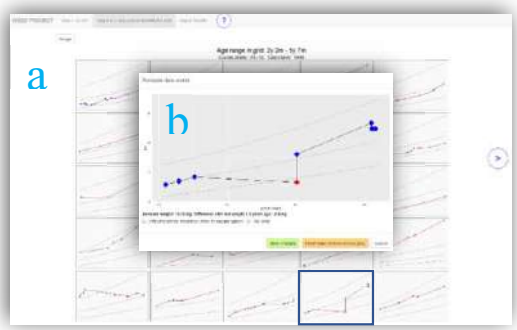


Figure 3 – Overview of the Main User Interface in One-step Annotation: a) Multiple Charts in Grid (5x5) and b) Individual Patient Chart Display as a Popup after User Selection.

Phase 2 – Iterative Refinement through Group Discussion

After the core functionality and user interface of the tool were created, the visual annotation tool was reviewed and iteratively refined in the DSAW meetings for several weeks. This phase was ended by the informal user testing with eight DSAW members. The user feedback was summarized in Table 2 in four categories. The overall feedback was positive, especially on the intuitiveness and the grid display of the charts, which confirmed our design to support rapid annotation. Improvements were made in annotation, interpretability, navigation, and system performance to produce the first stable version of the visual annotation tool.

Table 2 – User Feedback from Informal Testing in Phase 2

Category	Description	Action taken
Annotation	Points sometimes difficult to select	Radius for selection increased
	No easy way to reset annotation in a graph	Create a "remove all annotation" option
Interpret-ability	Get more detail about certain points	Hovering will display a point's weight and the time and weight difference with the previous one
Navigation	Instructions very limited	Tutorial mode and help button implemented
	Button to go to next page not intuitive	Button redesigned
System Performance	Large grids update very slowly	Mark chart instead of updating whole grid
	When new grid is loading, it seems that the app is frozen	Loading message now displayed
	Errors for accessing the app	With help of IT access now easier over network

Phase 3 – Experiment Workflow and Pilot Testing

In the pilot testing, five participants (two medical experts and three clinical informaticians) were recruited. Following the experiment protocol, each of them was assigned two datasets (225 charts per set) and provide feedback after completing the tasks. The average time to complete the first and the second dataset was 8.15 min (standard deviation, or SD: 2.8) and 6.3 min (SD: 2.1), respectively. There was a 2-minute decrease in completing the second dataset, which was due to the learning curve as reported by the participants, although the change was not statistically significant (p=0.097). The average SUS score was 78.5 (SD: 10.6), which confirmed the system to have above average usability. Additional feedback through the structured interview was positive, especially on the ease of use with intuitive user interface and fast learning curve. The workflow was confirmed to be reasonable and required no changes.

Discussion

In this paper, we reported the design and pilot evaluation of a web-based visual annotation tool to support the rapid annotation of weight errors. The tool was easy to use and has configurable settings (grid size and annotation flow), allowing users to optimize their annotation based on personal preferences. The intuitive design originated from the traditional growth charts and utilized a set of visual analytics techniques, providing expert users with a familiar yet enhanced chart review experience to identify pediatric weight errors.

At the time of completion of the current study, Conlen *et al*. published a paper at the Special Interest Group on Computer-Human Interaction in the Association for Computing Machinery (ACM CHI conference) about design principles for visual analytics in operations contexts [11]. This paper lists seven design principles that support and justify many of our design ideas of the visual annotation tool as discussed below. Principle 1 and 2 state that data must come with context and relationships between points must be shown. In our design, the adoption of digital growth charts with population percentiles provided a familiar experience for the annotators to interpret weights based on all other weight points in a chart (the context). Moreover, additional time and weight change information to the previous point can be viewed by hovering above a point of interest (the relationships). The context and relationships provided in our visual annotation tool were sufficient enough for experts to annotate weight errors in nearly all cases with high confidence, requiring no additional EHR data for interpretation, which may be hard to get and can slow down the annotation process. Principle 3 states that users should be able to travel through the data along the most important decisional dimension. In our design, the most important decisional dimension is the weight trend as a function of age in days and weight values. Given that abnormal weight patterns are rare but easily screened by medical experts, the individual patient charts were organized in grids so that users can easily page through the dataset to check the weight trends. Moreover, our design of displaying multiple instances at the same time in a resolution that preserves the overall pattern greatly increased the annotation speed, which is directly linked to Principle 7, where data should be shown at the resolution for the task at hand. In this two-level design (grid view versus individual charts), users can first identify suspicious patient charts and then annotate individual weight points as two distinct tasks. Principle 4 recommends that the user interface be flexible and able to adapt to the user's input. In our design, the grid size and annotation flow (1SA or 2SA) can be configured based on user preferences using the administration interface. Principle 5 indicates the need to surface issues early, which was done in our development process through the Agile Manifesto. Principle 6 requires that no data are removed, even if they are not used. By separating our input database (weight data) from the annotation database, it is ensured that only relevant input data was retrieved and never got overwritten or changed.

In addition to our effective design, the pilot testing confirmed the smoothness of our workflow and prepared us for the formal user-centered evaluation. In the pilot testing, participants were able to annotate 225 charts in less than 9 minutes. As reported by the participants, learning curve was quick since the participants were able to spend less time (2 minutes less) annotating the same amount of records. Extrapolating the results, 1500 charts may be annotated in 60 minutes, which indicates the achievement of our design goal to support rapid collection of expert-annotated weight errors.

One limitation in the pilot evaluation is the inability to validate the accuracy of annotation given that there was no "gold standard" (or ground truth) of weight errors. However, the gold standard may not be required in this case since we are focusing on the clinical importance of the errors, that is, whether abnormal weights can cause patient harm and subsequently require clinician's attention in busy clinical routines. We believe a reference standard generated by medical experts is sufficient for the development of an advanced algorithm to detect weight errors. Another limitation is the lack of formal user testing with a larger sample size. We have planned the next phase of the evaluation where 40 medical professionals will be recruited to further validate the visual annotation tool. Since our pilot testing has confirmed the smoothness of the experimental workflow, we are ready to move on to this next phase.

Once the tool is formally evaluated, we will use it to collect at least 15,000 patient charts as the reference standard, upon which machine learning algorithms will be trained to detect weight errors. The large amount of annotated patient charts enables the discovery of interesting and hidden patterns from the data using machine learning techniques, rather than predefined rules or thresholds. Our future work also involves developing a visual annotation tool as a generalized self-service tool for annotating time-series data since the concepts and process in the current scenario can be applicable to the datasets that are imbalanced (very few positives) and have temporality as a key dimension.

Conclusions

In this paper, we presented our design and pilot evaluation of a visual annotation tool to rapidly collect expert-annotated weight errors for the training of machine learning algorithms. The design and evaluation were rolled out in three phases, namely, Design of core functionality and user interface, Iterative refinement through group discussion, and Experimental workflow and pilot testing. The results showed positive user feedback and confirmed the achievement of our design goal to support effectiveness and efficient expert annotation. The pilot experiment prepared us for the formal user-centered evaluation in a lab setting as the next step. We will continue this research investigation with the ultimate goal of developing an advanced clinical decision support tool to alert clinicians with erroneous weight values and subsequently prevent adverse events and patient harm in clinical practice.

Acknowledgements

We thank Ms. Catherine Xu for her review of this manuscript. This research project was supported by the 2018 Mary K. Logan Research Award that Dr. Danny T.Y. Wu received from the Association for the Advancement of Medical Instrumentation (AAMI) in the United States.

References

[1] J.C. Pham, J.L. Story, R.W. Hicks, A.D. Shore, L.L. Morlock, D.S. Cheung, G.D. Kelen, and P.J. Pronovost, National study on the frequency, types, causes, and consequences of voluntarily reported emergency department medication errors, *J. Emerg. Med.* **40** (2011) 485–492. doi:10.1016/j.jemermed.2008.02.059.

[2] M.L. Rinke, A.D. Shore, L. Morlock, R.W. Hicks, and M.R. Miller, Characteristics of pediatric chemotherapy medication errors in a national error reporting database, *Cancer*. **110** (2007) 186–195. doi:10.1002/cncr.22742.

[3] K.M. Hirata, A.H. Kang, G.V. Ramirez, C. Kimata, and L.G. Yamamoto, Pediatric Weight Errors and Resultant Medication Dosing Errors in the Emergency Department, *Pediatr. Emerg. Care*. (2017). doi:10.1097/PEC.0000000000001277.

[4] R.J. Kuczmarski, C.L. Ogden, S.S. Guo, L.M. Grummer-Strawn, K.M. Flegal, Z. Mei, R. Wei, L.R. Curtin, A.F. Roche, and C.L. Johnson, 2000 CDC Growth Charts for the United States: methods and development, *Vital Health Stat. 11.* (2002) 1–190.

[5] C. Daymont, M.E. Ross, A. Russell Localio, A.G. Fiks, R.C. Wasserman, and R.W. Grundmeier, Automated identification of implausible values in growth data from pediatric electronic health records, *J. Am. Med. Inform. Assoc. JAMIA*. **24** (2017) 1080–1087. doi:10.1093/jamia/ocx037.

[6] P.A. Hagedorn, E.S. Kirkendall, M. Kouril, J.W. Dexheimer, J. Courter, T. Minich, and S.A. Spooner, Assessing Frequency and Risk of Weight Entry Errors in Pediatrics, *JAMA Pediatr.* **171** (2017) 392–393. doi:10.1001/jamapediatrics.2016.3865.

[7] A.M. Walling, D. Tisnado, S.M. Asch, J.M. Malin, P. Pantoja, S.M. Dy, S.L. Ettner, A.P. Zisser, H. Schreibeis-Baum, M. Lee, and K.A. Lorenz, The quality of supportive cancer care in the veterans affairs health system and targets for improvement, *JAMA Intern. Med.* **173** (2013) 2071–2079. doi:10.1001/jamainternmed.2013.10797.

[8] M.D. Aldridge, and D.E. Meier, It is possible: quality measurement during serious illness, *JAMA Intern. Med.* **173** (2013) 2080–2081. doi:10.1001/jamainternmed.2013.9469.

[9] Manifesto for Agile Software Development, (n.d.). https://agilemanifesto.org/ (accessed April 1, 2019).

[10] J. Brooke, SUS - A quick and dirty usability scale, *Dev. Part Usability Eng. Programme Integr. Off. Syst. Dev. Digit. Equip. Co Ltd Read. U. K.* (n.d.) 7.

[11] M. Conlen, S. Stalla, C. Jin, M. Hendrie, H. Mushkin, S. Lombeyda, and S. Davidoff, Towards Design Principles for Visual Analytics in Operations Contexts, in: Proc. 2018 CHI Conf. Hum. Factors Comput. Syst., ACM, New York, NY, USA, 2018: pp. 138:1–138:7. doi:10.1145/3173574.3173712.

Address for Correspondence

Danny T.Y. Wu, PhD, MSI, Office: +1 (513) 558-6464

Email: wutz@ucmail.uc.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190345

Providing Comorbid Decision Support via the Integration of Clinical Practice Guidelines at Execution-Time by Leveraging Medical Linked Open Datasets

William Van Woensel[a], Samina Abidi[a], Borna Jafarpour[a], Syed Sibte Raza Abidi[a]

[a] *NICHE Research Group, Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia, Canada,*

Abstract

Clinical Decision Support Systems (CDSS) utilize computerized Clinical Practice Guidelines (CPG) to deliver evidence-based care recommendations. However, when dealing with comorbidity (i.e., patients with multiple conditions), disease-specific CPG often interact in adverse ways (e.g., drug-drug, drug-disease interactions), and may involve redundant elements as well (e.g., repeated care tasks). To avoid adverse interactions and optimize care, current options involve the static, a priori integration of comorbid CPG by replacing or removing therapeutic tasks. Nevertheless, many aspects are relevant to a clinically safe and efficient integration, and these may change over time—task delays, test outcomes, and health profiles—which are not taken into account by static integrations. Moreover, in case of comorbidity, clinical practice often demands nuanced solutions, based on current health profiles. We propose an execution-time approach to safely and efficiently cope with comorbid conditions, leveraging knowledge from medical Linked Open Datasets to aid during CIG integration.

Keywords:

Practice Guideline, Decision Support Systems, Clinical; Comorbidity

Introduction

Clinical Practice Guidelines (CPG) are disease-specific and evidence-based recommendations to guide diagnosis, prognosis, and treatment [1]. By computerizing CPG as *Computer Interpretable Guidelines (CIG)*, they can be utilized by *Clinical Decision Support Systems (CDSS)* to deliver evidence-informed care recommendations. While CDSS can effectively manage a single, complex condition, they often cannot cope with multiple illnesses in a single patient (i.e., comorbidity). Indeed, managing comorbid conditions in a patient is a complex problem, and requires avoiding adverse interactions (e.g., drug-drug or drug-disease), by tailoring or adding new therapeutic interventions; and optimizing care and resource utilization, by eliminating redundant investigations. Hence, the direct, concurrent application of multiple disease-specific CIG is considered neither a safe nor efficient option.

A possible approach for managing comorbidity involves reviewing existing, disease-specific CPG, and manually integrating their workflows to yield a single, clinically safe and resource-optimized clinical workflow. Examples include the England and Wales NICE (National Institute for Health and Care Excellence) CPG [2], which manages kidney disease combined with hypertension, lipids and hepatitis C; and Abidi [3] performed a manual, ontology-based integration of CIG for Congestive Heart Failure (CHF) and Atrial Fibrillation (AF). Nonetheless, this solution is considered impractical, since it is not feasible to apply it to every possible comorbidity. In line with his, a range of computer-based approaches have been proposed to semi-automatically integrate comorbid CIG [3-7]. These approaches tend to focus on the a priori integration of comorbid CIG to realize a single, static clinical workflow. Nevertheless, we observe that execution-time aspects driving *clinical workflow execution*—i.e., lab test outcomes, medical resource availability, care delays, and the dynamic patient's health profile—also affect the clinical safety and efficiency of *comorbid CIG integration*. Due to runtime delays, test outcomes or changes in the patient's health parameters, integration decisions may need to be applied, adapted or reverted. Hence, we argue that a comorbid CIG integration approach should incorporate the *execution-time flexibility to dynamically apply or revise previous integration decisions*, in light of evolving clinical profiles and operational aspects. Below, we introduce 3 running examples (based on illness-specific CPG) which will be used to illustrate our work:

- *OA-Diabetes comorbidity*: both osteoarthritis (OA) and diabetes guidelines recommend Non-Steroidal Anti-Inflammatory Drugs (NSAID), such as aspirin and ibuprofen. Since NSAID increase the risk of bleeding in OA patients, this type of drug should only be described once for a comorbid patient.

- *AF-CHF / COPD-PE*: many illnesses require imaging scans, ECG and other tests (e.g., spirometry) at a point-of-care or healthcare facility. With many of these illnesses often occurring comorbidly, such as AF-CHF and Chronic Obstructive Pulmonary Disease (COPD) – Pulmonary Embolism (PE), grouping these tests into a single visit will improve the patient's quality of life and reduce incurred costs.

- *HTN-Diabetes*: thiazide diuretics such as Methyclothiazide are often prescribed for treating hypertension (HTN). However, they may worsen glycemic control for diabetes patients. In regular cases, the dose of thiazide diuretics should be reduced; when creatinine clearance is low and volume control is needed, it should be replaced by loop diuretics.

We present a knowledge-driven, execution-time approach for integrating comorbid CIG into a clinically safe and efficient comorbid clinical workflow. To cope with execution-time aspects and clinical pragmatics, our approach supports (1) refining or reverting integration decisions as more data becomes available—such the outcomes of medical tests, unforeseeable runtime delays, or limited hospital resources; and (2) defining alternative CIG integration decisions, which are conditional on the patient's health profile and other clinical events. Our approach involves a *CIG Integration Framework*,

which employs a Semantic Web based approach to CIG representation, integration and execution. To support the identification of adverse interactions, we leverage various Linked Open Datasets (LOD), including the SNOMED CT ontology[1], the Drug-Drug Interactions Ontology (DINTO)[2], the Drug Ontology (DrOn)[3] and Bio2RDF DrugBank ontology[4].

Methods

Our approach supports a practitioner in (a) identifying the set of clinical tasks, called *CIG integration points*, which could result in loss of clinical safety or efficiency in comorbid sitations; (b) given these integration points, instantiating appropriate *CIG integration policies* that ensure clinical safety and efficiency of the integrated CIG. In particular, a *CIG integration policy* is instantiated to deal with a particular comorbid patient scenario: e.g., to cope with drug interactions in a HTN-Diabetes comorbidity, a policy is instantiated to adjust dosages of Methyclothiazide.

The CIG Integration Ontology (*CIG-IntO*) defines the core terms for representing, integrating and executing CIG. We note that the SNOMED CT taxonomy is integrated into our CIG-IntO ontology, with an extended mapping to other ontologies (DrOn, DINTO, DrugBank). As we will illustrate in the paper, by leveraging knowledge from these datasets, our CIG framework can aid the practitioner in discovering CIG integration points.

A CIG integration policy is responsible for coordinating the execution-time integration of comorbid CIG; which is done by *operating on a multi-level state machine*. At runtime, these state machines manage the lifecycles for integration points. We elaborate on these lifecycles, and the available integration operations, below. Next, we discuss the high-level semantics of our proposed CIG integration policies.

Clinical task lifecycles

Workflow-related task lifecycle

Fig. 1 shows the workflow-related clinical task lifecycle.

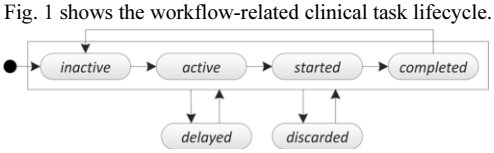


Figure 1– State machine for the CIG task workflow lifecycle.

During regular execution, a CIG task travels from *inactive* to *active* when it is next in line for execution. This may occur when its associated pre-conditions are satisfied and/or the previous task was completed. The practitioner selects an *activated* task for execution at their discretion, thus moving it to the *started* state. Finally, an *activated* task moves to the *completed* state when a practitioner marks it as complete, or when indicated by an external system (e.g., lab information system). Below, we exemplify the *CIG execution semantics* [8] (using OWL2 DL [9]) that move *inactive* tasks to the *active* state:

$$InactiveTask \cap TaskWithSatisfiedCondition \cap \exists hasPrevious.CompletedTask \subset \exists hasNewTaskState.Active \quad (1)$$

An *InactiveTask* with a satisfied pre-condition (*TaskWithSatisfiedCondition* class) and a prior, *completed* workflow task (*hasPrevious* relation), will be assigned the new *Active* state.

To coordinate the integration of comorbid CIG, an integration policy can apply *integration operations*, which move a clinical task to the *delayed* or *discarded* state, or *replace* a task in the CIG workflow. To ensure continuing clinical safety and efficiency in light of changing execution-time circumstances, our approach supports *reverting* these operations as well.

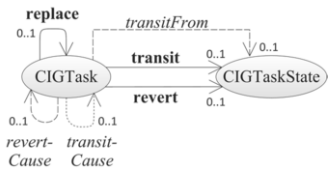


Figure 2– Integration operations on clinical tasks (CIG-IntO).

The *transit* relation indicates a transit operation on a CIG task (*CIGTask*) to a particular state (*CIGTaskState*). To support reverting the operation, the *transitFrom* property keeps the original *from* state. The *revert* operation is represented by the *revert* relation, indicating the state that the CIG task should be reverted to. The *replace* relation indicates the clinical task (subject) to be replaced and the replacement task (object). The causes for *transit* and *revert* operations are kept as well.

Outcome-related task lifecycle

Various outcome-related information, including medical test results or physician/patient choices, will only become available as the clinical workflow progresses. When faced with such new execution-time information, additional integration operations may be needed, while previous operations may need to be adapted or reverted. To that end, CIG tasks additionally feature an outcome-related lifecycle:

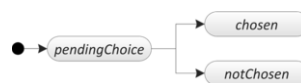


Figure 3– Integration operations on clinical tasks.

A clinical task is in the *pendingChoice* state when no choice has yet been made at the nearest preceding decision node; i.e., it is still unknown whether the task will be executed. After a choice is made, all tasks in the "chosen" branch travel to the *chosen* state, up until the next decision node; and tasks in the non-chosen branch(es) travel to the *notChosen* state. In other words, tasks in the *chosen* state are in line for execution, while *notChosen* tasks will not be executed.

CIG Integration Policies

We present an initial set of CIG integration policies to safely and efficiently integrate comorbid CIG at execution-time. We formalize their high-level semantics in terms of the CIG Integration Ontology (CIG-IntO) and First-Order Logic (FOL) rules. These integration semantics are informed by, and operate on, the multi-level state machine introduced previously.

EquivTasksPolicy

A practitioner creates an *EquivTasksPolicy* instance to cope with equivalent tasks: e.g., both OA and diabetes guidelines recommend NSAID, but this type of drug should only be prescribed once, since it increases risk of bleeding. To that end, the practitioner identifies the integration points and instantiates an *EquivTasksPolicy* as follows (using Turtle syntax[5] with *CIG-IntO* as default namespace):

```
:NSAID_equiv a :EquivTasksPolicy ;                    (2)
   :equivTask :task_OA_NSAID ;
```

[1] http://browser.ihtsdotools.org/
[2] https://bioportal.bioontology.org/ontologies/DINTO
[3] https://bioportal.bioontology.org/ontologies/DRON
[4] http://bio2rdf.org/
[5] https://www.w3.org/TR/turtle/

```
:equivTask :task_Diabetes_NSAID ;
```

Here, the *equivTask* relation indicates the equivalent tasks, i.e., prescribing the NSAID medications, as integration points. Below, we elaborate on how these integration points may be identified, and the concrete integration semantics.

- Finding equivalent tasks

In CIG-IntO, many workflow tasks are attached to SNOMED procedures; which, if applicable, is further linked to the prescribed medication (Turtle syntax, "sno" is the SNOMED namespace). For instance, in the OA CIG:

```
:task_OA_NSAID a :WorkflowTask ;                    (3)
    :involves :presc_OA_NSAID.
:presc_OA_NSAID a sno:Prescription; sno:drugUsed sno:NSAID.
```

Since the relevant workflow tasks from both the OA and diabetes CIG are linked to a *Prescription* (subclass of *Procedure*) of the same drug (*drugUsed*), the system flags the (potential) comorbidity issue to the practitioner, indicating the two clinical tasks as integration points. Based on their medical knowledge and expertise, the practitioner may then proceed by instantiating an appropriate CIG integration policy.

- CIG integration semantics

An informed execution-time choice should be made on which equivalent task will be kept, and which one(s) will be discarded. Simply discarding an arbitrary task at *design-time* may negatively affect CIG execution, as shown in Fig. 4. Here, task T_1 was discarded at design time (Fig. 4.a) since it would have been executed last under normal circumstances. However, due to execution-time delays, task T_1 is executed first; meaning that the following tasks, such as T_2, which rely on task T_1 or its equivalent (i.e., T_A) having been executed, need to be delayed until T_A is executed (Fig. 4.b).

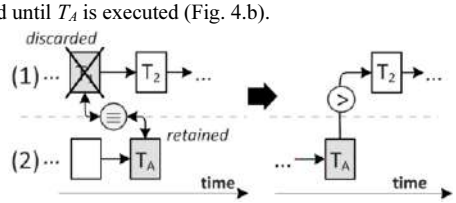


Figure 4– Integration operations on clinical tasks.

Our execution-time approach purposefully introduces a race condition; whichever task is reached first (i.e., *activated*) during execution will be retained. In Fig. 4, this means that T_1 will be retained instead of T_A, meaning that both workflows are able to proceed without delay.

These execution-time integration semantics can be represented using the following state-transition rule:

$$EquivTasksPolicy(p), intPt(p,t_1), intPt(p,t_2), t_1 \neq t_2, \quad (4)$$
$$ActiveTask(t_1), ChosenTask(t_2), NotDiscardedTask(t_2),$$
$$IncompleteTask(t_2), taskState(t_2,s)$$
$$\rightarrow transit(t_2,'discarded'), transitFrom(t_2,s), transitCause(t_2,t_1)$$

Once any integration point (*intPt)* is activated (t_1), any other integration point (t_2) part of a *chosen* branch, i.e., in line for execution, and not yet *discarded* or *completed*, will be moved to the *discarded* state (*transit* operation). The origin state (*transitFrom*) and cause (*transitCause*) are recorded as well. To simplify the formulation of integration semantics, we define complex OWL subclasses of *CIGTask*; e.g., *InactiveTask* class includes tasks with value *inactive* for property *state*.

ReplaceTasksPolicy

A *ReplaceTasksPolicy* instance is created to replace one or more tasks by safer and/or more cost-effective task(s). For instance, the thiazide diuretics prescribed for HTN may worsen glycemic control in case of HTN-Diabetes comorbidity. In clinical practice, solutions for such interactions are often nuanced, and depend on the patient's dynamic health profile. For normal cases, the dosage of thiazide diuretics should be reduced; but when creatinine clearance is low and volume control is needed, loop diuretics should be prescribed instead. This can be formulated as follows:

```
:thiazide_diuretics_diabetes a :ReplacePolicy ;          (5)
   :toReplace :thiazide_diuretics
   :replacement [
     :condition cond:default ;
     :element :low_dose_thiazide_diuretics ] ;
   :replacement [
     :condition [
       cond:operand :low_creatinine_clearance ;
       cond:conn cond:and ;
       cond:operand :volume_control_needed
     ] ;
     :element :loop_diuretics ] .
```

The *toReplace* relation indicates the integration point(s), i.e., the clinical task(s) to be replaced. The *replacement* relation indicates the conditional substitutions for the CIG task, marking (a) the condition (*condition* relation), which refers to the patient profile; and (b) the replacement task (*element* relation). The default substitution, i.e., with "default" condition, prescribes *low dose thiazide diuretics*; the second substitution prescribes *loop diuretics* in case *volume control is needed* and *low creatinine clearance* is observed.

Using our CIG framework, the practitioner can leverage LOD to identify such integration points. To that end, the *:thiazide_diuretics* task is linked to the related medication (e.g., see (3)). Based on this knowledge, there are multiple ways for the CIG framework find potential interactions:

- Finding drug-drug interactions (DDI)

Searching the DINTO and Bio2RDF DrugBank ontologies, the system discovers multiple DDI between Methyclothiazide and different types of insulin—if the patient was prescribed insulin for managing glucose levels, the system will flag these DDI to the practitioner together with the integration points. E.g., DrugBank encodes the DDI in question as follows:

```
drb:DB00232 dcterms:title "Methyclothiazide"@en ;        (6)
    drb:ddi-interactor-in db:DB00030_DB00232 .
drb:DB00030_DB00232 a drb:Drug-Drug-Interaction .
```

- Finding drug-illness interactions (DII)

The system finds that Methyclothiazide is a type of *Thiazide Diuretic* (SNOMED CT), which are categorized as *Hyperglycemia-associated Agents* (DrugBank); whereas Diabetes is a type of *Disorder of Glucose Metabolism* (SNOMED CT). Based on a DII knowledge source, which encodes adverse interactions between agents and disorders of that type, the system flags the DII and integration points to the practitioner.

- CIG integration semantics

When *multiple tasks* need to be replaced, as before, a choice needs to be made on which task to discard (see Fig. 4). In that case, a transition rule similar to (4) will be applied—which additionally ensures that other task(s) (i.e., t_2) are only discarded when a replacement was actually applied (see (7)). When only a *single task* is to be replaced (as in (5)), the *toReplace* task is simply discarded when it becomes active.

Based on the patient's current health profile, a substitution will be selected at execution-time, after which the *toReplace* task will be replaced in the CIG workflow (*NotApplied* ensures that a replacement is applied only once):

$$ReplaceTasksPolicy(p), intPt(p,t_1), ActiveTask(t_1), \quad (7)$$
$$replacement(p,r), condition(r,c), Satisfied(c),$$
$$NotApplied(r), element(r,t_2)$$
$$\rightarrow replace(t_1,t_2), applied(r,'true')$$

SimulTasksPolicy

To simultaneously execute certain clinical tasks, the practitioner creates an instance of *SimulTasksPolicy*. For instance, many comorbid illnesses involve imaging scans, ECG and

other tests (e.g., spirometry) at a point-of-care or healthcare facility; by grouping such medical tests together in time, we may improve the patient's quality of life and reduce healthcare costs. This can be encoded using *SimulTasksPolicy* as follows:

```
:simult_siteofcare_tests a :SimultTasksPolicy ;          (8)
  :simulTask [
    :involves [
      a sno:Evaluation_Procedure ;
      sno:placeOfTesting [ a sno:SiteOfCare ]
] ] .
```

In the above case, the policy aims to group in time any clinical task that *involves* any procedure of type of *Evaluation Procedure*, and with any *placeOfTesting* of type *SiteOfCare*. In SNOMED CT, medical tests such as ECG, MRI scans, spirometry are all subtypes of *Evaluation Procedure*; whereas *SiteOfCare* has subtype *Hospital*, among others. Based on the policy definition, the CIG integration framework identifies concrete integration points for the *SimulTasksPolicy*. Below, we exemplify such an integration point:

```
:task_AF_ECG a :WorkflowTask ;                           (9)
    :involves :proc_ECG_monitor .
:proc_ECG_monitor a sno:ECG_Monitoring ;
   sno:placeOfTesting sno:Hospital .
```

- *CIG integration semantics*

To group the identified integration points at execution-time, a *SimulTasksPolicy* will delay any activated *simulTask* until the other *simulTasks* have "caught up" (i.e., have also been activated)—at least, as long as it is clinically safe to do so.

In particular, the *max. (overall) delay* of the active task shows the max. period *the task may still be delayed* while safeguarding clinical safety. Hence, as long as the *max. delay* exceeds the (estimated) *activation time* of an inactive task, it will make sense, and not violate clinical safety, to delay the active task:

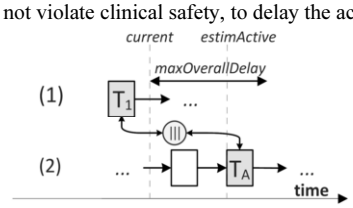


Figure 5– Example workflows for SimulTasksPolicy.

Here, the *max. overall delay* for T_1 exceeds the *estimated activation time* of T_A, meaning that T_1 can be safely delayed until T_A catches up. The following should hold: *current* + $maxOverallDelay_1 \geq estimActive_A$, which is formalized as follows:

$$maxOverallDelay(t_1, d_1), estimActive(t_2, a_2), t_1 \neq t_2, current + d_1 \geq a_2 \rightarrow delayCondition(t_1, t_2) \quad (10)$$

The *maxOverallDelay* property is calculated based on (a) the initial max. period that workflow tasks can be delayed while ensuring patient safety; and (b) any already incurred delays during execution, as well as the impact of discarded and replaced tasks. The *estimActive* property is calculated recursively from the prior task's *estimActive* time and duration.

The integration semantics are represented as follows:

$$SimulTasksPolicy(p), intPt(p, t_1), intPt(p, t_2), t_1 \neq t_2, ActiveTask(t_1), delayCondition(t_1, t_2), taskState(t_1, s), PendingOrChosenTask(t_2), InactiveTask(t_2) \rightarrow transit(t_1, 'delayed'), transitFrom(t_1, s), transitCause(t_1, t_2) \quad (11)$$

In case the *delayCondition* holds for an active task (t_1) and at least one other integration point (t_2), which is part of a branch with a pending decision or a chosen branch (*PendingOrChosenTask*), the active task will be delayed.

As the CIG execution progresses and more information becomes available, integration decisions may need to be adapted or reverted. When the delay condition no longer applies, the delay operation should be reverted:

$$SimulTasksPolicy(p), intPt(p, t_1), DelayedTask(t_1), transitCause(t_1, t_2), \neg(delayCondition(t_1, t_2)) \rightarrow revert(t_1, 'delayed'), revertCause(t_1, t_2) \quad (12)$$

In case the *delayCondition* no longer holds for a *delayed* task and the reason for the delay (*transitCause*), the delay operation will be reverted. Note that, at execution time, the delay condition could become valid again at any point; in that case, the *delayed* state transition (10) would be triggered again.

Secondly, when the operation's *transitCause*, i.e., reason for the delay, eventually becomes part of a non-chosen branch (*NotChosenTask*), it becomes pointless to delay the clinical task any further, and the delay operation will be reverted:

$$SimulTasksPolicy(p), intPt(p, t_1), DelayedTask(t_1), transitCause(t_1, t_2), NotChosenTask(t_2) \rightarrow revert(t_1, 'delayed'), revertCause(t_1, t_2) \quad (13)$$

Once the previously inactive task has caught up (i.e., is *activated*), task t_1 no longer needs to be delayed:

$$SimulTasksPolicy(p), intPt(p, t_1), DelayedTask(t_1), transitCause(t_1, t_2), ActiveTask(t_2) \rightarrow revert(t_1, 'delayed'), revertCause(t_1, t_2) \quad (14)$$

When multiple medical tests need to catch up, an active medical test will be delayed multiple times. By keeping the causes of task operations (*transit*/*revertCause* properties; Fig. 2), the integration algorithm reverts an operation *only once all its causes are reverted*; as discussed in the next section.

CIG Integration Algorithm

In this section, we discuss how the CIG integration algorithm applies the CIG task operations inferred by the integration semantics. We assume an underlying reasoner that is loaded with the rules representing the integration semantics. The CIG integration algorithm is called whenever dataset changes may lead the reasoner to fire one or more rules:

- Any tasks' state changes caused by the CIG execution algorithm (see *Workflow-related task lifecycle*).
- External events indicating execution-time delays (e.g., from scheduling software), which affects the estimated times (e.g., *maxOverallDelay, estimActive*).

The *performIntegration* function iteratively applies task operations and reverts until no more operations are inferred:

```
function performIntegration()                            (15)
  do
    op ← performTaskOperations()
    op ← op ∪ performTaskReverts()
  while (op ≠ ∅)
```

The *performTaskOperations* function calls the *performTransitOperations* and *performReplaceOperations function* (not shown). We show the former function below:

```
function performTransitOperations()                      (16)
  for each inferred transit operation:
    retract current state of task (= origin state)
    assert target state for task
    assert origin state(transitFrom),cause(transitCause)
end function
```

For each inferred *transit* operation, the function retracts the task's current state, and asserts its new state, keeping its origin state and transit cause. Below, we show *performTaskReverts*:

```
function performTaskReverts()                            (17)
  for each inferred revert operation:
    retract reverted cause (revertedCause)
    if all task transit causes ≠ ∅:
      retract current task state
      assert original target state (transitFrom)
end function
```

For each inferred *revert* operation, the "reverted" cause is retracted (e.g., see (11)-(13)). When no more *transit causes* are left, the task's current state is retracted and its original

state is asserted. We note that, by only reverting a *transit* operation once all its causes are reverted, we support scenarios where e.g., a medical test is delayed multiple times (see (10)).

The *CIG Client* (not detailed here) features a UI that visualizes the current state of comorbid CIG as a workflow, and allows clinicians to make decisions on task execution and integration.

Results

When deploying execution-time, comorbid CIG integration at a point-of-care, there is an expectation of timely recommendations. Hence, we evaluate our approach by measuring the performance of the CIG integration algorithm. To that end, we (a) modeled the comorbid CIG; (b) instantiated suitable CIG integration policies; and (c) executed the CIG integration algorithm to perform execution-time integration for a number of comorbidity scenarios.

We executed each experiment 10 times on a PC equipped with 8GB of RAM and an Intel® Core™ i7-3520 CPU. Table I shows the average performance results. Loading the comorbid CIG and integration policies took avg. ca. 270ms; calculating dynamic properties (e.g., *estimActive*) took avg. ca. 180ms.

Table 1 – CIG Integration Performance Results

Comorbid CIG Integration Scenario	performance (s)
EquivTasksPolicy	
(OA-diabetes) discard second NSAID	0.79
ConditionalReplacePolicy	
(HTN-Diabetes) in regular cases, reduce dosage; when creatinine clearance is low and volume control is needed, replace with loop diuretics.	0.94
SimulTasksPolicy	
group tests concurrently at health facility	0.9

Discussion

Table 1 shows that integrating comorbid CIG takes between 0.7 – 0.9s. We note that, since performance is mostly determined by the underlying reasoner (in our case, Apache Jena), selecting a more performant reasoner would improve performance. Nevertheless, we already consider these acceptable performance times for a consumer-grade PC.

To the best of our knowledge, the work by Anselma, Piovesan and Terenziani [4, 10] presents the first approach to focus on the temporal dimension for comorbid CIG integration. Analysis facilities allow a practitioner to analyse CIG for adverse interactions and identify solutions. To emulate our execution-time approach, the practitioner would need to frequently utilize these facilities *for any set of relevant actions*, and manually apply, adapt and revert integrations when needed. Wilk et al. [7] support a form of execution-time integration by repeating the integration process whenever new patient data becomes available. But the authors do not consider reverting integrations, and only supply two temporal revision operators. Zamborlini et al. [11] propose a rule-based approach that identifies repeated, alternative, and contradictory actions, and leverages external sources. In comparison, our work currently includes limited support for identifying adverse interactions by leveraging LOD.

Another aspect of our approach, not elaborated here due to space limitations, involves resolving conflicts between integration policies themselves. In future work, we aim to study more types of integration policies, and look into a global optimization scheme—currently, integration policies focus solely on their integration points, and disregard the global effect of their actions; e.g., a *SimulTasksPolicy* could delay a task that jeopardizes the success of more "important" policies.

Conclusions

We presented an execution-time approach and framework for comorbid CIG integration, which dynamically (a) applies CIG integrations, based on a priori defined integration policies and the latest execution-time data; (b) refines or reverts previous integrations, as more information becomes available at runtime. To provide solutions in line with clinical practice, alternative CIG integration decisions can be defined, which are conditional on the patient's up-to-date health profile. By leveraging knowledge from LOD, our framework helps practitioners in identifying adverse comorbid interactions, and formulating suitable integration policies.

References

[1] J. E. Brush, M. J. Radford, and H. M. Krumholz, "Integrating Clinical Practice Guidelines Into the Routine of Everyday Practice," *Crit. Pathways Cardiol. A J. Evidence-Based Med.*, vol. 4, no. 3, pp. 161–167, 2005. (3)

[2] National Institute for Health and Care Excellence (NICE), "Chronic kidney disease in adults," 2014.

[3] S. R. Abidi, "A Knowledge Management Framework to Develop, Model, Align and Operationalize Clinical Pathways to Provide Decision Support for Comorbid Diseases," Dalhousie University, Halifax, NS, 2010.

[4] L. Anselma, L. Piovesan, and P. Terenziani, "Temporal detection and analysis of guideline interactions," *Artif. Intell. Med.*, vol. 76, pp. 40–62, 2017.

[5] F. Real and D. Riaño, "Automatic combination of formal intervention plans using SDA* representation model," in *2007 conference on Knowledge management for health care procedures*, 2008, vol. Amsterdam, pp. 75–86.

[6] D. Riaño and A. Collado, "Model-Based Combination of Treatments for the Management of Chronic Comorbid Patients," in *Artificial Intelligence in Medicine SE - 2*, vol. 7885, N. Peek, R. Marín Morales, and M. Peleg, Eds. Springer Berlin Heidelberg, 2013, pp. 11–16.

[7] S. Wilk, M. Michalowski, W. Michalowski, D. Rosu, M. Carrier, and M. Kezadri-Hamiaz, "Comprehensive mitigation framework for concurrent application of multiple clinical practice guidelines,"*J.Biomed.Inform.*,vol.66,pp. 52–71, 2017.

[8] B. Jafarpour, S. S. R. Abidi, and S. R. Abidi, "Exploiting Semantic Web Technologies to Develop OWL-Based Clinical Practice Guideline Execution Engines," *IEEE J. Biomed. Heal. Informatics*, vol. 20, no. 1, pp. 388–398, 2016.

[9] P. Hitzler, M. Krötzsch, B. Parsia, P. F. Patel-Schneider, and S. Rudolph, "OWL 2 Web Ontology Language Primer (Second Edition)," 2012.

[10] L. Piovesan and P. Terenziani, "A Mixed-Initiative Approach to the Conciliation of Clinical Guidelines for Comorbid Patients," in *Knowledge Representation for Health Care*, 2015, pp. 95–108.

[11] V. Zamborlini, R. Hoekstra, M. da Silveira, C. Pruski, A. ten Teije, and F. van Harmelen, "Generalizing the Detection of Internal and External Interactions in Clinical Guidelines," in *9th International Joint Conference on Biomedical Engineering Systems and Technologies*, 2016, pp. 105–116.

Address for correspondence

William Van Woensel, Faculty of Computer Science, 6050 University Avenue, Dalhousie University, Halifax, Nova Scotia, Canada, B3H 1W5; E-mail: william.van.woensel@dal.ca.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190346

Proactively Guiding Patients Through ADL via Knowledge-Based and Context-Driven Activity Recognition

William Van Woensel, Samina Abidi, Syed Sibte Raza Abidi

NICHE Research Group, Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia, Canada

Abstract

Assisted Ambient Living (AAL) focuses on self-sufficiency, assisting disabled people to perform activities of daily living (ADL) by automating assistive actions in smart environments. Importantly, AAL provides opportunities for dynamically guiding patients with a cognitive decline through an ADL. Activity recognition is a pivotal task since it allows detecting when an ADL is started by recognizing its constituent activities. When dealing cognitive decline, activity recognition should also be able to detect when activities are performed incorrectly—e.g., performed out-of-order, at the wrong location or time, or with the wrong objects (e.g., utensils)—which is nevertheless not a common goal in activity recognition. Moreover, it should be able to cope with non-uniform ways of performing the ADL that are nevertheless correct. We present a novel knowledge-driven activity recognition approach, which employs semantic reasoning to recognize both correct and incorrect actions, based on the ADL workflow as well as associated environment context.

Keywords:
Artificial Intelligence, Activities of Daily Living

Introduction

Improving the quality of life for cognitively disabled patients, given the current trend of aging populations, reduces the burden on healthcare systems since patients would be allowed to remain self-sufficient as long as possible. Assisted Ambient Living (AAL) that utilizes "smart" environments outfitted with objects with sensing, actuating, computing, and communication features, is a promising solution to assists disabled people in performing daily living activities. "Smart" hardware is available off-the-shelf; whereas software frameworks [1] for interacting with smart environments, and ontologies (e.g., SOUPA [2], DogOnt [3]) for describing a context in smart homes, have been developed. Concurrently, consumer smartphones and smartwatches have become affordable and ubiquitous adding personal sensors (movement, activity, location) and communication capabilities to a smart home.

When dealing with cognitive decline, AAL supplies a unique opportunity to pro-actively assist patients: alerting them about overdue ADL, guiding them through an ADL while they perform it, and correcting their activities if needed. In this setting, accurate, comprehensive, and fault-tolerant activity recognition is paramount, as it enables the detection of ADL by recognizing their constituent activities, in a robust way that copes with incorrect actions performed by cognitively disabled patients. Data-driven and machine-learning based techniques have proven their usefulness in achieving high-accuracy activity recognition, but they require a large training dataset (cold-start problem), and the learned activity model cannot be reused for other individuals or environments. Knowledge-driven methods for activity recognition, using Semantic Web technologies, rely on semantically modeling the AAL environment and the activity's workflow [4]. Although this is initially complex and labor-intensive, knowledge-driven activity recognition is shown to offer greater flexibility and reusability to infer other ADL activities [5].

Research in knowledge-driven activity recognition, especially ADL recognition, has pursued two main issues: the representation of the ADL workflow to support recognition and resolving uncertainty in recognizing an activity due to noisy sensor input and non-uniform individual behavior. Knowledge-driven activity recognition methods typically model high-level ADL in terms of their constituent tasks [6], possibly including their sequential order [7]; or as a qualitative temporal model, based on Allen's temporal logic [8]. For instance, Chen et al. [6] perform activity recognition by utilizing semantic reasoning to classify unknown activities into a known, high-level activity class. Knowledge-based activity recognition methods have dealt with uncertainty by employing variations of ontological reasoning; such as log-linear DL [7], to assign a belief value about activity based on the sensor observations. In contrast, we aim to address uncertainty in the form of flexible, semantics-based ADL representations that, by design, deal with uncertainty caused by non-uniform individual behavior. Indeed, many correct ways typically exist to perform an ADL. An individual should be able to perform an ADL in their idiosyncratic way, as long as it is in line with the general workflow. Moreover, when dealing with cognitive decline, activity recognition should detect when necessary ADL is *not being performed*, or being performed *incorrectly* including actions that are out-of-order, at the wrong time or location, or utilizing the wrong objects (e.g., utensils). To the best of our knowledge, this goal is not addressed by current knowledge-driven approaches, and requires (a) built-in support by the activity recognition method; (b) comprehensive annotation of the ADL workflow with context information, including the needed objects (e.g., utensils, foodstuff), and an appropriate location and time; and (c) comprehensive context awareness in smart homes.

In this paper, we present a novel approach to computerizing complex ADL workflows, using an OWL ontology to represent ADL tasks and their temporal relations; and apply these constructs to perform knowledge-based activity recognition. Important goals include (a) dealing with uncertainty caused by non-uniform behaviors, i.e., idiosyncratic ways of doing things which are nevertheless correct; (b) coping with mistakes while executing ADL, as they are performed by individuals with cognitive decline; and (c) prompting to perform ADL at appropriate times and contexts. To support these goals, we utilize Semantic Web ontologies for annotating ADL workflows, such as the DogOnt Ontology [3]. We formally define the semantics of

our ADL workflow constructs, which are inspired by general (e.g., BPMN) and specialized (e.g., Clinical Practice Guideline) workflow languages. The paper is based on prior work [9], which we extend here with context-awareness, personalization towards the individual, and a more flexible ADL workflow notation and semantics.

Methods

Modeling Workflows for Activities of Daily Living

To model ADL workflows that meet our goals, we drew inspiration from formalisms for general-purpose workflows (e.g., UML activity diagrams, BPMN), computerizing Clinical Practice Guidelines (CPG), as well as high-level UI-design task models. Workflow languages typically include constructs to indicate start- and endpoints, and sequential and parallel (e.g., split and join) relations between tasks. CPG workflow languages typically support grouping multiple sub-tasks into high-level clinical tasks, as well as pre- and post-conditions. Also, tasks may be explicitly assigned to a patient or physician. In the UI design domain, Paterno et al. introduced ConcurTaskTree method for designing high-level UI task models, which similarly employs a hierarchical structuring of tasks and a set of temporal relations, among others [10]. In general, grouping subtasks into high-level or hierarchical tasks and the availability of diverse temporal relations suits our needs. ADLs are often decomposable into multiple levels of tasks, and temporal constraints may bind the correct performance of an ADL. At the same time, to cope with non-uniform behaviors, an ADL needs to encode the multitude of different ways in which activities can be performed ideally, without explicitly including all task permutations. Based on the state of the art, we proposed six workflow relations for modeling ADL:

- *Grouping tasks*: grouping lower-level tasks into a high-level, composite task that is complete once all of its (non-optional) constituent tasks are complete.
- *Order independent*: if no workflow relation is indicated between tasks, they may be executed in any order.
- *Sequential* ($T_1 >> T_2$): ordering the operands; the second task (T_2) cannot be begin before the first (T_1) is completed, where the completion of T_1 enables T_2 to take place. We present the following subtypes:
  - *Timeout-sequential* ($T_1 >>_{<timeout>} T_2$): once T_1 is completed, T_2 only starts once timespan passes.
  - *Conditional-sequential* ($T_1 >>_{[condition]} T_2$): once T_1 is completed, T_2 may only start once the condition is met.
- *Alternative* ($T_1 [] T_2$): either task can be chosen; once a choice is made, the other task can no longer be started. We further introduce the following subtype:
  - *Conditional-alternative* (T_1 [*condition*] T_2): if the condition is satisfied, T_1 is performed; else, T_2 is performed.
- *Iterative* ($T_{\{n, m\}}$): task is fulfilled between *n*, *m* times
  - *Optional* ([T]): task may be skipped (n=0, m=1).

In Figure 1, we show an example of ADL for making tea using a kettle. When a task was assigned to the smart home, it was annotated with "smart home" icon. The *StartFilling* composite task involved the *GetKettle* and *OpenTap* atomic tasks. These might be performed in any order; depending on individual's idiosyncrasies (and wastefulness!). Users might choose first to get the kettle or open the tap. Regardless, both tasks needed to be completed before moving on to *StopFilling*; and only after 3-5 seconds, or until the kettle is half-full (depending on the sensing/actuating capabilities of the smart home). Similarly, *StopFilling* task included the order-independent *PutKettleOnStove* and *CloseTap* atomic tasks. *TurnStoveOnMax* may be performed in any order (e.g., getting the kettle, turning on the stove, opening the tap). Although multiple degrees of freedom were introduced to allow idiosyncratic behavior, the workflow enforced a correct performance of the ADL. For example, an individual cannot get the kettle and put it on the stove without filling it first with water. We noted that *GetCup* and *GetTea* might be performed at any point, and were prerequisites for *PutTeaInCup*.

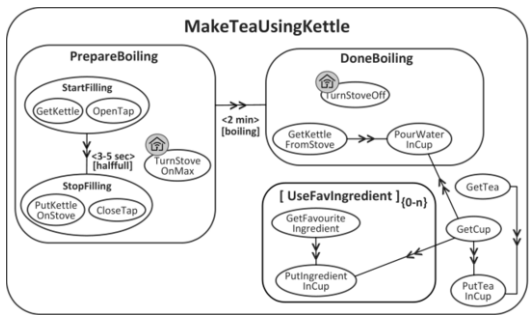


Figure 1– MakeTea ADL.

Once all *PrepareBoiling* subtasks were completed, after 2 minutes or once the water was boiling (depending on sensing capabilities), the individual could proceed with *DoneBoiling* subtasks. The task *GetKettleFromStove* was a prerequisite for *PourWaterInCup*, together with *GetCup* that must be completed first. The *GetTea* and *PutTeaInCup* might be performed at any point in the beginning; while the kettle was boiling; or after the water was poured in the cup. Moreover, individual might perform the optional up to n times that involved *GetFavouriteIngredient* and *PutIngredientInCup*. Finally, *TurnStoveOff* might be performed at any time.

Using our ADL ontology, the *MakeTeaWithKettle* ADL was encoded using RDF. Below, we discuss how this semantic ADL description can be extended with environment context.

Annotating Tasks with Environment Context

By associating ADL tasks with context, such as the required objects (e.g., utensils, foodstuff), and location and time where it should took place, the system can achieve the following:

- Guidance: inform the individual on which objects are needed, and where they are found (e.g., kitchen cupboard), together with the task's location and time.
- Troubleshooting: correlate the specified objects to the utilized objects with their time and location, and highlight mistakes.
- Prompting: based on the current task and time, prompt the individual to perform the (next) task.

Below, we illustrate how tasks from MakeTeaUsingKettle were annotated to support such features (Turtle notation, "do" represents DogOnt[1], "adl" is our custom ADL ontology):

adl:MakeTeaUsingKettle **adl:todoAt** time:around4pm ; **(1)**

[1] Since DogOnt was designed for home automation, we extended it with e.g., support for non-automated devices (e.g., kettle, water tap).

adl:requires [**a** *do:Kitchen*] ;
adl:GetKettle **adl:requires** [**a** *do:Kettle*] .
adl:PutKettleOnStove **adl:requires** [**a** *adl:Kettle*] .
adl:OpenTap **adl:requires** [**a** *do:WaterTap*] .
adl:CloseTap **adl:requires** [**a** *do:WaterTap*] .

These context annotations supported the adaptation to the current context and the individual's profile. Using blank nodes as existential variables, these context annotations represented a query over the environment context, including the location model (i.e., the layout of rooms and their contents), which is collected by the smart home sensors and the profile. Once resolved, these annotations would indicate concrete locations and objects, as well as the latest context state. For instance:

do:ikeaKettle **a** *do:Kettle* ; **(2)**
do:initiallyAt *do:kitchenCupboard1* ;
do:currentlyAt *do:stove1* .
do:kitchenCupboard1 **do:alwaysAt** *do:kitchen* ;
do:image *do:img1927202* .

In our case, the blank node expression in lines 2-3 (code 1) resolved the kettle activity; which specified *do:initiallyAt* in *do:kitchenCupboard1* and provided its latest detected location (*do:currentlyAt*) *do:stove1*. The cupboard was *do:alwaysAt* the kitchen, with an image illustrating its location. The detection of *do:currentlyAt* depended on the smart home's capabilities (e.g., a "smart" kettle with RFID tag that can be read by the stove). Using this concrete context, the system indicated necessity of a kettle for the second and third task (code 1) and the location of the kettle (*guidance*), alerted when the wrong item was used for boiling water (*troubleshooting*), and prompted the individual when tea-time is due (*prompting*).

ADL Workflow Semantics

Figure 2 shows the SmartAssist Ontology. Each *TemporalRelation* (e.g., sequential, alternative) had a left and right operands; whereas a *WorkflowTask* might act as the left or right operand for any number of temporal relations. A *Task* could be either an *AtomicTask* or *CompositeTask*. The former included tasks assigned to individuals (*PatientTask*) and the smart home (*SmartHomeTask*); and the latter included *ADL* themselves. The *TemporalRelation* class included a subclass for each type of temporal relation; with *SequentialRelation* and *AlternativeRelation* including subclasses for their conditional (and timeout) versions. Further, a *Task* could be associated with one or more *Context* (see code 1). To support the formal semantics of ADL workflows, each *Task* had an associated *State (inactive*, *active*, *started*, *completed*, or *error)*. These states and the transitions between them, are shown in the form of a state machine in Figure 3.

Each task was initially in the *inactive* state. A task became *active* when it was next in line for execution, as determined by the ADL workflow (multiple tasks may be *active* at the same time). Inversely, *inactive* meant that the task should not be executed at this time. The *completed* state indicated that the task was successfully executed, and was marked either by the individual or the smart home, depending on its sensing capabilities. In case an error was detected with regards to the task's execution, the state would transition to the *error* state. Orthogonal to these states, a task might also be in the *ready* state, meaning that their required-context had been met; or the *todo* state, meaning that their todo-context was met (i.e., at which the task should be done). For instance, *MakeTeaUsingKettle* (code 1) required the individual to be inside the kitchen (requires-context), and the task should be done at around 4 pm (todo-context); whereas *GetKettle* required first retrieving the kettle from the kitchen cupboard (requires-context). Although a task could be in multiple states, the *inactive* state was retracted once a task moved to the *active* state.

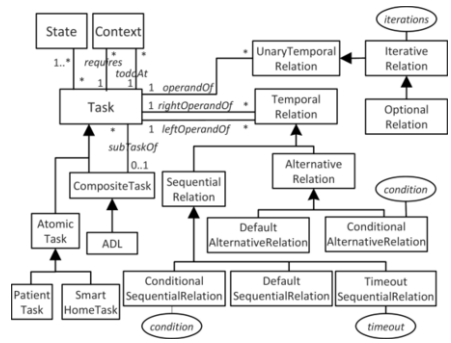


Figure 2– SmartAssist Ontology.

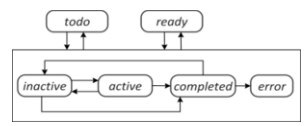


Figure 3– Task states and transitions between them.

The formal semantics of an ADL workflow was defined in terms of state transition rules that governed the transitions between task states. We listed them below using Description Logics (DL). To simplify the DL expressions, we assumed that each *Task* was assigned a type based on its current state. A task in the *completed* state was assigned the *Completed* type; a task in any other state was assigned *Incompleted*; operands for the *optional* relation were members of *Optional*.

(a) Any ↔ Ready

(a.1) $Task \cap \forall requiresContext.(Active) \subseteq Ready$

An ADL task was moved to the *ready* state once all its *required* context was active—e.g., in case the individual retrieved the *ikeaKettle* for making tea (see code 1, 2).

(b) Inactive → Active

Any ADL task for which any higher-level composing task (*subTaskOf* property) was *active,* was considered for transitioning to the *active* state (as indicated in the rules below).

(b.1) $Task \cap \forall subTaskOf.(Active) \cap \neg(\exists rightOperandOf.(TemporalRelation)) \subseteq Active$

Tasks not acting as the right operand for any temporal relation (b.1) were moved to the *active* state. For instance, *GetTea* and *GetCup* would become *active* once the ADL was *active*.

Note that this means that any ADL will always be activated as well, since they are not related to higher-level composing tasks, nor act as operands for temporal relations.

(b.2) $Task \cap \exists subTaskOf.(Active) \cap \forall rightOperandOf.($
$(SequentialRelation \cap \exists leftOperand.(Completed \cup Optional))$
$\cap (DefaultSequentialRelation \cup$
$(ConditionalSequentialRelation \cap \exists condition.(Satisfied))$
$\cup (TimeoutSequentialRelation \cap \exists timeout.(Satisfied)))$
$\subseteq Active$

Regarding the right-hand operands of *sequential* relations, for all incoming sequential relations, the left operand must be either *Completed* or *Optional*, before the task was moved to the *active* state. In case of the *conditional* or *timeout* version, the condition or timeout also needed to be satisfied,

respectively, before moving to *active*. For example, *PutTeaInCup* only moved to *active* once *GetCup* and *GetTea* were *activated*; and *StopFilling* only became *active* once the *StartFilling* task was *completed* and 3-5s had passed or "half-full" was satisfied (depending on the capabilities).

(b.3) $Task \cap \exists subTaskOf.(Active) \cap \exists operandOf.(DefaultAlternativeRelation) \subseteq Active$

Any operand of *default-alternative* directly moves to *active*.

(b.4) $Task \cap \exists subTaskOf.(Active) \cap \exists rightOperandOf.(ConditionalAlternativeRelation \cap \exists condition.(Unsatisfied)) \subseteq Active$

For the *conditional-alternative* version, in case its condition was not met, the right operand would move to *active*. Otherwise, the left operand task would simply remain active, due to the rule in (b.1) that activated all left- (or non-) operands.

(c) Active → Inactive

(c.1) $Task \cap Active \cap \exists leftOperandOf.(ConditionalAlternativeRelation \cap \exists condition.(Unsatisfied)) \subseteq Inactive$

In case the *conditional-alternative* relation's condition was not met, its left operand would move to *inactive*.

(d) Active → Complete

(d.1) $Task \cap Active \cap \forall operandOf(AlternativeRelation \cap \exists operand.(Completed)) \subseteq Completed$

An ADL task acting as operand for an *alternative* relation would be completed once one of the other operands were completed. The user might only execute one alternative task.

(d.2) $CompositeTask \cap \forall hasSubTask.(Completed) \subseteq Completed$

A composite task would be moved to the *completed* state once all its subtasks were *completed*. For instance, the *StopFilling* task would move to *complete* once its subtasks *PutKettleOnStove* and *CloseTap* were *completed*.

(e) Completed → Inactive

(e.1) $Task \cap Completed \cap Active \cap \forall transitiveSubTaskOf.(ADL \cup Completed) \subseteq Inactive$

Any *completed* task that is part of a *completed* ADL (*transitiveSubTaskOf* property) will move back to the *inactive* state.

(f) Completed → Error

(f.1) $Task \cap Inactive \cap Completed \subseteq ErrorTask$

Regarding erroneous performance, a task that was *completed* while *inactive* would move to the error state since it was not executed within ADL workflow. For instance, *CloseTap* task would move to the error state if it was *completed* in the *inactive* state and before completing *StartFilling* composite task.

(f.2) $Task \cap Completed \cap \neg Candidate \subseteq ErrorTask$

In case a *completed* ADL task was not *Ready*, i.e., its *requiredContext* had not yet been reached (a.1), it would be moved to the error state such as completing *GetKettle* with a different item (e.g., glassware).

(g) Active → Todo

(g.1) $Task \cap Active \cap todoAtContext.(Active) \subseteq TodoTask$

An *active* task with active *todoAt* context would move to the *todo* state—e.g., *MakeTeaWithKettle* at around 4 pm.

Note that, for brevity, we did not consider cases where a task acts as the right operand for different types of temporal relations (it is not needed in our example). This required correlating the number of *active* state transitions to the number of types of incoming temporal relations.

Knowledge-Driven Activity Recognition

Based on the ADL workflows and detected user actions (i.e., atomic tasks), a semantic reasoning process applied the proposed task state transition rules to recognize the current states of ADL activities. In doing so, we realized *knowledge-driven activity recognition*, where the start and completion of ADL were recognized, based on the execution of their constituent, lower-level tasks; and erroneously performed tasks were marked (*error* state), based on workflow relations and associated context. Importantly, it supported non-uniform behavior during ADL completion, allowing individuals to perform the ADL in an idiosyncratic way as long as they adhere to the core ADL constraints. Our approach dealt with multiple, simultaneous ADL when the individual perform constituent tasks in any interleaved way that adheres to the individual ADL.

To detect the atomic actions that drove knowledge-driven activity recognition and detected current context, we relied on sensor data processing techniques, as elaborated by Van Woensel et al [9]. Each time an atomic action or new context was detected, the semantic reasoning process was executed. Activity recognition results were passed to any top-level component (e.g., knowledge-driven pro-active assistance).

Knowledge-Driven, Pro-active Assistance

We supplied three features for pro-active assistance:

Guidance

Assistive acts were issued to guide a patient through their daily ADL routines. Once the start of an ADL was recognized, guidance kept the patient apprised of the current progress in the workflow. For example, when a patient performed a cooking activity, assistive acts provided information about the activity's progression (current subtask) and its current state (instructions, location of utensils and ingredients, etc.) on nearby devices (e.g., TV, smartphone, tablet). If the smart home was connected to devices, some tasks could even be performed automatically, such as the *TurnStoveOnMax* and *TurnStoveOff* tasks (Figure 1). The system could offer guidance by leveraging context annotations such as showing that a kettle was needed for a task (code 1) and the location of kettle (code 2).

Troubleshooting

Troubleshooting involved alerting the individual about detected errors during ADL execution and explaining the correct workflow. Our novel activity recognition process detected when a task was inconsistent with the ADL workflow (rule f.1). Further, our system detected when a task was done in the wrong context (e.g., wrong item, location; rule f.2). An error would also be raised when an activated task (e.g., *GetKettle*) was not completed within a reasonable time.

Prompting

When an ADL was overdue, as inferred from its associated context (see code 1; rule g.1), the system would prompt the individual with increased urgency until the ADL is carried out—ensuring their daily routine continues as expected. Corresponding to this increasing urgency, we applied an *evolving* notification lifecycle [11], where interactions (e.g., icons, audio & haptic feedback) increased in obtrusiveness overtime, together with frequency, until the complete.

Results

We implemented a system prototype to evaluate the feasibility of our activity recognition approach. Uutilizing the Hermit

reasoner (v. 1.3.8.4) for reasoning over the DL transition rules:

- Ontology loading: utilizing the Hermit API to load the ontology, state transition rules and ADL workflow.
- Initialization: running semantic reasoning after loading.
- Detected action: performing the semantic reasoning step after detecting an atomic action or new context.

Using this prototype, we performed a preliminary evaluation of our activity recognition method, executing 10 simulated scenarios for the *MakeTeaWithKettle* ADL (Figure 1), where the individual performed different permutations of 13 atomic tasks (5 correct, 5 incorrect performances). We ran the experiments on a Lenovo Think-Pad T530 laptop running Windows 7, with an Intel Core i7-3520M CPU (2.90 GHz) and 8Gb of RAM. We ran each scenario 10 times and retrieved the average performance results of each operation (Table 1). The prototype was able to properly detect each of the incorrectly performed actions in the simulated scenarios.

Table 1– Activity Recognition Performance

Operation	Average processing time (ms)
Ontology loading	31
Initialization	299
Detected action	374

Discussion

Table 1 shows that the semantic reasoning process, which underlies our knowledge-driven activity recognition, takes between 0.3 and 0.4 seconds. While performance is mostly determined by the reasoner (in this case, Hermit), these seem acceptable performance times for a consumer-grade PC.

In contrast to data-driven activity recognition, which train an activity model based on a collected dataset, knowledge-driven activity recognition relies on an a priori designed knowledge artifact. A subset of knowledge-driven works applies logical reasoning to infer high-level activities from detected atomic actions. Chen et al. [5] performed activity recognition by classifying an unknown instance, with sensor observations attached as properties, as an activity class with constraints defined on properties (e.g., *hasContainer*) and values (e.g., cup). This approach is unable to cope with tasks being performed incorrectly, or in the wrong temporal sequence, since this would lead to an incorrect classification. Helaoui et al. [6] represented activity models using log-linear DL, and proceeded similarly to Chen et al. In case of incompatibilities due to noisy sensor observations, the most probable activity is inferred based on confidence values. However, their approach only considers a *sequential* temporal relation. Okeyo et al. [8] incorporates composite, interleaving and concurrent task relations based on Allen's temporal logic. However, the approach does not support conditional, alternative, or optional relations, and does not focus on dealing with incorrect actions.

Future work involves studying how to deal with uncertainty resulting from faulty sensor observations (e.g., utilizing Probabilistic Description Logics [12]). We also aim to investigate automatic generation of ADL using automated planning, based on the required context of tasks. Further, we aim to utilize our method to determine non-compliance to health activities in behavioral self-management.

Conclusions

We proposed a novel approach to knowledge-driven activity recognition, based on a formal semantics in the form of state transition rules. Properties of our method include: (a) coping with uncertainty, caused by non-uniform behaviors; and (b) robustness towards mistakes during ADL performance. Based on recognized activities, a pro-active assistance process provides features such as guidance, i.e., keeping the patient appraised of their progress; troubleshooting, i.e., correcting the patient's actions, if needed; and prompting, i.e., informing the patient when an ADL is required. We further linked ADL with environment context. Future work involves accounting for more types of variation in which people complete their tasks—e.g., heating water using the microwave instead and personalizing the utilized modalities for pro-active assistance.

References

[1] M.A. Razzaque, M. Milojevic-Jevric, A. Palade, and S. Clarke, Middleware for Internet of Things: A Survey, *Ieee Internet of Things Journal* **3** (2016), 70-95.

[2] H. Chen, F. Perich, T. Finin, and A. Joshi, SOUPA: Standard ontology for ubiquitous and pervasive applications, *Proceedings of Mobiquitous* 2004 (2004), 258-267.

[3] D. Bonino and F. Corno, DogOnt - Ontology Modeling for Intelligent Domotic Environments, *Semantic Web - Iswc* 2008 **5318** (2008), 790-803.

[4] G. Azkune, A. Almeida, D. Lopez-de-Ipina, and L.M. Chen, Extending knowledge-driven activity models through data-driven learning techniques, *Expert Systems with Applications* **42** (2015), 3115-3128

[5] L.M. Chen, J. Hoey, C.D. Nugent, D.J. Cook, and Z.W. Yu, Sensor-Based Activity Recognition, *Ieee Transactions on Systems Man and Cybernetics Part C-Applications and Reviews* **42** (2012), 790-808.

[6] L.M. Chen, C.D. Nugent, and H. Wang, A Knowledge-Driven Approach to Activity Recognition in Smart Homes, *Ieee Transactions on Knowledge and Data Engineering* **24** (2012), 961-974.

[7] R. Helaoui, D. Riboni, and H. Stuckenschmidt, A probabilistic ontological framework for the recognition of multilevel human activities, in: *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, ACM, Zurich, Switzerland, 2013, pp. 345-354.

[8] G. Okeyo, L.M. Chen, and H. Wang, Combining ontological and temporal formalisms for composite activity modelling and recognition in smart homes, *Future Generation Computer Systems-the International Journal of Escience* **39** (2014), 29-43.

[9] W. Van Woensel, P.C. Roy, and S.S.R. Abidi, Achieving Pro-Active Guidance of Patients through ADL via Knowledge-Driven Activity Recognition and Complex Semantic Workflows, Proc. 10th Int. Conf. Semant. Web Appl. Tools Heal. Care Life Sci **2042** (2017).

[10] F. Paternò, *Model-based design and evaluation of interactive applications*, Springer, London, 1999.

[11] W. Van Woensel, P.C. Roy, S.R. Abidi, and S.S.R. Abidi, A Mobile and Intelligent Patient Diary for Chronic Disease Self-Management, *Medinfo 2015: Ehealth-Enabled Health* **216** (2015), 118-122.

[12] T. Lukasiewicz, Expressive probabilistic description logics, *Artificial Intelligence* **172** (2008), 852-883.

Address for correspondence

William Van Woensel, E-mail: william.van.woensel@dal.ca.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190347

IEC 62304 Ed. 2: Software Life Cycle Standard for Health Software

Alpo Värri[a], Patty Kranz-Zuppan[b], Richard de la Cruz[c]

[a]Tampere University, Tampere, Finland
[b] Medtronic, plc, Mounds View, Minnesota, USA
[c] Silver Lake Group, Inc. (SLGI), Minnetonka, Minnesota, USA

Abstract

The quality of software is high in medical devices due to the strict regulatory requirements and their implementation in the software development processes through the use of the IEC 62304 standard. The goal of this standard revision project was to extend the scope of the standard to all health software and also to bring the requirements of the 12 year old standard back to the state-of-the-art including provisions for cybersecurity. The joint IEC/SC62A and ISO/TC215 project team revised the standard and adapted its risk management, usability, and security requirements to serve both the medical device industry and the overall health software industry. The resulting second version of the standard has gone through a multistage global voting process to achieve a consensus of the requirements to serve both these communities. The resulting standard has potential to have a major impact on the quality of software used in health care globally.

Keywords:
Standards; Health Care Technology; Software Engineering

Introduction

The quality of software in medical devices is very important because a failure in medical device software can have fatal consequences to the patients in some cases. A classic example of this is the case of the Therac-25 radiation therapy machine in the 1980's [1]. The weaknesses in the production processes of the software to this device resulted in undetected programming errors which led to massive radiation overdoses in at least six accidents and three dead patients. It is a well-known fact in the software engineering field that the testing of the software does not alone ensure sufficient quality of the software, but the production process of the software needs to be of high quality, as well. Few, if any other documents have a greater impact on how the software development processes have been arranged in the medical device industry than the IEC standard 62304 Medical device software – Software life cycle processes [2]. This standard is recognized by the Food and Drug Administration in the USA. It is also on the list of harmonized standards under Directive 93/42/EEC for Medical devices in the European Union; although, it does not necessarily cover all the requirements of the Directive 2007/47/EC for medical devices [3]. The medical device industry applies this standard because it may be the easiest way to demonstrate the conformity of the company's medical device software production process with the regulatory requirements. The regulatory requirements suggest that the standard is typically used in conjunction with the quality system standard ISO 13485 [4], the risk management standard ISO 14971 [5], and the medical device usability standard IEC 62366-1 [6], which are harmonized standards as well. The standard has a three class software safety classification A, B, and C of which class C represents the highest risk class software.

The growing needs to improve the quality of all the software used in health care has led to the idea to extend the scope of the 62304 standard beyond medical devices to all health software. The term health software is defined to include all software used for managing, maintaining, or improving the health of individual persons, or the delivery of care, still also including the software in medical devices. In the regulatory meaning, the medical device software can be embedded to a device or it can be a medical device as such, so-called Software as a Medical Device (SaMD). Software with a health purpose runs as a service, e.g. in a cloud service, could be a medical device if it fulfills the definition of a medical device, but if not, it is still covered by the definition of health software. This independence of the hardware platform means that this definition and the new version of the 62304 standard cover the popular mobile health apps.

After the approval of the revision proposal, the work on the second edition of the 62304 standard was (re)started in October 2014. The revision work was carried out in a project team of the Joint Working Group seven (JWG7), which is a co-operation working group between IEC subcommittee 62 A *Common aspects of electrical equipment used in medical practice* (IEC/SC62A) and ISO technical committee 215 *Health Informatics* (ISO/TC215) because both committees are interested in contributing to this work. The project team consisted of experts nominated by the national standards bodies of the member countries of these standardization committees.

The health informatics standards are not very well-known by the health informatics research community, partly because the standards are relatively expensive. For this reason, it is important to introduce the new version of the 62304 standard here also to the scientific community, because health informatics researchers may later face it anyway if they want to commercialize their research results.

Methods

Design Specification

The work began by defining the target of the second version of the standard. The highlights of the planning were:

- The scope of the standard would be extended from medical device software to cover all health software
- The standard would remain a life cycle standard in contrast to being a product standard
- The software maintenance process requirements would remain in the standard
- The same software safety classification would remain in the standard and the required rigor to the safety class C software would be preserved
- Significant changes to the requirements should be avoided unless there is a compelling motivation for the change
- The new version of the standard should be applicable to fulfill the regulatory requirements of medical device software

The expansion of the scope to all health software expands the user base of the standard significantly to new audiences. The designers of non-medical device health software have often not applied equally strictly controlled processes as the medical device manufacturers and the required level of rigor may come as a surprise to these companies when their customers begin to require the use of 62304 ed. 2 in the software production.

Keeping the standard as a life cycle standard instead of a product standard leaves out two main activities of the standard. A product standard such as the IEC 82304-1 [7] contains also software use requirements and the validation that the use requirements are met but these are out of the scope of this standard.

The need to keep essential changes minimal comes from the high costs that major changes could cause to the industry when they would have to change their processes and train the personnel for the changes. Major changes might also make it more difficult for the regulatory bodies to accept the standard as a way to fulfill the regulatory requirements, particularly if the changes are considered to result in less safe software than before.

Literature Review

When the revision was started, a literature review was carried out as well. For a standardization project, other related standards are the most relevant literature, because a standard can refer to other standards normatively if the other standards contain material that is not feasible to repeat in the current standard. As the list of potentially relevant standards spans a few pages, only a small subset of them is mentioned here. An earlier version of the ISO/IEC/IEEE 12207 [8] standard was used in the drafting the first version of 62304. Its latest developments were checked to maintain a certain level of compatibility to general software development standards. The IEC 60601-1 standard [9] contains requirements for medical device software and they were compared to the requirements in 62304. The risk management subclauses of 62304 were checked against ISO 14971 [5] and its draft versions under development. A similar comparison was carried out with ISO 13485 [4]. ISO 90003 [10] was checked as an alternative to ISO 13485 for health software which is not classified as a medical device. In addition to the harmonized usability standard IEC 62366-1 [6], the corresponding medical device standard IEC 60601-1-6 [11] was also considered.

The IEC 61508-3 functional safety of programmable electronic systems standard [12] was checked for comparison although its field of application is not in the medical domain. The ISO/IEC TR 29110 series [13] was identified to contain recommendations for software life cycle management in small organisations but its requirements were not sufficiently strict for high risk medical device software production. SWEBOK V3 [14] was used as the latest state-of-the-art document for software engineering.

During the course of the work, it became apparent that the cybersecurity issues in medical devices and health software in general became more and more important. A cybersecurity problem in a medical device can develop into a patient safety problem too, if the problem is in a safety critical device. Previously, when medical devices were isolated to their own networks, the cybersecurity issues could be rather safely ignored by the medical device community, but in today's interconnected world, their addressing has become mandatory. Regulatory cybersecurity guidance for manufacturers of medical devices has been published and already revised in the USA [15] and corresponding guidance is expected from the European Commission as well. Moreover, the health software which is not embedded in a medical device is typically at risk of cybersecurity attacks as personal medical data is more valuable to cybercriminals than credit card numbers [16].

A large number of security-related standards have been identified by the project team members. The NIST Cybersecurity Framework [17] can be taken as the starting point for improving security in software development but other alternatives exist too. The project team investigated the applicability of several security standards. The ISO 27799 health informatics security standard [18] refers to the general ISO/IEC 27002 standard [19] but they are not very much related to the software life cycle processes. The IEC/ISO 80001 series of standard documents [20] and particularly TR 80001-2-2 [21] and TR 80001-2-8 [22] are also relevant because they inform the health delivery organizations how to manage security risks with networked medical devices which run software developed according to 62304. These standards introduce security requirements to software which need to be addressed in the software development life cycle. The IEC 62443 series of standards, particularly the IEC 62443-4-1 [23] from the industrial automation field is a good reference source for secure development life cycle requirements. The Microsoft Security Development Lifecycle model [24] is an alternative security-aware software design model from the general software development field. Finally, the ISO/TR 17791 [25] contained a survey of standards for enabling safety in health software; thus, the project team had a good comprehension of the existing literature to be considered.

Working Method

At the outset of the work, the project team had the draft amendment [26] of the 62304 standard in its disposal. It updated parts of the first version of the 62304 without the scope extension to health software. The project team began to introduce changes to the document based on the design specification and also on the feedback received from the field in the application of the 62304.

The work was carried out in a series of meetings. In between the meetings, the core project team members formulated new paragraphs of text to the subsections to be revised. The project team has produced three drafts of the standards, which have been circulated for voting in the both standardization committees IEC/SC62A and ISO/TC215. The national standardization bodies have collected the comments of each member country, and they have been delivered to the project team for handling. The project team is responsible of

addressing each one of the comments and either (partially) approve the comment or reject it with sufficient motivation. For example, the most recent vote resulted in about fifty pages of comments to be considered, indicating great interest in the contents of the standard. The governing working group needs to accept the project team's disposition of comments before the draft standard can proceed to the next stage. When all the comments have been addressed, the standard draft is sent to the Final Draft International Standard (FDIS) vote, which decides the approval of the document as an international standard. This is the normal working method of the international standardization organizations like IEC and ISO.

Results

At the time of writing, the latest draft addresses around 95 per cent of the received comments in agreement of the stakeholders. The scope has been extended to all health software and the necessary changes relating to this change have been implemented to the standard. The draft 62304 second edition standard covers now the following stages for the software development process:

1. Software risk management process
2. Software development planning
3. Software requirement analysis
4. Software architectural design
5. Software detailed design
6. Software unit implementation
7. Software integration and integration testing
8. Software system testing
9. Software release
10. Software configuration management
11. Software problem resolution

The software maintenance process is similar to the development process, but the stages 2 and 3 have been replaced with the establishment of the software maintenance plan and the problem and modification analysis.

Both the old and new versions of the 62304 require the use of a quality management system in the software production. The old version of the standard required that the manufacturer applied a risk management process complying with ISO 14971 [5]. The new version requires the conformity to ISO 14971 only when a software failure can contribute to a hazardous situation, which in turn can lead to injury or damage to the health of people, or damage to property or the environment. The new risk management process must also manage risks associated with security. The new version does not make any security standard mandatory, but it suggests some that can be used.

The new version makes the requirement to apply a usability engineering process explicit. The use of the IEC 62366-1 [6] is not mandatory but it is given as an example of how to demonstrate the conformity to the usability design requirement.

The software safety classification still has the three classes, A, B, and C. The assignment of the software to these classes has be clarified. The procedure can be seen in Figure 1.

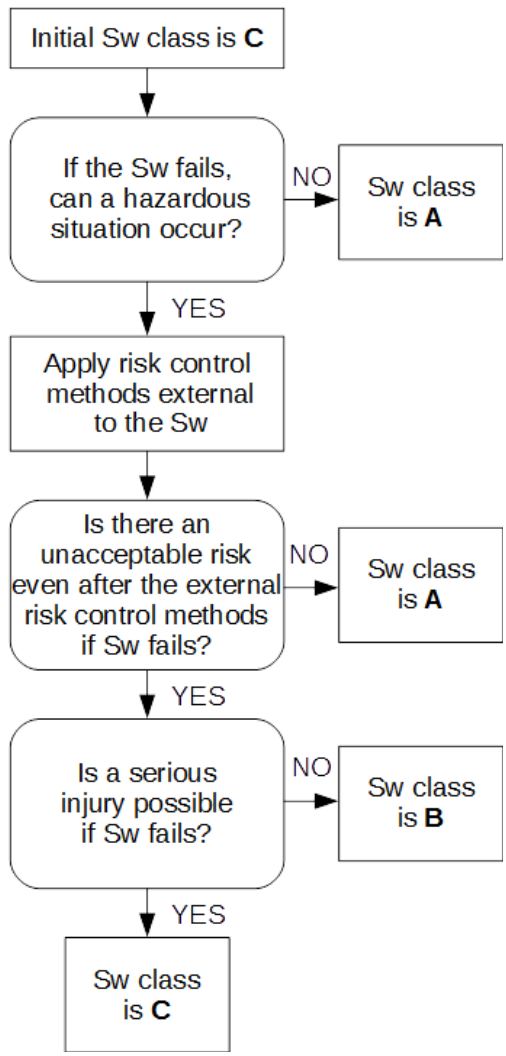


Figure 1. The Procedure to Assign the Safety Class to Software.

Legacy software is software that was produced before the new version of the standard was published, and it was therefore not possible to produce the software according to this standard. The new version has provisions how to demonstrate the compliance of the legacy software with respect to the new version.

The changes in the requirements for the software development process are relatively minor. There are a few additional requirements for the software system test record contents, but otherwise the changes are not very significant.

Similarly, the changes in the requirements for the software maintenance process are few. The new version requires that the change requests are analyzed also with safety class A software, which was not a requirement before.

The software risk management process has new requirements. More potential causes of contributions to hazardous situations must now be identified, including those related to cybersecurity.

The changes to the software configuration management process are again minor. The same applies to the changes in the software problem resolution process.

Over half of the document contains rationale and guidance in the implementation of the standard. Much of the text has remained the same or has been revised for additional clarity. The extended risk management sections now include new material which explains IT risks too. The relationships between the software safety classification and risk management are explained in more detail as well. The risk management of legacy software contains also additional new information.

The informative Annex C about the relationship to other standards has been partially rewritten. New standards are introduced to the table of standards to be considered. Accordingly, the bibliography now contains references to updated sources.

Discussion

The new version of the 62304 software life cycle processes standard extends its scope to all software production for health purposes while maintaining the current user base of medical device software producers. It remains to be seen, how widely the new audiences begin to use the standard voluntarily, because its implementation to the processes of a software company can be a significant effort. This is particularly the case with small software companies that have not produced regulated software before. Larger software companies with established quality systems may have an easier task in adjusting their processes.

The benefits of beginning to apply the revised 62304 standard to a company which has not applied a standard-based process before are the following:

1. The planning, programming, testing, and the documentation of the software becomes more controlled
2. As a result of the above, the final product will have fewer errors
3. The company will have a well-structured process to handle software updates, planned or corrective updates
4. The reputation of the company improves because the audited proof of high-quality software production processes is not so easy to achieve

For the customer, the benefit is that the products produced under the state-of-the-art software production standard are likely to be of higher quality than those produced without this methodology.

When the health software buying organizations hear about the completion of the second edition of the 62304 standard with its extended scope, they may decide to use it in procurement in order to improve the quality of the software they buy. When this happens, software companies wishing to stay competitive need to begin to apply this standard. The national regulators of health software in developed countries may also recognize their opportunity to improve the health software quality in their countries by requiring conformity to this standard. This would ensure the success of this standard in the field.

The additional requirements regarding cybersecurity do not lengthen the standard much, but their inclusion demands changes in the organizations applying the standard. The many publicly well-known cybersecurity incidents have probably triggered most members of the industry to react to the threats even without the requirements from this standard.

As the new standard has not yet passed the final vote, all the details of the standard have not been frozen. There is still discussion about the requirements regarding risk management, but the discussion will converge to a conclusion which the majority of the national standards bodies can approve.

The update of a standard causes consequences in other standards, as well. At least documents like IEC 82304-1 [7] and IEC/TR 80002-1 [27], and IEC/TR 80002-3 [28] need to be partly revised.

It is interesting to note how little direct impact the scientific literature actually has in the final version of this kind of a standard. The standard preparation work needs to consider other closely related standards more than the scientific works. Additionally, comments from the national ISO member country votes need to be considered in order to reach consensus. In this process, new radical ideas from the research community can easily get lost.

Conclusions

This paper presented the goals, methods, and results of the IEC 62304 Software life cycle processes standards for health applications. The revision of the standard brings the standard to the state-of-the-art in software production including provisions to cybersecurity and extends its potential user base to all health software producers. Thus, it has potential to have a major impact to the quality of software used in health care.

Acknowledgements

The authors want to thank all the 62304 edition 2 project team members for their contributions to the work. The support of the Finnish Standards Association (SFS) to the travel expenses to the project team meetings is also acknowledged.

References

[1] Therac-25, http://en.wikipedia.org/wiki/Therac-25 [accessed March 30, 2019]
[2] IEC 62304:2006 Medical device software – Software life cycle processes, https://www.iso.org/standard/38421.html [accessed November 22, 2018]
[3] European Commission, Harmonized Standards, Medical Devices, http://ec.europa.eu/growth/single-market/european-standards/harmonised-standards/medical-devices_en [accessed November 22, 2018]
[4] ISO 13484:2016 Medical devices -- Quality management systems -- Requirements for regulatory purposes, https://www.iso.org/standard/59752.html [accessed November 22, 2018]
[5] ISO 14971:2007 Medical devices -- Application of risk management to medical devices, https://www.iso.org/standard/38193.html [accessed November 22, 2018]
[6] IEC 62366-1: 2015 Medical devices -- Part 1: Application of usability engineering to medical devices, https://www.iso.org/standard/63179.html [accessed November 22, 2018]
[7] IEC 82304-1 Health Software -- Part 1: General requirements for product safety,

https://www.iso.org/standard/59543.html, [accessed November 22, 2018]
[8] ISO/IEC 12207:2017 Information technology – Software life cycle processes, https://www.iso.org/obp/ui/#iso:std:iso-iec-ieee:12207:ed-1:v1:en [accessed November 22, 2018]
[9] IEC 60601-1:2005, Medical electrical equipment – Part 1: General requirements for basic safety and essential performance, IEC 60601-1:2005/AMD1:2012.
[10] ISO/IEC 90003:2014, Software engineering – Guidelines for the application of ISO 9001:2008 to computer software, https://www.iso.org/obp/ui/#iso:std:iso-iec:90003:ed-2:v1:en [accessed November 22, 2018]
[11] IEC 60601-1-6:2010+AMD1:2013 CSV, Medical electrical equipment – Part 1-6: General requirements for basic safety and essential performance – Collateral standard: Usability, https://webstore.iec.ch/publication/2596 [accessed November 22, 2018.
[12] IEC 61508-3, Functional safety of electrical/electronic/programmable electronic safety-related systems – Part 3: Software requirements.
[13] ISO/IEC TR 29110-1:2011 Software engineering -- Lifecycle profiles for Very Small Entities (VSEs) -- Part 1: Overview, JTC1/SC07, http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html [accessed November 10, 2014]
[14] P. Bourque, R.E. Fairley, Eds., Guide to the Software Engineering Body of Knowledge Version 3.0, SWEBOK V3, https://www.computer.org/web/swebok/v3 [accessed November 22, 2018]
[15] FDA, Content of Premarket Submissions for Management of Cybersecurity in Medical Devices, Draft Guidance for Industry and Food and Drug Administration Staff, USA 18.10.2018,https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM623529.pdf [accessed November 22, 2018]
[16] C. Humer, J. Finkle, Your medical record is worth more to hackers than your credit card. Reuters Technology News, 24.9.2014, https://www.reuters.com/article/us-cybersecurity-hospitals/your-medical-record-is-worth-more-to-hackers-than-your-credit-card-idUSKCN0HJ21I20140924 [accessed November 22, 2018]
[17] NIST Cybersecurity Framework, version 1.1, April 2018, USA, https://doi.org/10.6028/NIST.CSWP.04162018 [accessed November 22, 2018]
[18] ISO 27799:2016 Health informatics -- Information security management in health using ISO/IEC 27002, https://www.iso.org/obp/ui/#iso:std:iso:27799:ed-2:v1:en [accessed November 22, 2018]
[19] ISO/IEC 27002:2013 Information technology - Security techniques - Code of practice for information security controls, https://www.iso.org/obp/ui/#iso:std:iso-iec:27002:ed-2:v1:en [accessed November 22, 2018]
[20] IEC 80001-1:2010 Application of risk management for IT-networks incorporating medical devices -- Part 1: Roles, responsibilities and activities, https://www.iso.org/obp/ui/#iso:std:iec:80001:-1:ed-1:v1:en [accessed November 22, 2018]
[21] IEC 80001-2-2:2012 Application of risk management for IT-networks incorporating medical devices -- Part 2-2: Guidance for the communication of medical device security needs, risks and controls, https://www.iso.org/obp/ui/#iso:std:iec:tr:80001:-2-2:ed-1:v1:en [accessed November 22, 2018]
[22] IEC 80001-2-8:2016 Application of risk management for IT-networks incorporating medical devices -- Part 2-8: Application guidance-- Guidance on standards for establishing the security capabilities identified in IEC 80001-2-2, https://www.iso.org/obp/ui/#iso:std:iec:tr:80001:-2-8:ed-1:v1:en [accessed November 22, 2018]
[23] IEC 62443-4-1:2018 Security for industrial automation and control systems - Part 4-1: Secure product development lifecycle requirements, https://webstore.iec.ch/publication/33615 [accessed November 22, 2018]
[24] Microsoft Corporation, What is the Security Development Lifecycle? https://www.microsoft.com/en-us/sdl/default.aspx [accessed November 22, 2018]
[25] ISO/TR 17791:2013 Health informatics -- Guidance on standards for enabling safety in health software, https://www.iso.org/obp/ui/#iso:std:iso:tr:17791:ed-1:v1:en [accessed November 22, 2018]
[26] IEC 62304:2006/Amd.1:2015 Medical device software -- Software life cycle processes AMENDMENT 1, https://www.iso.org/obp/ui/#iso:std:iec:62304:ed-1:v1:amd:1:v2:en,fr [accessed November 22, 2018]
[27] IEC TR 80002-1:2009 Medical device software - Part 1: Guidance on the application of ISO 14971 to medical device software, https://webstore.iec.ch/publication/7488, [accessed November 22, 2018.
[28] IEC/TR 80002-3:2014 Medical device software - Part 3: Process reference model of medical device software life cycle processes (IEC 62304), https://webstore.iec.ch/publication/7489 [accessed November 22, 2018]

Address for Correspondence

Alpo Värri, Tampere University, Korkeakoulunkatu 3, FI-33720 Tampere, Finland, Alpo.Varri@tuni.fi.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190348

Enhancing Guideline-Based Prescribing and Personalized Medication Scheduling

Charles Wachira, Samuel Osebe, William Ogallo, Aisha Walcott-Bryant

IBM Research - Africa, Nairobi, Kenya

Abstract

Poor communication of medication administration instructions is a preventable cause of medication nonadherence. The Universal Medication Schedule (UMS) framework improves adherence by providing a simplified set of dose timing rules. However, this framework does not readily generalize to individuals with varying daily routines. We propose a point-of-care solution for enhancing guideline-based electronic prescribing and personalizing dose schedules. We describe a JSON-based approach to encode and execute standard treatment guidelines to support electronic prescribing as well as an algorithm for optimizing medication administration schedules based on a patient's daily routine. We evaluated the structure and accuracy of our JavaScript Object Notation (JSON) formalism focusing on Kenya's hypertension treatment guidelines. Our experiments compare the medication schedules generated by our algorithm with those generated by pharmacists. Our findings show that treatment guidelines can be efficiently represented and executed using the JSON formalism, and that different medication administration schedules can be generated automatically and optimized for patients' daily routines.

Keywords:

Drug Administration Schedule, Clinical Decision Support, Medication Adherence

Introduction

Despite the known benefits and advancements in treatments, medication nonadherence remains one of the most significant healthcare challenges. The World Health Organization estimates that 50% of patients on long-term therapies are nonadherent, with direct consequences on patient health outcomes and increased healthcare costs [1]. Although the causes of medication nonadherence are multifactorial, patient misunderstanding of prescription drug label instructions is an important contributor to the situation, and patients who use multiple medications and with lower literacy levels are disproportionally affected [2–4].

Research has shown that the manner in which prescriptions are written and in which medications labeled are important underlying factors responsible for patient misunderstanding of their medication instructions [5–8]. Vagueness and unnecessary complexity of prescriptions result in variability of pharmacy interpretations of the prescriptions and subsequently makes it difficult for patients to know when to take their medications [6], and many dispensing labels emphasize drug characteristics rather than dosing instructions [7]. For example, in a study of 446 United States veterans, only 42% were able to correctly interpret a prescription with the directions "take one tablet daily with meals" [9].

In 2008, the United States' Institute of Medicine (IOM) introduced the Universal Medication Schedule (UMS) concept to simplify medication administration instructions for patients and their caregivers with the goal of improving patient understanding and medication adherence primarily to solid dosage forms used daily [10]. More than 90% of multi-drug regimens are dosed four or fewer times per day [11], and as a result, the UMS standardizes administration instructions to provide explicit timing with four standard dose periods (morning, noon, evening, bedtime). Evidence supporting the use of UMS instructions indicates that patients are 33% more likely to interpret medication instructions correctly when using the UMS [12]. Since the inception of the framework, researchers have investigated its potential for adoption in other settings such as the United Kingdom [13], and have expanded it to non-solid dosage forms [14] with considerable promise. However, there is a scarcity of literature on generalizing the framework to accommodate persons with different routines.

We posit that application of the UMS without the promotion of rational prescribing, and also considering daily routines of patients, may not sufficiently solve the problem of nonadherence and inaccurate use of medications. We believe that interfacing guideline-based prescribing with customized medication schedules based on the UMS would result in better understanding of medication instruction and lead to improved adherence. Our approach supports recent advances in health information technology decision-making aids, which have the potential to transform prescribing and dispensing practices by intelligently incorporating prescribing guidelines and medication scheduling.

We developed a point-of-care solution for enhancing guideline-based electronic prescribing and personalizing dose schedules. This involves generating electronic prescriptions based on patient parameters and existing care guidelines recommendations for some common chronic diseases (Hypertension, Diabetes, and HIV), and subsequently predicting the most appropriate dose periods, given the patient's drug regimen and routine (e.g., meals schedule and sleep schedule). The output of this process is an optimized medication schedule for guideline-recommended drug regimens. The solution can be integrated with a medication adherence reminder system to further enhance the appropriate use of medications.

In this paper, we describe the development of a guideline implementation and execution engine using the JSON formalism as well as an algorithm for dosing schedule optimization. We also provide the preliminary intrinsic evaluation of the accuracy and performance of the solutions and discuss our plan for their extrinsic evaluation.

Methods

Guideline Implementation and Execution

In this sub-section, we describe the approach used to encode and execute treatment guidelines (available as text-based flowcharts) using the JavaScript Object Notation (JSON) data structure format. Clinical practice guidelines are a critical component of care delivery because they standardize care across different settings, making it possible to sustain the quality of care in disparate contexts. Such contexts may be defined by several factors including demographic, diagnostic, environmental, and individual patient-level factors. Standard treatment guidelines provide evidence-based knowledge about treatment regimen options for patients with particular diseases and comorbidities. In personalized medication scheduling, clinical guidelines are critical for reducing variations in prescribing and dispensing of medication, as well as determining medication administration instructions that would be optimized to better fit the patient's daily routine. Furthermore, implementing standard treatment guidelines electronically would help meet the information needs for the care provider at the point-of-care, and also provide a means of data collection for determining guideline deviation in the care delivery process.

Guideline Encoding

The encoding process involves breaking down the guideline into individual steps, determining the data needed to execute the guideline, and writing the logic that consumes the data. First, we began by analyzing the guideline to determine and extract all the data items, referred to as parameters, that are needed to execute the guideline. These were then represented as a JSON data structure that specified the description of each data item (e.g., "age") and the data item's data type (e.g., "integer"). All data items are required upfront before execution begins. We preferred this approach over requesting data from users at every step as the user experience would be affected. Second, we manually extracted the steps in the guideline flowchart and represented these as nodes. Each node contains the following components:

1. Identity: Unique identifier for the node.
2. Description: Human readable information on what the node does.
3. Children: A list of all the child node identifiers, represented by set $N = \{n_0, \ldots, n_n\}$.
4. Type: Node type, represented by set $T = \{t_0, t_1\}$ where t_0 is a condition type and t_1 is a recommendation type. A node can be of type t_0, t_1 or both.
5. Action: A set of procedures that need to be taken by a caregiver, $A = \{a_0, \ldots, a_n\}$, where a_n can be null.
6. Function: A pointer to a machine executable logic. This logic is typically written by the encoding expert using a machine-readable language.
7. Fired: A flag that is set from false to true when the node is visited. Its role is to provide information on the traversal that led to a specific recommendation.

Third, we represented the guideline as a network graph $G = \{N, E\}$ with the set of nodes N, and the edges, E, where an edge $e \in E$ is represented by the relationship $e(n_m, n_n)$ where n_m is the parent and n_n is the child. Each node is a group of key-value pairs represented as JSON objects. We selected the network graph data structure over the tree data structures because it supports connected subgraphs and multiple parents per child node, which is common in flowcharts for treatment guidelines. We do not, however, support traversal from a child to its immediate parent because of the nature of this problem. For example, for nodes $N = \{n_0, n_1, n_2\}$ where n_n is the child of n_{n-1}, the edges $e(n_1, n_0)$ and $e(n_2, n_1)$ would not be allowed while traversing the network.

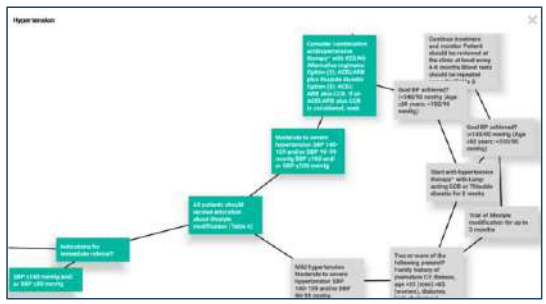


Figure 1 – Guideline snippet showing nodes activated based on input parameters

Guideline Execution

A sample guideline snippet is shown in Figure 1. Guideline execution begins with the root node. Once parameters are fed to the root node, the associated logic is retrieved and executed using a subset of the parameters. The root node function includes skipping logic for procedures that are long term. For example, when monitoring a condition for three months, the execution will skip from the root node to the node corresponding to the third month, all factors held constant. When a node is visited, its logic is retrieved from the logic repository. In this implementation, logic is represented as Java methods whose identifiers are similar to the node identifiers, providing a mapping between a node and the logic functions. Each node has only one logic function. Once the logic is retrieved, it is executed to determine the action to perform. This action can either be a traversal to one of the child nodes or a termination of the execution. If the action is a termination, the node involved returns recommendations which correspond to a procedure to be conducted.

Dose Schedule Optimization

We describe our approach that extends the Universal Medication Schedule (UMS) [10] to enable the provision of customized medication administration instructions to persons with different daily routines. The UMS is crucial in simplifying dose schedules. It is equally critical to consider patient routines in order to improve their medication experiences. However, manually determining the right schedule for the patient may be cognitively tasking and time-consuming. We propose a dose-schedule algorithm to automatically determine and rank dose schedules given patient regimens and routines. This would help dispensing staff develop more suitable medication administration plans for their patients.

Algorithm Description

Our dose-schedule algorithm computes the most optimum dose periods given a patient drug regimen and routine. The output of this algorithm is an optimized schedule for the patient's drug regimen. Each patient has a single optimized dosing schedule that can be integrated with a medication adherence reminder system or a patient health wallet to enhance appropriate use of medications. The inputs to the algorithm are as follows:

1. Prescription Information: The algorithm requires medication administration instructions such as dose frequency, dose quantity, total pills prescribed, recommended time of day, and meal restriction (e.g., "with a meal", "without a meal", "with or without a meal") for each drug in the patient's prescription.
2. Dose Intervals: The algorithm requires the dose intervals for medications given more than once a day. Based on the UMS, we set acceptable dose interval as 8 to 12 hours for medications administered two times a day, 5 to 8 hours for medications administered three times a day, and 5 to 6 hours for medications administered four times a day.
3. Patient's routine: This includes the time to wake up, breakfast time, lunch time, dinner time, and sleep time to the nearest hour.
4. Active Medication Schedule: If the patient has an active medication schedule, the algorithm uses this information along with the patient's new prescription, to generate a new medication schedule.

The dose schedule optimization process consists of several steps. These are described as follows:

1. Compute the total number of hours that the patient is awake based on sleep and wake times inputs.
2. For each drug in the prescription, compute all possible dose periods given the dosing interval and the output of (1).
3. Apply a meal constraint to the output of (2) i.e., for each drug in the prescription, check the meal restriction and constrain dose periods to hours that support this direction of use.
4. For each drug, constrain the output of (3) with any administration instructions that specify a specific time of day for taking a drug. For example, drugs which are recommended to be taken at bedtime will be confined to bedtime hours.
5. Repeat steps (2) to (4) for all drugs in the prescriptions.
6. Compute all possible combinations of the output of (5) along with present active medication schedule. This step is accomplished using recursion.
7. For each combination n in (6), compute scores based on the number of dose periods N_{dp}, number of pills per dose period N_p, and the distribution of interval durations between dose periods T_{int}. In this paper, these scores were obtained as follows:
   - Dose period score: The dose period score $S_{dp,n}$ for each combination n is calculated as the normalization of the number of dose periods $N_{dp,n}$ using the smallest score $N_{dp,min}$ and largest score $N_{dp,max}$. Each $S_{dp,n}$ is between 0 and 1, where larger values are preferable. In this formulation, combination of medications having the higher number of dose periods are more heavily penalized.

     $$S_{dp,n} = \frac{N_{dp,n} - N_{dp,min}}{N_{dp,min} - N_{dp,max}} + 1 \qquad \text{(eq1)}$$

   - Pill score: The pill score $S_{p,n}$ for each combination n is calculated as the normalization of the number of pills $N_{p,n}$ using the smallest score $N_{p,min}$ and the largest score $N_{p,max}$. Each $S_{p,n}$ is between 0 and 1 where larger values are preferable. In this formulation, combination of medications having the higher number of pills are more heavily penalized.

     $$S_{p,n} = \frac{N_{p,n} - N_{p,min}}{N_{p,min} - N_{p,max}} + 1 \qquad \text{(eq2)}$$

   - Interval Score: The coefficient of variation $C_{v,n}$ for the distribution of the duration of intervals between dose periods $T_{int,n} = \{t_{int,n,0}, \dots, t_{int,n,i}\}$ for each combination n is calculated as the ratio of standard deviation and the mean of the distribution. The interval score $S_{cv,n}$ for each combination n is then calculated as the normalization of the coefficient of variation $C_{v,n}$ using the smallest coefficient of variation $C_{v,min}$ and the largest coefficient of variation $C_{v,max}$. Each $S_{cv,n}$ is between 0 and 1, where larger values are preferable. In this formulation, combination of medications having the lower coefficient of variation are more heavily penalized.

     $$S_{cv,n} = \frac{C_{v,n} - C_{v,min}}{C_{v,max} - C_{v,min}} \qquad \text{(eq3)}$$

8. The scores in (7) are weighted by a) minimizing the number of dose periods per day, b) minimizing the number of pills per dose period, c) maximize time intervals between dose periods.

   $$S_n = W_1 * S_{dp,n} + W_2 * S_{p,n} + W_3 * S_{cv,n} \qquad \text{(eq4)}$$

   Where W_n is a weight and $W_1 + W_2 + W_3 = 1$
9. The weighted scores S_n for each combination n from eq4 are ranked and used to select the medication schedule (combination) with the best score.

We implemented the algorithm using Java. The inputs to the algorithms were collected using a mobile app that we built to automate the data entry process, while the prescriptions and existing schedules were queried from the patient's electronic health record. The output of the algorithm is shown in Figure 2.

Experiments and Results

To evaluate the JSON-based guideline system, we encoded a treatment protocol for the management of hypertension among adults in primary care that was endorsed by the Kenya Ministry of Health and the Healthy Heart Africa project [15]. This resulted in a total of 14 nodes covering non-hypertensive, mild hypertension, and chronic hypertension. 50% of the nodes were terminal nodes, meaning they resulted in therapy recommendations. Patients who have been on the therapy get reviewed every 4 to 6 months. Accuracy was determined by selection of a parameter set that covered all the seven terminal nodes. This was then fed into the guideline for execution and both the recommendations and the execution path were evaluated against human results. Lastly, we compared the attributes of the JSON formalism with those of popular guideline encoding platforms.

The guideline execution took 1ms on average. This is accounted by the fact that the guideline logic comprised of simple comparison arithmetic. This execution resulted in the correct execution path and recommendations for all the parameters tested. Table 1 shows the comparison of GLIF 3 [16], PROforma [17], GLIDES [18], and the JSON formalism presented in this paper. The guideline format compares the

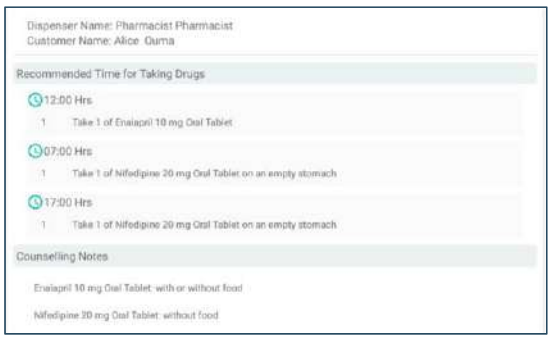


Figure 2 – Example output of a personalized dosing schedule

representations such as XML, RDF, and JSON which can be used interchangeably, and for which there exists parsers for mapping one structure onto another. Both GLIF 3 and PROFORMA have graphical tools for encoding guidelines. This, however, requires an expert such as a knowledge engineer. GLIDES approach consists a hybrid of a graphical tool and natural language approach where a free text guideline is mapped onto an XML structure [19].

To assess the performance of the dose schedule optimization algorithm, we began by generating synthetic prescriptions from existing hypertension and HIV clinical guidelines and drug labels. This process involved specifying the medication instructions for single medications. Each instruction consisted of the clinical drug, dose, frequency, and food restriction tuples e.g., "Amlodipine 10 mg Oral Tablet", "1 tablet (10 mg)", "once daily", "with or without food". Next, using information from the guidelines, we generated all possible combinations of single medications that formed known regimens for the treatment of hypertension or HIV. Examples of this include 1 Calcium Channel Blocker (CCB) + 1 Angiotensin Receptor Blocker (ARB) for the management of hypertension and 2 Nucleotide Reverse Transcriptase Inhibitors (NRTIs) + 1 Non-nucleoside Reverse Transcriptase Inhibitor (NNRTI) for HIV treatment. At the same time, we created random patient routines, each describing different sleep times, wake times, breakfast times, lunch times, and dinner times. These two sets of data (regimens and routines) were used to randomly select 100 regimen-routine combinations that were processed by the algorithm. The output of this process was 100 optimized dose schedules along with the time taken to generate each. We evaluated the average time taken to generate each optimized regimen, and compared the output of the algorithm to the output generated by two pharmacists for a random sub-sample of 23 prescriptions. The two pharmacists worked together and did not make changes to the prescriptions.

The mean dose optimization time was 498.4 ms (IQR: 453 ms to 508 ms). Of the 23 randomly selected prescriptions, the algorithm and the pharmacists had 18 common optimized regimens, implying that the algorithm is approximately 78% accurate against human experts. Inspection of the 46 individual medication instructions in the 23 prescriptions revealed that 43 instructions (93%) were correctly determined by the algorithm. This suggests that the algorithm performs better when optimizing dose schedules for individual drugs but its performance deteriorates when combining information from individual drugs to form a regimen.

Table 1. A Comparison of different guideline formalism by Format, Encoding Methodology, Execution Approach, and API Availability

	JSON	GLIF 3	PROforma	GLIDES
Format	JSON	RDF	XML	XML
Encoding	Programmatic Interface Encoding	Graphical Modelling	Graphical Modelling	Graphical & Natural Language
Execution	Programmatic Execution	GLEE Execution (GLEE)	Execution Engine (Arezzo)	Execution Engine (CDS Systems)
API	Available	Not Available	Not Available	Not Available

Discussion

This study describes an approach for implementing standard treatment guidelines for electronic prescription generation and subsequently personalizing the medication administration instructions, through a dose-schedule optimization algorithm, in the prescription based on the patient's routine. To the best of our knowledge, our study is the first to attempt to formalize standard treatment guidelines as JSON data structures. This study showed that the rate-limiting step, which is the recursive dose scheduling optimization step, takes less than a second for a single prescription optimization process. This implies that the solutions proposed in this study could be adopted at the point-of-care to enhance prescribing and dispensing as well as improve the patient's medication adherence.

Several approaches have been used for creating computerized guidelines for clinical decision support including the Arden Syntax, GLIF, and PROforma [20]. JSON formalism is a programmatic approach where functions containing the guideline logic are written using an object-oriented programming language. Consequently, it can be executed by any system that can interpret the object-oriented language chosen. These factors and the fact that it is available via an application programming interface make it easily scalable once the encoding has been done. The other guideline approaches require specialized execution and sharing platforms.

There is a dearth of literature on using technology to optimize medication administration schedules as a means of promoting medication adherence in the outpatient setting. The review by Choi et. al. [21] reveals that whereas many mobile applications have been developed in an attempt to improve medication adherence, a majority consist of similar alerts and drug information access functions. A similar review by Ahmed et al. [22] found that three strategies utilized by developers of mobile applications for medication adherence include reminders (e.g., alerts and SMS), behavioral approaches (e.g., patient tracking and gamification), and patient education. Interestingly, Stawarz [23] demonstrated that adherence to long-term therapies could be improved by taking into consideration the patient's context such as daily routine events, locations, and objects. However, currently available reminder applications do not support patients in adopting daily routines, while behavioral applications tend to ignore the importance of contextual cues [23]. Accordingly, it is critical that personalized medication scheduling solutions such as the one proposed in this study are developed and studied to complement existing approaches for enhancing medication adherence.

Conclusion and Future Work

We presented a point-of-care solution that combines guideline-based electronic prescribing and personalized medication scheduling. This was achieved through a JSON-based approach for encoding and executing clinical guidelines, and also by extending the Universal Medication Schedule to fit different patient mealtime and sleep routines. The findings of this work suggest potential in the combination of guideline-based prescribing and the customization of drug administration instructions to improve medication adherence.

A key limitation of this work is the lack of adequate evaluation of the proposed solutions, and the lack of real data from a clinical setting for the evaluation. We intend to address these in a pilot study with real patients and care providers. Our plan for the pilot includes rigorous evaluation of the accuracy, acceptability, and completeness in the transformation of clinical care guidelines concepts into the JSON formalism. We will also compare the performance of this approach to existing formalisms. We plan to compare the accuracy and similarity of the dose schedules produced by our algorithm to those produced by a committee of human experts. Furthermore, we would like to extend the functionality of our tool to detect and react to patients who deviate from, or violate, their medication schedules. Natural Language Processing (NLP) techniques can be adapted to make this possible, and this technique can also aide in the processing of guidelines. Finally, we will add checks for drug-drug interactions in the prescription orders to improve prescription safety.

References

[1] WHO | Adherence To Long-Term Therapies: Evidence for Action [Internet]. WHO. [cited 2018 Nov 24]. Available from: http://www.who.int/chp/knowledge/publications/adherence_report/en/

[2] Wolf MS, Davis TC, Curtis LM, Bailey SC, Knox JP, Bergeron A, et al. A Patient-Centered Prescription Drug Label to Promote Appropriate Medication Use and Adherence. J Gen Intern Med. 2016 Dec;31(12):1482–9.

[3] Davis TC, Wolf MS, Bass PF 3rd, Thompson JA, Tilson HH, Neuberger M, et al. Literacy and misunderstanding prescription drug labels. Ann Intern Med. 2006 Dec 19;145(12):887–94.

[4] Miller TA. Health literacy and adherence to medical treatment in chronic and acute illness: A meta-analysis. Patient Educ Couns. 2016 Jul;99(7):1079–86.

[5] Winterstein AG, Linden S, Lee AE, Fernandez EM, Kimberlin CL. Evaluation of consumer medication information dispensed in retail pharmacies. Archives of Internal Medicine [Internet]. 2010 Aug 9;170(15):1317–24.

[6] Wolf MS, Shekelle P, Choudhry NK, Agnew-Blais J, Parker RM, Shrank WH. Variability in pharmacy interpretations of physician prescriptions. Med Care. 2009 Mar;47(3):370–3.

[7] Shrank WH, Agnew-Blais J, Choudhry NK, Wolf MS, Kesselheim AS, Avorn J, et al. The variability and quality of medication container labels. Arch Intern Med. 2007 Sep 10;167(16):1760–5.

[8] Shrank WH, Avorn J. Educating patients about their medications: the potential and limitations of written drug information. Health Aff (Millwood). 2007 Jun;26(3):731–40.

[9] Trettin KW, Narus E. Implementation of a VA Patient-centered Prescription Label. Advances in Human Aspects of Healthcare. 2012;429.

[10] Hernandez LM. Standardizing medication labels: Confusing patients less: Workshop summary. National Academies Press; 2008.

[11] Benner JS, Chapman RH, Petrilla AA, Tang SSK, Rosenberg N, Schwartz JS. Association between prescription burden and medication adherence in patients initiating antihypertensive and lipid-lowering therapy. Am J Health Syst Pharm. 2009 Aug 15;66(16):1471–7.

[12] Davis TC, Federman AD, Bass PF 3rd, Jackson RH, Middlebrooks M, Parker RM, et al. Improving patient understanding of prescription drug label instructions. J Gen Intern Med. 2009 Jan;24(1):57–62.

[13] Kenning C, Protheroe J, Gray N, Ashcroft D, Bower P. The potential for using a Universal Medication Schedule (UMS) to improve adherence in patients taking multiple medications in the UK: a qualitative evaluation. BMC Health Serv Res. 2015 Mar 11;15:94.

[14] Bailey SC, Wolf MS, Lopez A, Russell A, Chen AH, Schillinger D, et al. Expanding the Universal Medication Schedule: a patient-centred approach. BMJ Open. 2014 Jan 10;4(1):e003699.

[15] Health Heart Africa. Protocol for the identification and management of hypertension in adults in primary care [Internet]. [cited 2018 Nov 25]. Available from: http://www.jkuat.ac.ke/departments/hospital/wp-content/uploads/2015/08/Hypretension-Protocol.pdf

[16] Boxwala AA, Peleg M, Tu S, Ogunyemi O, Zeng QT, Wang D, et al. GLIF3: a representation format for sharable computer-interpretable clinical practice guidelines. Journal of Biomedical Informatics [Internet]. 2004 Jun 1 [cited 2019 Mar 28];37(3):147–61.

[17] Sutton DR, Fox J. The syntax and semantics of the PRO forma guideline modeling language. Journal of the American Medical Informatics Association. 2003;10(5):433–43.

[18] Shiffman R, Michel G, Kashyap N, Dixon M. A Systematic and Replicable Approach to Knowledge Formalization.

[19] Shiffman RN, Michel G, Krauthammer M, Fuchs NE, Kaljurand K, Kuhn T. Writing clinical practice guidelines in controlled natural language. In Springer; 2009. p. 265–80.

[20] De Clercq PA, Blom JA, Korsten HH, Hasman A. Approaches for creating computer-interpretable guidelines that facilitate decision support. Artificial intelligence in medicine. 2004;31(1):1–27.

[21] Choi A, Lovett AW, Kang J, Lee K, Choi L. Mobile applications to improve medication adherence: existing apps, quality of life and future directions. Advances in Pharmacology and Pharmacy. 2015;3(3):64–74.

[22] Ahmed I, Ahmad NS, Ali S, Ali S, George A, Saleem Danish H, et al. Medication Adherence Apps: Review and Content Analysis. JMIR Mhealth Uhealth. 2018 Mar 16;6(3):e62.

[23] Stawarz KM. Towards better medication adherence apps: Preventing forgetfulness by facilitating the formation of routine-based remembering strategies. 2017;

Corresponding Address:

Charles Wachira, charles.wachira1@ibm.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190349

Meta-Analysis of the Sensitivity of Decision Support Systems in Diagnosing Diabetic Retinopathy

Priscyla Waleska Simões[a], Maitê Gabriel dos Passos[b], Laura Lopes Amaral[a], Diego Garcia[c], Ronaldo Borges Vicente[b], Larissa Letieli Toniazzo de Abreu[b], Jéssica Katelyn de Siqueira Vieira[a], Maria Marlene de Souza Pires[d], Eros Comunello[e], Luciane Bisognin Ceretta[f], Claudia Maria Cabral Moro Barra[c], Harki Tanaka[a], Deborah Ribeiro de Carvalho[c], Tiago Ribeiro de Oliveira[a], Patrícia Duarte Simões Pires[b]

[a]Engineering, Modeling and Applied Social Sciences Center (CECS), Biomedical Engineering Graduate Program (PPGEBM), Universidade Federal do ABC (UFABC), Alameda da Universidade s/n, 09606-045 São Bernardo do Campo, SP, Brazil
[b]Universidade do Extremo Sul Catarinense, Brazil
[c]Health Technology Graduate Program, Pontifícia Universidade Católica do Paraná, Brazil
[d]Medical Sciences Graduate Program, Universidade Federal de Santa Catarina, Brazil
[e]Vision Lab, Applied Computing Graduate Program, Universidade do Vale do Itajaí, Brazil
[f]Public Health Graduate Program, Universidade do Extremo Sul Catarinense, Brazil

Abstract

Diabetic Retinopathy (DR) is one of the most common microvascular complications presenting by patients diagnosticated with diabetic diseases. Uncontrolled hyperglycemia may manifest as visual impairment and blindness. The early detection of DR is essential to minimize the risk and consequence of visual diminishing. The standard gold diagnoses tool relies on different imaging modalities and requires a judgment of expert photographers, which are not available in most of the primary care centers or remote location. In that scenario, an automate or semiautomated DR screening systems can contribute to improving the accuracy of the diagnostic. Thus, we performed a Systematic Review and Meta-Analysis to evaluate the Decision Support Systems (DSS) in diagnosing DR. The overall Diagnostic Odds Ratio was 73.15 (95%CI: 37.54-142.50), sensitivity was 97.70 (95%CI: 97.50-97.90) and specificity was 90.30 (95%CI: 90.00-90.60). Our results corroborate with the concept of usefulness of DSSs in early diagnosis, screening and preliminary evaluation of suspicious images of DR.

Keywords:
Decision Support Systems, Clinical. Meta-analysis. Diabetic Retinopathy.

Introduction

The World Health Organization (WHO) indicate that Chronic diseases such as Diabetes Mellitus (DM) are experiencing a fast increase in rates worldwide. In 1997, were diagnosticated 124 million people with diabetes, where 97% corresponded to Type 2 [1]. The number of individuals already exceeded 317 million in 2012 and could reach 300 million in 2015. The prevalence of type 2 diabetes mellitus (DM2) increased in the last 25 years [2, 3].

Diabetic Retinopathy (DR) is one of the most common microvascular complications of the diabetic disease and has been responsible for the majority of cases of visual impairment and blindness in adults between 20 and 74 years [4]. The prevalence of patients with complication reaches more than 40% in the age group over 40 years [5]. According to the Center for Disease Control and Prevention, about 4.2 million Americans have Retinopathy In the United States, of which 655,000 have a high risk of vision loss [6].

The symptoms of DR are nonspecific in the early stages and relies on regular and precise screening for early detection. The delayed diagnostic and consequent delayed treatment will be ineffective to avoid the progress of the loss vision [7, 8, 9]. However, the actual standard gold diagnosis tool relies on different imaging modalities and requires a judgment of expert photographers, which are commonly not available in most of the primary care centers or remote location.

Techniques in image processing and segmentation of retinal vasculature are necessary for an accurate interpretation of acquired patient images. In that sense, health Informatics approaches, as Decision Support Systems (DSS), can be defined as an intervention that can offers health professionals with additional clinical knowledge and specific information which has the potential to increase the accuracy of care decisions [10].

In order to reduce the DR numbers, health studies are regularly carried out. Previous studies have investigated methods to support the diagnosis and screening of Diabetic Retinopathy [11, 12].

Some studies have shown that Decision Support Systems, when used as reminders, improve preventive care, increase diagnostic accuracy, clinical performance [13], influence clinical decision-making [14, 15] and significantly improve decision quality [16, 17].

Despite all the investment, the accuracy of Decision Support Systems in diagnosing Diabetic Retinopathy still an unclear matter. Therefore, in this research, we aimed to provide evidence on the subject, which may collaborate in the development of strategies for early detection of peripheral diabetic neuropathy, and contribute to more effective glycemic control. We evaluated the accuracy of Decision Support Systems in diagnosing Diabetic Retinopathy through a Systematic Review and Meta-analysis.

Materials and Methods

The search strategy was carried out by means of an exhaustive search in the following 10 databases, including Medline via

Pubmed, Embase, and Gray Literature, by relevant publications made from 1970 to 2018.

The databases were searched using the descriptors presented in the Medical Subject Headings (MeSH), in the Emtree dictionary provided by Embase, and synonyms, and included the following terms, "Diabetic Retinopathy", and synonyms, which were associated with evaluation technology "Clinical decision support system" (and synonyms).

The "*" symbol was also used, and its syntax presents small differences in the database used to retrieve all variations with suffixes of the source words. The terms above were combined using the Boolean operators "AND," "OR" and "NOT."

The search was limited to studies in humans, but there was no language restriction. Reference lists of all primary studies retrieved were verified. Besides, references cited in relevant studies were also verified. The authors of articles that published studies with incomplete information were contacted.

Four researchers carried out the initial search of the abstracts and titles identified from the research strategy in the databases mentioned above independently. Disagreements on the inclusion or exclusion of each study were resolved by consensus.

Diagnostic accuracy studies were used to support the diagnosis of Diabetic Retinopathy (target condition) through the Decision Support Systems (test under evaluation). Thus, the diagnostic test under evaluation consisted of the result provided by the Decision Support Systems (positive or negative).

We analyzed studies that considered patients with Diabetic Retinopathy and evaluated by the gold standard (fundus examination or funduscopy). And we included patients with type 1 diabetes mellitus or type 2 diabetes mellitus, and were excluded patients with any other specific types of diabetes

Data were extracted from the included studies and the year of publication, country and continent (from the research), type of Diabetes Mellitus, study design, results of the system (and respective diagnostic categories), and demographic data such as age and sex, as well as the number of individuals with Diabetes, number of individuals with Diabetic Retinopathy and number of DSS correct answers.

In addition to the previously mentioned information, it was verified whether the gold standard (fundoscopy) was used as a confirmatory examination, besides the description of the techniques associated to DSS, techniques used in the processing of the images and resolution of the images used. Four reviewers independently abstracted the data cited above. Disagreements were resolved by consensus.

Each reviewer calculated the pre-test probability of Diabetic Retinopathy among subjects with diabetes, sensitivity, and specificity, and Odds Ratio Diagnostic (DOR) [18] of the primary studies. Studies that did not present the necessary data for the construction of the 2 x 2 contingency table were excluded. The 2x2 contingency table was constructed in each selected study, for which all gold standard and DSS results were classified as either positive or negative. Sensitivity, specificity and DOR were calculated.

In the studies in which only one of the cells in the 2 x 2 contingency table presented the value 0 (zero), the value 0.5 was added in all cells in order to do the calculations feasible, however, in those cases where the value 0 (zero) occurred in more than two cells, exclusion of the study was performed in the analysis [19].

The evaluation of the methodological quality was performed according to the Quality Assessment of Diagnostic Accuracy 2 (QUADAS-2) tool [20, 21].

In order to produce a combined pooled estimate of the of the studies, a meta-analysis was developed in R software version 3.6.0 (Comprehensive R Archive Network, http://cran.r-project.org/) [22], and Review Manager (RevMan) version 5.3.5 (developed at The Nordic Cochrane Center, Copenhagen, Denmark) [23].

Bivariate analysis was used to calculate combined estimates of sensitivity, and specificity, with 95% confidence intervals (CI) to estimate the summary values presented in the meta-analysis [24, 25, 26]. The measures were summarized by the DerSimonian and Laird random effect model (which takes into account the presence of heterogeneity in the studies) and we calculated all global means weighted with 95% CI) [27]. The graphical representation of the results was made using forest plots [26].

The heterogeneity in the different studies was analyzed by Cochran Q test, I^2 inconsistency test, and the variation between the studies by τ^2 [28].

Due to the identified heterogeneity, sensitivity analyses were performed to identify the associated co-factors responsible for heterogeneity. Potential cofactors associated with heterogeneity were also analyzed using meta-regression. Thus, as the results showed heterogeneity, it was decided not to elaborate the Summary Receiver Operating Characteristic (SROC) [29]. The Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines were followed in reporting (http://www.prisma-statement.org/). The pilot study with initial results was reported in Medical Informatics Europe Conference (MIE) [30].

Results

Primarily, 581 citations were identified in the databases searched as potentially employed the DSS in diagnosing Diabetic Retinopathy. Based on the titles and abstracts we retrieved 85 complete articles from the initial 581 citations. In the end, 18 primary studies met the meta-analysis inclusion criteria [8, 31-47] and were considered eligible for our Meta-analysis.

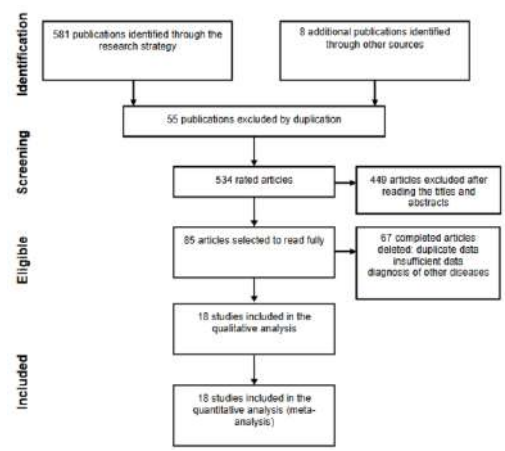


Figure 1– Flowchart of paper selection

Thirteen studies described the image quality used as input to the DSS. We found 2 studies [35, 44] labeled as poor resolution inputs and 11 considered with high resolution inputs [8, 31, 32, 33, 34, 37, 38, 39, 40, 41, 42].

Diabetic Retinopathy was found in 11446 cases (pre-test probability of 24.33%), while 35,608 (75.67%) presented normal images. Table 1 presents the pooled information of all studies

included in the meta-analysis, including True Positive (TP), False Positive (FP), False Negative (FN), True Negative (TN), Sensitivity, Specificity and Diagnostic Odds Ratio (DOR).

Table 1 – Pooled contingency table

Study - Year	DSS				Sensitivity (CI 95%)	Specificity (CI 95%)	DOR (CI 95%)
	Gold Standard Positive		Gold Standard Negative				
	TP	FN	FP	TN			
Sinthanayothin 2003	227	56	143	345	0.802 (0.751-0.847)	0.707 (0.664-0.747)	9.678 (6.81-13.74)
Usher 2004	253	14	206	230	0.790 (0.736-0.838)	0.695 (0.649-0.738)	8.510 (5.95-12.16)
Kahai 2006	71	0	24	51	1.000 (0.949-1.000)	0.680 (0.562-0.783)	300.6 (17.87-505.60)
Abramoff 2008	318	60	2632	4679	0.844 (0.803-0.879)	0.636 (0.625-0.647)	9.372 (7.07-12.24)
Acharya 2009	53	11	4	23	0.865 (0.765-0.933)	0.824 (0.566-0.962)	25.44 (6.68-96.91)
Skevofilakas 2010	17	0	1	37	1.000 (0.805-1.000)	0.974 (0.862-0.999)	875.0 (33.91-2.2E4)
Jaafar 2011	103	7	0	9	0.982 (0.903-1.000)	0.906 (0.807-0.965)	327.0 (53.33-2005.0)
Noronha 2012	120	0	3	117	1.000 (0.970-1.000)	0.975 (0.929-0.995)	809.1 (413.4-1.8E5)
Saleh 2012	68	8	1	21	0.895 (0.803-0.953)	0.955 (0.772-0.999)	115.5 (19.05-700.3)
Jaya 2015	177	11	1	11	0.944 (0.888-0.977)	0.907 (0.817-0.962)	144.3 (50.25-414.4)
Mukti 2015	107	1	19	3	0.991 (0.949-1.000)	0.136 (0.029-0.349)	12.86 (1.78-92.48)
Vijila 2015	143	7	12	138	0.953 (0.906-0.981)	0.920 (0.864-0.958)	211.9 (83.24-539.9)
Vimala 2017	31	3	2	24	0.912 (0.763-0.981)	0.923 (0.749-0.991)	88.20 (16.00-486.1)
Mansour 2018	9316	0	258	25552	1.000 (1.000-1.000)	0.990 (0.989-0.991)	1.8E6 (1.1E5-2.9E7)
Powal 2018	13	4	6	19	0.765 (0.501-0.932)	0.760 (0.549-0.906)	9.00 (2.24-36.06)
Romero-Aroca 2018	192	46	173	700	0.807 (0.751-0.855)	0.802 (0.774-0.828)	16.71 (11.65-23.98)
Sandhu 2018a	72	5	2	18	0.935 (0.855-0.979)	0.900 (0.683-0.988)	97.55 (20.10-473.5)
Sandhu 2018b	38	2	4	36	0.950 (0.831-0.994)	0.900 (0.763-0.972)	124.9 (24.94-625.5)
Total	**11181**	**265**	**3459**	**32149**	**0.977 (0.975-0.979)**	**0.903 (0.900-0.906)**	**73.15 (37.54-142.5)**
					97.7% (97.5%-97.9%)	**90.3% (90.0%-90.6%)**	

The overall sensitivity was 97.7% (95% CI: 97.5%-97.9%) and the specificity was 90.3% (95% CI: 90.0%-90.6%), as shown in the Figure 2 and Figure 3.

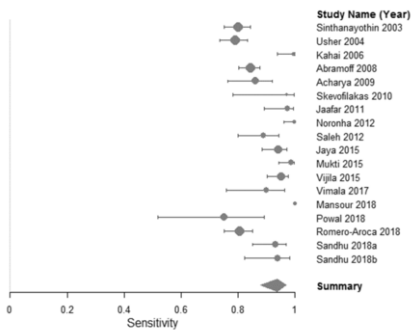


Figure 2 – Forest plot of the sensitivity

The overall DOR showed in Table 1 and Figure 4 (ln(DOR)) was 73.15 (95% CI: 37.54-142.50) which means that considering the use of DSS the probability of a positive result among subjects with Diabetic Retinopathy is 73.15 times greater than the probability of a positive result in individuals with normal retinas images.

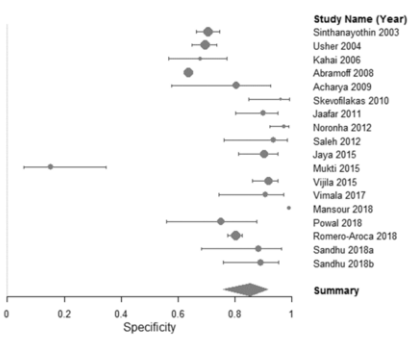


Figure 3 – Forest plot of the specificity

The heterogeneity was medium (I^2=68.53%; P<0.001; τ^2=1.456) [28].

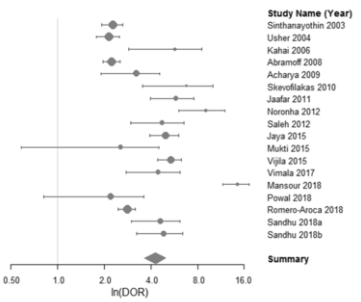


Figure 4 – Forest plot of the DOR

We performed several subgroups and sensitivity analyzes due to the heterogeneity presented. Though, the heterogeneity remained high. Based on our meta-regression analysis including clinical and technological co-factors, we did not observe evidence that any cofactors represented the cause of the observed heterogeneity.

Discussion

The American Diabetes Association along with the American Academy of Ophthalmology formulated Guidelines for Diabetic Retinopathy screening with annual Eye Fund examination for qualify patients [48, 49, 50, 51]. The early detection has the great potential to prevent visual loss since the retinal disease has a pre-symptomatic state [48, 49, 50, 51].

The agreement between ophthalmoscopy and photographs can vary from 61.9% to 86.3%; showing that digital images have the sensitivity of 78% and specificity 86% [52]. When comparing images through the stereoscopic, the ophthalmoscopy presents a sensitivity of 34% and specificity of 100% when compared with the reference standard – stereoscopic [52].

The DSSs based on images may be superior to fundoscopy in the classification of lower levels of Diabetic Retinopathy. It may present as an adjuvant method in diagnosis such as technologies have not been as effective in visualizing other pathologies related to Diabetic Retinopathy, such as cataracts and glaucoma [8].

Despite the limitation of medium heterogeneity, found among the studies included in our meta-analysis, the accuracy of decision support systems was high in support of the diagnosis of diabetic retinopathy. Although, the subgroup analyses and sensitivity revealed that the cofactors presented in the studies could not explain the heterogeneity origin.

Conclusion

The use of Decision Support Systems in diagnosing Diabetic Retinopathy has the potential to noninvasively improve the quality of the management of diabetes patients and contribute to the reduction of cases of visual impairment and blindness. The analysis by QUADAS-2 revealed the high methodological quality of included studies. The high sensitivity and medium specificity points towards to the interpretation of an improvement in the accuracy promote by the DSSs when compared with the high rates of false positives found. The medium heterogeneity observed in the study could not be explained by the cofactors presented. Finally, our results corroborate with the concept of usefulness of DSSs in early diagnosis, screening and preliminary evaluation of suspicious images of Diabetic Retinopathy.

Through the above, we suggest that our findings be interpreted with caution and that future researchs investigate the causes of the presented heterogeneity.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001, and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Brazil.

References

[1] Amos AF, McCarty DJ, Zimmet P. The rising global burden of diabetes and its complications: estimates and projections to the year 2010. *Diabetic Medicine*. 1997;14(S5).

[2] Panagiotoglou T, Tsilimbaris M, Ginis H, Karyotakis N, Georgiou V, Koutentakis P, Pallikaris I. Ocular rigidity and outflow facility in nonproliferative diabetic retinopathy. *J Diabetes Res.* 2015;2015:141598.

[3] King H, Aubert RE, Herman WH. Global burden of diabetes, 1995–2025: prevalence, numerical estimates, and projections. *Diabetes Care*. 1998;21(9):1414-31.

[4] Yoon K-H, Lee J-H, Kim J-W, Cho JH, Choi Y-H, Ko S-H, et al. Epidemic obesity and type 2 diabetes in Asia. *The Lancet.* 2006;368(9548):1681-8.

[5] Ip MS, Domalpally A, Sun JK, Ehrlich JS. Long-term effects of therapy with ranibizumab on diabetic retinopathy severity and baseline risk factors for worsening retinopathy. *Ophthalmology*. 2015;122(2):367-74.

[6] Gao X, Gauderman WJ, Marjoram P, Torres M, Chen YD, Taylor KD, Rotter JI, Varma R. Native American ancestry is associated with severe diabetic retinopathy in Latinos. *Invest Ophthalmol Vis Sci.* 2014 Sep 24;55(9):6041-5.

[7] Hex N, Bartlett C, Wright D, Taylor M, Varley D. Estimating the current and future costs of Type 1 and Type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. *Diabet Med*. 2012 Jul;29(7):855-62.

[8] Saleh MD, Eswaran C. An automated decision-support system for non-proliferative diabetic retinopathy disease based on MAs and HAs detection. *Comput Methods Programs Biomed*. 2012 Oct;108(1):186-96.

[9] Wu J, Geng J, Liu L, Teng W, Liu L, Chen L. The relationship between estimated glomerular filtration rate and diabetic retinopathy. *J Ophthalmol*. 2015;2015.

[10]Berner ES. *Clinical Decision Support Systems*.New York. Springer. 2007.

[11]Mookiah MR, Acharya UR, Chua CK, Lim CM, Ng EY, Laude A. Computer-aided diagnosis of diabetic retinopathy: a re-view. *Comput Biol Med*. 2013 Dec;43(12):2136-55.

[12]Piyasena MMPN, Murthy GVS, Yip JLY, Gil-bert C, Peto T, Gordon I, Hewage S, Kamalakannan S. Systematic review and me-ta-analysis of diagnostic accuracy of detection of any level of diabetic reti-nopathy using digital retinal imaging. *Syst Rev*. 2018 Nov 7;7(1):182.

[13]Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *Bmj*. 2005;330(7494):765.

[14]Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux P, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *Jama*. 2005;293(10):1223-38.

[15]Hunt J, Siemienczuk J, Gillanders W, LeBlanc B, Rozenfeld Y, Bonin K, et al. The impact of a physician-directed health information technology system on diabetes outcomes in primary care: a pre-and post-implementation study. *J Innov Health Inform*. 2009;17(3):165-74.

[16]Sintchenko V, Coiera E, Iredell JR, Gilbert GL. Comparative impact of guidelines, clinical data, and decision support on prescribing decisions: an interactive web experiment with simulated cases. *JAMA.* 2004;11(1):71-7.

[17]van Wyk J, van Wijk M, Moorman P, Mosseveld M, van der Lei J, editors. Cholgate-a randomized controlled trial comparing the effect of automated and on-demand decision support on the management of cardiovascular disease factors in primary care. *Proceedings: a conference of the American Medical Informatics Association/ AMIA Annual Fall Symposium AMIA Fall Symposium*; 2003.

[18]Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol*. 2003;56(11):1129-35.

[19]Deeks JJ, Bossuyt PM, Gatsonis C (ed). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0*. The Cochrane Collaboration, 2013. [Available from: http://srdta.cochrane.org/]

[20]Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-36.

[21]Schueler S, Schuetz GM, Dewey M. The revised QUADAS-2 tool. *Ann Intern Med*. 2012;156(4):323.

[22]Doebler P. M*eta-Analysis of Diagnostic Accuracy R Package*; 2017 [Available from: https://cran.r-project.org/web/packages/mada/mada.pdf].

[23]Collaboration TC. The Cochrane Collaboration; 2019 [Available from: http://www.cochrane.org/].

[24]Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, et al. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med*. 1994;120(8):667-76.

[25]Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-90.

[26]Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM. Systematic Reviews of Diagnostic Test Accuracy. *Ann Intern Med*. 2008;149(12):889-97.

[27]DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-88.

[28]Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. *Br Med J*. 2001;323(7305):157.

[29]Deeks JJ, Morris JM. Evaluating diagnostic tests. *Baillieres Clin Obstet Gynaecol*. 1996;10(4):613-30.

[30]Simões PW, Passos MG, Abreu LLT, Garcia D, Barra CMAM, Garcia A, Pires MMS, Martins PJ, Ceretta LB. Accuracy of decision support systems for diagnostic diabetic retinopathy. *Eur J Epidemiol*. 2016; 31: 177-178.

[31]Sinthanayothin C, Kongbunkiat V, Phoojaruenchanachai S, Singalavanija A, editors. Automated screening system for diabetic retinopathy. Image and Signal Processing and Analysis, 2003, *ISPA 2003 Proceedings of the 3rd International Symposium on;* 2003: IEEE.

[32]Usher D, Dumskyj M, Himaga M, Williamson T, Nussey S, Boyce J. Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening. *Diabet Med*. 2004;21(1):84-90.

[33]Kahai P, Namuduri KR, Thompson H. A decision support framework for automated screening of diabetic retinopathy. *Int J Biomed Imaging*. 2006;2006:45806.

[34]Abràmoff MD, Niemeijer M, Suttorp-Schulten MS, Viergever MA, Russell SR, Van Ginneken B. Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes. *Diabetes Care*. 2008;31(2):193-8.

[35]Acharya U, Lim C, Ng E, Chee C, Tamura T. Computer-based detection of diabetes retinopathy stages using digital fundus images. Proceedings of the Institution of Mechanical Engineers, Part H: Eng Med. 2009;223(5):545-53.

[36]Skevofilakas M, Zarkogianni K, Karamanos BG, Nikita KS, editors. A hybrid Decision Support System for the risk assessment of retinopathy development as a long term complication of Type 1 Diabetes Mellitus. Engineering in Medicine and Biology Society (EMBC), 2010 *Annual International Conference of the IEEE*; 2010: IEEE.

[37]Jaafar HF, Nandi AK, Al-Nuaimy W. Decision support system for the detection and grading of hard exudates from color fundus photographs. *J Biomed Opt*. 2011Nov;16(11):116001.

[38]Noronha K AU, Nayak KP, Kamath S, Bhandary SV. Decision support system for diabetic retinopathy using discrete wavelet transform. Proceedings of the Institution of Mechanical Engineers, Part H: *Eng Med*. 2013;227(3):251-61.

[39]Jaya T, Dheeba J, Singh NA. Detection of hard exudates in colour fundus images using fuzzy support vector machine-based expert system. *J Digit Imaging*. 2015;28(6):761-8.

[40]Vijila SA, & Rajesh RS Classification of Hard Exudates in Fundus Images using Support Vector Machine Classifier. *Res J Pharm Biol Chem Sci*. 2015;6(3):440-445.

[41]Mukti FA, Eswaran C, & Hashim N. Detection and classification of diabetic retinopathy anomalies using bag-of-words model. *J Med Imaging Health Inform*. 2015;5(5):1009-1019.

[42]Vimala GS A G, & Mohideen K. Diabetic retinopathy screening system: A validation analysis with multiple fundus image databases. *Biomed Res*. 2017;28(4): 1598-1603.

[43]Mansour RF. Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy. *Biomed Eng Lett*. 2018;8(1):41-57.

[44]Porwal P, Pachade S, Kokare M, Giancardo L, & Mériaudeau F. Retinal image analysis for disease screening through local tetra patterns. *Comput Biol Med*. 2018;102:200-210.

[45]omero-Aroca P, Valls A, Moreno A, Sagarra-Alamo R, Basora-Gallisa J, Saleh E, Baget-Bernaldiz M, Puig D. A Clinical Decision Support System for Diabetic Retinopathy Screening: Creating a Clinical Support Application. *Telemed J E Health*. 2019 Jan;25(1):31-40.

[46]Sandhu HS, Eladawi N, Elmogy M, Keynton R, Helmy O, Schaal S, El-Baz A. Automated diabetic retinopathy detection using optical coherence tomography angiography: a pilot study. *Br J Ophthalmol*. 2018 Nov;102(11):1564-1569.

[47]Sandhu HS, Eltanboly A, Shalaby A, Keynton RS, Schaal S, & El-Baz A. Automated Diagnosis and Grading of Diabetic Retinopathy Using Optical Coherence Tomography. *Invest Ophthalmol Vis Sci*. 2018;59(7):3155-3160.

[48]Singer DE, Nathan DM, Fogel HA, Schachat AP. Screening for diabetic retinopathy. *Ann Intern Med*. 1992;116(8):660-71.

[49]Fong DS, Aiello L, Gardner TW, King GL, Blankenship G, Cavallerano JD, et al. Retinopathy in Diabetes. *Diabetes Care*. 2004;27(suppl 1):s84-s7.

[50]Group ETDRS. Early Treatment Diabetic Retinopathy Study Group: photocoagulation for diabetic macula edema. *Arch Ophthalmol*. 1985;106:1796-806.

[51]Vujosevic S, Benetti E, Massignan F, Pilotto E, Varano M, Cavarzeran F, et al. Screening for diabetic retinopathy: 1 and 3 nonmydriatic 45-degree digital fundus photographs vs 7 standard early treatment diabetic retinopathy study fields. *Am J Ophthalmol*. 2009;148(1):111-8.

[52]Lin DY, Blumenkranz MS, Brothers RJ, Grosvenor DM, Group TDDS. The sensitivity and specificity of single-field nonmydriatic monochromatic digital fundus photography with remote image interpretation for diabetic retinopathy screening: a comparison with ophthalmoscopy and standardized mydriatic color photography. *Am J Ophthalmol*. 2002;134(2):204-13.

Address for correspondence

Prof PhD Priscyla Waleska Simões – pritsimoes@gmail.com

Biomedical Informatics

Engineering, Modeling and Applied Social Sciences Center (CECS)

Universidade Federal do ABC (UFABC)

Alameda da Universidade s/n, 09606-045

São Bernardo do Campo, SP, Brazil

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190350

Generating a Health Information Technology Event Database from FDA MAUDE Reports

Ethan Wang[a], Hong Kang[b], Yang Gong[b]

[a] College of Natural Sciences, The University of Texas at Austin, Austin, TX, USA
[b] School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA

Abstract

Patient safety events (PSEs), or medical errors, are major impediments to healthcare system safety. Health information technology (HIT) is expected to promote quality of care. Nonetheless, HIT also creates unintended consequences that concern patient safety consolidating a high-quality database of HIT events is essential to understanding their nature. Previous studies demonstrated the potential to use FDA Manufacturer and User Facility Device Experience (MAUDE) database to extract HIT events. In this study, we utilized classic and CNN models to extract HIT events from MAUDE. Both individual and combined models were evaluated on the test set, where the best model identified HIT events with ~90% accuracy and achieved a ~.87 f1 score. This model was capable of identifying HIT events in an HIT-exclusive database and serving as a quality and error check tool during event reporting. Moreover, the strategy of HIT event identification may scale in developing other PSE subtype-specific databases.

Keywords:

Patient Safety; Medical Errors; Information Storage and Retrieval

Introduction

Patient safety events (PSEs) are defined as "any event or action that leads to or has the potential to lead to a worsened patient outcome related to the event or action" [1]. Many PSEs can be attributed to medical errors - unintentional and preventable adverse effects due to poor care. Medical errors are the third leading cause of death in the United States, responsible for approximately 251,000 deaths annually [2]. In response, the Institute of Medicine published "*To Err is Human*," a paradigm-shifting report calling on states to provide public, standardized, and mandatory reporting systems [3]. Nonetheless, current practices of incident reporting are imperfect. While report collection has been heavily emphasized, little effort has been directed to analyzing reported events [4]. This is especially troubling because identification of common themes is crucial in identifying relevant issues to improve patient safety.

The adoption of health information technology (HIT) is critical in reducing medical errors [5]. The U.S. Department of Health and Human Services defines HIT as "the electronic systems health care professionals – and increasingly, patients – use to store, share, and analyze health information." HIT includes, but is not limited to, electronic health records (EHRs), personal health records (PHRs), and electronic prescribing systems [5]. HIT-enabled care yields improved data collection, data availability, cost efficiency, coordination of care, and risk analysis, thus reducing extraneous care and medical errors [6]. Hugely beneficial to healthcare, reliance on HIT is not without its faults. HIT unavailability, malfunction, and improper use may increase the likelihood of adverse events. Listed in the top 10 technology-hazards for healthcare by the Emergency Care Research Institute and comprising one sixth of all PSEs, reducing HIT errors will maximize patient safety [7, 8].

Collecting HIT events is the first step in identifying causes and preventing recurrence. The Agency of Healthcare Research and Quality (AHRQ) has spearheaded this task and has standardized PSE reporting, including HIT event reporting, in the Common Formats (CF). However, because the CF includes HIT as a set of contributing factors rather than as an event category, very few HIT event reports were archived [9]. As a result, most HIT events remain uncategorized or insufficiently categorized under only broader categories such as device or medical/surgical supply. In addition, many reporters leave fields blank due to the extensiveness and challenges in knowledge of patient safety reporting. These limitations make data collection on HIT events difficult. Therefore, an HIT-exclusive event database is in an urgent need.

The FDA MAUDE database is an extensive, online, and public resource with great potential for extracting HIT events. Updated once a month, the MAUDE searchable database contains a wealth of information, including manufacturer information, adverse event information, device information, and patient outcomes [10]. Although only an estimated 0.46~0.69% [11] of MAUDE reports are HIT-related, MAUDE contains over 6 million total reports from mandatory reporters (i.e. importers, manufacturers, and device user facilities) and voluntary reporters (i.e. patients and healthcare professionals) as of Aug 2018. This proportion produces challenges in the identification of HIT events [12]. The key problem is that standard classification algorithms are less effective on imbalanced datasets and that data pre-classification is necessary for suitable training [13]. This study expands on our previous study that generated a database containing 97% HIT events by using term frequency-inverse document frequency (TF-IDF), and biterm topic modeling (BTM) [14].

In this study, to improve on the low F1 score of our prior study, we upgraded the identification method by applying four classic classification algorithms (i.e., Logistic Regression, SVM, Bernoulli Naïve Bayes, and Random Forest), and a convolutional neural network (CNN). Both combined models and individual models were assessed to identify HIT events from MAUDE reports for database generation. An HIT specific database, the product of our model, will be helpful in identifying, preventing, and learning from HIT events. The identification strategy may also be scalable to other PSE subtypes.

Methods

Data pre-processing

A pre-filtered dataset of MAUDE reports was expert-reviewed to determine if the reports were HIT-related [11]. We then randomly categorized the reports to achieve a standard 70/10/20 training/validation/test set split.

Tokenization and word embedding

For the CNN model, Keras API with tensorflow backend was used to tokenize the unstructured text fields of each MAUDE report, and tokenized sequences were then padded to a maximum sequence length of 1,000 words [15]. For most deep learning text-classification models, word representation through feature vector classification is essential in maximizing accuracy [16]. Due to the absence of a large unsupervised training set to initialize word vectors, we used publicly available pre-trained word vectors provided by Global Vectors for Word Representation (GloVe). Based on word co-occurrence, GloVe word vectors are highly effective in characterizing word similarity. Our model used trainable GloVe vectors with a dimensionality of 50 [17].

For the classic models, the unstructured text fields of each MAUDE report were interpreted into a TF-IDF matrix using the scikit-learn python package [18]. TF-IDF weights word importance by calculating an inverse proportion of word frequency in a particular document and the percentage of all documents in which the word appears. TF-IDF is calculated through the following formula where $w_{i,j}$ is the TF-IDF weight, $tf_{i,j}$ is the number of occurrence of term i in document j, and N is the total number of documents.

$$w_{i,j} = tf_{i,j} * log(\frac{N}{df_i})$$

The words with higher TF-IDF weights are rarer, signifying greater importance in the document [19].

Classic model architecture

Following TF-IDF word embedding, four popular machine learning algorithms –SVM, Bernoulli naïve Bayes (BNB), random forest (RF), and logistic regression (LR) – were constructed using the sci-kit learn python package to classify the weighted text [18].

CNN model architecture

The CNN architecture was inspired by Kim's work [20]. We used Keras API with tensorflow backend to build our model [15]. After GloVe word embedding, we applied three 1D convolutional layers with kernel size 2, 3, and 4 respectively. Each 1D convolutional layer has 256 filters and was followed by a max-pooling layer to produce 256 features. To minimize overfitting, we applied 30% dropout to the max-pooled layer. The three outputs were then concatenated and flattened into a single layer that was fed to a dense layer of 1024 units with rectified linear units (relu) activation followed by 50% dropout. The output layer had one unit and was activated by a sigmoid function. Hyperparameters were tuned according to the model performance on the validation set.

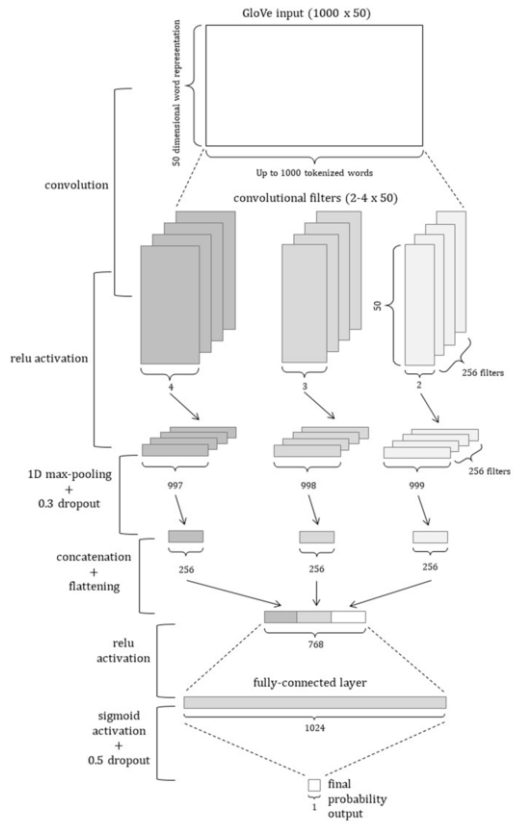


Figure 1: Illustration of CNN architecture for text classification. The figure depicts the three convolutional filter sizes (2, 3, and 4), max-pooling, concatenation, and final output layers used to classify a single document.

Combined model architecture

Among the five individual models, three with the best performance were applied to produce combined models. In each combined model, the probabilities of individual model were averaged to compute the probabilities of the combined model. If the averaged probability that the event is an HIT event equals or exceeds 50%, the combined model guesses 1 (HIT event), and if not, the model guesses 0 (non-HIT event).

Results

Dataset statistics

The dataset, containing 5,588 MAUDE reports, was generated by applying a keyword filter to recent nine-year MAUDE data (2008-2016) [11]. The dataset was relatively balanced between HIT - 2,192 (39.2%) and non-HIT – 3,396 (60.8%) PSEs. For model training, this balanced proportion is preferable to the 0.46~0.69% HIT event proportion in the MAUDE database as a whole [11]. In addition, we extracted 18,132 individual tokens from the 5,588 total reports.

CNN hyperparameter tuning

We tuned numerous CNN hyperparameters including batch size, learning rate, dropout, kernel size, max-pooling pool size, embedding dimensions, maximum token number, and

maximum sequence length. Increasing maximum token number and sequence length improved model accuracy. Thus, a maximum sequence length of 1,000 and maximum token number of 20,000 were used to incorporate as much training data as possible. Multiple dropout layers of 30-50% performed optimally, and max-pooling the product of each convolutional filter to an output of length 1 outperformed smaller max-pooling pool sizes. We found that batch size and learning rate had a negligible effect on model performance. Increasing embedding dimensions beyond 50 failed to improve model performance while increasing training time as well.

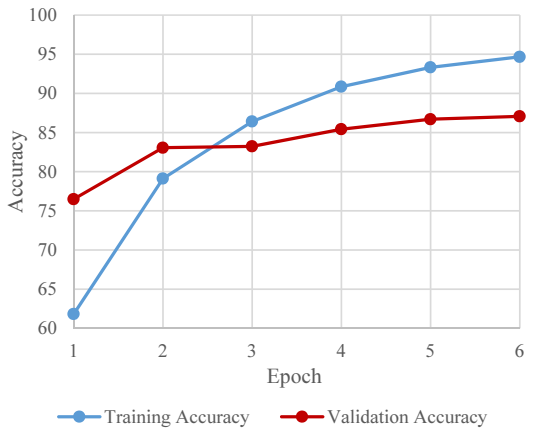


Figure 2: Training and validation accuracy of the CNN model over 6 epochs

While training the CNN model, we found that the rate at which validation accuracy improves decreases significantly around the 2th epoch, and fails to improve consistently past the 6th epoch (Figure 2). Moreover, training accuracy continues to improve past the 6th epoch, signifying overfitting of the training data. We found that the CNN model performed best on the test set when trained for 6 epochs.

Individual and combined model performance

Table 1: Accuracy, Precision, Recall, and F1-score of each individual and combined model on test set

Methods	Accuracy	Precision	Recall	F1
SVM+LR+CNN	.9012	.8796	.8606	.8700
SVM+LR	.8965	.8838	.8413	.8621
SVM+CNN	.8901	.8595	.8534	.8564
LR	.8895	.8828	.8366	.8591
LR+CNN	.8864	.8582	.8438	.8509
SVM	.8860	.8584	.8584	.8584
CNN	.8836	.8463	.8522	.8493
BNB	.8529	.8033	.8440	.8232
RF	.8467	.8690	.7186	.7885

*SVM = support vector machine, LR = logistic regression, CNN = convolutional neural network, BNB = Bernoulli naïve Bayes, RF = random forest

As shown in Table 1, combined models generally outperformed individual models by all metrics. LR+CNN was an exception to this trend. LR, the best performing individual model, outperformed LR+CNN. Among the other three combined models, SVM+LR+CNN outperformed SVM+CNN by all metrics and outperformed SVM+LR on all metrics except precision. Differences in the classification algorithms behind individual models may explain why the combined models performed better. One model may misclassify reports accurately classified by another model. Thus, averaging probability predictions can help mitigate the erroneous predictions of individual models by factoring in other models as well.

Effect of data size on model performance

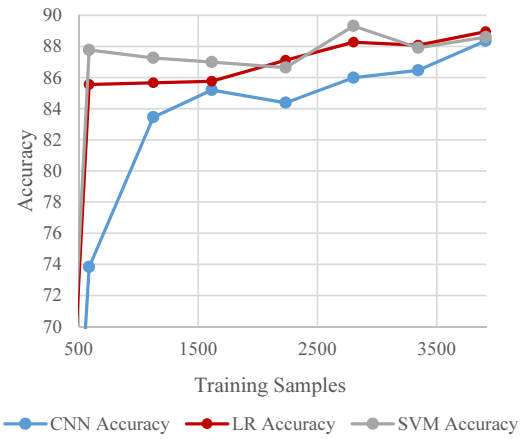


Figure 3: Improvement in test accuracy as number of training samples increases

Given a larger dataset, we believe that the CNN model and combined models involving the CNN will improve. Shown in Figure 3, CNN test accuracy improves steadily as the number of training samples increases. In contrast, classic machine learning methods such as logistic regression and SVM are less sensitive to training sample count

Discussion

Related work

Beyond the FDA MAUDE database, analysis of 48 PSE databases revealed significant technological limitations (i.e. insufficient interoperability, lack of report quality validation) that limit data entry quality. Providing embedded validators, enabling feedback review, and improving interoperability are areas of potential database improvement [21].

Substantial precedent exists in the using machine learning models to extract PSE subcategories. A hybrid CNN and bidirectional long short-term memory (LSTM) autoencoder model has been used in the detection of bleeding events from electronic health record notes with an overall F-score of 0.938 [22], and an artificial neural network (ANN) was used to identify variables which were analyzed via logistic regression to classify fall events from a nursing PSE database from a Taiwan medical center [23].

Growing and using an HIT event database

In our previous study, the models were trained on a small dataset with 490 MAUDE reports (312 HIT, 178 Non-HIT), and reached up to 88.6% accuracy, performing similarly to the classic models in this study [14]. However, we did not have an individual test set in the previous study due to the limited data size, which made the model suffer from severe overfitting. In addition, to obtain high precision, we traded off recall in the previous models. In this study, we used an upgraded dataset

from 9 years (2008-2016) of keyword-filtered MAUDE reports while the previous study only used reports from 2015 [11, 14]. We also upgraded the classifiers and proposed the SVM+LR model for HIT event identification. With an accuracy of ~90% and f1-score of ~.87, the SVM+LR+CNN model outperforms previous models and holds promise in growing an HIT event database.

To extract contextually relevant themes and topics regarding HIT events, we can use the database to process a high quantity of HIT reports at a time. Natural language processing (NLP) techniques may be the best method for identifying connections between HIT events. Such connections will provide experts with the information necessary to address the root causes of HIT events. Developing effective solutions to prevent future HIT events is the final step of this process. As HIT becomes increasingly influential in healthcare, utilization of the constantly updated HIT database will help ensure that HIT functions accordingly to ensure patient safety.

Importance of an HIT event database

The inappropriate use, design, and implementation of HIT are profoundly detrimental to patient safety. However, identifying and fixing relevant HIT issues is challenging without an organized database of HIT events. Many seemingly insignificant events may signify deeper issues, which, without a broader database, are difficult to identify. A database will allow for the detection of broader trends indiscernible from individual documents. Identifying such trends is crucial to fixing HIT issues and preventing future accidents. The fact that the nature of HIT events differs from PSEs as a whole further necessitates the accumulation of an HIT event database. For instance, contributing factors to HIT events such as equipment/device function and data availability do not exist for non-HIT events. The prevalence of certain contributing factors differs between PSE subtypes as well. While poor staff communication is a contributing factor in an estimated 40.0% of HIT events, it only factors into 24.8% of non-HIT events [24]. The uniqueness and impact of HIT mandates the HIT event database construction.

Probability outputs as a reporting aid

Screening the quality of new entries into PSE databases is another means of improving event reporting. Our model can interpret the unstructured text of PSE reports during reporting and continually update the reporter on the probability that the report is HIT or non-HIT. This yields numerous benefits. First, if the reporter writes a report description but has difficulty categorizing the report, our model can help inform their decision on whether or not to check the HIT-related box. Conversely, if the reporter erroneously labels a report, probability predictions may better inform the reporter on what label the report should be. This functionality will help reporters avoid labeling/entry mistakes. Finally, if the reporter is certain that the PSE is clearly HIT or non-HIT but unsure of how to write the report, the model's probability prediction could help guide the reporter into writing narrative text that best coincides with the report category. The overarching benefits of improved report descriptions, fewer mistakes, and more accurate labeling are that HIT-event extraction will be easier in the future, and more high-quality data will be available for analysis.

Tracking HIT changes

In 2015 alone, the U.S. Patent and Trademark Office granted 17,596 patents for new medical devices [25]. HIT is constantly changing, and we must adjust our model to captures such changes. As new PSE and HIT events are added to the MAUDE database, we can update our model. Alternatively, we could completely retrain our model on new data each year to avoid incorporating out of date HIT events. Meticulously choosing a training set that balances the inclusion new HIT events and exclusion of irrelevant, old HIT events is key to constructing a model reflective of the present nature of HIT.

Challenges of identifying HIT events with machine learning

NLP is limited in its capability to account for word context, sentence meaning, and typographical errors [26]. Generalized NLP and GloVe word embedding provide substantial information on individual word frequency and meaning, but cannot discern differences due to word context. Numerous factors such as individual writing style, interactions between different document sections, and accidental errors cannot be accounted for by NLP and machine learning models. Manual review may remain a useful supplement in the classification of reports poorly identified by NLP and machine learning.

Limitations

Missing, duplicated, and non-standardized entries were common in the MAUDE database. Out of 1,103 medical device reports, 64 reports with duplicated information and 6 reports referring to multiple patients were found. The investigation also identified the varying quality and completeness of reports is another issue [27]. Despite its flaws, the MAUDE database is the only publicly accessible database suitable for the extraction of HIT events.

Future work

Optimizing and improving our model

Deducing the quantity and precision of HIT reports desired for database construction is crucial. Because our database must be HIT-exclusive, false positives are more unfavorable than false negatives. Precision takes precedence over recall, and increasing the threshold needed to label an event as HIT beyond 50% without significantly reducing recall would be optimal.

Other deep learning models may perform better than classic models or the CNN. We will apply a recurrent neural network (RNN) for HIT event classification. Bidirection, hierarchical models, and attention may also improve RNN performance [28]. Recurrent CNN (RCNN) also offers potential for HIT event classification [29].

Application to other categories of PSEs

While PSE databases already exist for certain PSE subtypes such as medication error, two additional PSE subtypes that lack a specific database are patient falls and administrative error [30]. Applying a keyword filter to the MAUDE database can help isolate training data for database generation and probability outputs with the proposed SVM+LR+CNN model.

Conclusion

Our strategy to extract HIT events from MAUDE reports for an HIT event database construction bests prior models by all significant performance metrics. Database generation is an essential first step in identifying themes, causes, and solutions to HIT events. Moreover, our model has great potential in improving the quality and accuracy of HIT event reporting.

Acknowledgements

This project is supported by UTHealth Innovation for Cancer Prevention Research Training Program Post-Doctoral Fellowship and Summer Undergraduate Fellowship (Cancer Prevention and Research Institute of Texas grant # RP160015),

and by the Agency for Healthcare Research & Quality (1R01HS022895). The Titan Xp GPU used for this research was donated by the NVIDIA Corporation.

References

[1] Patient Safety Event, in, EMS Voluntary Event Notification Tool (E.V.E.N.T.), 2018.
[2] J.G. Anderson, K. Abrahamson, Your Health Care May Kill You : Medical Errors, *Studies in Health Technology and Informatics* **234** (2017), 13-17.
[3] L.T. Kohn, J.M. Korrigan, M.S. Donaldson, To err is human: building a safer health system, in, U.S. Institute of Medicine, November 1999.
[4] C. Macrae, The problem with incident reporting, *The British Medical Journal Quality Safety* **25** (2016), 71-75.
[5] B. Chaudhry, J. Wang, S. Wu, M. Maglione, W. Mojica, E. Roth, S. C. Morton, P. G. Shekelle, Systematic Review : Impact of Health Information Technology on Quality, Efficiency, and Costs of Medical Care, *Annals of Internal Medicine* **144** (2006), 742-752
[6] P.F. Cipriano, K. Bowles, M. Dailey, P. Dykes, G. Lamb, M. Naylor, The importance of health information technology in care coordination and transitional care, *Nursing Outlook* **61** (2013), 475-489
[7] Top 10 health technology hazards for 2015, in, Emergency Care Research Institute (ECRI), November 2014.
[8] K.C. Cheung, W. van der Veen, M. L. Bouvy, M. Wensing, P.M. van der Bemt, P.A. de Smet, Classification of medication incidents associated with information technology, *Journal of American Medical Informatics* **21** (2014), 63-70.
[9] Common Formats – Scope and Reporting, in, Agency for Healthcare Research and Quality.
[10] Manufacturer and User Facility Device Experience Database (MAUDE), in, U.S. Food & Drug Administration, June 2018.
[11] H. Kang, J. Wang, B. Yao, S. Zhou, Y. Gong, Toward Safer Health Care : A Review Strategy of FDA Medical Device Adverse Event Database to Identify and Categorize Health Information Technology (HIT) Related Events. *JAMIA Open (submitted)* (2018).
[12] F. Magrabi, M.S. Ong, W. Runciman, and E. Coiera, Using FDA reports to inform a classification for health information technology safety problems, *Journal of American Medical Informatics Association* **19** (2012), 45-53.
[13] N. Japkowicz, Learning from Imbalanced Data Sets, in, AAAI Technical Report, 2000, pp. 10-15
[14] H. Kang, Z. Yu, Y. Gong, Initializing and Growing a Database of Health Information Technology (HIT) by using TF-IDF and Biterm Topic Modeling, in : *AMIA Annual Symposium*, San Francisco, CA, USA, 2018, 1024-1033.
[15] F. Chollet, *Deep Learning with Python*, Manning Publications, Shelter Island, NY, November 2017.
[16] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, P. Kuksa, Natural Language Processing (Almost) from Scratch, *Journal of Machine Learning Research* **12** (2011), 2493-2537.
[17] J. Pennington, R. Socher, C.D. Manning, GloVe : Global Vectors for Word Representation, in : *Empirical Methods in Natural Language* Processing, Doha, Qatar, 2014, pp. 1532-1543.[18] F. Pedregosa, G. Varoquax, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research* **12** (2011), 2825-2830.
[19] J. Ramos, Using TF-IDF to Determine Word Relevance in Document Queries, in : *The First Instructional Conference on Machine* Learning, Piscataway, NJ, USA, 2003.
[20] Y. Kim, Convolutional Neural Networks for Sentence Classification, in: *Empirical Methods in Natural Language Processing*, Doha, Qatar, 2014.
[21] Y. Gong, H. Kang, X. Wu, L. Hua, Enhancing Patient Safety Event Reporting, *Applied Clinical Informatics* **8** (2017), 893-909.
[22] R. Li, B. Hu, W. Liu, F. Liu, F. Cunningham, D.D. McManus, H. Yu, Detection of Bleeding Events in Electronic Health Record Notes Using Convolutional Neural Network Models Enhanced with Recurrent Neural Network Autoencoders : Deep Learning Approach, *JMIR Medical Informatics* **7** (2019).
[23] T. Lee, C. Liu, Y. Kuo, M. Mills, J. Fong, C. Hung, Application of data mining to the identification of critical factors of patient falls using a web-based reporting system, *International Journal of Medical Informatics* **80** (2011), 141-150.
[24] R. Mardon, L. Olinger, M. Szekendi, T. Williams, E. Sparnon, K. Zimmer, Health Information Technology Adverse Event Reporting: Analysis of Two Databases in, The Office of the National Coordinator for Health IT, November 25, 2014.
[25] Medical Devices : All Classified Utility Patents (OR/XR), in, U.S. Patent and Trademark Office Patent Technology Monitoring Team Report, December 31, 2015.
[26] I. Zuckerman, D. Litman, Natural Language Processing and User Modeling : Synergies and Limitations, *User Modeling and User-Adapated Interaction* **11** (2001), 129-158.
[27] J.M. Sandberg, I. Gray, A. Pearlman, R.P. Terlecki, An Evaluation of the Manufacturer and User Facility Experience database that inspired the United States Food and Drug Administration's Reclassification of transvaginal mesh, *Investigative and Clinical Urology* **59** (2018), 126-132.
[28] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, E. Hovy, Hierarchical Attention Networks for Document Classification, *Conference of the North American Chapter of the Association of Computational Linguistics : Human Language Technologies*, New Orleans, Louisiana, 2016.
[29] S. Lai, L. Xu, K. Liu, J. Zhao, Recurrent Convolutional Neural Networks for Text Classification, *Twenty-Ninth AAAI Conference on Artificial Intelligence*, Austin, TX, USA, 2267-2273.
[30] Z.R. Wolf, *Patient Safety and Quality : An Evidenced-Based Handbook for Nurses*, Agency for Healthcare Research and Quality, Rockville, MD, March 2008.

Address for correspondence

Yang Gong, MD, PhD

7000 Fannin St Suite 600, Houston, TX 77030, USA

Yang.Gong@uth.tmc.edu

Phone: +1 713 500 3547, Fax: +1 713 500 3907.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190351

Using Electronic Health Records and Machine Learning to Predict Postpartum Depression

Shuojia Wang[a,b], Jyotishman Pathak[a], Yiye Zhang[a]

[a] Weill Cornell Medicine, Cornell University, New York, New York, USA
[b] School of Public Health, School of Medicine, Zhejiang University, Zhejiang, China

Abstract

Postpartum depression (PPD) is one of the most frequent maternal morbidities after delivery with serious implications. Currently, there is a lack of effective screening strategies and high-quality clinical trials. The ability to leverage a large amount of detailed patient data from electronic health records (EHRs) to predict PPD could enable the implementation of effective clinical decision support interventions. To develop a PPD prediction model, using EHRs from Weill Cornell Medicine and NewYork-Presbyterian Hospital between 2015-17, 9,980 episodes of pregnancy were identified. Six machine learning algorithms, including L2-regularized Logistic Regression, Support Vector Machine, Decision Tree, Naïve Bayes, XGBoost, and Random forest were constructed. Our model's best prediction performance achieved an AUC of 0.79. Race, obesity, anxiety, depression, different types of pain, antidepressants, and anti-inflammatory drugs during pregnancy were among the significant predictors. Our results suggest a potential for applying machine learning to EHR data to predict PPD and inform healthcare delivery.

Keywords:

Depression, Postpartum; Electronic Health Records; Machine Learning

Introduction

Postpartum depression (PPD) is a nonpsychotic depressive episode that begins one year within childbirth[7]. The prevalence of PPD is reported to be 13% in high-income countries [21] and 15% in low and middle-income countries[9]. PPD is one of the most frequent and serious maternal morbidities after delivery [26]. It not only interferes with mothers' emotional wellbeing [5], but is also associated with infant morbidity, and children's poorer cognitive and behavioral skills later in life [28].

Despite the serious adverse consequences of PPD, there is a lack of consensus and evidence on PPD screening and treatment from high-quality clinical trials [22]. A number of key predictors have been identified from previous meta-analysis studies, including a history of psychiatric illness, prenatal depression, stressors and illness during pregnancy, poor social support, poor self-esteem, and lower socioeconomic status [1, 23]. Few studies found a significant association between prescription drug use during pregnancy with PPD [20].

Although risk factors were reported from previous studies, effective interventions against them and identification of at-risk women still need further evaluations [22]. A number of screening and preventative measures were proposed previous studies with varying outcomes [22]. For example, prior PPD prediction studies were prospective studies conducted with small sizes [14,26]. Features used in these predictive models often included questionnaires measuring psychological statuses such as demographics, education level, self-esteem, and social support, but not current diagnoses and medications. A commonly used questionnaire for perinatal depression screening is the Edinburgh Postnatal Depression Scale (EPDS) [10,26], although its effectiveness in screening has been questioned in previous studies [27].

The current knowledge gap on PPD contributes to substantial variations across clinical practices in screening and information collection [22]. Addressing these challenges, it has been pointed out by the US Preventive Services Task Force (USPSTF) that electronic health records (EHRs)-based tools may be considered in implementing PPD-related interventions [22]. EHR data are routinely collected and contain a detailed history of health and health services utilization [8]. Moreover, models developed using EHR data can be potentially integrated within the EHR system as clinical decision support (CDS), allowing effective screening for expectant mothers at risk of developing PPD.

In this study, we propose that machine learning algorithms can be applied to EHR data, containing information from the full three trimesters of pregnancy period to delivery, to construct a predictive model for PPD outcome. We performed a pilot study using six machine learning algorithms featuring longitudinal clinical information and patients' socio-demographic characteristics. The overarching goal of this study is to demonstrate that machine learning models can be used to predict PPD, and to carefully evaluate the risk factors identified from EHR data.

Methods

Data

In this study, we used EHRs from Weill Cornell Medicine and NewYork-Presbyterian Hospital from 2015 to 2017 as the data source. The study data are represented using Observational Medical Outcomes Partnership (OMOP) common data model [19] and include patient socio-demographics, timestamped outpatient and inpatient diagnoses, and timestamped medication prescriptions.

Study population

Pregnant women with a fully completed antenatal care procedure at the hospital and with a singleton birth were included in the study. The exclusion criteria were: (1) those with unknown pregnancy length of gestational weeks, (2) those

with missing information from at least one trimester during pregnancy.

Outcomes and Predictors

Clinical assessment of PPD was used as the outcome in this study. The main outcome was defined based on the Statistics Canada [3] and International Classification of Diseases, 10th Revision (ICD-10-CM) codes O99.3 and O99.34 as well as their ICD-9-CM equivalents for a diagnosis of PPD within 12 months after childbirth. We considered patients' birthdate, race, maternal status, average body mass index (BMI), gestational week, and delivery type as time-independent predictors, and medication prescriptions and diagnoses at each clinical visit as the time-dependent predictors.

Age was calculated as baseline age at the first visit of prenatal care. Marital status was extracted from unstructured clinical notes and categorized as single (unmarried, divorced and windowed), married, and unknown. Race groups included White, Asian, American Indian or Alaska Nation, Black or African American, Other combinations not described, and Unknown. We included Native Hawaiian, Other Pacific Islander, and other racial combinations as "Other combinations not described." Gestational weeks were computed using the delivery dates and gestational checkup weeks. The specific trimester of medication prescription and diagnoses were identified by the time interval between each event and delivery. We defined trimester of pregnancy as follows: first trimester (0-12 weeks), second trimester (13-28 weeks), and third trimester (29 weeks- gestation). All diagnoses were represented as Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT) codes. Medication and dosage were standardized by Anatomical Therapeutic Chemical (ATC) Classification System.

In order to perform variable selection for the prediction model, we first selected the variables above the median frequency for all variables. Then, univariate logistic regression (LR) analyses were performed, in which factors with p-values below 0.05 were assigned as potential predictive factors.

Prediction model

In this study, six machine learning models, including Support Vector Machine (SVM), Random Forest (RF), Naïve Bayes, L2-regularized LR, Extreme Gradient Boosting (XGBoost) and Decision Tree were used to build PPD prediction models. We evaluated each model's performance using the area under the receiver-operator curve (AUC) in 10-fold cross validation. All machine learning and statistical analyses were performed with R version 3.4.3 (R Foundation for Statistical Computing, Vienna, Austria).

SVM is a classifier which transforms input data into a multidimensional hyperplane using kernels to discriminate two classes [12]. RF is an ensemble learning method that operates by constructing a multitude of decision trees and outputting the class that is voted by a majority of the trees [4]. Naïve Bayes classifier uses the Bayes Theorem to predict membership probabilities for each class by assigning a class with the highest probability as the most likely class [33]. LR is a regression model with a binary dependent variable [2]. L2-regularized LR tunes and generalizes the model in order to balance the bias-variance trade off [32]. XGBoost is a scalable tree boosting algorithm which trains a sequence of models to minimize errors made by existing models [24]. Lastly, Decision Tree predicts class membership by inferring decision rules from the training data. For all models, we applied an oversampling method to the training data as our outcome was imbalanced (see Table 1). Oversampling is a popular method in dealing with class imbalance problems, which changes the training sets by repeating instances in the minority training set [17].

Feature Importance

In order to provide more interpretability to our model, we examined the association of predictors categorically with PPD. We compared models with different feature compositions (see Table 3). First, we examined the temporality of the features by grouping medication prescribed by trimesters. Then, we examined feature categories by building models with socio-demographic information only, medication information only, diagnostic information only, and medication combined diagnostic information. Lastly, we built a model with only variables selected using the univariate LR. The Pearson correlation of variables was tested to prevent multicollinearity. If the correlation coefficients were greater than 0.8, variables were combined. Only variables whose associations with PPD were statistically significant were selected. Odds ratios with 95% confidence intervals (CI) and p-values are presented in Table 4.

Results

Demographics

Table 1 shows the characteristics of pregnant women with and without PPD. Results are presented as the mean ± standard deviation for continuous variables and N (%) for categorical variables. A p-value less than 0.05 is considered statistically significant in statistical analyses. Among the studied population, 9,980 episodes of pregnancy were identified. There was a significant difference in age between two groups using a student t-test. The mean age was 33.92 (SD 4.51) years old in non-PPD group and was 34.36 (SD 4.61) years old in the PPD group. The pre-pregnancy average BMI in PPD group is higher than that in non-PPD group. There were significant differences in race between the PPD and non-PPD groups using a Fisher exact test. The number of single mothers is higher in the PPD group than non-PPD group (23.15% vs. 15.96%).

Table 1- Baseline Characteristics of Pregnant Women.

Variables	Non-PPD	PPD
N	9211	769
Age, years*	33.92 ± 4.51	34.36 ± 4.61
Pre-pregnancy BMI, kg/m^2*	23.61 ± 4.41	23.93 ± 4.99
Race*		
White	4801(52.12)	478(62.16)
Asian	1455(15.80)	62(8.06)
American Indian or Alaska Nation	30(0.33)	3(0.39)
Black or African American	492(5.34)	45(5.85)
Other combinations not described	1067(11.58)	90(11.70)
Not known	1366(14.83)	93(12.09)
Marital Status*		
Single	1470(15.96)	178(23.15)
Married	4610(50.05)	416(54.10)
Not known	3131(33.99)	175(22.76)
Cesarean section*		
No	8352(90.67)	679(88.29)
Yes	859(9.33)	90(11.70)

**Significant statistical difference found between two groups.*

Prediction Model Performance

Table 2– Prediction Results.

Machine learning technique	AUC	Sensitivity	Specificity
SVM	0.79	0.894	0.580
L2 LR	0.78	0.887	0.594
RF	0.78	0.959	0.391
Naïve Bayes	0.78	0.867	0.616
XGBoost	0.77	0.915	0.527
Decision Tree	0.69	0.986	0.386

In this study, prior to variable selection, 256 variables were extracted including socio-demographic characteristics, disease diagnoses, and medications across 3 trimesters. We then identified 98 potential predictors using univariate LR analyses. Among the selected variables, 71 variables were diagnoses and 22 were medications. Results from the 6 machine learning models using all 98 predictors are shown in Table 2. AUC for different classifier was the highest with SVM (0.79), followed by L2-regularized LR (0.78), RF (0.78), Naïve Bayes (0.78), XGBoost (0.77), and the lowest was 0.69 for the Decision Tree. We further computed the sensitivity of different models. The Decision Tree had the highest sensitivity (98.6%), followed by the RF (95.9%), XGBoost (91.5%), SVM (89.4%), LR (88.7%) and naïve Bayes (86.7%). The specificity was highest for naïve Bayes (61.6%), followed by L2-regularized LR (59.4%), SVM (58.0%), XGBoost (52.7%) and Decision Tree (38.6%).

Feature Importance

Using SVM, the best performing model, we investigated model performance using different feature compositions, presented in Table 3. The AUC for the model using only 1st, 2nd and 3rd trimester information was 0.66, 0.64, and 0.65, respectively. The AUC for the model with variables in both 1st trimester and 2nd trimester, 2nd trimester and 3rd trimester was 0.69 and 0.72, respectively, both of which lower than the complete feature set. In addition, the AUC for the model with only demographic variables was 0.60. The AUCs for the diagnoses model or medication classes model were 0.72 and 0.65, respectively. When we combined medications and diagnoses together, the AUC increased to 0.76.

Table 3– Prediction Results in Different Variable Combinations.

Predictors	AUC	Sensitivity	Specificity
Trimesters			
1st	0.66	0.855	0.428
2nd	0.64	0.831	0.424
3rd	0.65	0.867	0.424
1st +2nd	0.69	0.908	0.307
2nd +3rd	0.72	0.854	0.524
Categories			
Demographic	0.60	0.551	0.609
Diagnose	0.72	0.850	0.560
Medication	0.65	0.882	0.389
Diagnose+ Medication	0.76	0.875	0.577
Logistic-selected	0.76	0.892	0.588

The univariate LR identified 26 predictors out of the 98 predictors whose associations with PPD have significant and meaningful odds ratios for PPD (Table 4). The AUC using 26 important features were lower than using the whole features. None of the reduced models performed as well as the full model (Table 3).

Table 4– Association between Predictors and PPD.

Variables		OR(95%CI)	*P*
Marital status			
Single		REF	
Married		0.82(0.65,1.04)	0.096
Not known		0.58(0.45,0.76)	<0.001
Race			
White		REF	
Asian		0.54(0.39,0.74)	<0.001
American Indian or Alaska Nation		0.54(0.07,2.34)	0.475
Black or African American		0.68(0.43,1.04)	0.084
Other combinations not describe		0.82(0.62,1.07)	0.158
Declined		0.95(0.71,1.25)	0.713
Diagnose			
Anxiety	1st	10.49(6.22,17.75)	<0.001
Depressive disorder	1st	18.58(9.73,35.93)	<0.001
Mental disorder	1st	4.04(1.34,12.02)	0.013
Obesity	1st	1.75(1.03,2.85)	0.031
Threatened miscarriage	1st	1.72(1.15,2.51)	0.007
Abnormal weight gain	2nd	2.84(1.22,5.98)	0.010
Anxiety	2nd	4.08(2.27,7.28)	<0.001
Depressive disorder	2nd	4.35(1.30,15.71)	0.021
Diarrhea	2nd	2.78(1.28,5.54)	0.006
Mental disorder	2nd	6.87(2.36,20.41)	<0.001
Premature labor	2nd	5.20(1.92,12.75)	0.001
Muscle pain	2nd	3.74(1.35,9.05)	0.006
Vomiting of pregnancy	2nd	2.43(1.02,5.47)	0.037
Anxiety	3rd	9.52(5.67,16.00)	<0.001
Abdominal pain	3rd	1.71(1.08,2.62)	0.018
Backache	3rd	3.68(1.67,7.41)	0.001
Hypertensive disorder	3rd	3.24(1.49,6.76)	0.002
Mental disorder	3rd	2.84(1.29,6.16)	0.009
Major depression, single episode	3rd	5.49(2.30,12.78)	<0.001
Palpitations	3rd	2.38(1.010,5.022)	0.033
Medication			
Antidepressants	1st	13.47(7.58,24.13)	<0.001
Antidepressants	2nd	10.84(4.86,25.08)	<0.001
Antidepressants	3rd	24.21(12.39,49.20)	<0.001
Anti-inflammatory agents	2nd	16.64(1.73,158.16)	0.009

**SNOMED code: anxiety (48694002/197480006/21897009/ 247808006/198288003), depressive disorder (35489007/ 94631000119100), mental disorder (74732009/267320004/ 199257008/199261002), obesity (414916001/171000119107/ 415530009/238136002), threatened miscarriage (54048003/ 73790007/75933004), abnormal weight gain (237288003), diarrhea (62315008), premature labor (282020008/6383007/ 49550006), muscle pain (68962001), vomiting of pregnancy (90325002/422400008), backache (161891005), hypertensive disorder (38341003), major depression, single episode*

(70747007/36923009/79298009/15639000/251000119105/43085200 1/76441001), palpitations (80313002).

In Table 4, amongst factors related to diagnoses during pregnancy, obesity, anxiety, depressive disorder, and mental disorder, threatened miscarriage in 1st trimester; abnormal weight gain, anxiety, depressive disorder, diarrhea, mental disorder, premature labor, muscle pain, vomiting in 2nd trimester; and anxiety, abdominal pain, backache, hypertensive disorder, mental disorder, palpitations, major depression, and single episode in 3rd trimester were found to be associated with increased odds of PPD. Among medications, the use of antidepressants during pregnancy, and anti-inflammatory agents in 2nd trimester were associated with increased odds of PPD. Among the predictors, hormone use had no associations with PPD, despite their mention in previous literature [30].

Discussion

In this study, we employed 6 machine learning models to predict PPD using EHR data. Experimental results demonstrated the feasibility of our approach for PPD risk prediction based on information available during prenatal care in an EHR. We found several disease diagnoses and medications during pregnancy that potentially contribute to the prediction of PPD.

The performances of the model using variables in one specific trimester only, both 1st and 2nd trimesters or both in 2nd and 3rd trimesters were not as good as using variables in whole prenatal care. Thus, our findings potentially suggest that screening should consider health and health service utilization throughout the pregnancy period. When we separated the variables into demographic variables, diagnoses variables, and medication variables, disease diagnoses had the best performance in predicting PPD. Although, using the combination of disease diagnoses with medication improved the performance in predicting PPD than using disease diagnoses alone. Again, more comprehensive information provides more improved prediction performance.

Our data set included multiple features in EHR data. Although SVM had the best performance, the difference across the performance of SVM, L2-regularized LR, RF, Naïve Bayes, and XGBoost was minimal, although differences existed with respect to sensitivity and specificity. The AUC of the decision tree model was the lowest compared with the other five models. This may be explained by the tendency of the decision tree to depend on single variables in generating decision rules [13].

Several associations found in our study are consistent with previous studies. These include race [16] as demographics and threatened abortion [6], prenatal mental disorder [15], in particular, diagnoses such as depression disorder, anxiety, and single episode major depression, backache in the 3rd trimester, and muscle pain in 2nd trimester [6]. Pain is often expressed as symptoms of depressive disorder [6]. Thus, our results indicate that there is a strong need for perinatal interventions to focus on expectant mothers' mental health as prevention for PPD.

Correspondingly, we found that antidepressant use across three trimesters is a strong predictor of PPD. The treatment of women with depression or other psychiatric diseases during pregnancy or postpartum is a complex clinical challenge [25]. A previous study reported that women who had antidepressant treatment during pregnancy were less likely to report postnatal depressive symptoms, compared with the nonmedicated counterpart [18]; however, discontinued antidepressant medications during pregnancy was a risk factor for PPD [11]. On the other hand, antidepressants used during pregnancy might be associated with gestational hypertension and preeclampsia [31]. We did not find analgesic as an independent risk factor for PPD, although its association was reported in previous literature [29].

We recognize that this is a pilot study and there are certain limitations in our current work. First, since we combined medications by ATC classes, we were not able to differentiate drug use by specific dosage levels. In future studies using larger data sets, we will identify and combine different sources of disease diagnoses, and consider the dose-response relationship with medications and PPD. Second, the machine learning methods used in this study are standard methods, and the oversampling method used to handle our imbalance data may have contributed to overfitting and impacted model performance [17]. More advanced machine learning methods such as the neural network models will be used to improve AUC in future work. Third, since we used EHRs from a single health system, our data may miss mothers diagnosed with PPD outside of our health system, as well as information on those who were seen by clinicians outside of our health system before their pregnancy. Future studies will try to leverage multi-site dataset to minimize missing and erroneous data points. Lastly, in this study, we aimed to predict PPD rather than evaluating the causal relationships between variables in the pregnancy period and PPD. Future studies will use causal inference methods to control for potential and time-dependent confounders.

Conclusions

In this pilot study, we demonstrate promising PPD prediction results using a machine learning approach with information on patient demographics, diagnoses, and medications available from EHRs. Our goal is to create an accurate PPD prediction model to identify risk factors for PPD and facilitate effective screening of mothers who may require early intervention for PPD using an EHR. We envision that the model may be integrated with the EHR system for a provider-facing CDS or with a mobile or web platform to be used as a patient-facing CDS in a future phase of the study.

Acknowledgements

Authors have no financial interests to disclose. We would like to acknowledge Alison Hermann, MD, Scott Breitinger, MD, Kelsey Power, PhD, and Olivier Elemento, PhD at WCM who provided valuable feedback for the study. This study was funded in part by the Walsh McDermott Scholarship, R01 MH105384, P50 MH113838 and China Scholarship Council.

References

[1] C.T. Beck, Predictors of postpartum depression: An update, *Nurs Res* **50** (2001), 275-285.
[2] C.M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag, 2006.
[3] L. Blais, S.I. Salah Ahmed, M.F. Beauchesne, A. Forget, F.Z. Kettani, and K.L. Lavoie, Risk of Postpartum Depression Among Women with Asthma, *J Allergy Clin Immunol Pract* **7** (2019), 925-933.e922.
[4] L. Breiman, Random Forests, *Machine Learning* **45** (2001), 5-32.
[5] S. Brummelte and L.A. Galea, Postpartum depression: Etiology, treatment and consequences for maternal care, *Horm Behav* **77** (2016), 153-166.

[6] F.T. Burgut, A. Bener, S. Ghuloum, and J. Sheikh, A study of postpartum depression and maternal risk factors in Qatar, *J Psychosom Obstet Gynaecol* **34** (2013), 90-97.
[7] J.L. Cox, D. Murray, and G. Chapman, A controlled study of the onset, duration and prevalence of postnatal depression, *Br J Psychiatry* **163** (1993), 27-31.
[8] R.S. Evans, Electronic Health Records: Then, Now, and in the Future, *Yearb Med Inform* **Suppl 1** (2016), S48-61.
[9] J. Fisher, M. Cabral de Mello, V. Patel, A. Rahman, T. Tran, S. Holton, and W. Holmes, Prevalence and determinants of common perinatal mental disorders in women in low- and lower-middle-income countries: a systematic review, *Bull World Health Organ* **90** (2012), 139G-149G.
[10] E. Franck, M.A. Vanderhasselt, L. Goubert, T. Loeys, M. Temmerman, and R. De Raedt, The role of self-esteem instability in the development of postnatal depression: A prospective study testing a diathesis-stress account, *J Behav Ther Exp Psychiatry* **50** (2016), 15-22.
[11] S.H. Friedman and P.J. Resnick, Postpartum Depression: An Update, *Women's Health* **5** (2009), 287-295.
[12] E. Gultepe, J.P. Green, H. Nguyen, J. Adams, T. Albertson, and I. Tagkopoulos, From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system, *J Am Med Inform Assoc* **21** (2014), 315-325.
[13] M. Hassan and R. Kotagiri, A new approach to enhance the performance of decision tree for classifying gene expression data, *BMC proceedings* **7** (2013), S3-S3.
[14] K.L. Honey, P. Bennett, and M. Morgan, Predicting postnatal depression, *J Affect Disord* **76** (2003), 201-210.
[15] Y.K. Kim, J.W. Hur, K.H. Kim, K.S. Oh, and Y.C. Shin, Prediction of postpartum depression by sociodemographic, obstetric and psychological factors: a prospective study, *Psychiatry Clin Neurosci* **62** (2008), 331-340.
[16] K.B. Kozhimannil, C.M. Trinacty, A.B. Busch, H.A. Huskamp, and A.S. Adams, Racial and ethnic disparities in postpartum depression care among low-income women, *Psychiatr Serv* **62** (2011), 619-625.
[17] X. Liu, J. Wu, and Z. Zhou, Exploratory Undersampling for Class-Imbalance Learning, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **39** (2009), 539-550.
[18] A. Lupattelli, M.J. Twigg, K. Zagorodnikova, M.E. Moretti, M. Drozd, A. Panchaud, A. Rieutord, R.G. Juraski, M. Odalovic, D. Kennedy, G. Rudolf, H. Juch, and H. Nordeng, Self-reported perinatal depressive symptoms and postnatal symptom severity after treatment with antidepressants in pregnancy: a cross-sectional study across 12 European countries using the Edinburgh Postnatal Depression Scale, *Clin Epidemiol* **10** (2018), 655-669.
[19] R. Makadia and P.B. Ryan, Transforming the premier perspective hospital database into the observational medical outcomes partnership (OMOP) common data model, *EGEMS (Wash DC)* **2** (2014), 1110.
[20] J.R. Murphy, S. Paul, A.L. Dunlop, and E.J. Corwin, Maternal peripartum antibiotic exposure and the risk of postpartum depression, *Res Nurs Health* **41** (2018), 369-377.
[21] M.W. O'Hara and A.M. Swain, Rates and risk of postpartum depression—a meta-analysis, *International Review of Psychiatry* **8** (1996), 37-54.
[22] E. O'Connor, C.A. Senger, M.L. Henninger, E. Coppola, and B.N. Gaynes, Interventions to Prevent Perinatal Depression: Evidence Report and Systematic Review for the US Preventive Services Task ForceUSPSTF Evidence Report: Interventions to Prevent Perinatal DepressionUSPSTF Evidence Report: Interventions to Prevent Perinatal Depression, *JAMA* **321** (2019), 588-601.
[23] N.K. Ozcan, N.E. Boyacioglu, and H. Dinc, Postpartum Depression Prevalence and Risk Factors in Turkey: A Systematic Review and Meta-Analysis, *Arch Psychiatr Nurs* **31** (2017), 420-428.
[24] L. Pan, G. Liu, X. Mao, H. Li, J. Zhang, H. Liang, and X. Li, Development of Prediction Models Using Machine Learning Algorithms for Girls with Suspected Central Precocious Puberty: Retrospective Study, *JMIR medical informatics* **7** (2019), e11728-e11728.
[25] J.L. Payne and S. Meltzer-Brody, Antidepressant use during pregnancy: current controversies and treatment strategies, *Clinical obstetrics and gynecology* **52** (2009), 469-482.
[26] D. Pham, G. Cormick, M.M. Amyx, L. Gibbons, M. Doty, A. Brown, A. Norwood, F.M. Daray, F. Althabe, and J.M. Belizan, Factors associated with postpartum depression in women from low socioeconomic level in Argentina: a hierarchical model approach, *J Affect Disord* **227** (2018), 731-738.
[27] J. Shakespeare, F. Blake, and J. Garcia, A qualitative study of the acceptability of routine screening of postnatal women using the Edinburgh Postnatal Depression Scale, *Br J Gen Pract* **53** (2003), 614-619.
[28] J. Smith-Nielsen, A. Tharner, M.T. Krogh, and M.S. Vaever, Effects of maternal postpartum depression in a well-resourced sample: Early concurrent and long-term effects on infant cognitive, language, and motor development, *Scand J Psychol* **57** (2016), 571-583.
[29] T. Suhitharan, T.P. Pham, H. Chen, P.N. Assam, R. Sultana, N.L. Han, E.C. Tan, and B.L. Sng, Investigating analgesic and psychological factors associated with risk of postpartum depression development: a case-control study, *Neuropsychiatr Dis Treat* **12** (2016), 1333-1339.
[30] K. Takehara, Y. Tachibana, K. Yoshida, R. Mori, N. Kakee, and T. Kubo, Prevalence trends of pre- and postnatal depression in Japanese women: A population-based longitudinal study, *J Affect Disord* **225** (2018), 389-394.
[31] N. Zakiyah, L.F. ter Heijne, J.H. Bos, E. Hak, M.J. Postma, and C.C.M. Schuiling-Veninga, Antidepressant use during pregnancy and the risk of developing gestational hypertension: a retrospective cohort study, *BMC Pregnancy Childbirth* **18** (2018), 187.
[32] X. Zhang, B. Hu, X. Ma, and L. Xu, Resting-State Whole-Brain Functional Connectivity Networks for MCI Classification Using L2-Regularized Logistic Regression, *IEEE Trans Nanobioscience* **14** (2015), 237-247.
[33] Z. Zhang, Naïve Bayes classification in R, *Annals of translational medicine* **4** (2016), 241-241.

Address for correspondence

Corresponding author: Yiye Zhang
Email: yiz2014@med.cornell.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190352

Expanding Virtual Reality to Teach Ultrasound Skills to Nurse Practitioner Students

Elizabeth Weiner[a], Jeffry Gordon[a], Susanna Rudy[b], Ryan McNew[a]

[a] Frist Nursing Informatics Center, Vanderbilt University School of Nursing, Nashville, TN, USA,
[b] Vanderbilt University School of Nursing, Nashville, TN, USA

Abstract

This team has had prior successful production efforts in the use of Second Life to implement a virtual reality world into both nursing education and practice. Our current efforts center around a virtual reality application that provides learners with the opportunity to master competencies in the use of ultrasound technologies, which have surfaced in many nurse practitioner certifications as a necessary skill. Using the authoring tool of CenarioVR™, and capturing video footage of a nurse practitioner using the ultrasound equipment, students can don a VR headset for an immersive experience. This ultrasound equipment is expensive, and allowing learners to learn skills in a virtual environment poses less risk to both the equipment and the patients.Future production efforts in augmented and mixed reality applications are described to engage others in cutting edge use of VR.

Keywords:

Virtual reality; education, nursing; ultrasonography

Introduction

Healthcare educators, both in academic and practice settings, continue to seek the most realistic simulations for their students. With the rising cost of healthcare equipment, along with an increased attention to patient safety, these simulations have taken on even more importance. Clinical placement sites do not always have the opportunities to practice certain skills, and have become more difficult to obtain along with seasoned preceptors.

In this article we will describe the development of an immersive virtual reality simulation to teach nurse practitioner students how to effectively use ultrasound equipment. This competency has recently been added for certification in many of the nurse practitioner specialties in the United States.

Simulation is an educational technique which aims to mimic elements of the real world to help achieve specific goals related to learning [1]. When using the term "simulation," most nursing faculty and students think of the typical simulation environment comprised of lower level task trainers to high fidelity simulators. This type of simulation has been around for the last 40 years, but is expensive, difficult to schedule, and takes up physical space [2]. Simulation teaching methods have expanded to include peer-to-peer learning, standardized patients, screen-based computer simulations, and virtual reality.

Virtual reality (VR) is a high-level computer-based interactive and multimedia technology that offers virtual objects and environments that represent actual environments [3]. Virtual simulations have been used to train users in various skills such as bronchoscopy, colonoscopy, laparoscopy, and endoscopic procedures, and in disaster triage [4]. It is fully immersive, which tricks your senses into thinking you are in a different environment or world apart from the real world [5]. The unique contribution of VR is its ability to provide spherical full motion video. The user is placed in the center of a sphere with the capability to look up, down, right, and/or left. This is in contrast to 360° video where the user can only look left and right. VR can also be contrasted to 3D video, typically on a flat plane, which gives a stereoscopic view with depth perception of the scene.

Three of the authors of this article have completed projects using the VR world of Second Life (SL). SL is a 3D virtual world, created by its residents. The world is driven by the interactions of real-world individuals and their avatars. Thus, for every avatar one encounters in SL, there is a live person somewhere in the world who is directing that avatar's actions, emotions, words, dress, etc.

The SL environment was built as part of prior funded grants from the U.S. Department of Health and Human Services Administration (Grants #D80HP11271 and #U1KHP1296). These grant activities have been reported in prior international informatics conferences [6,7,8]. The purposes of these grants were to provide nursing faculty the opportunity to manage clinical simulations while advancing in their own competency and proficiency levels of simulation management while in the SL virtual world. Synchronous sessions were provided by master teachers, after an orientation to SL had been completed by the users. Initial findings demonstrated that faculty participants thought that the programs met or exceeded their learning needs and would be directly applicable to their teaching. Project barriers were the time spent in development of this complex environment and the technology support needed to successfully run the developed scenarios. Those users who were technophobic had difficulties in learning to move their avatars as needed, which distracted them from the overall learning experience. On a more positive note, the replicated Eskind Diabetes Center created in SL was used as the basis for nurse practitioners to provide care to diabetes patients while both were avatars. Finally, this patient care example was presented as a method to better engage patients in their own care.

SL has been used by other nursing educators over the last decade. Trangenstein, Weiner, Gordon and McNew completed a literature search that reviewed 29 nursing educational virtual environments and discovered that nursing educators used SL as explorers and developers [9]. Explorers used SL for a support group, networking,and uncovering health-related sites. The developers were associated with land ownership and were involved in distance learning and simulation activities. Seven unique simulations for nursing education were identified. This type of virtual simulation environment has been generally well received and evaluations of its used in educational settings has been positive [10].

Augmented reality overlays digital information on real-world elements [5]. Pokémon Go is a widely known example. In augmented reality applications, the real world is central but enhanced with other digital details, thus supplementing your reality or environment. Mixed reality brings together real world and digital elements [5]. Mixed reality allows the user to see and immerse oneself in the world around the user while interacting with a virtual environment. It combines the ability to be in a real world enhanced by an imaginary place or image therefore offering an experience that is not possible in the real world alone.

The purpose of this article is to describe the development process used in an immersive VR simulation aimed to teach ultrasound skills to nurse practitioner students. While evaluation data will not be available until the actual presentation, the structure of the evaluation will be discussed. In addition, future applications using both augmented reality and mixed reality will be described.

Methods

Ultrasound has been described as the stethoscope of the 21st century. Evans and Evans demonstrated that with some training, health care professionals can successfully use the ultrasound machine to identify common pathologies [11]. They also claim it can shorten patient waiting time, lower costs, improve diagnoses, and reduce the pressure on other radiography lists.

Portable ultrasound technology is increasingly used in the diagnosis and management of critically ill patients (See Figure 1). As an easily accessible and frequently used bedside tool, ultrasound has the potential to improve patient safety; however if untrained providers use the equipment, the result could be harmful [12].

Figure 1 – Ultrasound Machinee

The charge then becomes to educate large numbers of students on how to successfully setup and use the ultrasound machine on various pathologies. This project was selected for several reasons. First, the typical ultrasound machine demonstration works best for the few students in the front and poorly for those in the back unable to see everything that is transpiring. Such an approach is not scalable to the numbers who need the training. Secondly, many of our nursing specialties want to teach how to use the device but because of the expense, we only have two machines to split among all of the specialties. Each machine was purchased with university discounts but with accessories averaged $85,000 US. These machines have sensitive electrodes, so moving the equipment around the building is also not a desirable option. By simulating the machine and watching an expert complete the ultra sound examination, the students save valuable time when using the actual machine because they will have already been taught how to use it. A seasoned nurse practitioner who has mastered these skills and uses them frequently becomes their expert role model. Students receive consistent quality controlled instruction and learn how to use the machine effectively and safely without risk to patient. Finally, the topic itself is highly visual, making Immersive VR a highly appropriate tool to teach how to setup and use the ultrasound machine.

Format of the Project

The ultrasound application lends itself to VR video development. It is very visual and the technology supports training users how to use the ultrasound device.

The application was divided into two components. The first component simulates the setup and configuration of the machine. The student can select their phenomena of interest, identify the correct probe to use, set the depth, brightness, contrast, and color desired for the correct exposure (see Figure 2). Finally, the student can enter in the patient information and have that stored in a database just as it is on the real device.

Figure 2 – Positioning the Probe

Data elements are entered in the simulation on a standard computer screen, identical to the screens of the actual device itself.

Once the ultrasound machine is configured properly and the appropriate data entered, the student enters the second component. They are told to download the Immersive VR application, launch it on their cell phone, and insert the phone into the VR headset. When the student puts on the headset, he or she is positioned in the middle of the sphere, called a scene, directly in the room looking over the patient so that the student can see everything the clinician is doing in a virtual 3dimensional experience. By looking around, the student can see where and how the clinician is placing the probe.

The student can view actual output of the ultrasound projected above the patient's head simply by looking up above the patient's head. This technique also amplifies the output for easier viewing (see Figure 3).

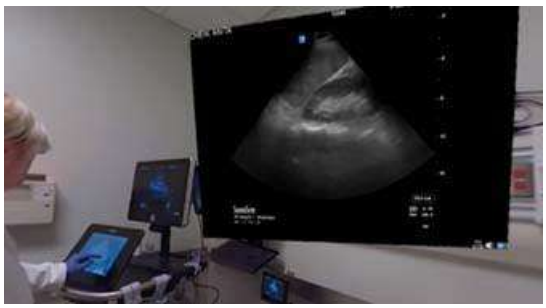


Figure 3 – Ultrasound Output

The student can also view the facial expressions and non-verbal communication of the patient and clinician. The student has the illusion of actually being in the room with only the clinician and the patient. Unlike group viewing situations where views might be obstructed by other students and the output too small to see by everyone, this method provides for a quality-controlled experience.

The student can pause the scene, rewind it, and turn his or her head to view the actual ultrasound output in real time corresponding to the probe positioning and movement (see Figure 4).. At various points in the exam, the instructor can pause and ask a question or even ask the student the next position for the probe.

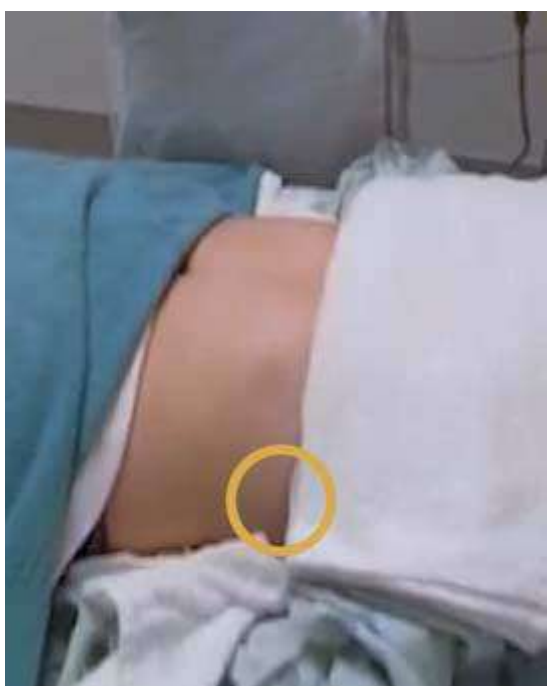


Figure 4 – Identifying Probe Position

Student responses can be recorded and stored in a learning management system (LMS) for later review and performance evaluation. Feedback to the student can be textual, oral, or video branched to another sphere.

Development Tools Selected

A software authoring environment needed to be chosen. CenarioVR is an easy to use immersive VR development SaaS (Software as a Service) tool, developed by the Trivantis Corporation, that allows the developer to create photo and video spheres and incorporate them into a learning environment. The product can be played on a computer screen in HTML5 using mouse controls for movement or placed on a small mobile device that is inserted into a VR headset (such as Google cardboard, Daydream, or Oculus Go) for the full immersive experience (see Figure 5). When used immersively, the student has the sense of actually being in the room participating directly with the activities. The video can be paused, rewound, or branched to other spheres, called "scenes". Artifacts such as the actual ultrasound output can be inserted anywhere in the sphere and reflect the output of the actual exam in real time on the video sphere. Questions can be asked, and answers recorded and shared to the learning management system (LMS) through xAPI or SCORM. The students can then be graded on the quality of their work in the environment.

Figure 5 – VR Headsets

The video was collected using a spherical camera. The GoPro Fusion camera was selected along with Fusion Studio software on the PC (see Figure 6). Control of the camera was initiated through the GoPro phone application. The camera was positioned close to the patient, opposite from the clinician. Care was taken to make sure the clinician's hand positioning was fully visible to the camera and not eclipsed by the patient's body. A portable DVR was attached via HDMI to the ultrasound device to capture the video output from the ultrasound device.

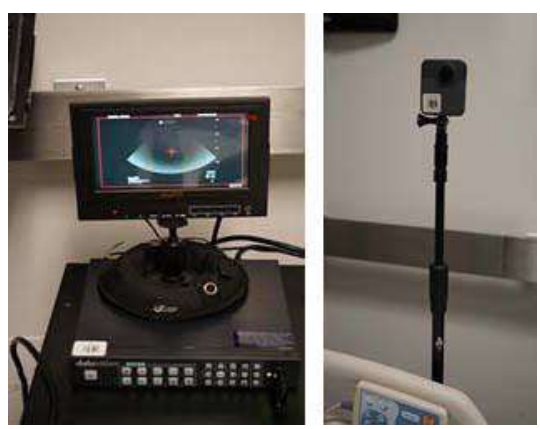


Figure 6 – The DVR Video Capture Device and GoPro Fusion Camera

The spheres were assembled into equirectangular format and downloaded through the Fusion studio application, then compressed in the software application, Handbrake. The resulting scenes were brought into Adobe Premiere on the main track. The video of the output of the ultrasound was downloaded onto the PC and placed into a separate track in Premiere. It was aligned so that the video of the output matched with what the clinician was doing on the patient.

Using the VR tools in Premiere, the ultrasound output video was sized large to view and positioned over the head of the patient. It was as if a very large screen monitor was displaying the output over the head of the patient. It should be noted that placing video inside a sphere in Premiere requires a minimum of 1GB of video ram. We selected an Alienware 15 i7 laptop with 32GB of memory for its graphics processing capabilities (see Figure 7).

Figure 7 – Video Processing Using Alienware 15 i7

At specific points the video of the sphere was paused and questions asked. Successful selection of the answer started the video again.

Results

The project will be implemented with acute care nurse practitioner students during the 2019 spring semester. A short video of the use of the project will be shown during the conference as a video.

Evaluation results will also be presented. Participating students will be asked to complete the National League for Nursing's Simulation Design Scale and the Student Satisfaction and Self –Confidence in Learning [13]. Both tools have had prior testing with Chronback's alpha measures of .92 for presence of features, .96 for the importance of features, .94 for satisfaction, and .87 for self-confidence [13].

In addition, an evaluation checklist will be developed based on the Ultrasound Guidelines developed by the American College of Emergency Physicians [14]. The checklist created by Patrawall et al., [12] will be consulted in evaluating competencies for critical care ultrasound.

Discussion

This work is original due to the fact that ultrasound training is not available in a format that offers a scalable solution. Academic and practice settings suffer in their ability to be able to provide enough training equipment that can be used for the amount of time needed for skills mastery in an area that is becoming more important for nurses to administer correctly. Once students or practitioners have passed the competencies using the VR solution, the time to use an actual machine to demonstrate mastery should be substantially reduced.

It is anticipated that there will be positive results from the evaluation data submitted by students. It has been difficult due to the number of students needing to use the machine to have such an opportunity. In addition, one-on-one faculty supervision is not always available, and group teaching lends itself to situations where viewing is not feasible by all in the room, and students may miss important details such as placement of the probes. An ultrasound simulation allows for both individual use, but also allows for the repetition that might be needed for some students.

Additional innovations now seem feasible to implement within our facility. The Vanderbilt University School of Nursing has just completed a $23.6 million building expansion. The result is a new five-story building; an atrium connecting all School of Nursing building; a state-of-the-art simulation lab; student service office;, a virtual classroom; technology advanced classrooms, conference, offices, and seminar rooms; a green roof terrace; and outdoor green space.

An augmented reality way-finding application is currently being developed. Once users see the interactive maps on the large video walls in the atrium, additional wayfinding graphics can be transferred to their smartphones so that they can follow directions to their destinations. Rather than having to record directory details, the directions will be transferred directly to their phones.

A second planned application will be for a mixed reality project using the Microsoft HoloLens. Authoring with Unity, we plan to build holographic images of hearts so that various cardiac abnormalities and treatments can be explored among faculty and students as they engage with this digital content and the world around them.

Conclusions

VR offers one solution to expand the teaching-learning environment. Any strategy that allows learners to practice without added stress and with increased confidence can only lead to increased quality outcomes. In a time when clinical placements are becoming more difficult, the chance to practice skills in a safe environment that leads to skill mastery is an added plus. This ultrasound application has the potential to meet all of these requirements. As other clinical specialties require the use of these skills, this VR application will become even more popular.

The development team has also been able to extrapolate two other augmented and mixed reality applications that could benefit our programs. Innovation has become an expected requirement for our students, so faculty and staff must continue to think of useful innovations.

Acknowledgements

The authors wish to thank Allie Saxon for her willingness to be the patient for this demonstration. In addition, JoEllen Holt (Skills and Simulation Lab Director) and Diana Kelley (Simulation Lab Nurse) were most helpful in preparing and participating in the project. Recordings were made in the new Simulation Lab Facility which opened in January of 2019.

References

[1] J. Williams, D. Jones, and R. Walker, Consideration of using virtual reality for teaching neonatal resuscitation to midwifery students, *Nurse Educ Pract* **31** (2018), 126-129. doi:10.1016/j.nepr.2018.05.016.

[2] S. Cooper, R. Cant, J. Porter, F. Bogossian, L. Mckenna, S. Brady, and S. Fox-Young, Simulation based learning in midwifery education: a systematic review, *Women Birth* **25** (2012), 64-78.

[3] E. G. İsmailoğlu and A. Zaybak, Comparison of the effectiveness of a virtual simulator with a plastic arm model in teaching intravenous catheter insertion skills, *CIN* **36** (2) (2018), 98-105. doi:10.1097/CIN.0000000000000405.

[4] C. Foronda, C. M. Alfes, A. J. Kleinheksel, D. A. Nelson, J. M. O'Donnell, and J. T. Samosky, Virtually nursing: emerging technologies in nursing education, *Nurse Educ* **42** (1) (January/February 2017), 14-17. doi:10.1097/NNE.0000000000000295.

[5] Virtual reality vs. augmented reality vs. mixed reality, (2018). https://www.intel.com/content/www/us/en/tech-tips-and-tricks/virtual-reality-vs-augmented-reality.html?wapkw=virtual+reality (accessed November 8, 2018).

[6] E. Weiner, R. McNew, P. Trangenstein, and J. Gordon, Using the virtual reality world of Second Life to teach nursing faculty simulation management. In C. Safran, H. Marin, and S. Reti (eds.), *Proceedings from MEDINFO 2010: Partnerships for Effective eHealth Solutions, 13th World Congress on Medical and Health Informatics*, (2010), 615-619. doi:10.3233/978-1-60750-588-4-615.

[7] E. Weiner, T. Elasy, N. Lorenzi, R. McNew, L. Shackleford, K. Wolff, S. Martin, and C. Benson, Extending nurse practitioner care using the virtual reality world of Second Life*, Proceedings from NI2012: 11th International Congress on Nursing Informatics*, (2012), 437-441.

[8] E. Weiner, P. Trangenstein, R. McNew, and J. Gordon. Using the virtual reality world of Second Life to promote patient engagement, *Stud Health Technol Inform* **225** (2016), 198-202. doi:10.3233/978-1-61499-658-3-198.

[9] P. A. Trangenstein, E. E. Weiner, Gordon, J. S., and R. McNew, An analysis of nursing education's immersion into Second Life, a multi-user virtual environment (MUVE), *Stud Health Technol Inform* **160** (2010), 644-647. doi: 10.3233/978-1-60750-588-4-644.

[10] S. Metcalfe, Second Life patient scenarios: enhancing the diversity of the nursing profession, *Creat Nurs* **22**(3) (2016), 166-170. doi: 10.1891/1078-4535.22.3.166.

[11] T. Evans and C. Evans, Is ultrasound the stethoscope of the 21st century?, *Int J Surg* **18** (2015), 238-244.

[12] P. Patrawalla, L. A. Eisen, A. Shiloh, B. J. Shah, O. Savenkov, W. Wise, L. Evans, P. Mayo, and D. Szyld, Development and validation of an assessment tool for competency in critical care ultrasound, *J Grad Med Educ* 7(4) (2015), 567-573. doi:10.4300/jgme-d-14-00613.1.

[13] National League for Nursing, NLN/Laerdal research study instruments, (2018). https://sirc.nln.org/mod/page/view.php?id=88 (accessed November 13, 2018).

[14] Ultrasound guidelines: emergency, point-of-care and clinical ultrasound guidelines in medicine, *Ann Emerg Med* **69**(5) (2017), e27-e54. doi:10/1016/j.annemergmed.2016.08.457.

Address for correspondence

Elizabeth (Betsy) Weiner, Vanderbilt University School of Nursing, 461 21st Ave. S., Nashville, TN 37240-1104, USA; E-mail: Betsy.Weiner@Vanderbilt.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190353

Evaluating the Validity of a Knowledge-Based System for Proactive Knowledge Transfer for Caregiving Relatives

Dominik Wolff[a], Marianne Behrends[a], Thomas Kupka[a] and Michael Marschollek[a]

[a] *Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover, Germany*

Abstract

The evaluation of the validity of a knowledge-based system is of great importance during its development. It determines whether the system represents the experts' knowledge correctly. This is highly important, but also particularly difficult, if the expert knowledge is not explicit, but only implicit and tacit. In the following the validity's evaluation of a system for education of caregiving relatives is presented. To evaluate the system's knowledge delivery strategy against the experts' opinion, several fictious characters were created. The evaluation revealed inconsistencies in the knowledge base. After resolving these, the experts' opinion is represented to a large extent by the system. Nevertheless, the used evaluation approach is not capable of detecting all inconsistencies. Therefore, a strong need of a system's learning capacity to integrate feedback from a larger group of real caregiving relatives exists. In addition, a rule-based component, representing disease specific knowledge, should be implemented.

Keywords:

Knowledge Bases, Nursing Informatics, Evaluation Studies

Introduction

Approaches in the field of artificial intelligence divide into two main fields: soft computing and symbolic approaches. While the data driven soft computing approaches use existent data to learn how to solve a problem, the implementation of knowledge driven symbolic artificial intelligence systems involves the explicit formalization of the subject-specific experts' knowledge. This enables artificial intelligence engineers to model the knowledge in a machine interpretable representation language. The necessary cooperation within an interdisciplinary team during the explication and formalization process is a challenging task for multiple reasons. One reason is, that the practitioners of interdisciplinary studies lack a coherent, defensible sense of their purpose [1]. This often manifests itself in communication problems. Even more, the experts' knowledge ordinarily is not explicit, but implicit and tacit. "Compared to explicit knowledge, tacit knowledge is subjective, personal, and context-specific. It is difficult to describe, examine, and use" [2], which results in an even harder explication and formalization process. To ensure that the developed knowledge-based system represents the experts' knowledge and behaves in the intentional way, a sophisticated evaluation must be performed. Therefore, the evaluation of the validity of a knowledge-based system is of great importance during its development, because it determines whether the system represents the experts' knowledge correctly [3].

In [4], a wider methodology for the evaluation of knowledge-based systems is introduced. Here, a comprehensive approach is presented, which covers not only the validity of the system used, but also the functionality and impact of the complete system. The approach structures into four phases. In the first phase, early prototypes are created using rapid prototyping or requirements specifications. The second evaluation phase estimates the systems validity. Therefore, the inference mechanism, as well as the knowledge base must be evaluated. The knowledge base can be analyzed in a static or dynamic way. Static analysis refers to procedures where no system execution is required, such as internal consistency checks. On the other hand, dynamic analysis incorporates system execution to detect incorrect parts and presents them to the experts. Furthermore, the knowledge base needs to be evaluated against a set of cases different from those used during the system's development, which act as a kind of gold standard. Here, multiple methods like traditional error rate, top down correlation or receiver operated characteristics could be used. The third phase evaluates the system's functionality, which incorporates user interaction, user interface and reasoning efficiency, but is not limited to these. In the last phase, the systems impact is evaluated. Therefore, the outcome of the system's application is compared to the current state. Depending on the system's purpose, various measures can be applied, such as benefits of the expenses.

This approach is used to evaluate a knowledge-based system for education of informal caregivers, which are mostly caregiving relatives. For many relatives the care situation involves extensive burden [5], which is caused, among other things, by a lack of knowledge. Against this background, a system for personalized education of caregiving relatives is developed. The methodology of the system was introduced and three prototypes were evaluated in [6]. Here, statistical approaches are utilized to calculate scores for each knowledge resource to be delivered, based on experts' weighting of profile items. The aim of our research work for this paper was the evaluation of our system's validity. Therefore, the results of the second evaluation phase, which covers the validity of the system, will be presented and discussed. Since the third approach from [6], which utilizes dynamic weights, performed best during prototype testing, it is only further evaluated.

Methods

Before testing with real human users, the system's knowledge delivery strategy on which the validity depends must be evaluated. As stated above, the validity evaluation involves two steps: the evaluation of the inference mechanism and the evaluation of the knowledge base. In particular, the inference mechanism consists of the mathematical formula for calculating the score, while the knowledge base consists of the experts' weightings of profile items for all 82 knowledge resources. Typically, the inference algorithm could be evaluated using software

validation techniques [7]. But these limit the evaluation process to models, where an explicitly expressed requirement or gold standard can be formulated and are not applicable to validate the expatiating process itself. The method described in [6] was created to expatiate the tacit experts' knowledge and therefore cannot be evaluated against an explicit formalization alone. Here, the validity of both, inference mechanism and knowledge base, depend on each other and can hardly be observed individually. In consequence, good validity can only be achieved by good results in both components. Considering that a completely new statistical approach was used to formalize the tacit knowledge and not a representation language, an automated static evaluation of the knowledge base is not possible. Hence, a dynamic analysis was performed. In a first step, a selection of 21 knowledge resources, including important and unimportant knowledge resources, are evaluated against an expert's ordering. For evaluating the entire knowledge base, the total number of 82 knowledge resources must be considered. In this setting, experts are not able to order all knowledge resources for multiple test settings. Accordingly, the calculated orderings were submitted to the experts for evaluation.

Evaluation against experts ordering

A fictious character was created by the experts in nursing science. Johanna is 74 years old and intensively cares for her 76 years old husband Peter. Peter has been suffering from the consequences of a stroke for 2 years and is hemiplegic. Due to hemiplegia, he can no longer walk without support, cannot dress alone and needs support in tasks of daily living. He is already fallen several times. For Johanna it is a matter of course to take care of her relatives. She supports intuitively and has little information about the disease progression, possible "care techniques" and possible support services. She feels overwhelmed in many things and therefore suffers from sleep deprivation. Johanna has become increasingly isolated and receives no support from third parties, such as outpatient nursing services. For the evaluation, the experts selected a subset of 21 knowledge resources, divided into five highly important, three less important and 9 unimportant knowledge resources. The ordering of the system is compared with the order obtained by the experts. For the system's ordering knowledge resources with a score greater than 0.5 are regarded as important. All others are considered as unimportant.

Evaluation of the complete knowledge base

For the evaluation of the knowledge base, it is not sufficient to regard only a subset of knowledge resources. However, ordering the total amount of 82 knowledge resources by the experts is a very time-consuming task and cannot be performed for multiple test settings. Therefore, a dynamic evaluation process is performed in which the experts judge a given ordering calculated by the system. Thus, the authors created 7 fictious caregivers by regarding the weighting of one or multiple thematically very proximal knowledge resources as the caregiver's profile. As a result, the used knowledge resources will be top ranked in the system's ordering. The system's ordering was cut at a threshold of 0.5. Everything below was excluded for a simulation of the later system behavior, in order to not disturb the later real users with unimportant knowledge resources. Additionally, resources not fitting the disease of the care recipient were excluded as well. In order to qualify, whether the experts are just marking everything as correct, unimportant knowledge resources were interspersed to the correct orderings (without the experts' awareness). This results in three completely important orderings, two completely unimportant ones and two of mixed importance. These orderings were evaluated by the experts for a second time approximately four weeks later, yet with the knowledge, that unimportant resources were interspersed. In the following, the descriptions and orderings to be evaluated by the experts are presented. The unimportant resources are marked.

Klaus and Gerda

Klaus cares for his 75 years old mother Gerda. He himself is 53 years old and works as an architect. Gerda suffered a stroke one year ago. Since then, she is partly aggressive towards him and insults him very harshly. Such behavior has never occurred before the stroke. Furthermore, Gerda is frequently sad and depressed. This imposes a great burden on Klaus. Due to the burden he is commonly exhausted. Klaus wishes to be able to withdraw from the care situation but feels ashamed for such thoughts.

1. Changes in personality after a stroke
2. Detecting and preventing depression
3. How does the sense of taste change with age?

Charlotte and John

John (67) cares for his demented partner Charlotte. She is 70 years old. Charlotte needs help with dressing and other activities. She is overextended with everyday decisions, but concurrently cannot admit that she is dependent on help. This makes John sorrowful and enraged. When her children come to visit, John is embarrassed about the way Charlotte is dressed. This often leads to an argument.

1. Clothing
2. Bladder training (unimportant)
3. Psychological changes in dementia
4. Restlessness - sources and handling (unimportant)

Mareike and Matthias

Mareike is 60 years old and works as an office clerk. She takes care of her demented spouse Matthias. Due to Matthias disorientation at night, Mareike sleeps poorly. Matthias is capable of changing his position in bed independently, which frequently awakens Mareike. She feels emotionally and physically exhausted due to the care.

1. How does sleep behavior change with age?
2. Suggestions for a restful sleep
3. Sources of sleeping problems
4. Coping with stress
5. Restlessness in the evenings
6. Dealing with nocturia
7. Detecting and preventing depression
8. Kinaesthetic - definition and relief in care
9. Exhaustion

Hannes and Jana

The pensioner Hannes takes care of his partner Jana. He does not get enough sleep because he must take care of Jana constantly, even at night. Jana is temporally and locally disoriented and frequently walks through the apartment without a destination. Two months ago, Jana was diagnosed with dementia.

1. Physical activity (unimportant)
2. Course of dementia (unimportant)

3. Kinaesthetic - definition and relief in care (unimportant)
4. Financial relief for day and night care (unimportant)
5. Structuring the toilet calls (unimportant)
6. Finger food (unimportant)

Alina and Charlie

The 42 years old Charlie has the feeling that his mother is increasingly dependent on support due to problems in communicating her needs. He also has to cut her food into small pieces. Recently, Alina accidentally caused a grease fire while cooking. Charlie was barely able to prevent the worst. His mother is unaware of this incident's severity.

1. Drinking – when, how much and what?
2. How does the sense of taste change with age?
3. Changes in motion patterns (unimportant)
4. Biography work used correctly (unimportant)
5. Swallowing and swallowing disorders

Anton and Marie

Anton is cared by his 40-year-old daughter Marie, as he needs help with basic everyday tasks. Anton is demented, often disoriented and struggles with multi-step actions. Additionally, he has difficulties communicating his needs. Marie supports him with his meals.

1. Handling incontinence material (unimportant)
2. Changes in motion patterns (unimportant)
3. Dealing with nocturia (unimportant)
4. Assistive devices - everyday relief (unimportant)

Dieter and Helga

55-year-old Dieter cares for his mother Helga. Helga needs assistance with everyday tasks such as washing, positioning, dressing and undressing. She is able to take meals herself, yet has become urinary incontinent and has been increasingly dependent on help since the last six months. Dieter wishes for support.

1. Assistive devices - everyday relief
2. Constructional changes
3. What is fecal incontinence?
4. Handling incontinence material

Results

Evaluation against experts ordering

Comparing the orderings obtained by the experts and by the system (Table 1), seven knowledge resources were assigned to different levels of importance. Five unimportant knowledge resources in the experts' ordering are rated with a score above 0.5 and therefore are regarded as important by the system. These are *What is a tendency to run?*, *Dealing with refusal to eat*, *Finger food*, *Tips for washing demented persons* and *Dealing with nocturia*. One important knowledge resource from the experts ordering (*Communication with people after a stroke*) is regarded as unimportant by the system and *Fall prevention* actually is not within the systems ordering, since it is not represented in the knowledge base. The ordering of unimportant knowledge resources is meaningless, since these knowledge resources are not displayed to the user. Regarding this, the two orderings have a Kendall rank correlation coefficient of 0.31, where the missing knowledge resource *Fall prevention* was excluded to obtain orderings of the same length. This indicates a weak positive correlation between the two orderings.

Table 1 - Orderings by importance obtained by the experts and by the system

Experts' ordering	Systems ordering
Fall prevention	Kinaesthetic - definition and relief in care (1.0)
Kinaesthetic - definition and relief in care	The concept of Bobath (1.0)
The concept of Bobath	Restriction of movement after a stroke (1.0)
Restriction of movement after a stroke	Tips for washing demented persons (1.0)
What is urinary incontinence?	Kitchen aids after a stroke (0.833)
Structuring the toilet calls	What is fecal incontinence? (0.833)
What is fecal incontinence?	Drinking – when, how much and what? (0.833)
Drinking – when, how much and what?	Constructional changes (0.833)
Communication with people after a stroke	Dealing with nocturia (0.75)
Detecting and preventing depression	Dealing with refusal to eat (0.722)
Kitchen aids after a stroke	Structuring the toilet calls (0.685)
Constructional changes	What is urinary incontinence? (0.667)
What is a tendency to run?	Finger food (0.633)
Wandering of demented persons	Detecting and preventing depression (0.567)
Dealing with refusal to eat	What is a tendency to run? (0.561)
Finger food	Restlessness in the evenings (0.25)
Tips for washing demented persons	Changes in motion patterns (0.21)
Biography work used correctly	Wandering of demented persons (0.167)
Restlessness in the evenings	Biography work used correctly (0.0)
Dealing with nocturia	Communication with people after a stroke (0.0)
Changes in motion patterns	

Note: Borderlines inside the table mark changes in the importance and numbers in brackets show the system's calculated scores

Evaluation of the complete knowledge base

Table 2 shows the evaluation results of the complete knowledge base. A knowledge resource is identified as critical and therefore requires further investigation if at least two out of three experts reviewed it not as supposed (false positives and false negatives) in the same evaluation cycle. For Klaus and Gerda this applies to *How does the sense of taste change with age?* in both evaluations. For Mareike and Matthias in both evaluations, it is *Kinaesthetic - definition and relief in care* and *Detecting and preventing depression*. For the characters Hannes and Jana *Course of dementia* and *Financial relief for day and night care*

are the critical knowledge resources in both evaluations and *Finger food* is critical in the first evaluation. *Biography work used correctly* is critical in the first evaluation, *Drinking – when, how much and what?* and *Swallowing and swallowing disorders* in the second evaluation of the characters Alina and Charlie. For Anton and Marie *Changes in motion patterns* and *Assistive devices - everyday relief* are critical in both evaluations. The knowledge resource *What is fecal incontinence?* is critical in both evaluation cycles for the personas Dieter and Helga. Furthermore, the knowledge resource *Constructional changes* needs to be regarded as critical in the second evaluation. Especially critical are the knowledge resources *Course of dementia* (Hannes and Jana) and *What is fecal incontinence?* (Dieter and Helga), as they are reviewed the opposite way by all three experts in both evaluation cycles. Moreover, *How does the sense of taste change with age?* (Klaus and Gerda) and *Financial relief for day and night care* (Hannes and Jana) are reviewed the opposite way by all three experts in one evaluation cycle and by two experts in the other evaluation cycle. The knowledge resource *Biography work used correctly* from the characters Alina and Charlie, which is unimportant from the systems point of view, is noticeable. In the first evaluation cycle, all three experts assessed it as important. While in the second evaluation cycle, all experts rated it as unimportant.

Table 2 – Experts' review of the characters. The sign + stands for reviewed as important and the sign - stands for reviewed as unimportant

		Expert #1		Expert #2		Expert #3	
Evaluation cycle		**#1**	**#2**	**#1**	**#2**	**#1**	**#2**
Klaus and Gerda	**1.**	+	+	+	+	+	+
	2.	+	+	+	+	+	+
	3.	+	-	-	-	-	-
Charlotte and John	**1.**	+	+	+	+	+	+
	2.	-	-	-	-	-	+
	3.	+	+	+	+	+	+
	4.	-	-	-	-	+	-
Mareike and Matthias	**1.**	+	+	-	-	+	+
	2.	+	+	+	+	+	+
	3.	+	+	+	+	+	+
	4.	+	+	+	+	+	-
	5.	+	+	+	+	+	-
	6.	+	+	-	-	+	+
	7.	+	+	-	-	-	-
	8.	-	-	-	+	-	-
	9.	+	+	+	+	+	+
Hannes and Jana	**1.**	-	-	-	+	-	-
	2.	+	+	+	+	+	+
	3.	-	-	-	-	-	-
	4.	+	+	+	+	+	-
	5.	-	+	-	-	-	-
	6.	+	+	+	-	-	-
Alina and Charlie	**1.**	+	+	+	-	+	-
	2.	+	+	+	-	+	-
	3.	+	-	-	+	+	-
	4.	+	-	+	-	+	-
	5.	+	+	-	+	+	-
Anton and Marie	**1.**	-	-	-	-	-	-
	2.	+	-	+	+	-	+
	3.	-	-	-	-	-	-
	4.	+	+	-	-	+	+
Dieter and Helga	**1.**	+	+	-	+	+	-
	2.	+	+	-	-	+	-
	3.	-	-	-	-	-	-
	4.	+	+	+	+	+	+

For characterizing the experts, table 3 shows important metrics. No ordering was reviewed perfectly by the experts. The high sensitivity (0.905) and lower specificity (0.5) of expert #1 shows, that she reviews overly positive. She detects the majority of the important knowledge resources but misses the unimportant ones. The opposite occurs for expert #2. She reviews in a more negative way, as the sensitivity (0.619) is smaller and the specificity (0.714) is higher. Expert #3 reviews in a positive way, but not as positive as expert #1 (sensitivity: 0.81, specificity: 0.571).

Table 3 - Metrics for the experts' overall results from the first evaluation cycle

	Expert #1	**Expert #2**	**Expert #3**
Sensitivity/recall	0.905	0.619	0.81
Specificity	0.5	0.714	0.571
Precision	0.731	0.765	0.739
F1-Score	0.809	0.684	0.773

The metrics' results for the second evaluation cycle, where the experts were informed about the existence of interspersed unimportant knowledge resources, are presented in Table 4. Expert #1 shows approximately the same sensitivity and precision, but a better specificity. For expert #2, almost the same results can be observed in both evaluation cycles. The largest difference between the two cycles occurs for expert #3. The sensitivity drops from 0.81 to 0.476 and the specificity rises from 0.571 to 0.714. This indicates that expert #3 identifies fewer important knowledge resources, but more unimportant ones and therefore reviews in a more negative way.

Table 4 - Metrics for the experts' overall results from the second evaluation cycle

	Expert #1	**Expert #2**	**Expert #3**
Sensitivity/recall	0.857	0.667	0.476
Specificity	0.643	0.714	0.714
Precision	0.783	0.778	0.714
F1-Score	0.818	0.718	0.571

Discussion

Knowledge resources with different classifications of importance were identified by the evaluation against the experts' ordering. These were handed to and examined by the expert group. They revised every knowledge resource with the exception of *Communication with people after a stroke*. As this resource includes basic knowledge, which is important for all caregivers caring for someone suffering from a stroke, it was excluded from this provision strategy and included in a group of basic knowledge. This ensures that all affected caregivers are provided with this basic knowledge. Changes in the weighting haven been made to *Dealing with refusal to eat* and *Dealing with nocturia*. The remaining resources were classified as unimportant by the experts, as they are merely of interest to people with dementia and the disease is not part of the profile data used. The experts' knowledge about the delivery of disease specific knowledge resources is explicit. For that reason, inclusion of the disease in this profile data set is not target-oriented. Yet, it should be implemented as a rule-based system. The entire knowledge base was reviewed by the experts with regard to disease specific knowledge delivery. The order after revision and excluding resources that are not relevant due to the disease is shown in Table 5. By the revision process and excluding resources by disease, the knowledge provision is improved and thus the Kendall rank correlation coefficient increases to 0.53. Even more, the unimportant knowledge resources are filtered out by the score's threshold of 0.5.

Table 5 - Orderings by importance after revision and excluding by disease

Experts ordering	Systems ordering
Fall prevention	Kinaesthetic - definition and relief in care (1.0)
Kinaesthetic - definition and relief in care	The concept of Bobath (1.0)
The concept of Bobath	Restriction of movement after a stroke (1.0)
Restriction of movement after a stroke	Fall prevention (0.833)
What is urinary incontinence?	Kitchen aids after a stroke (0.833)
Structuring the toilet calls	Constructional changes (0.833)
What is fecal incontinence?	Drinking – when, how much and what? (0.833)
Drinking – when, how much and what?	What is urinary incontinence? (0.815)
Detecting and preventing depression	What is fecal incontinence? (0.792)
Kitchen aids after a stroke	Structuring the toilet calls (0.685)
Constructional changes	Detecting and preventing depression (0.567)
Dealing with refusal to eat	Dealing with nocturia (0.455)
Dealing with nocturia	Dealing with refusal to eat (0.388)

The small amount of knowledge resources regarded as important by the system for the seven fictious characters indicates that the weightings of the different knowledge resources are quite disjunctive. The approach for the evaluation of the complete knowledge base revealed further inconsistencies, which are examined by the experts. Particularly critical are resources that have been incorrectly evaluated by all three experts in both evaluation cycles. Until now, all inconsistencies were resolvable by revising the knowledge base and it can therefore be assumed that the inference mechanism, more precisely the mathematical formula, works convincingly. With the shown approaches many inconsistencies were found. Nevertheless, even the second approach cannot completely validate whether the knowledge base corresponds to the experts' tacit and implicit knowledge. This would require a much larger and more heterogeneous selection of characters. Even more, tables 2, 3 and 4 exhibit that the experts' opinion is not completely stable over time. For these reasons, an evaluation against the experts' opinion is only expedient up to a certain point. To tackle this problem, the system should be extended with a learning capability and then be trained with data derived from a large group of real caregiving relatives. Furthermore, the system needs to be extended by a rule-based component, representing explicit knowledge about disease specific knowledge transfer.

Conclusions

The evaluation process described has shown some inconsistencies in the knowledge base. After resolving these, it can be assumed that both the inference mechanism and the knowledge base represent the implicit and tacit experts' knowledge to a good extent. Nevertheless, not all inconsistencies could be found in the knowledge base. Therefore, a strong need for a learning capability exits. Additionally, disease specific knowledge transfer could improve the system.

Acknowledgements

The evaluated knowledge-based system is part of the project "Mobile Care Backup" (MoCaB)[1], which is funded by the German Federal Ministry of Education and Research (grant number 16SV7472).

References

[1] T. Benson, Five Arguments against Interdisciplinary Studies, *Issues in Integrative Studies* **1** (1982), 38–48.

[2] D.N. Ford, and J.D. Sterman, Expert knowledge elicitation to improve formal and mental models. *Syst Dyn Rev* **14** (1998), 309–340.

[3] D.E. O'Leary, Methods of Validating Expert Systems, *Interfaces* **18** (1988), 72–79.

[4] K. Clarke, R. O'Moore, and et al., A methodology for evaluation of knowledge-based systems in medicine, *Artif Intell Med* **6** (1994), 107–121.

[5] C.J.M. Schölzel-Dorenbos, I. Draskovic, and et al., Quality of Life and Burden of Spouses of Alzheimer Disease Patients, *Alzheimer Dis Assoc Disord* **23** (2009), 171–177.

[6] D. Wolff, M. Behrends, and et al., Personalized Knowledge Transfer for Caregiving Relatives, *Stud Health Technol Inform* **247** (2018), 780–784.

[7] IEEE, *IEEE Standard for System and Software Verification and Validation,* IEEE, Piscataway, NJ, USA 2012.

Address for correspondence
Dominik Wolff. Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hanover Medical School Carl-Neuberg-Str. 1. D-30625 Hannover, Germany Dominik.Wolff@plri.de

[1] http://mocab-projekt.de/

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190354

Decision Support Tools for Drugs Prescription Process in a Hospital in Argentina

Valeria Zapico, Luciana Rubin, Soledad Díaz, Laura Gambarte, Romina Rebrij, Carlos Otero, Daniel Luna

Department of Health Informatics, Hospital Italiano de Buenos Aires, Ciudad de Buenos Aires, Argentina

Abstract

While medications can improve the health of patients, the prescription process is complex and prone to errors.

The structured medical order entry systems (CPOE) with clinical decision support (CDS) are increasingly implemented to improve patient safety, however the organizations that decide to implement them will have several challenges: understanding which classes of CDS can admit their systems, ensure that clinical knowledge is adequate and design tools for proper monitoring.

We share our experience of over ten years of development and implementation of clinical decision support tools during drugs prescription process and tools that have allowed us to monitor them correctly.

Keywords:

Decision Support Systems, Clinical; Knowledge Bases; Computerized Physician Order Entry

Introduction

A key process to improve drug safety is to implement computerized physician order entry systems (CPOE) that enable the incorporation of clinical decision support systems (CDS). With the implementation of electronic health records (EHR) throughout the world, the use of these CDS is increasing and enables the access to relevant patient information[1,2]. Health information technologies associated with a reduction in medication errors and adverse events[3-5] , improved patient care[6] and improved profitability for organizations[7] are considered potentially transformative and are being increasingly implemented [8,9].

When incorporating CDS into the drug prescription process, organizations should choose whether to develop knowledge bases (KB) or use commercial databases. Given that the resources and experience necessary for its development and implementation are available only to some large academic centers, most organizations choose to purchase KB from commercial suppliers [10].This action poses some challenges: a) high variability among the KB [11,12] for which some authors suggest that more than one should be used [13] b) the commercial KB tends to be very inclusive and place more emphasis on the amplitude of the coverage than on the clinical relevance or severity of the adverse events [7] and c) in some cases KB are not customizable and even when they are, most organizations lack the resources to do so effectively [14].

Many of the CDS that have shown utility are those that have been developed by the same institutions that use them, they have the capacity to elaborate knowledge, test interventions before implementing them and monitor its effectiveness over time [15].

In Argentina, the lack of infrastructure and the digital immaturity of health institutions are factors that have determined that CPOEs have not yet reached a sufficient degree of maturity and even the institutions that have it have not implemented CDS.

The aim of this work is to describe the components of a CDS fully developed in Hospital Italiano de Buenos Aires as well as the monitoring tools used to monitor them.

Methods

Hospital Italiano de Buenos Aires (HIBA) is a teaching hospital founded in 1853 in Buenos Aires, Argentina that works as a health network with 25 ambulatory centers, and 150 medical offices distributed along Buenos Aires city and its suburbs. HIBA has 750 beds and admits 45000 patients annually. The hospital handles 3 million outpatients each year. Since 1998 the Department of Health Informatics has designed, developed, and implemented a hospital information system (HIS) that manages health and administrative information from data capture to analysis. The HIS includes a web based, problem oriented, and patient centered EHR with a CPOE that allows the electronic prescription of drugs, connected with the pharmacy system, and integrated with the nurse care process.

The first CDS that were incorporated into the CPOE were: 1) duplicate prescription alert and 2) drug-drug interaction alert. In this case, in addition to considering the prescribed drug and those already existing in the pharmacological indications module, the CDS needs to access a KB, and since at that time there were no companies in the Argentine market that provided them in the Hospital, the decision was made to create a source of knowledge structured and controlled by international standards.

As a result of the validation of different sources of information, a reliable, up-to-date and Spanish-language KB was created with the main purpose of providing classified, structured and evaluated information on drug interactions to nourish the CDS and improve the safety of the drugs prescription process. After 6 years of implementation of the CDS, a process of analysis and redesign of the KB was initiated (see drug-drug interaction alert), followed by a redesign of the alerts using user-centered design techniques (UCD)[16]. At the same time, it was decided to add pharmacological alerts for drugs - allergy, teratogenic risk, maximum dose, etc., so adaptations and improvements were done in the KB to support and in some cases increase the specificity of the alerts (see Monitoring tools).

Currently a multidisciplinary team works to give continuity to this CDS project, conformed by a physician specialist in Health Informatics responsible for the coordination of the team, pharmacologists responsible for gathering the evidence and defining the content that is loaded in the KB, medical career students in charge of the data entry, pharmacists that create the functional requirements, test and authorize the

deployment of the necessary changes in the KB and two developers, one in charge of the maintenance of the BCF and another one of the pharmacological alerts that the CDS generates.

Results

A description of the CDS (chronologically ordered) implemented in the HIBA's EHR and the tools used for its monitoring are named below.

Alert for Duplicate Indication

This alert informs the physician that he is prescribing a drug that was already prescribed, a situation that generally occurs when several physicians provide care to the same patient or when prescriptions coexist for a drug to be administered in different ways. These unintentional duplications are the cause of toxicity adverse events that can be prevented [15]. However, since there are also situations in which duplication is intentional and valid (such as the indications of analgesics at regular intervals and additional doses in the presence of pain), when the HIBA implemented CPOE it did so with an interruptive alert that appears as a window emergent in the center of the prescription screen on which the prescriber must click "Accept" without having to justify continuing with the prescription.

Drug-Drug Interaction Alert

The frequent commercialization of new medicines, increasingly older populations with multiple comorbidities and polypharmacy with different specialists treating the same patient and where each of them may not be familiar with all the medications that their patients receive [16,17] are cited as factors that contribute to the frequency of interactions between drugs. These represent, not only a common cause of adverse reactions to medications but also of therapeutic failures. Regardless of the outcome, an important feature of drug interactions is that they are often preventable [17] in this way information systems that could retrieve all of the medications a patient is using to detect and remind doctors of IDDs are believed increasingly necessary to improve patient safety [18].

The redesign of HIBA's KB in 2006 allowed the creation of new functionalities that would later be taken into account when displaying alerts as: the categorization of active or passive according to whether they would manifest themselves actively interrupting the prescription workflow or not, requires monitoring feature to inform the prescribing physician about the clinical or laboratory parameters to be controlled when deciding to continue with the indication of the drugs and to evaluate the use of alternative drugs so that the alert promotes action through recommendations for alternative treatment [19].

With these changes implemented in the KB and after the redesign of the alerts mentioned above, it was determined that only drug-drug interactions with severity C, D or X, according to Lexicomp classification[20], would generate alerts: for drug-drug interactions severity C, the alerts would be non-disruptive blue light, for drug-drug interaction severity D, alerts would be disruptive orange and for interactions drug drug severity X, alerts would be non-disruptive red (see Figure 1).

Drug allergy alert

The verification of drug allergy is a very important feature in terms of patient safety. This type of intervention presents an alert when a physician prescribes a medication for which the patient has an allergy documented in their EHR [21]. It is very important that the patient's allergy data is stored in their EHR before prescribing medications. However, in some cases, it is not possible to know the allergic status of the patient at the time he enters the hospital. For this reason and in order to increase the number of registered allergic patients in the EHR, a circuit has been designed with the pharmacy service that reminds the physicians to register the allergic state of every patient who contacts the hospital.

When a physician prescribes a drug that matches the allergic patient status, an irruptive alert appears. The user has the option of cancelling the prescription and accept the alert or continuing the indication and ignore the alert, in the latter case they are requested to select a structured justification or enter one in free text (see monitoring tools).

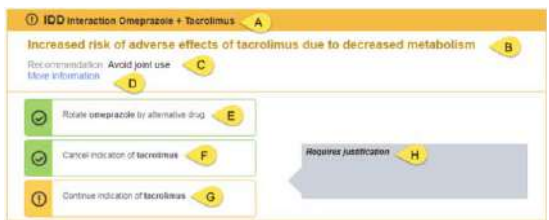


Figure 1 – IDD Alert Interface. (A) Interaction between Omeprazole and Tacrolimus severity D (B) risk of interaction; (C) main recommendation for the prescribing physician; (D) link to more information; (E) suggestion to rotate Omeprazole, already indicated, by Pantoprazole or Ranitidine that do not interact with Tacrolimus; (F) "cancel indication of Tacrolimus"; (G) "continue indication of Tacrolimus"; (H) required justification

Alert for Maximum Dose

Many preventable adverse drug reactions (ADRs) are caused by dosage errors in the prescription process. For this reason, it is proposed that CDSS linked to KB that contain the maximum recommended therapeutic doses (MRTD) can prevent the prescription of excessive doses and their consequences [22]. A particularity of the MRTD is that they can vary according to the route of administration [23], the age of the patients [24] and the unit of time (dose, day or accumulated) and these characteristics must be registered in the KB. It is important to note about this alert, that the selection of the generic is not enough to trigger it; the physician must select the dose so that it is contrasted with the KB record. Figure 2 shows the interface of an alert for an exceeded dose of Acetaminophen for an adult patient, which includes an action-oriented button to adjust posology.

Teratogenic Drug Alert in Pregnancy

There are drugs that are highly teratogenic and should never be administered to a woman who is or may be pregnant; many others who are relatively contraindicated to a greater or lesser extent[15]. Since 2015 when a physician prescribes a teratogenic risk drug in the HIBA's CPOE to a pregnant woman, an alert is triggered. The alert's input are: a) the existence of a pregnancy problem registered in the patient's EHR and b) the teratogenic risk of the prescribed drug documented in the KB. As in the case of the drug-allergy alert, this CDS also depends on elements present in the EHR problems' list[25] that are usually incomplete or have outdated information [26]. To solve the lack of pregnancy status registration, it was decided that the opening of an obstetric chart would automatically load the problem "Pregnancy" in the EHR of the patients. A set of controlled terms (derived

from delivery, caesarean section, abortion or puerperium) generates the automatic resolution of the problem when they are created in the system by physicians, increasing the specificity of this alert.

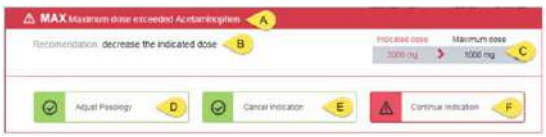


Figure 2 – Alert Interface for Maximum Dose (A) Maximum exceeded dose of Acetaminophen; (B) recommendation to reduce the dose; (C) comparison between indicated dose, 2000 mg, and maximum dose, 1000 mg; (D) adjust posology leads the user to modify the prescribed dose (E) "cancel indication of Acetaminophen"; (F) "continue indication of Acetaminophen" in this case the user must justify choosing between structured options or free text.

Pharmacogenomic Alerts

Since 2014 when the physician prescribes a drug, the execution of the process begins, the inference engine, which retrieves the genetic information from the EHR, infers if there is any risk to the patient in case the drug that is being prescribed is supplied. If there is a risk, the inference engine issues an alert that informs the user about the risk, allowing him to cancel or continue with the prescription and also allows him to send a query through an email to the Clinical Pharmacology section [27].

Frequent Posologies

Frequent posologies are defined as complete orders of pre-written medications that include dosage, route and frequency of administration. The supply of these lists of posologies improve the users system acceptance and reduces errors [15,28,29].

For the development of frequent posologies implemented in the CPOE of the HIBA in 2015, an analysis of all the medications that had been prescribed during the previous year and had a rate of use greater than 80%. This group defined the posologies that would later be reviewed by a group of pharmacists. Each generic could have more than one recommended dose but not more than 3 and the physician always has the opportunity to create his own posology if he considered, or even use a recommended one and edit it. The information on each drug is contained in the KB discriminated by generic and by age group.

Drug-Food Interaction Alert

It is important not to forget the interactions between drugs and foods that can inadvertently modify the effect of drugs. Most clinically relevant interactions are caused by food-induced changes in the bioavailability of the drug (changes in absorption or metabolism) that correlates with the clinical effect of most drugs[30,31].

In 2016, this CDS was implemented to warn physicians about these interactions. The alert is generated when a medicine that is being indicated has a record of contraindication with some food in the KB. The alert it is not interruptive, and it was designed only to raise awareness of the possible interaction in prescribers.

Drug-Laboratory Test Interaction Alert

To minimize the risk that certain medications may have, it is recommended to perform laboratory tests before their administration and in many cases also at regular intervals during a treatment. It is known that not taking into account these physiological parameters can result in morbidity and even mortality for the patients [32,33]. For this reason in 2016 it was decided to develop and implement this CDS (see monitoring tools).

The first laboratory results contemplated for the creation of the alert were potassium and calcium, due to their relevance.

The drug-laboratory interaction (DLI) alert inputs are: a) a result of serum measurement out of range; or b) a blood calcium measurement result out of range, and b) the prescription of a drug that can decrease more these electrolytes, in the case of values lower than the minimum (see Figure 3) or increase them (in the case of values greater than maximum) or that without having the potential to affect the results, its pharmacological effect can be modified (for example, the case of patient with a result of measurement of Potassium less than 3.5 millimole / liter and prescription of Digoxin).

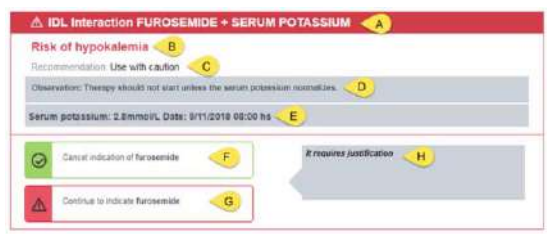


Figure 3 – Alert Interface by DLI. (A) DLI between Furosemide and the test Serum potassium measurement; (B) risk of interaction; (C) (D) recommendation and observation for the prescribing physician; (E) result of serum potassium measurement and date of it (F) "cancel indication of furosemide"; (G) "continue indication of furosemide"; (H) required justification

Tools Developed for CDS Monitoring

Monitoring is often included as a success factor for CDS [6] because it permits to iteratively refine the interventions implemented and thus improve their use and impact. In the HIBA the process of monitoring is covered through boards, reports and review of RME.

Dashboards

Dashboards enable a global view of what happens with the alerts. A general dashboard shows users the distribution of alerts by type, behavior of the prescribers (continue, cancel, etc) and justifications to continue. Also, five specific dashboards (one for each alert) are used to represent the singularities, so, for example, the alerts for allergy-drug show the most frequent allergens that have generated alerts; and the IDD dashboard shows the drug drug pairs that have been prescribed.

Reports

Although the dashboards shows what is happening in a given domain, it is not within their scope to show what happens in each specific situation. Specific reports enable the team to identify, for each alert activated, the prescribing user, their medical specialty, the drug with the intention of being indicated, the action taken after the appearance of the alert, the reason why it was ignored and the patient in which the drug was prescribed. Also for IDD alerts, the reports show the drugs that form each pair and for DLI the result of the value of Potassium or Calcium at the moment the alert was generated.

Revision of Electronic Medical Records

Based on the information provided by the reports, the review of medical records helps the team to search for adverse effects (in ignored alerts), false positives, patterns of use by specialty, and unintended consequences. From these monitoring actions the team was able to realize that in the development of the alert by DLI it had not been accounted for the route of administration of the indicated drug, reason that generated many false positives and for which there was the need to add a field in the KB to record this data. And from the analysis of free text justifications for the allergy alert we noticed that 30% of the cases were due to indications from the Allergy service to submit the patients to desensitization tests, which is why this reason was added as a structures option.

Discussion

This paper describes the pharmacological CDS embedded in HIBA's homegrown HIS.

Undoubtedly, the decision to incorporate CDS during prescription is a strategy used by organizations that seek to increase safety in the quality of care. Being able to have a full integration between the CPOE, the EHR and CDS is the best scenario that allows a comprehensive approach to improve the quality of care provided [34,35]. The biggest challenge when implementing CDS is to achieve and maintain a correct balance between useful alerts and overexposure [36] and a good way to achieve this is to keep the KB updated to ensure that the records contained in them are correct, consistent, complete and current [37]. The commitment is not trivial and requires substantial resources [38], in our case it has taken many years of a great team work to achieve a truly multidisciplinary approach. However, even when the clinical content is adequate, the lack of completeness and / or updating of the data in the medical records does not allow for the specificity that we expect from the CDS.

On the other hand, the only way to make decisions is through the use of information, which is why the data provided by our monitoring tools are essential to know if the interventions are accepted by users, if their actions are modified and finally, if our interventions are impacting the quality of care and the health of our patients.

Conclusions

Healthcare organizations wishing to implement integrated CDS to CPOE and HCEs should be prepared to take on three major challenges: understand what kinds of CDS their systems can support, manage knowledge and design tools that allow them to know how it works.

References

[1] W.E. Haefeli, and H.M. Seidling, [Electronic decision support to promote medication safety], *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. **61** (2018) 271–277.

[2] Y. Zenziper Straichman, D. Kurnik, I. Matok, H. Halkin, N. Markovits, A. Ziv, A. Shamiss, and R. Loebstein, Prescriber response to computerized drug alerts for electronic prescriptions among hospitalized patients, *Int. J. Med. Inform.* **107** (2017) 70–75.

[3] D.C. Radley, M.R. Wasserman, L.E. Olsho, S.J. Shoemaker, M.D. Spranca, and B. Bradshaw, Reduction in medication errors in hospitals due to adoption of computerized provider order entry systems, *J. Am. Med. Inform. Assoc.* **20** (2013) 470–476.

[4] E. Ammenwerth, P. Schnell-Inderst, C. Machan, and U. Siebert, The effect of electronic prescribing on medication errors and adverse drug events: a systematic review, *J. Am. Med. Inform. Assoc.* **15** (2008) 585–600.

[5] T.K. Nuckols, C. Smith-Spangler, S.C. Morton, S.M. Asch, V.M. Patel, L.J. Anderson, E.L. Deichsel, and P.G. Shekelle, The effectiveness of computerized order entry at reducing preventable adverse drug events and medication errors in hospital settings: a systematic review and meta-analysis, *Syst. Rev.* **3** (2014) 56.

[6] E. Yoshida, S. Fei, K. Bavuso, C. Lagor, and S. Maviglia, The Value of Monitoring Clinical Decision Support Interventions, *Appl. Clin. Inform.* **9** (2018) 163–173.

[7] R.M. Reichley, T.L. Seaton, E. Resetar, S.T. Micek, K.L. Scott, V.J. Fraser, W. Claiborne Dunagan, and T.C. Bailey, Implementing a Commercial Rule Base as a Medication Order Safety Net, *J. Am. Med. Inform. Assoc.* **12** (2005) 383–389.

[8] D. Blumenthal, and M. Tavenner, The "Meaningful Use" Regulation for Electronic Health Records, *N. Engl. J. Med.* **363** (2010) 501–504.

[9] Z. Morrison, A. Robertson, K. Cresswell, S. Crowe, and A. Sheikh, Understanding Contrasting Approaches to Nationwide Implementations of Electronic Health Record Systems: England, the USA and Australia, *J. Healthc. Eng.* **2** (2011) 25–42.

[10] S. Phansalkar, A. Wright, G.J. Kuperman, A.J. Vaida, A.M. Bobb, R.A. Jenders, T.H. Payne, J. Halamka, M. Bloomrosen, and D.W. Bates, Towards meaningful medication-related clinical decision support: recommendations for an initial implementation, *Appl. Clin. Inform.* **2** (2011) 50–62.

[11] A.I. Vitry, Comparative assessment of four drug interaction compendia, *Br. J. Clin. Pharmacol.* **63** (2007) 709–714.

[12] L.M. Wang, M. Wong, J.M. Lightwood, and C.M. Cheng, Black box warning contraindicated comedications: concordance among three major drug interaction screening programs, *Ann. Pharmacother.* **44** (2010) 28–34.

[13] P.L. Smithburger, S.L. Kane Gill, N.J. Benedict, B.A. Falcione, and A.L. Seybert, Grading the Severity of Drug-Drug Interactions in the Intensive Care Unit: A Comparison Between Clinician Assessment and Proprietary Database Severity Rankings, *Ann. Pharmacother.* **44** (2010) 1718–1724.

[14] A. Wright, G.J. Kuperman, A.J. Vaida, A.M. Bobb, R.A. Jenders, T.H. Payne, J. Halamka, M. Bloomrosen, D.W. Bates, and S. Phansalkar, Towards Meaningful Medication-Related Clinical Decision Support: Recommendations for an Initial Implementation, *Appl. Clin. Inform.* **02** (2011) 50–62.

[15] G.J. Kuperman, A. Bobb, T.H. Payne, A.J. Avery, T.K. Gandhi, G. Burns, D.C. Classen, and D.W. Bates, Medication-related clinical decision support in computerized provider order entry systems: a review, *J. Am. Med. Inform. Assoc.* **14** (2007) 29–40.

[16] D.R. Luna, D.A. Rizzato Lede, C.M. Otero, M.R. Risk, and F.G.B. de Quirós, User-centered design improves the usability of drug-drug interaction alerts: Experimental comparison of interfaces, *J. Biomed. Inform.* **66** (2017) 204–213.

[17] Y. Böttiger, K. Laine, M.L. Andersson, T. Korhonen, B. Molin, M.-L. Ovesjö, T. Tirkkonen, A. Rane, L.L. Gustafsson, and B. Eiermann, SFINX-a drug-drug interaction database designed for clinical decision support systems, *Eur. J. Clin. Pharmacol.* **65** (2009) 627–633.

[18] Website, (n.d.). The Leapfrog Group website http://www.leapfroggroup.org/ Accessed January 31, 2017 (accessed November 21, 2018).

[19] A.J. Grizzle, M.H. Mahmood, Y. Ko, J.E. Murphy, E.P. Armstrong, G.H. Skrepnek, W.N. Jones, G.P. Schepers, W.P. Nichol, A. Houranieh, D.C. Dare, C.T. Hoey, and D.C. Malone, Reasons provided by prescribers when overriding drug-drug interaction alerts, *Am. J. Manag. Care*. **13** (2007) 573–578.

[20] Lexi-Interact Data Fields, (n.d.). http://webstore.lexi.com/Information/Product-Information/Lexi-Interact-Fields (accessed November 25, 2018).

[21] G.J. Kuperman, T.K. Gandhi, and D.W. Bates, Effective drug-allergy checking: methodological and operational issues, *J. Biomed. Inform.* **36** (2003) 70–79.

[22] H.M. Seidling, A. Al Barmawi, J. Kaltschmidt, T. Bertsche, M.G. Pruszydlo, and W.E. Haefeli, Detection and prevention of

prescriptions with excessive doses in electronic prescribing systems, *Eur. J. Clin. Pharmacol.* **63** (2007) 1185–1192.

[23] C.E. Bekes, Medication errors and adverse drug events in an intensive care unit: Direct observation approach for detection, *Yearbook of Critical Care Medicine*. **2007** (2007) 316–317.

[24] I.H. Bartelink, C.M.A. Rademaker, A.F.A.M. Schobben, and J.N. van den Anker, Guidelines on paediatric dosing on the basis of developmental physiology and pharmacokinetic considerations, *Clin. Pharmacokinet.* **45** (2006) 1077–1097.

[25] A. Wright, E.S. Chen, and F.L. Maloney, An automated technique for identifying associations between medications, laboratory results and problems, *J. Biomed. Inform.* **43** (2010) 891–901.

[26] H.C. Szeto, R.K. Coleman, P. Gholami, B.B. Hoffman, and M.K. Goldstein, Accuracy of computerized outpatient diagnoses in a Veterans Affairs general medicine clinic, *Am. J. Manag. Care*. **8** (2002) 37–43.

[27] B.R. Goldspiel, W.A. Flegel, G. DiPatrizio, T. Sissung, S.D. Adams, S.R. Penzak, L.G. Biesecker, T.A. Fleisher, J.J. Patel, D. Herion, W.D. Figg, J.J.L. Lertora, and J.W. McKeeby, Integrating pharmacogenetic information and clinical decision support into the electronic health record, *J. Am. Med. Inform. Assoc.* **21** (2014) 522–528.

[28] T.S. Lesar, Factors Related to Errors in Medication Prescribing, *JAMA*. **277** (1997) 312.

[29] T.K. Gandhi, S.N. Weingart, A.C. Seger, J. Borus, E. Burdick, E.G. Poon, L.L. Leape, and D.W. Bates, Outpatient prescribing errors and the impact of computerized prescribing, *J. Gen. Intern. Med.* **20** (2005) 837–841.

[30] R. Bushra, N. Aslam, and A.Y. Khan, Food-drug interactions, *Oman Med. J.* **26** (2011) 77–83.

[31] N. Corti, and A.B. Taegtmeyer, [Clinically important food-drug interactions: what the practitioner needs to know], *Praxis* . **101** (2012) 849–855.

[32] R.L. Howard, A.J. Avery, P.D. Howard, and M. Partridge, Investigation into the reasons for preventable drug related admissions to a medical admissions unit: observational study, *Qual. Saf. Health Care*. **12** (2003) 280–285.

[33] Y.-F. Chen, A.J. Avery, K.E. Neil, C. Johnson, M.E. Dewey, and I.H. Stockley, Incidence and possible causes of prescribing potentially hazardous/contraindicated drug combinations in general practice, *Drug Saf.* **28** (2005) 67–80.

[34] G.J. Kuperman, R.M. Reichley, and T.C. Bailey, Using commercial knowledge bases for clinical decision support: opportunities, hurdles, and recommendations, *J. Am. Med. Inform. Assoc.* **13** (2006) 369–371.

[35] Website, (n.d.). The Leapfrog Group website http://www.leapfroggroup.org/. (accessed November 24, 2018).

[36] N.R. Shah, A.C. Seger, D.L. Seger, J.M. Fiskio, G.J. Kuperman, B. Blumenfeld, E.G. Recklet, D.W. Bates, and T.K. Gandhi, Improving Acceptance of Computerized Prescribing Alerts in Ambulatory Care, *J. Am. Med. Inform. Assoc.* **13** (2006) 5–11.

[37] D.F. Sittig, A. Wright, L. Simonaitis, J.D. Carpenter, G.O. Allen, B.N. Doebbeling, A.M. Sirajuddin, J.S. Ash, and B. Middleton, The state of the art in clinical knowledge management: an inventory of tools and techniques, *Int. J. Med. Inform.* **79** (2010) 44–57.

[38] H. Tilson, L.E. Hines, G. McEvoy, D.M. Weinstein, P.D. Hansten, K. Matuszewski, M. le Comte, S. Higby-Baker, J.T. Hanlon, L. Pezzullo, K. Vieson, A.L. Helwig, S. -M. Huang, A. Perre, D.W. Bates, J. Poikonen, M.A. Wittie, A.J. Grizzle, M. Brown, and D.C. Malone, Recommendations for selecting drug-drug interactions for clinical decision support, *Am. J. Health. Syst. Pharm.* **73** (2016) 576–585.

Address for correspondence

Luciana Rubin

Department of Health Informatics, Hospital Italiano de Buenos Aires

Juan Domingo Perón 4190, C1181ACH.

Ciudad de Buenos Aires, Argentina

email: luciana.rubin@hospitalitaliano.org.ar

This page intentionally left blank

III. Enabling Precision Medicine and Public Health

This page intentionally left blank

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190356

A Dashboard for Latent Class Trajectory Modeling: Application in Rheumatoid Arthritis

Beatrice Amico[a], Arianna Dagliati[b,c], Darren Plant[d], Anne Barton[d,e], Niels Peek[b,e], Nophar Geifman[b]

[a] *Department of Computer Science, University of Verona, Verona, Italy*
[b] *Centre for Health Informatics, University of Manchester, Manchester, United Kingdom*
[c] *Manchester Molecular Pathology Innovation Centre, University of Manchester, United Kingdom*
[d] *Division of Musculoskeletal & Dermatological Sciences, University of Manchester, United Kingdom*
[e] *NIHR Manchester Biomedical Research Centre, University of Manchester, United Kingdom*

Abstract

A key trend in current medical research is a shift from a one-size-fit-all to precision treatment strategies, where the focus is on identifying narrow subgroups of the population that would benefit from a given intervention. Precision medicine will greatly benefit from accessible tools that clinicians can use to identify such subgroups, and to generate novel inferences about the patient population they are treating. We present a novel dashboard app that enables clinician users to explore patient subgroups with varying longitudinal treatment response, using latent class mixed modeling. The dashboard was developed in R Shiny. We present results of our approach applied to an observational study of patients with moderate to severe rheumatoid arthritis (RA) on first-line biologic treatment.

Keywords:

Precision Medicine, Medical Informatics Applications, Rheumatoid Arthritis

Introduction

Complexity and variability of patients' trajectories in response to treatment poses significant challenges in many medical fields, especially in those requiring long-term care [1-4]. While it is possible to define high-level clinical phenotypes, the prediction of outcomes with respect to interventions is complicated by the high variability in responses [5-7]. One of the main trends in current medical research is a shift from a one-size-fit-all to precision approaches, where the focus is on identifying narrow subgroups of the population who would benefit from a given intervention. In order to achieve this goal of precision medicine, the analysis of longitudinal data in clinical research is becoming increasingly important. Longitudinal analytics methods, their exploitation in the context of clinical decisions, and their translation into clinical practice through accessible tools, represents a potential for enabling precision healthcare [8].

The plethora of rich data routinely collected in clinical practice and in clinical trials captures underlying information that could allow for identification of specific subgroups of patients that may in turn predict if these will benefit from specific treatments. Despite this rapid growth in available data, and advancement in machine learning methods, the application of these in medical research and routine clinical practice is still very difficult and impeded by several factors. One of these issues revolves around the limited involvement of clinicians in the discovery process and a missed link between data-driven discovery and their application in real environments. The development of informatics applications that can introduce data-driven discoveries directly into clinical practice is a current unmet necessity in medical informatics. Precision medicine will greatly benefit from accessible tools that clinicians can use to identify groups of patients that respond differently to therapies, and to generate novel inferences about the patient population they are treating. Source data and visual analytics can improve diseases' management by enacting the implementation of the learning health care system cycle: the introduction of clinical data in outcomes research, together with the translation of research findings into care, can support decision-making with the realization of precision medicine [6].

In rheumatoid arthritis, clinicians already have some classification criteria but are developed using subjective thresholds and typically applied to one or two follow-up time-points. Our approach aims at removing subjectivity and making use of all available data. One approach for subgroup discovery in longitudinal data is latent class mixed modeling (LCMM), a type of latent class analysis that is increasing in popularity as a powerful method for discovering meaningful and differing subgroups with homogeneous patterns of change over time [9-11]. Here we revise the LCMM analysis framework proposed in [12], where authors detail the methodological steps for off-line analysis, adding the necessary steps to allow on-the-fly analysis by clinicians. More specifically, we developed a dashboard tool that allows clinicians to perform several key steps of an LCMM analysis themselves; for example, the clinician can refine and redefine data-driven models, identify which of these models are clinically plausible and relevant, test specific hypotheses, and ultimately translate the results into clinical practice.

We present results of our approach applied to an observational study of moderate to severe rheumatoid arthritis (RA) patients about to commence treatment with a biologic drug. RA is a chronic, systematic inflammatory joint disease of autoimmune nature [13]. RA is a heterogeneous disease that is classified using a set of clinical factors. Patients with similar clinical features in early disease may go on to experience a very different disease course or response to medication [14]. Recently, the treatment of RA was improved by the introduction of biologic disease-modifying antirheumatic drugs which target elements of the immune system, these are typically reserved for those with an inadequate response to non-biologic disease modifying antirheumatic drugs (DMARDs) [13,15]. Clinical responses and efficacy of biologics vary largely among different individuals. Based on experience and observations made in clinic, although defined in a non-methodical fashion, response criteria are accepted (EULAR classification in non/intermediate/poor-response) [16], and clinicians agree on stratifying patients with respect to the response to the therapy as: primary non-responders which never responded or failed to respond within the first 3 months, secondary non-responders who are patients that initially responded but then loose response after 3 months, and good responders. Secondary noon-response

is seen in a significant minority of primary responders. Reasons for this could include: patients stop taking drug (i.e. feel better and are less motivated), patients develop anti-body against the therapeutic (immunogenicity), as the TNF/TNF pathway is brought under control a second inflammatory pathway may kick in and cause the disease to flare. In RA, a precision medicine approach will allow better understanding of which patients respond to specific therapies within a given time window, as well as improved disease monitoring. In order to understand the underlying mechanisms of response, well powered biomarker discovery studies are needed. A first step is to this process is to better define the phenotype so that biomarkers can be contrast across meaningful patient groupings and include all patients across multiple time-points.

Here we describe an accessible tool to perform LCMM analysis in RA. The goal is to provide clinicians and medical researchers with the possibility to automatically identify different responders to biological treatments over time. The tool relies on a dashboard-based approach to translate the data-driven retrieved models into medical inference and practice [17,18].

Methods

Latent Class Trajectory Modeling

We applied LCMM with the goal of identifying subgroups of RA patients with distinct responses to different types of biological treatment over time. The implementation of the dashboard is based on the framework proposed in [12], which includes eight steps: 1: definition of a scoping model; 2: refinement of the number of classes; 3: refinement of the model structure on the basis of fixed-effects through random-effect specification; 4: model adequacy assessment; 5: graphical presentations; 6: classes discrimination; 7: clinical characterisation and plausibility; and, 8: sensitivity analysis. Our developed dashboard implements the framework as an interactive tool aimed to give the possibility to clinicians to carry out, independently, each step of the analysis. In particular, clinicians can perform steps related to graphical representation, classes' discrimination and clinical plausibility, independently from the first steps of the analysis. Indeed, for this application we initially performed steps 1 through 4 to retrieve a *favoured model*, which we used as the default model in the dashboard. Dashboard's users can exploit this favoured model to perform the last steps of the analysis (i.e. visualize classes trajectories and compare clinical characterises in the classes), but they can also change the model parameters (e.g. the number of classes, fixed-effects) and re-run the whole analysis. Below we describe in detail the steps that we implemented in the dashboard.

Step 1: Scoping Model

We used a maximum likelihood approach to fit the model through the 'lcmm' function from the R package lcmm [19]. The function estimates mixed-effect models and latent class mixed-effect models for different types of longitudinal outcomes. We built models for the entire cohort, using all the possible combination of the available variables. While the scoping model is based on the entire cohort, the dashboard provides the option to stratifying patients on the basis of the class of treatments and perform the analysis on these subsets separately.

Step 2: Number of Classes

We tested the scoping model to determine the optimal number of classes K: 1-10 number of classes. The lcmm function provides Akaike (AIC) and Bayesian information criterion (BIC) as model fit indices. The K number of classes chosen was primarily based on BIC as suggested in [12]. AIC was considered to confirm or clarify the empirical solution. The dashboard presents as a default parameter the number of classes with the lowest BIC and gives the option of changing this number from a fixed range (1-10 classes).

Step 3 Model Refinement

We further refined the model using the model with the lowest BIC derived in step 2, considering the possibility of using linear or quadratic specification of time (days) as the random-effect, logarithm transformed values of our outcome measure, and linear or quadratic link functions to model the longitudinal outcome. The dashboard allows to perform the LCMM analysis with linear or quadric time, transforming or not the outcome and with different link functions.

Step 4 Model Adequacy Assessment

For each subject, we calculated the posterior probability of being assigned to each trajectory class and exclusively assigned the individual to the class with the highest probability. Average maximum posterior probability of assignments above 70% in all classes was considered acceptable. We ensure that each class includes at least 10% of the initial population, otherwise the model is discharged.

Steps 1-4 allow a favoured model structure to be selected using the lowest BIC value and satisfactory values from the model adequacy assessments. The favoured model is used to set the default parameters. Although, as already specified, the dashboard allows to modify the parameters and perform the analysis from scratch.

Step 5 Graphical Presentations and Class Separation

We assessed the design choices according to the clinical relevance for the specific application in RA. Severity RA is estimated using the Disease Activity Score in 28 joints (DAS28). It is an index which combines different scores: the count of the 28 swollen joints and 28 tender joints, the C-reactive protein (CRP) which levels rise in blood in response to inflammation, and the Visual Analog Scale (VAS) of patient's general health. The results of the model are visualized with the DAS28 trajectories of classes over time as well as with all four DAS28 components' individual trajectories. Differently from the original LCMM framework, where classes' discrimination was assessed by degrees of separation, this step is embedded in the graphical representation of the trajectories to simplify the tool usability by final users. The trajectories of the subjects belonging to each discovered latent class are represented with locally weighted scatter plot smoothing method (LOESS) and include confidence intervals. This step allows clinicians to independently clarify the meaning of the presented solutions and to revise the assumptions in performing latent class mixed modeling that might be context dependent.

Step 6 Clinical Characterisation and Plausibility

To assess the clinical meaningfulness of the resulting trajectories/classes, the dashboard allows for comparison of relevant clinical characteristics in the discovered latent classes through a graphical representation with violin plots (for continuous variables) and bar plot (for categorical variables).

Cohort and variable description

The data we used to perform the analysis were derived from a prospective cohort study, BRAGGSS (Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate). BRAGGSS is a UK based study including over 50 recruiting centres. Patients participate in the BRAGGSS study are followed for 12 months, during which they are assessed at baseline and three follow-up visits, at 3, 6 and 12 months [20]. From the entire cohort, we selected 1,531 patients with at least one measure of DAS28 recorded. The baseline characteristics of the total patient population are shown in Table 1. Considering the

different types of treatment, patients are stratified to three groups: biological drugs, biological drugs plus DMARDs (excluding Methotrexate), and finally biological drugs plus DMARDs (including Methotrexate). Methotrexate (MTX) is the first-line therapy for RA.

Table 1 – Cohort characteristics

Treatment group	Only Bio	Bio and DMARD	Bio and MTX
Number of subjects	275	248	1006
Gender (%)			
Male	21	21	26
Female	79	79	73
Age at baseline	65,5	65,2	62,7
(Mean and SD)	12,6	11,2	12,6
DAS 28 at baseline	4,19	4,38	4,18
(Mean and SD)	1,22	1,23	1,24
BMI at baseline	29,8	29,2	30,0
(Mean and SD)	19,7	10,8	15,2

Dashboard Implementation

The dashboard was built in response to the interest of clinicians and medical researchers for having a tool that they can independently use to identify different response trajectories of DAS28 during the first 12 months of treatment with biological drugs. To assess the usability of the tool and the clarity of the analysis to end-point users, we organized biweekly meetings with RA specialist clinicians, during which we discussed each step of the analysis and design of the dashboard according to clinicians' needs.

Dashboard Architecture

We implemented the dashboard using Shiny [https://shiny.rstudio.com]. Shiny is an open source R package to build interactive web application on the basis of R scripts. Shiny applications have two components, a user interface (UI) object and a server function, which are passed as arguments to the shinyApp function that creates a Shiny app object from this UI/server pair. The user interface object controls the layout and appearance of the application (in our case the Dashboard graphical user interface (GUI)). The server function contains the instructions to run the analyses (The Analysis Engine that performs the LCMM analysis). Finally, the shinyApp function creates Shiny objects from an explicit UI/server pair.

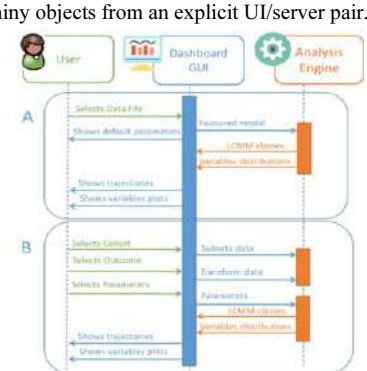


Figure 1 – UML Sequence Diagram.

Dashboard Functionalities

Figure 1 describes the system's functionalities along with the main actions of a user, dashboard GUI and analysis engine. Panel A shows the actions needed to run the favoured model, panel B depicts the actions needed to refine and re-run the model. The first user's action is to select a properly formatted text file (see Supplementary), the default parameters shown in the dashboard GUI are the ones used to run and present the favoured model. The analysis engine perform the LCMM on the basis of these parameter, assign each subject to the most probable class and sends back the results to the dashboard as DAS28 trajectories and variables values in the classes. The dashboard GUI shows the graphs of the trajectories and variables distributions. The user can re-run the whole analysis selecting a different cohort of patients (i.e. only patients treated with a specific drug regiment), outcome (i.e. log transformed DAS28) and different parameters, for example a different number of classes, as shown in Figure 1, panel B. The last set of actions to needed perform the analysis and present the results are then identical as in the approach in presented in panel A.

Results

LCMM model

In RA, clinicians expect to encounter three types of patients: primary non-responders (flat trajectories), secondary non-responders (trajectories with decreasing trends followed by increasing trends), and good responders (trajectories with decreasing trends). The favoured model selected within the first four steps of the analysis identified four classes. This is the default model shown by the dashboard, which has been discussed by clinicians on the basis of different clinical characteristics and DAS28 components. Our LCMM analysis was able to recognize a fourth group of responders: secondary good responders (trajectories with increasing trends followed by decreasing trends)

Dashboard Interface

Figures 2 and 3 depict the Dashboard GUI's results pane: DAS28 trajectories and their components in Figure 1, and demographic and clinical values distributions in Figure 2.

The main page of the Dashboard contains a panel where the user can choose the input dataset. The "View Data" button allows to visualize the dataset on which we are going to run the analysis. If the user doesn't modify any parameter, the favourite model parameters are set by default. The user starts the analysis by pressing the "Best model" button and receives a message when the analysis is complete. The same panel allows users to select a different subset of the dataset with the dropdown menus "Choose the outcome" and "Choose the treatment group", and to preview the filtered and transformed dataset. The rest of the dropdown menus and check boxes allow to set the lcmm parameters: covariates of the model, time effect, number of classes and link function. Pressing the "Start the analyses" button, the lcmm analysis starts. If all the required variables have been correctly selected, the user receives feedback that the analysis has started, otherwise he/she receives an error message which advises to complete the form in the correct way.

The first panel (Figure 2) visualizes the trajectories of the Das28, and their components. The second panel is accessible via the "Compare classes" tab and shows the variables distributions. Continuous variables: age, BMI and the health assessment questionnaire disability index (HAQ score), are presented with violin plots; categorical ones are presented with bar plots. By clicking the "Reset form" button, the user can reset the dashboard to the initial state and restart the analysis.

Qualitative Assessment of the Tool

The qualitative assessment of the dashboard's functionalities and usability has been carried out through 13 meetings where we discussed the analyses, functionalities and interface of the dashboard. We organised bi-weekly meetings with clinicians and separately, 4 meetings with a technical group composed of

experts in medical informatics. The first bi-weekly meetings consisted in preliminary sessions where we defined the key functionalities for the dashboard. First of all, we clarified the interest of clinicians in having a tool to perform this type of analysis. Secondly, we discussed expectations from the dashboard. Over several weeks, we refined the dashboard according to the feedback we received. For example, one request was to be able to visualize the trajectories of DAS28 and its components, especially for VAS and the count of tender joints. This is because RA has significant implications for patient quality-of life and increased psychological symptoms. Depression and anxiety have implications for disease activity primarily due to their influence on tender joints and patient global assessment. In addition, the capability to visualize the distribution of the type of biological drugs in each sub-group was also requested. This is because anti–tumor necrosis factor (anti-TNF) is the most common biologic therapy, but sometimes patients fail to respond on this initial anti-TNF therapy. For clinicians it is important to visualize the distribution of the type of biologic drug in each subgroup, in order to better assess the trajectories of DAS28. When we finalized the tool, we presented it at our internal meeting. We received positive feedback in terms of usability and the results of the lcmm analysis. The main refinement suggested was to add the option of visualizing the number of patients for each class, in order to create awareness in the classification that we obtained from the analysis; this has now been implemented within the dashboard.

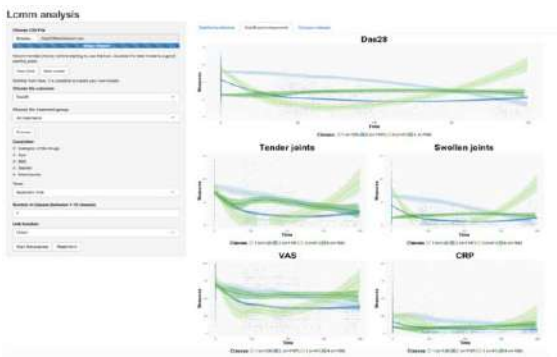


Figure 2 – Latent Classes' Trajectories: Das28 and its Components

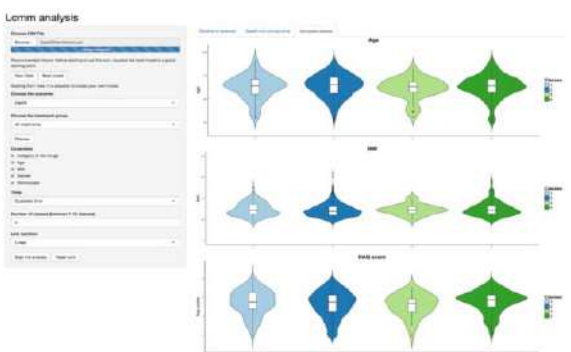


Figure 3 – Variable Distribution in the Latent Classes

Discussion

We present here a dashboard for performing latent class mixed-modeling analysis in the context of RA. The presented results and application case study are based on data from the BRAGGS multi-center study. To the best of our knowledge, this the first example of an accessible tool that enables lcmm analyses to be carried out, along with graphical representation of the results, specifically intended for a clinician and medical research user base. The dashboard provides clinicians the possibility of generating novel hypotheses regarding treatments responses, identify subgroups of patients that respond differently over time, and to assess results by comparing the mined groups with demographic and clinical features. The dashboard presents an initial, but crucial step towards enabling the translation of data-driven approaches into medical practice bettering the definition of phenotypes and supporting precision medicine.

Alongside the aforementioned contributions, several limitations of the study need to be recognised. While the graphical comparison of clinical characteristics and covariate serves to assess the clinical plausibility of the discovered classes, they might be biased. Additional methods like multinomial regression or corrections for measurement error in the classification of individuals to reduce bias could reduce potential biases. The dashboard has been evaluated by qualitative assessments through meetings with clinicians and medical informatics experts. While this approach contributes to implement a tool that responds to real-world and practical needs, a structured and quantitative usability validation is likely to further benefit the development and implementation of such a tool. Once the usability of the tool is assessed, further technical and methodological enhancements can be applied. In future versions, it may be beneficial to create a direct link between the dashboard and the BRAGGS database. Discriminative and sensitivity analysis can also be added as final steps for analyses. The BRAGGS dataset includes data on adherence to treatments and proteomics analysis performed on a subset of the cohort. Future work could focus on including these and other data, as an additional approach for subgroup validation and biomarker discovery.

Conclusions

While the initial goal of developing a tool to perform lcmm for precision medicine is situated within the scope of the current dashboard, the functionality scope can be extended. The application of lcmm analyses has shown promising results in several medical fields [9,10], and the system can be extended, generalised, and applied in any of these. Furthermore, this approach would also facilitate public and patient involvement/engagement in medical research. Accessible results will encourage involvement in study design and interpretation/dissemination of study findings through various patient groups. While more work is needed to determine the real impact of the use of such systems in medical practice, dashboard frameworks [17,18] can work as a bridge between different clinical fields, machine learning approaches and precision medicine, creating the possibility of real improvement in clinical research and practice.

Acknowledgements

This work was supported by the Medical Research Council and the Engineering and Physical Sciences Research Council grant MR/N00583X/1 "Manchester Molecular Pathology Innovation Centre (MMPathIC): bridging the gap between biomarker discovery and health and wealth. Support was also provided by the NIHR Manchester Biomedical Research Centre and Versus Arthritis (grant ref 20385).

References

[1] A. Giannoula, A. Gutierrez-Sacristán, Á. Bravo, F. Sanz, and L. I. Furlong, "Identifying temporal patterns in patient disease trajectories using dynamic time warping: A population-based study," *Sci. Rep.*, vol. 8, no. 1, 2018.

[2] A. B. Jensen, P. L. Moseley, T. I. Oprea, S. G. Ellesøe, R. Eriksson, H. Schmock, P. B. Jensen, L. J. Jensen, and S. Brunak, "Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients," *Nat. Commun.*, 2014.

[3] G. Defossez, A. Rollet, O. Dameron, and P. Ingrand, "Temporal representation of care trajectories of cancer patients using data from a regional information system: An application in breast cancer," *BMC Med. Inform. Decis. Mak.*, 2014.

[4] D. W. Joyce, A. A. Kehagia, D. K. Tracy, J. Proctor, and S. S. Shergill, "Realising stratified psychiatry using multidimensional signatures and trajectories," *J. Transl. Med.*, 2017.

[5] R. R. Edwards, R. H. Dworkin, W. Wen, and D. Yarnitsky, "Patient phenotyping in clinical trials of chronic pain treatments: IMMPACT recommendations," *Pain*. 2016.

[6] J. D. Tenenbaum, P. Avillach, Z. Zhao, and R. R. Freimuth, "An informatics research agenda to support precision medicine: seven key areas," *J. Am. Med. Informatics Assoc.*, vol. 1, no. 919, p. ocv213, 2016.

[7] S. S. Gambhir, T. J. Ge, O. Vermesh, and R. Spitler, "Toward achieving precision health," *Science Translational Medicine*. 2018.

[8] N. A. Goldenberg, S. R. Daniels, P. M. Mourani, F. Hamblin, A. Stowe, S. Powell, D. E. Ford, and J. Ellen, "Enhanced Infrastructure for Optimizing the Design and Execution of Clinical Trials and Longitudinal Cohort Studies in the Era of Precision Medicine," *Journal of Pediatrics*. 2016.

[9] N. Geifman, H. Lennon, and N. Peek, *Patient stratification using longitudinal data - Application of latent class mixed models*, vol. 247. 2018.

[10] N. Geifman, R. D. Brinton, R. E. Kennedy, L. S. Schneider, and A. J. Butte, "Evidence for benefit of statins to modify cognitive decline and risk in Alzheimer's disease," *Alzheimer's Res. Ther.*, vol. 9, no. 1, 2017.

[11] N. Geifman, R. E. Kennedy, L. S. Schneider, I. Buchan, and R. D. Brinton, "Data-driven identification of endophenotypes of Alzheimer's disease progression: Implications for clinical trials and therapeutic interventions," *Alzheimer's Res. Ther.*, vol. 10, no. 1, 2018.

[12] H. Lennon, S. Kelly, M. Sperrin, I. Buchan, A. J. Cross, M. Leitzmann, M. B. Cook, and A. G. Renehan, "Framework to construct and interpret latent class trajectory modelling," *BMJ Open*, vol. 8, p. 20683, 2018.

[13] J. S. Smolen and D. Aletaha, "Rheumatoid arthritis therapy reappraisal: Strategies, opportunities and challenges," *Nature Reviews Rheumatology*. 2015.

[14] C. A. Wijbrandts and P. P. Tak, "Prediction of Response to Targeted Treatment in Rheumatoid Arthritis," *Mayo Clinic Proceedings*. 2017.

[15] J. S. Smolen, R. Landewé, M. De Wit, and D. Van Der Heijde, "EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update," *Ann. Rheum. Dis.*, 2017.

[16] M. Schiff, R. Goldblum, and J. Tesser, "SAT0144 Can rheumatoid arthritis (RA) incomplete secondary responders to TNF-alpha attain an efficacious and safe response by switching to certolizumab pegol?:," *Ann. Rheum. Dis.*, 2013.

[17] A. Dagliati, L. Sacchi, V. Tibollo, G. Cogni, M. Teliti, A. Martinez-Millana, V. Traver, D. Segagni, J. Posada, M. Ottaviano, G. Fico, M. T. Arredondo, P. De Cata, L. Chiovato, and R. Bellazzi, "A dashboard-based system for supporting diabetes care," *J. Am. Med. Informatics Assoc.*, vol. 25, no. 5, pp. 538–547, 2018.

[18] A. Franklin, S. Gantela, S. Shifarraw, T. R. Johnson, D. J. Robinson, B. R. King, A. M. Mehta, C. L. Maddow, N. R. Hoot, V. Nguyen, A. Rubio, J. Zhang, and N. G. Okafor, "Dashboard visualizations: Supporting real-time throughput decision-making," *J. Biomed. Inform.*, vol. 71, pp. 211–221, 2017.

[19] C. Proust-Lima, V. Philipps, and B. Liquet, "Estimation of Latent Class Linear Mixed Models : the new package lcmm," *arXiv*, no. 503.00890v2 [stat.CO], pp. 1–13, 2016.

[20] D. Plant, J. Bowes, C. Potter, K. L. Hyrich, A. W. Morgan, A. G. Wilson, J. D. Isaacs, and A. Barton, "Genome-wide association study of genetic predictors of anti-tumor necrosis factor treatment efficacy in rheumatoid arthritis identifies associations with polymorphisms at seven loci," *Arthritis Rheum.*, 2011.

Address for correspondence

Arianna Dagliati,

Manchester Molecular Pathology Innovation Centre

Citylabs 1.0

Nelson Street, Manchester M13 9NQ, UK

arianna.dagliati@manchester.ac.uk

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190357

Knowing Patients Better After a Stroke and Secondary Prevention

G.A. Escobar-Rodríguez[a], R.J. Pérez-Esteban[b], F.P. Pérez-León[a], J. Moreno Conde[a], A. Palomino-García[b], C. Parra-Calderón[a], M.D. Jiménez Hernández[b]

[a]Technological Innovation Group, Virgen del Rocío University Hospital, Seville, Spain.
[b] Department of Neurology, Virgen del Rocío University Hospital, Seville, Spain.

Abstract

The habits and lifestyles are the fundamental factors in the control of cardiovascular risk. Patients who have had a cerebrovascular accident (CVA) have a high risk of having a new event with similar characteristics. The exponentially growing success, penetration and adherence of the new communication technologies, based on applications (APPs), allows to use them to obtain information and influence the risk factors. We propose that empowering patients in their disease can make a more efficient management of it. For this reason, we designed and developed a system which integrates a mobile application and a web application. This system also makes use of peripheral devices to monitor patients and allow the automatic acquisition of information to enable the characterization of this kind of patients in relation to habits and lifestyle. At the same time, the system can also empower these patients with their disease to do secondary prevention.

Keywords:
Stroke; Secondary Prevention; Medical Informatics Applications.

Introduction

The modifiable cardiovascular risk factors (CVRF) are determinants of the major scourge in morbidity/mortality and global disability that generates cardio and cerebrovascular disease [1]. Globally, it is known that patients with ischemic stroke have a high risk of recurrence or the appearance of other cardiovascular events [2]. Keeping the CVRF controlled is essential for reducing the risk of having a cerebrovascular event delaying and even preventing the onset [3]. After an ischemic stroke, the risk of recurrence during the first year is approximately 10% and subsequently, 5% annually. In addition, the risk of presenting a coronary disease is estimated at 6% during the first year and then at an annual 4.6% after a first episode of stroke [4, 5]. The exponentially growing success, penetration and adherence that new mHealth communication technologies, based on applications (APPs) and devices such as wearables and others, allows us to use them as a means to obtain information and influence cardiovascular risk factors. Several studies have been published to demonstrate efficacy in promoting small but important changes in patient behavior and improvements in health status [6, 7]. In addition, the potential benefit for public health is very great given the ubiquity of such devices [8].

We propose in our hypotheses that it is possible to prevent stroke-related events by empowering patients to make more efficient management of their disease. At the same time, we think that we can better know this type of patients by means of knowing their habits and lifestyles better, through a data model that characterizes them and applying new technologies.

To achieve these objectives, we have designed and developed a system composed of two applications, an APP for patients and a Web Application for researchers and clinicians. In addition, we have also integrated some peripheral devices to monitor patient activity and improve the data visualization and follow-up. All peripheral devices used are available in the market at low cost. On the other hand, we have also improved the access to information material and patient training.

The main objective of this paper is to present the technological results prepared for the study focused on the secondary prevention of cerebrovascular events.

Methods

The methodology to achieve technological developments is based on formal software engineering techniques that ensure the suitability of the tools developed to meet the established requirements. In this sense, an iterative and incremental agile project methodology will be used, differentiating the following phases:

a) Requirements analysis: the technical requirements of the project are identified and defined through bidding techniques applied to the health area. The objective is to have a system requirements document with the specifications of the all proposed technological solutions. A working group consisting of a doctor and an engineer gather the functional requirements of the system. These requirements will be analyzed by a multidisciplinary group (clinicians and technologists) of 20 researchers. In the last phases, patients will be included in the process to validate the requirements.

b) Design: the user-centered design paradigm will be applied during this phase. The use of co-design techniques involves the end users in the development process and guarantees the adequacy and acceptance of the product to the end user.

c) Implementation: After the analysis and design phases, the implementation of the system will begin, using an agile methodology based on the development of prototypes and the establishment of goals. Since it is a very long and complex phase, we focus on the technologies used in each module to meet the requirements established above.

d) Testing: the evaluation of the prototypes mentioned above makes it possible to identify new requirements and correcting or improving the possible deficiencies in the design of the solution. To achieve this, unit tests of the created software will be performed. After checking the correct functioning of each component, integration tests will be carried out, where the correct flow of information between the components of the proposed infrastructure will be checked. In addition, in the

next phase the developed tools will be made available to the end users to check if the tools work correctly.

The next phase is the pre-pilot. In this phase, patients will be recruited into the examination room. These patients must have recently had a stroke. The objective of this phase is to make the last adjustments and checks.

Afterwards, a pilot phase will be carried out. The study we are doing in this project is a prospective study. There will be two groups of patients, a control group and an intervention group. At recruitment, they will be provided with the devices and the designed protocol will be applied to them. Patients will be assigned to the groups randomly.

Finally, we will analyze the collected data using the developed system.

Results

All the phases of the described methodology have been carried out including the developments and the tests. In the following sections we explain the results in more detail.

a) Requirements analysis: the technical requirements of the project were identified and defined. The required solution was based on (1) a distributed platform formed by a backend, with 2 differentiated front-ends (the clinician's website and the patient's application), (2) integrated with the Hospital Information System (HIS) (to authenticate healthcare personnel or to get clinical constants) and (3) with other external devices used by the patient; such as a quantifier wristband (to measure his physical activity) and a bluetooth blood pressure meter. (4) The web had to be modular, scalable to the estimated number of users and allow the patient monitoring through the collection of data, scales and constants. Regarding the app, (5) it had to be responsive (accessible from any device) and multiplatform (used in android and iOS mobiles), (6) accessibility adapted to patients who have suffered from stroke , (7) it should alert about the next patient consultation, (8) generate personalized notifications, (9) collect patient behavior constants through forms, and (10) allow the clinician to modify the multimedia content of the app through a web interface in order to achieve patient empowerment. Using this data, (11) the web interface should allow an analysis of the patient's stored data for the extraction of the healthcare professional conclusions. (12) As well as, it should be available from the professional's computer of the recruitment and patient follow-up centers. Furthermore, (13) physicians had not to intervene in the review of patients with the aim of not influencing their decisions.

b) Design: The result of applying the techniques mentioned in this section of the methodology is described in the following figure (figure 1).

Figure 1: Infrastructure design

In this figure, the following modules are differentiated:

- Healthcare web application: application used by clinicians. This application will allow the patient recruitment, information registration and follow-up through scheduled reviews, thanks to the integration with the patient's social and medical history (BDU, in Spanish, demographics user database). In addition, this application stores the patient's records and allow to see the patient's evolution and to carry out the subsequent analysis of the information. Finally, it has been developed and implemented under the regulations established by the Information and communication technologies service of the Virgen del Rocío Hospital in Seville.
- Remote devices: set of peripheral or wearable devices that allow the patient to be tracked. A Fitbit quantifier wristband (from which physical activity, step and arterial frequency are obtained) and a blood pressure monitor are used to store patient data and incorporate them into the study through the use of web services based on their respective clouds. A chromecast is also used to facilitate the visualization of audiovisual content on the patient's television, allowing access to education and risk factors information associated with suffering a new stroke.
- Mobile app: application used by patients at home, to track the disease and improve knowledge about stroke in order to prevent a new cardiovascular event. The patients can visualize their profile, see and add the cardiovascular risk factors associated with the stroke they have, access educational content on the CVRF associated with the patient, make weekly forms for monitoring their disease, visualize their evolution through graphs and visual color codes, and manage their next reviews through alarms. In addition, the application uses artificial intelligence algorithms to give personalized recommendations and advice to the patient. These advices are about the key elements of your illness and are shown using intelligence notifications or when the forms are made.

c) Implementation: The results obtained after applying the techniques mentioned in methodology section and after developing prototypes have been the following.

- Healthcare professional web application: a clinical trials and research studies management platform known as ITC-Bio has been used for the development of this interface. This platform is developed through a framework based on MEAN STACK, which includes the following technologies: MongoDB, Express, AngularJS and Node.js. The platform makes use of a service gateway developed using the Mirth Connect service bus for communications between its components. This bus is a free software initiative, whose objective is to allow the electronic registration of clinical information of patients or the advanced analysis of information, among others. For the storage of clinical information, a PostgreSQL database is used (relational and also based on free code). In addition, within the framework of the PrevenCer project, a module has been developed within the ITC-Bio platform for alarm management, as well as for the analysis and monitoring of the information recorded by patients through their mobile applications.

- Remote devices: a module has been developed that allows the integration of the data collected by the FitBit® wristbands. To do this, the Mirth Connect communications bus has been used to obtain the patient's desired variables from the Fitbit Cloud (through its native APIs). For example, pulse, physical activity, etc.
- Patient mobile application: the IONIC framework has been used for the development of the mobile application. This framework allows generate multiplatform hybrid applications (they can be used in both iOS and Android). For this purpose, the following technologies are used: HTML/CSS/JS, TypeScript, Angular 4 Framework and Apache Cordova (to access the native capabilities of the mobile phone). In addition, it has been designed using styles and libraries such as Bootstrap, which allow the application to be responsive (adapted to other devices such as tablets). For the integration of the mobile application with the health professional's web application, the service platform has been used, developed with the Mirth Connect integration bus, through the JSON information exchange and interoperability standard. Finally, for patient monitoring, several devices (wearables) have been used that are activated by Bluetooth from the application (e.g. the Fitbit® wristband), or use external services to integrate (e.g., Youtube for Chromecast®).

All the information is stored in the Hospital databases within a secure network. Secure communication protocols are used for the exchange of information based on standards.

d) Testing: unit tests of the created software have been carried out. After checking the correct functioning of each component, integration tests have been carried out, where the correct flow of information between the components has been checked.

All the components of the system have been designed for the target population with special emphasis on communication techniques, vocabulary, graphic interface and dynamics adapted to this type of group. The technology infrastructure, the APP and desktop application is set up. The devices are available and ready to start the study. The maximum market price for all devices is approximately € 200 per patient, according to the CVRF and the conditions of each patient. Then, the pre-piloting phase is beginning with patients recruited in consultation.

Figure 2: Patient application screenshot

In Figure 2, it is possible to see how the developments have been focused on the end users, elderly Spanish-speaking patients with motor and cognitive sequelae. We show two screenshots. One of them shows the main menu with six buttons: (1) My profile: it contains the data of the patients. (2) Education: it contains the information about the risk factors customized for the patient, (3) My evolution: it contains the evolutions of the risk factors for the patient, (4) Registry: it is where the patient can register their information. (5) My calendar: it picks up the patient's appointments. (6) Information: it contains the information of the project. The other screenshot in about the evolution. It is possible to visualize the evolution of risk factors.

Discussion

We present the project that we are running between January 2017 and December 2020. The objective is to perform the secondary prevention of stroke and obtain a data model that helps to better understand patients who have had a stroke. We have made the design and development of a system that supports this study through low cost devices and available for most of the public or for health services.

We believe that by obtaining the data model of this type of patients, it is possible to design more efficient secondary prevention measures and improve adherence to treatment.

We have found a few limitations of the study. Patients can obtain training and uncontrolled information through the apps of the devices delivered in the study.

These patients should not install the wristbands or the APP tension meter in order not to use other means of training and information not controlled in the study.

Perspectives

Afterwards, the pilot phase will begin. A total of 160 patients will be recruited and monitored for one year. We will provide the APP for the patients from the intervention group, as well as, the FitBit® wristbands, the iHealth® tensiometer and the chromecast®. For the patients from the control group we will provide only the FitBit® wristbands, iHealth® tensiometer. These patients should not install the bracelets or the APP tension meter in order not to use other means of training and information not controlled in the study.

Once the pilot phase is over, we will analyze the collected data using the developed system (ITC-Bio) which includes a data analysis module.

Once the study is finished, if the behavior of the patients is modified and adherence to healthy lifestyles and medication improves with the training and information introduced in the patients through the means of support. This will help better control modifiable risk factors and will delay the onset of new cardio and cerebrovascular events.

We will publish the conclusions obtained from the study.

Conclusions

We have built a system consisting of an APP, a desktop application and some devices adapted to the end users. We have created an infrastructure capable of integrating the whole system. In this way, we will support a research on secondary prevention of personalized stroke through the monitoring of patients, creating at the same time a data model and allowing the final analysis of the study. The tests carried out on the entire system have been satisfactory. We are currently starting

the study. We hope that by the end of 2020 we can present the results of the study.

Acknowledgements

This research was funded in part by the project (nº PS-0094-2016) "Sistema Monitorizador para la Adquisición de Datos y Prevención Secundaria del Ictus" by subvencionado por la call for biomedical R&D and health sciences (Topic: Public Health Research projects) Ministry of Health of the Andalusian regional government. Thanks also to the Platform for Innovation in Medical Technologies and Health (ITEMAS Platform, codes PT13/0006/0036 and PT17/0005/0036) of Carlos III Institute of Health, co-funded by FEDER, European Regional Development Funds (European Union). Many thanks to all the researchers for their useful contributions and suggestions. We are especially grateful to the researchers: María Auxiliadora Caballero García, José Antonio Rivas González, Jose Luís García Garmendia, Luis Castilla Guerra, Mercedes Romera Tellado, María del Carmen Fernandez Moreno, Ana Moreno Verdugo, Massimo Mangiola Conte, Raimundo Caro Quesada, Flora Maria Villarrasa Clemente, Serafín Moro Muñoz, José Manuel López Suárez, Pedro José Moreno Alvarez, Antonio Javier Ramos Guerrero, David Sánchez Pardo.

References

[1] Furie KL, Rost NS. Overview of secondary prevention of ischemic stroke [Monograph on the internet]. Waltham (MA): UpToDate; 2015 [cited MAR 8 2015]. Available in: http://www.uptodate.com/

[2] Kaplan RC, Tirschwell DL, Longstreth WT, Manolio TA, Heckbert SR, Lefkowitz D, et al.Vascular events, mortality, and preventive therapy following ischemic stroke in the elderly. Neurology. 2005;65:835-842.

[3] Zhang Y, Tuomilehto J, Jousilahti P, Wang Y, Antikainen R, Hu G. Lifestyle factors and antihypertensive treatment on the risks of ischemic and hemorrhagic stroke. Hypertension 2012;60:906–912.

[4] Kaplan RC, Tirschwell DL, Longstreth WT, Manolio TA, Heckbert SR, Lefkowitz D, et al.Vascular events, mortality, and preventive therapy following ischemic stroke in the elderly. Neurology. 2005;65:835-842.

[5] Jonathan F. Easton, Christopher R. Stephens, Maia Angelova. Risk factors and prediction of very short term versus short/intermediate term post-stroke mortality: A data mining approach. Computers in Biology and Medicine 54 (2014) 199–210.

[6] Stephens J, Allen JK, Dennison Himmelfarb CR. "Smart" coaching to promote physical activity, diet change, and cardiovascular health. J Cardiovas Nurs. 2011; 26(4):282-4.

[7] Michel Klein, Nataliya Mogles, Arlette van Wissen, Intelligent mobile support for therapy adherence and behavior change, Journal of Biomedical Informatics. Oct 2014. 51:137-151.

[8] Heval Mohamed Kelli,Bradley Witbrodt, and Amit Shah: The future of mobile health applications and devices in cardiovascular health: Euro Med J Innov. 2017 Jan; 2017: 92–97.

Address for correspondence

Germán Antonio Escobar Rodríguez
german.antonio.er@gmail.com
Avda. Manuel Siurot S/N - 41013 Sevilla
Hospital Universitario Virgen del Rocío
Servicio Andaluz de Salud
Phone: +34 649 10 36 92

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190358

Modeling Chronic Obstructive Pulmonary Disease Progression Using Continuous-Time Hidden Markov Models

Guido Antonio Powell[a], Aman Verma[a], Yu Luo[a], David Stephens[b], David Buckeridge[a]

[a] Department of Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada,
[b] Department of Mathematics and Statistics, McGill University, Montreal, Quebec, Canada

Abstract

Understanding the progression of chronic diseases, such as chronic obstructive pulmonary disease (COPD), is important to inform early diagnosis, personalized care, and health system management. Data from clinical and administrative systems have the potential to advance this understanding, but traditional methods for modelling disease progression are not well-suited to analyzing data collected at irregular intervals, such as when a patient interacts with a healthcare system. We applied a continuous-time hidden Markov model to irregularly-spaced healthcare utilization events and patient-level characteristics in order to analyze the progression through discrete states of 76,888 patients with COPD. A 4-state model allowed classification of patients into interpretable states of disease progression and generated insights about the role of comorbidities, such as cardiovascular diseases, in accelerating severe trajectories. These results can improve the understanding of the evolution of COPD and point to new hypotheses about chronic disease management and comorbidity.

Keywords:

Latent multi-state models, chronic obstructive pulmonary disease, disease progression.

Introduction

Chronic diseases typically progress slowly over many years. For example, patients with chronic obstructive pulmonary disease (COPD) may progress from mild to very severe disease over the span of more than a decade [1]. Understanding how, and to the extent possible why, patients progress through their disease is crucial for improving disease detection, management and prognosis. For example, inappropriate care at early stages of the disease can accelerate progression to more severe stages.[2,3] At a clinical level, the GOLD criteria define stages of COPD disease (based on lung function assessments) and identify appropriate care by disease stage.[4] However, in the absence of accessible clinical and physiological data for the whole COPD population, health system managers are unaware of the distribution of disease severity and cannot tailor interventions to alter their trajectories.

Fortunately, the data needed to model disease progression exist in the form of the vast amount of individual patient data stored in longitudinal health records, such as healthcare administrative databases. Despite their value, these data are not direct observations of the underlying processes that drives disease progression, but instead they record indirect measures such as healthcare encouters, diagnositc codes and drugs. As a result, inferential methods are needed to model disease progression as a function of these observeables.

Several factors complicate the modeling of disease progression. First, patient data are recorded only when services are delivered, resuling in irregularly-spaced observations that differ in granularity between patients or across one patient's course of illness. Disease trajectories, in terms of the rate of progression and the profile of health service used, will also vary widely across patients, further creating challenges in modeling progression for groups of patients. Moreover, for large segments of patient trajectories, records may be sparse due to infrequent visits, or records may be incomplete (e.g., censored data or care provided outside the health system). Finally, the comorbid conditions experienced by most patients are an important driver of both observed healthcare use and progression in the underlying disease of focus.

Research that has used data from electronical health records to model chronic diseases such as COPD has tended to simplify or collapse the data over time (e.g. 30-day readmission [5]). Other researchers have opted to use regularly spaced data from relatively small and selected clinical cohorts, allowing them to model disease progression as time-to-event [6,7]. In modeling the heterogeneity in COPD progession, researchers have clustered clinical measures cross-sectionally [8] or over-time as latent trajectories from smaller regularly-spaced trial data [9].

Previous disease progression research using mutli-state modeling has typically made use of known disease status or known transition rates [10–13] rather than modeling the latent disease progression as a stochastic process. Recent research by Wang et al. [14] has developed a novel approach to modeling COPD progression using a hidden Markov model (HMM), which assumes a latent Markov process governing an evolution of observed events. These events are defined by latent groupings of COPD-related and comorbidity diagnostic codes (ICD-9) measured in 90-day bins.

To model the heterogeneous, sparse, non-equidistant, and incomplete longitudinal observations in a large cohort of COPD patients we propose using a continuous-time hidden Markov model (CTHMM) under a generalized linear modeling framework with patient-level covariates of comorbidities [15]. Such multi-state models can describe patient status over time as a discrete-time realization of a continuous-time Markov process, while accommodating irregular spacing of observations [16]. We are interested in understanding how a CTHMM can be used to infer latent disease progression from an observed evolution in complex healthcare utilization data.

We use a CTHMM on data from a large cohort of COPD patients to model trajectories of the chronic disease and

examine the role comorbid conditions in the progression. Our objective is to generate hypotheses about the evolution of the illness and how the management of COPD may be improved.

Methods

Data

A cohort of COPD patients was selected from an open, dynamic cohort sampled in 1998 as a 25% random sample drawn from people registered with the provincial health insurance agency (*Régie de l'assurance maladie du Québec*) in the census metropolitan area of Montreal[17]. At the start of every following year, 25% of births and new residents arrived to the area in the past year were sampled to maintain a representative cohort. Patients were followed until they died or moved out of the region. The administrative records included outpatient diagnoses and procedures submitted through billing claims, and procedures and diagnoses from hospital records.

Using established case-definitions based on diagnostic codes [18,19], we enrolled 76,888 COPD patients with an incident event (ICD-9 491x, 492x, 496x; ICD-10 J41-J44) occurring after a minimum of two years at risk and followed patients until December 2014. Comorbid diseases were determined at baseline (upon entry in the larger dynamic cohort) using established case-definitions for 9 diseases: acute myocardial infarction (AMI), asthma, cancer, congestive heart failure (CHF), diabetes, hypertension, ischemic heart disease (IHD), mental illness, and stroke). Patient age and sex were also retained for model covariates.

When a patient has an ecounter with the health system, the data recorded are indirect measures of their underlying disease state. For this study, we used data captured during an outpatient visit with a general practitioner (GP), an outpatient visit with a specialist (SPEC) (coded as respirologist or internist), visits to an emergency department (ED), and hospitalizations (HOSP). Health services used prior to the incident COPD event were not considered.

Model description

To model disease progression within the COPD cohort, we used a continuous-time hidden Markov model (CTHMM). Our model assumes the observed events are generated from hidden states that we interpret as belonging to states of disease severity, by computing the probability of different healthcare utilization events with known levels of severity.[20] The observable data in this CTHMM were healthcare utilization events, which were classifed into four mutually exclusive types (GP, SPEC, ED, HOSP). If two events occured on the same day, then we used the most severe event, where HOSP was the most severe, follwed by ED, SPEC, and GP.

Given an observed event, the probability of each of the four observables was determined by a multinomial model, using GP as the reference event. For each state, there was a set of parameters β that specified the multinomial model of the four healthcare utilization outcomes. Each state's multinomial model was specified as a set of three logistic regression models of the probability of specialist, ED, or hospital visit versus the probability of a GP visit. Each logistic regression model was an intercept-only model, meaning that the probability of each healthcare utilization outcome was the same for any patient in the same state. Fixing the utilization probability within states facilitates the interpretion (or labeling) of different states, strictly based on differences in the severity of patient need across states.

Because the observed values (healthcare utilization events) were irregularly-spaced in time, we modeled the transitions between hidden states in continuous time. Also, as patients may have extended periods without any healthcare visits, we allowed for transitions between states even if there were no observed healthcare utilization events. The transition rate between states was specified by the transition rate matrix Q, whose parameters were fit as part of the model, where each row and column corresponded to a hidden state. If X_t is the state membership at time t, then $q_{ij} = \lim_{\Delta t \to 0} \frac{P(X_{t+\Delta t}=j|X_t=i)}{\Delta t}$ where $i \neq j$. The diagonal entries of Q are computed so that the rows of Q sum to zero. Q was calculated separately for each patient using the vector of their baseline covariates $\mathbf{W}$, such that $\log(q_{ij}) = \mathbf{W}^{\mathrm{T}}\xi_{ij}$ where $i \neq j$.

We fixed the transition matrix to allow transitions only to one adjacent subsequent state (a Markov chain), except in one branching state with two transitions. The parameter set ξ_{ij} for all other state pairs were fixed to zero. By restricting the transitions in this manner we aimed to model a set of alternative disease progression trajectories following an earlier Markov chain progression. We compared the goodness-of-fit of a 4-state model (as an initial state transitioning to a state that branches to two absorbing states) to three models with similar topologies but additional states: 5-states (one state before the branching state), 6-states (one state before each absorbing state), and 7-states (a combination of the 5- and 6-state additions). We compared models using the Bayesian information criteria (BIC), mean posterior probabilities of predicted states, as well as criteria from guidelines on model fitting in HMM concerning the interpretability and usefulness of additional states. [21]

Model fit

Our model required us to fit three sets of parameters. π specified the probability of starting in each state, ξ specified the relationship of the patient's baseline covariates to the transition rate matrix, and β specified the probability of each observable (healthcare utilization) for each state.

To fit our model, we used expectation-maximization (EM) [22], an iterative algorithm in which previous parameter values are used to compute new values. The EM algorihm requires starting values for the model parameters. For π, we specified equal probability of starting in any state, the β parameters were randomly drawn from a normal distribution with a mean of 0 and a standard deviation of 1, and the ξ parameters were drawn from a uniform distribution between 0 and 1, except for the parameter for age, which was scaled by the maximum age in the data. We scaled Δt to be expressed in years.

At each iteration of EM, we used the forward-backward algorithm [23,24], along with the current parameter values for π, β, and ξ, to compute, for each consecutive pair of observations within each patient, the probability of starting and ending in each state pair. From these state pair membership probabilities, we computed the probability of state membership at all observed times for all patients.

We then computed new estimates for each of our parameters. We computed a new value for π using the mean probability of state membership at the first observed time point across all patients, and computed new values for β by fitting a multinomial model for each state, weighting each observation by the probability of state membership at that time. Finally, we computed new values for ξ by using a nested EM procedure [25,26]. We stopped the algorithm when the sum of the

difference in norms between the previous and new parameter values was below 0.05.

Results

Figure 1 (A) describes the β and 5-year, non-diagonal Q coefficients of a 4-state model. A 4-state model was chosen over 5-, 6-, and 7-state Markov chain models despite decreases in BIC in latter models (4-state, 5853716; 5-state, 5767363; 6-state, 5721712; 7-state, 5668360). This decision was based on the 4-state model's higher mean posterior probabilities of predicted states (0.93 vs 5-state, 0.91; 6-state, 0.90; 7-state, 0.87) and its better interpretability of distinct states, allowing for better understanding of predictors of transition in disease progression modeling and for the aim of hypothesis generation. We interpreted state 1 as representing a controlled state with high probability of GP visits (86%). Patients in state 2, interpreted as a state of higher disease severity, had a higher probability of ED visits (51%) and hospitalizations (22%). State 3 was interpreted as a stage of more specialized disease management (40% specialist visits and 41% GP services), compared to state 4 in which patients have a combination of higher GP use (65%) but also higher ED visits (23%).

Patients typically transitioned slowly (16% over 5 years) from state 1 to state 2, while transitions away from state 2 were more accelerated (57% to state 3 and 29% to state 4 over 5 years). The complements of these rates represent the probability of remaining in states 1 (84%) and 2 (14%) over 5 years. Transitions from state 1 to 3 or 4 only occurred through state 2.

Figure 1 (B) shows four illustrative patients trajectories of observed health service use along with predicted states at observation time based on patients' posterior probability. .

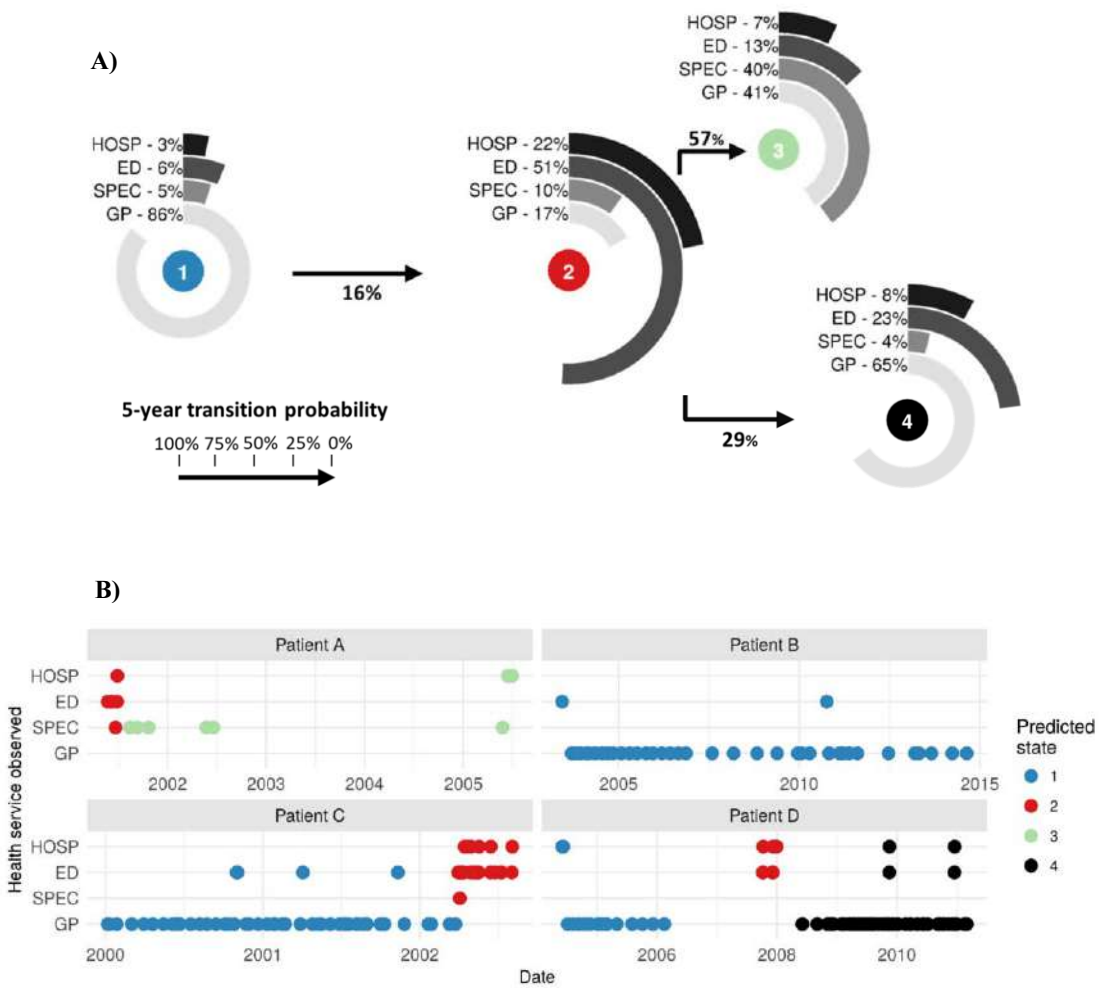


Figure 1 ***A)*** *Diagram of 4-state model of COPD progression based on healthcare utilization. Radial bars represent the probability of each health service (HOSP = hospitalization, ED = emergency department visit, SPEC = specialist visit, GP = general practicioner visit) within a state (labeled 1 to 4), while distance (horizontally) between each state represents the transition probability over 5 years (more distant states represent lower transition probability).* ***B)*** *Observed health service use trajectories of four illustrative patients. Health service types are plotted on separate lines, while patients' predicted state (from1 to 4) at observation time is represented by the dot color*

Table 1 describes the marginal associations between patients' baseline covariates and 5-year transition rates. Noteworthy associations include hypertension, IHD, and stroke increasing the probability of transitioning to and remaining in state 2 (RR ≥ 1.3 and ≥ 1.5, respectively), while cancer is the only covariate associated with higher transition probabilities from state 2 to 3 (1.3). No covariates have large effects on transition probabilities from state 2 to state 4.

Table 1– Risk ratio of transition between states after 5 years by patient characteristics.

Start state	1		2		
End state	1	2	2	3	4
Variable					
Age (10 years)	0.8	1.3	1.0	0.8	1.1
Sex (Male)	0.9	1.2	1.2	0.9	1.0
AMI	1.0	1.0	1.0	1.0	1.0
Asthma	0.9	1.2	1.3	1.1	0.9
Cancer	1.0	1.0	0.9	1.3	0.9
CHF	1.0	1.1	1.2	0.9	1.0
Diabetes	0.9	1.1	0.9	1.0	1.0
Hypertension	0.8	1.8	2.1	0.7	0.9
IHD	0.9	1.4	1.5	0.9	0.9
Mental illness	0.9	1.1	1.1	1.0	1.0
Stroke	0.9	1.3	1.6	0.6	1.1

Discussion

We applied a continuous-time hidden Markov model to healthcare utilization data for a cohort of COPD patients to model their unobserved, latent, disease progression. The model allowed us to learn trajectories through underlying disease states from heterogeneous, sparse and incomplete observations. We found that the model provided interpretable results, allowing us to generate hypotheses about how healthcare utilization evolves at different stages of disease progression.

The discrete states suggest slow progression from a controlled state of the disease (state 1) to one that requires acute care (state 2) and is typically more transient. Patients will mostly transition rapidly from that acute state to one where their disease is managed with primary and specialized care (state 3). However our model specification allowed us to identify an alternative trajectory from state 2 to state 4 where patients are potentially managed inadequately (lower specialist care and higher ED visits than state 3). This pattern suggests possible disparities in patient management (state 3 vs state 4) following hospitalizations and emergency care (state 2) as a window for targeting further investigations into COPD trajectories.

Including patient covariates in the model allowed us to identify predictors of less favourable disease progression patterns. Having a history of different cardiovascular diseases (CHF, hypertension, IHD, and stroke) or of asthma; being older, and being male predicted more severe trajectories, either by precipitating transition towards a state requiring more hospitalizations and ED visits (state 2) or by increasing the probability of remaining in that state. A higher probability of remaining in state 2 typically occurred at the expense of transitions to more specialized care in state 3. We interpret this finding as suggesting that comorbid conditions may increase the risk of repeated acute care in COPD patients and reduce the rate of progression to a state with specialized management. Patients with a history of cancer, however, were more likely to transition to state 3, possibly due to their greater need for specialized treatment. Further analysis can validate these covariates as being modifiers of disease progression by investigating to what extent they predict similar healthcare use patterns outside of a COPD cohort.

These results demonstrate that with little specification beyond model topology (allowed transitions) and number of states, electronic health data can provide new insights and hypotheses on disease progression and management. At a clinical level, such models can improve prognoses by matching patients to typical trajectories of a sub-population. At a health systems level, decision makers can identify health service profiles (e.g. access to specialist care) that are predictive of slower disease progression.

An important aspect of this type of analysis is to weigh multiple criteria in model selection, such as interpretability and relative contribution of additional states [21], especially if the predictors of state transitions are also meaningful. Future research can explore the effect of specifying topologies allowing for transitions back to earlier states, more than one branching state, or higher order Markov chains (states depending on *n* previous states), as well as additional transition covariates such as prescribed drugs and treatments.

Certain limitations of our approach must be considered. Despite our longitudinal focus, covariates of transition probabilities could only be analyzed in terms of their respective baseline measures, a limitation of CTHMMs. Time-varying measures of comorbidities and other covariates would better reflect the evolving impact of patient characteristics on disease progression. In addition, classification of both comorbid status and inclusion in the COPD cohort was based on deterministic case definitions that ignore potential misclassification. An improved approach could learn a probablistic classification of both comorbidities and the modeled disease in a manner similar to how the model learned disease states.

Conclusions

By modeling complex longitudinal observations generated by latent disease states, we have created opportunities for future research in understanding the progression of COPD with potential applications to other chronic diseases.

Acknowledgements

This research was supported by an operating grant from the Canadian Institutes of Health Research.

References

[1] GOLD Executive Committee, Pocket guide to COPD diagnosis, management and prevention: A guide for health care professionals, (2015). http://www.goldcopd.org/uploads/users/files/GOLD_Pocket_2015_Feb18.pdf.

[2] J.J. Soler-Cataluña, M.A. Martínez-García, P. Román Sánchez, E. Salcedo, M. Navarro, and R. Ochando, Severe acute exacerbations and mortality in patients with chronic obstructive pulmonary disease, *Thorax*. **60** (2005) 925–931. doi:10.1136/thx.2005.040527.

[3] S. Spencer, P.M.A. Calverley, P.S. Burge, and P.W. Jones, Impact of preventing exacerbations on deterioration of health status in COPD, *Eur. Respir. J.* **23** (2004) 698–702.
[4] C.F. Vogelmeier, G.J. Criner, F.J. Martinez, A. Anzueto, P.J. Barnes, J. Bourbeau, B.R. Celli, R. Chen, M. Decramer, L.M. Fabbri, P. Frith, D.M.G. Halpin, M.V. López Varela, M. Nishimura, N. Roche, R. Rodriguez-Roisin, D.D. Sin, D. Singh, R. Stockley, J. Vestbo, J.A. Wedzicha, and A. Agustí, Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary, *Am. J. Respir. Crit. Care Med.* **195** (2017) 557–582. doi:10.1164/rccm.201701-0218PP.
[5] R. Sharif, T.M. Parekh, K.S. Pierson, Y.-F. Kuo, and G. Sharma, Predictors of Early Readmission among Patients 40 to 64 Years of Age Hospitalized for Chronic Obstructive Pulmonary Disease, *Annals ATS*. **11** (2014) 685–694. doi:10.1513/AnnalsATS.201310-358OC.
[6] H. Müllerova, D.J. Maselli, N. Locantore, J. Vestbo, J.R. Hurst, J.A. Wedzicha, P. Bakke, A. Agusti, and A. Anzueto, Hospitalized Exacerbations of COPD: Risk Factors and Outcomes in the ECLIPSE Cohort, *Chest*. **147** (2015) 999–1007. doi:10.1378/chest.14-0655.
[7] J. Vestbo, W. Anderson, H.O. Coxson, C. Crim, F. Dawber, L. Edwards, G. Hagan, K. Knobil, D.A. Lomas, W. MacNee, E.K. Silverman, R. Tal-Singer, and on behalf of the E. Investigators, Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE), *European Respiratory Journal*. **31** (2008) 869–873. doi:10.1183/09031936.00111707.
[8] M. Paoletti, G. Camiciottoli, E. Meoni, F. Bigazzi, L. Cestelli, M. Pistolesi, and C. Marchesi, Explorative data analysis techniques and unsupervised clustering methods to support clinical assessment of Chronic Obstructive Pulmonary Disease (COPD) phenotypes, *J Biomed Inform*. **42** (2009) 1013–1021. doi:10.1016/j.jbi.2009.05.008.
[9] D.E. Stull, I. Wiklund, R. Gale, G. Capkun-Niggli, K. Houghton, and P. Jones, Application of latent growth and growth mixture modeling to identify and characterize differential responders to treatment for COPD, *Contemporary Clinical Trials*. **32** (2011) 818–828. doi:10.1016/j.cct.2011.06.004.
[10] P.K. Andersen, and N. Keiding, Multi-state models for event history analysis, *Stat Methods Med Res*. **11** (2002) 91–115.
[11] L. Meira-Machado, J. de Uña-Alvarez, C. Cadarso-Suárez, and P.K. Andersen, Multi-state models for the analysis of time-to-event data, *Stat Methods Med Res*. **18** (2009) 195–222. doi:10.1177/0962280208092301.
[12] J.M. Lange, R.A. Hubbard, L.Y.T. Inoue, and V.N. Minin, A joint model for multistate disease processes and random informative observation times, with applications to electronic medical records data: A Joint Model for Multistate Disease Processes and Random Informative Observation Times, *Biometrics*. **71** (2015) 90–101. doi:10.1111/biom.12252.
[13] N. Bartolomeo, P. Trerotoli, and G. Serio, A multistate model to evaluate COPD progression integrating drugs consumption data and hospital databases, *Epidemiology, Biostatistics and Public Health*. **12** (2015). http://ebph.it/article/view/11145 (accessed November 19, 2016).
[14] X. Wang, D. Sontag, and F. Wang, Unsupervised Learning of Disease Progression Models, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, NY, USA, 2014: pp. 85–94. doi:10.1145/2623330.2623754.
[15] Y. Luo, D.A. Stephens, A. Verma, and D.L. Buckeridge, Latent Multi-State Models For Non-Equidistant Longitudinal Electronic Health Records, (n.d.).
[16] J.D. Kalbfleisch, and J.F. Lawless, The Analysis of Panel Data Under a Markov Assumption, *Journal of the American Statistical Association*. **80** (1985) 863–871. doi:10.2307/2288545.
[17] A. Shaban-Nejad, M. Lavigne, A. Okhmatovskaia, and D.L. Buckeridge, PopHR: a knowledge-based platform to support integration, analysis, and visualization of population health data, *Ann. N.Y. Acad. Sci.* (2016). doi:10.1111/nyas.13271.
[18] C. Blais, S. Jean, C. Sirois, L. Rochette, C. Plante, I. Larocque, M. Doucet, G. Ruel, M. Simard, P. Gamache, D. Hamel, D. St-Laurent, and V. Emond, Quebec Integrated Chronic Disease Surveillance System (QICDSS), an innovative approach, *Chronic Dis Inj Can*. **34** (2014) 226–235.
[19] L. Lix, J. Ayles, S. Bartholomew, C. Cooke, J. Ellison, V. Emond, N. Hamm, H. Hannah, S. Jean, S. LeBlanc, J.M. Paterson, C. Pelletier, K. Phillips, R. Puchtinger, K. Reimer, C. Robitaille, M. Smith, L. Svenson, K. Tu, L. VanTil, S. Waits, and L. Pelletier, The Canadian Chronic Disease Surveillance System: A model for collaborative surveillance, *1*. **3** (2018). doi:10.23889/ijpds.v3i3.433.
[20] A.S. Gershon, G.C. Mecredy, S.D. Aaron, P.G. Camp, K. Tu, P. Hernandez, and T. To, Development of quality indicators for chronic obstructive pulmonary disease (COPD): A modified RAND appropriateness method, *Canadian Journal of Respiratory, Critical Care, and Sleep Medicine*. **3** (2019) 30–38. doi:10.1080/24745332.2018.1476030.
[21] J. Pohle, R. Langrock, F.M. van Beest, and N.M. Schmidt, Selecting the Number of States in Hidden Markov Models: Pragmatic Solutions Illustrated Using Animal Movement, *JABES*. **22** (2017) 270–293. doi:10.1007/s13253-017-0283-8.
[22] A.P. Dempster, N.M. Laird, and D.B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)*. **39** (1977) 1–38.
[23] L.E. Baum, and J.A. Eagon, An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology, *Bull. Amer. Math. Soc.* **73** (1967) 360–363.
[24] L.E. Baum, and G.R. Sell, Growth transformations for functions on manifolds., *Pacific J. Math.* **27** (1968) 211–227.
[25] M. Bladt, and M. Sørensen, Statistical Inference for Discretely Observed Markov Jump Processes, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*. **67** (2005) 395–410.
[26] P. Metzner, I. Horenko, and C. Schütte, Generator estimation of Markov jump processes based on incomplete observations nonequidistant in time, *Phys Rev E Stat Nonlin Soft Matter Phys*. **76** (2007) 066702. doi:10.1103/PhysRevE.76.066702.

Address for correspondence

Guido Antonio Powell, email: guido.powell@mail.mcgill.ca

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190359

A New Approach to Compare the Performance of Two Classification Methods of Causes of Death for Timely Surveillance in France

Yasmine Baghdadi[a], Alix Bourrée[a,b], Aude Robert[c], Grégoire Rey[c], Anne Gallay[a], Pierre Zweigenbaum[b], Cyril Grouin[b], Anne Fouillet[a]

[a] *Santé publique France, Saint-Maurice, France*
[b] *LIMSI, CNRS, Université Paris-Saclay, F-91405 Orsay, France*
[c] *Epidemiology Center on Medical Causes of Death (Inserm-CépiDc), Kremlin-Bicêtre, France*

Abstract

Timely mortality surveillance in France is based on the monitoring of electronic death certificates to provide information to health authorities. This study aims to analyze the performance of a rule-based and a supervised machine learning method to classify medical causes of death into 60 mortality syndromic groups (MSGs). Performance was first measured on a test set. Then we compared the trends of the monthly numbers of deaths classified into MSGs from 2012 to 2016 using both methods. Among the 60 MSGs, 31 achieved recall and precision over 0.95 for either one or the other method on the test set. On the whole dataset, the correlation coefficient of the monthly numbers of deaths obtained by the two methods were close to 1 for 21 of the 31 MSGs. This approach is useful for analyzing a large number of categories or when annotated resources are limited.

Keywords:

Machine learning, sentinel surveillance, cause of death

Introduction

Syndromic surveillance is "a real-time (or near real-time) collection, analysis, interpretation, and dissemination of health-related data to enable the early identification of the impact (or absence of impact) of potential human or veterinary public health threats that require effective public health action" [1]. Mortality is the indicator traditionally used to measure the severity of the impact of an event on a population and is of fundamental significance to health surveillance. As part of the French syndromic surveillance system SurSaUD, timely mortality surveillance is carried out by Santé publique France, the French public health agency [2].

Mortality surveillance is based on death certificates which consist of an administrative part and a medical part based on a model developed by the World Health Organization (WHO) [3]. The medical part is divided into two sections. First, the sequence of morbid events is reported in Part I in descending order from immediate to underlying causes of death (4 fields). Secondly, the unrelated but contributory causes of death are reported in Part II (2 fields).

The medical part consists of free-text fields where physicians write down one or more causes of death, with one or more words which are frequently misspelled. The French Epidemiology Center on Medical Causes of Death (Inserm-CépiDc), the institution responsible for the French national mortality database, then codes the medical causes of death. The coding follows WHO rules, using the 10th revision of the International Classification of Diseases (ICD-10). It can take between 6 and 24 months for the coded medical causes of the death to be made available to epidemiologists. Efforts at automating this coding task to reduce this delay have been made through several shared tasks during the CLEF eHealth campaigns [4,5].

Since 2007, an Electronic Death Registration System (EDRS) allows physicians to certify deaths electronically. The added value of EDRS for reactive mortality surveillance is to send free-text causes of death within a few minutes after the validation of the death certificate. A pilot study demonstrated that E-death certificates constitute valuable data for real-time mortality surveillance [6].

Timely mortality surveillance is based on the routine monitoring of syndromic groups [7]. A Mortality Syndromic Group (MSG) is defined as a cluster of medical causes of death (pathologies, syndromes, or symptoms) that meets the objectives of early detection and impact assessment of public health events. Definitions of MSGs are based on a list of ICD-10 codes, each code belonging to a unique MSG. We defined one hundred MSGs: sixty MSGs consist of acute or severe causes of death and are monitored for early alert purposes, and forty contain chronic diseases. All are used to evaluate the impact of public health events. They span twenty different topics including cardiac and circulatory conditions, respiratory conditions, cancer, infectious diseases, digestive conditions, genito-urinary conditions, symptoms, poisoning, and injuries.

Since e-death certificates are entered in free-text format, the automatic classification of medical causes of death into MSGs requires the use of Natural Language Processing (NLP). In a previous study, we implemented and evaluated two methods to classify causes of death into MSGs: a rule-based method and a Support Vector Machine (SVM) method. Two models were developed using the SVM method: one was based on surface features and the other was a hybrid model combining surface features and features obtained through the rule-based method. We evaluated the performance of the methods on seven out of one hundred MSGs (influenza, lower respiratory diseases, asphyxia and abnormal respiration, acute respiratory diseases, sepsis, chronic digestive diseases, and chronic endocrine diseases). These MSGs were chosen to illustrate the variation within three main dimensions: their epidemiological relevance to mortality surveillance, their frequency of occurrence in death certificates, and the relatedness of their definitions (e.g., acute vs chronic diseases). The results showed that the rule-based method and the hybrid model displayed high performance with F-measures over 0.94 for all seven MSGs. The SVM model achieved lower performance compared to the other two models with F-measures varying from 0.91 to 1 for all MSGs except for chronic digestive diseases (0.88).

The objectives of the present study were to 1) analyze the performance of the rule-based method and the hybrid model to classify causes of death into sixty additional MSGs, and to

2) compare the trends in the monthly numbers of certificates classified into MSGs from 2012 to 2016 using both methods. This second analysis was also performed on the seven MSGs from the previous study.

Materials and methods

1. Corpus characteristics

Data consisted of the medical section of e-death certificates received routinely between 2012 and 2016 (203,797 certificates). E-death certification has been gradually rolled out, recording 5% of the national mortality in 2012 and 12% in 2016. These data were anonymized and contained both administrative and demographic information as well as free-text medical causes of death. To evaluate the performance of the classification methods, we used 4,500 annotated death certificates. We define the term "entity" as an expression of a cause of death on a field of the medical part of the death certificate.

2. Training and test splits

Using random sampling without replacement, the dataset was split into two parts: a training set using death certificates from 2012 to 2015 and a test set using death certificates from 2016. Three thousand five hundred certificates were annotated for the training set and one thousand for the test set. For this annotation task, each annotator had a detailed definition of each MSG based on lists of ICD-10 codes. Two annotators assigned MSGs to each field of the medical part of death certificates in both subsets. The two annotated subsets were compared to each other. Errors were discussed between the two annotators and corrected if necessary. The agreement rate was 0.90 on the test set. Final annotated subsets represent the ground truth against which the methods tested were evaluated.

3. Data preprocessing

The data and dictionary were preprocessed by applying 6 steps: conversion of all characters to lowercase, removal of diacritics, standardization of compound words (i.e. the replacement of hyphens by spaces to obtain two distinct words), replacement of punctuation (except commas, simple quotes, and semicolons) by a space, stop words removal, and removal of multiple spaces in the text and at its beginning and end. The preprocessing steps aimed to obtain a narrower set of more relevant terms.

4. Rule-based method

The rule-based method consisted of assigning MSGs to all entities from death certificates based on an ad-hoc dictionary provided by the Inserm-CépiDc. This dictionary is composed of entities found in previous death certificates with their corresponding ICD-10 codes. The principle of the rule-based method is to assign an MSG based on ICD-10 codes to an entity found in a death certificate when it matches an entity found in the dictionary. Our method relied on four steps : 1) entity normalization through the removal of unnecessary words, replacement of synonyms, acronym expansion, etc., based on more than 900 rules; 2) text segmentation based on comma and semi-colon punctuation marks, prepositions, a few causal expressions, and factorization of coordinated elements (e.g., "hepatitis A and B" is replaced by entities "hepatitis A" and "hepatitis B"); 3) spelling correction using a Levenshtein distance computed with words from the dictionary; and 4) dictionary matching. This method aimed to find entities from the dictionary on death certificates, from the longest span to the shortest one, without overlap. Assignment of an MSG was attempted after each step.

5. Machine learning method

Among the supervised machine learning methods, we trained a linear SVM classifier [8, 9] using the default hyperparameter (C parameter equal to 1). Since it may be possible to associate 0, 1, or more MSGs to each field of a death certificate, we performed a multi-label classification using the one-versus-all strategy [8]. The SVM classifier was trained for all MSGs. We implemented a hybrid model using a combination of surface features and the MSGs previously predicted using the rule-based method as its features. We selected bag-of-words unigrams (set of unique words for each example in the training set), bag-of-words bigrams (set of unique sequences of two consecutive words for each example in the training set), and character trigrams (providing a degree of robustness to spelling errors) as surface features. Previous studies show that those surface features provide the best classification performance on French mortality data [9]. This classifier was implemented using the scikit-learn library in Python [10].

6. Evaluation metrics on the test set

Both methods were previously trained and developed on the training set. This study focused on the analysis of MSGs that were found in similar proportion and with at least three mentions in the training set, test set, and data from 2012 to 2016. We analyzed the performance of the two methods on sixty MSGs belonging to eighteen topics.

Three evaluation metrics were considered: precision, recall, and F-measure. In order to simplify the reporting of the results we defined three groups of MSGs depending on the performance of the methods on the test set:

- Group 1 contained MSGs with similar performance for both methods. We defined similar performance as having the three evaluation metrics belonging to the same performance level. The performance levels of each evaluation metric were defined using four categories: $\geq$0.95, [0.90-0.95), [0.85-0.90), and <0.85.
- Group 2 contained the remaining MSGs with similar performance for either the rule-based method or the hybrid model.
- Group 3 contained the remaining MSGs with heterogeneous performance for each method. We defined heterogeneous performance as the precision and recall belonging to different performance levels.

The cut offs for the performance levels were based on the objectives of the mortality surveillance system. A surveillance system implemented for the timely detection of outbreaks must be based on methods with high recall and precision [11].

7. Comparison of the evolution of the monthly numbers of death certificates from 2012 to 2016 using the two methods

We applied the rule-based method and the hybrid model on the entire 2012-2016 dataset in order to verify how the performance measured on the test set would translate to the whole set. All of the causes of death contained in the certificates were classified using both methods. We then counted the number of death certificates by MSGs and by month for both methods (nb_E-death$_{RBM,month}$ and nb_E-death$_{Hybrid,month}$ for each MSG). The comparison of the monthly numbers of MSGs was based on:

1. Visual analysis of the monthly trends of each MSG. We defined 3 patterns: a) the monthly trends of an MSG using the rule-based method and the hybrid model overlap; b) the existence of a small difference between the 2 curves; c) the existence of a large difference between the 2 curves.

2. The calculated difference (Δ_i) of the monthly numbers of death certificates between the two methods expressed as a proportion for the MSG i (formula below) where m corresponds to each month from 2012 to 2016 (m=60 months in total) and i refers to each analyzed MSG.

1. $$\Delta_i = \frac{\frac{\sum_{m1}^{m60}\left(\text{nb E-death}_{RBM,i,m} - \text{nb E-death}_{Hybrid,i,m}\right)}{60}}{\frac{nbEdeath_{RBM,i} + nbEdeath_{Hybrid,i}}{2}}$$

3. The calculated correlation coefficient of nb_E-deathRBM,i and nb_E-deathHybrid,i.

2. This comparison was performed first for the seven MSGs of the previous study and then for only the MSGs with performance (3 metrics) of either method over 0.95.

8. Comparison of the monthly trend of MSGs to an external data source

For MSGs (such as "influenza" or "low acute respiratory infections" [12]) which are expected to have a similar temporal pattern to those of winter infectious epidemics, we visually compared the patterns of the monthly numbers of death certificates and the monthly numbers of attendances of the pathology. For this, we collected from the SurSaUD system the monthly numbers of attendances in emergency departments (EDs) with clinical diagnoses of "influenza" and "low acute respiratory infections" from 2012 to 2016. This analysis focused on two of the seven MSGs of the previous study for which an external data source was available.

Results

1. Analysis of performance of the rule-based method and hybrid model on the test set

1.1 MSGs with similar performance for both methods (Group 1)

Among the twenty-two MSGs of Group 1, performance (3 metrics) of the methods was over 0.95 for 19 MSGs (Table 1). Those excellent performance values are due to distinct features (consistency of entities, useful cues, and context for entity identification) depending on the MSG.

Table 1: Number of MSGs for which performance metrics belong to a performance level; France 2016 data.

	Group 1	Group 2	
Performance level	Both methods	Rule-based method	Hybrid model
≥0.95	19	6	6
[0.90-0.95)	0	1	2
[0.85-0.90)	1	0	1
<0.85	2	3	1

They belonged to nine different topics (Figure 1): 1-Cardiac and circulatory conditions (4 MSGs/11 MSGs analyzed in this topic), 2-Cancer (1/1), 3-Respiratory conditions (3/4), 4-General symptoms (3/9), 5-Infectious diseases (1/6), 6-Nutritional and Endocrine conditions (1/2), 7-Nervous system conditions (4/4), 10-Mental and behavioral disorders (1/3), and 15-Digestive conditions (1/1). Two MSGs had performance below 0.85 ("5-Bacterial infections" and "16-Other undefined and unspecified causes of death").

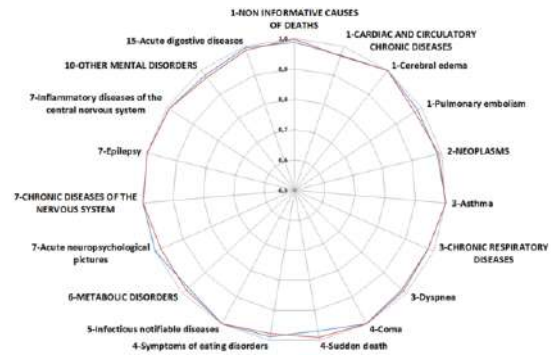


Figure 1: F-measures of the rule-based method (blue) and the hybrid model (red) for nineteen MSGs of Group 1 with performance over 0.95; France 2016 data.

1.2 MSGs with similar performance for either one or the other method (Group 2)

Eighteen MSGs belonged to Group 2. Similar performance (3 metrics) was observed solely with the rule-based method for 8 MSGs. Similar performance was also observed solely with the hybrid model for eight other MSGs. Similar performance belonging to different levels were provided by the two methods for 2 MSGs.

Performance of either the rule-based method or the hybrid model was over 0.95 for a majority of the MSGs in this group (12/18) (Figure 2). Performance of the rule-based method or the hybrid model was also comprised between [0.90-0.95) for three MSGs, while lower performance (below 0.85) was measured for four MSGs (Table 1).

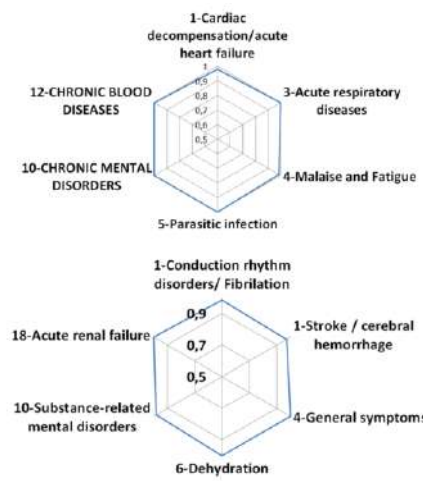


Figure 2: F-measure of the rule-based method (upper part) or hybrid model (lower part) for MSGs with performance (3 metrics) over 0.95 for one of both methods; France 2016 data.

1.3. MSGs with heterogeneous performance (Group 3)

Group 3 contained twenty MSGs with heterogeneous levels of precision and recall. While the rule-based method achieved a higher recall than the hybrid model for more than half of the MSGs (11/20), the rule-based method achieved either higher or lower precision than the hybrid model for six MSGs and seven MSGs, respectively (Table 2). The precision of both methods was similar for seven MSGs and recall was similar for three MSGs.

Table 2: Number of MSGs of Group 3 with performance level (PL) obtained using the rule-based method over, below or equal to the PL obtained using the hybrid model; France 2016 data.

	Precision	Recall
$PL^{a}_{RBM} > PL_{Hybrid}$	6	11
$PL_{Hybrid} > PL_{RBM}$	7	6
$PL_{RBM} = PL_{Hybrid}$	7	3

[a]PL: Performance level

2. Comparison of the monthly numbers of death certificates classified into MSGs by the two methods

2.1 Seven MSGs of the previous study

For five out of seven MSGs evaluated in the previous study, the trends in the monthly numbers of death certificates using the rule-based method overlapped the trends of the monthly numbers of death certificates using the hybrid model. The monthly trends were closely correlated for the two other MSGs. They all had small Δ_i. (Table 3).

Table 3: Number of MSGs according to the levels of $\Delta_i$and the visual pattern of the monthly numbers of death certificates. France-2012-2016

		Visual patterns		
	Δ_i	Overlap	Good correlation	Low correlation
7 MSGs	Small	5	2	0
19 MSGs* Group1	Small	7	6	1
	Intermediate	0	1	1
	High	0	0	3
12 MSGs* Group 2	Small	5	3	0
	Intermediate	0	2	0
	High	0	0	2

*MSGs with performance over 0.95

Small: Δ_i varying from -5% to 5%; Intermediate: Δ_i varying from [-10% to -5%] or to [5% to 10%]; High: Δ lower than -10% or over +10%

2.2 MSGs of Group 1 with performance over 0.95

Correlated trends were also observed for 7/19 MSGs in Group 1 with performance over 0.95 in the test set. The Δ_i of those MSGs ranged from -5.0% to 5.0% (Table 3) and their correlation coefficient was very close to 1. A good visual correlation was also observed between the monthly variations of deaths certificates for 7/19 other MSGs with performance over 0.95, even if a gap was noticed between the monthly numbers of deaths certificates provided by the two methods (Table 3). These seven MSGs also had Δ_i varying from -5.0% to 5.0%, except for "7-Acute neuropsychological pictures" for which $\Delta_i$was 5.9% (Table 3). Correlation coefficients of these MSGs were over 0.99.

A poor visual correlation between the trends of the monthly numbers of deaths using the two methods was observed for five other MSGs which achieved performance over 0.95 for both methods in the test set. For these MSGs, Δ_i were mostly lower than -10.0% or over 10.0% (Table 3).

2.3 MSGs of Group 2 with performance over 0.95

Overlapping trends of the monthly numbers of deaths provided by the two classification methods was observed for 5/12 MSGs of Group 2 with performance of one or the other method over 0.95 (Table 3). For these MSGs, Δ_i ranged from -3.9% to 1.7% and the correlation coefficient was 1.

Five MSGs also had a good visual correlation between the trends of the monthly numbers of deaths of the two methods (Table 3). For 3 MSGs, Δ_i was small with a correlation coefficient over 0.99.

Poor visual correlation of the trends was observed for two MSGs ("10-Chronic mental disorders" and "5-Parasitic infections"). The Δ were 12.5% and 24.3% and coefficient correlation were 0.98 and 0.96 respectively.

3. Comparison of the monthly trends of MSGs to an external data source

The monthly trends of the number of death certificates including a mention of "influenza" was very close for the two methods. Compared to the trends of the monthly numbers of ED attendance for influenza, seasonality and peaks were concomitant even if magnitudes were different. We noted that the number of deaths increased only during winters, the epidemic period of influenza in France (Figure 3). Likewise, the same observations were valid for the MSG "Low acute respiratory infections".

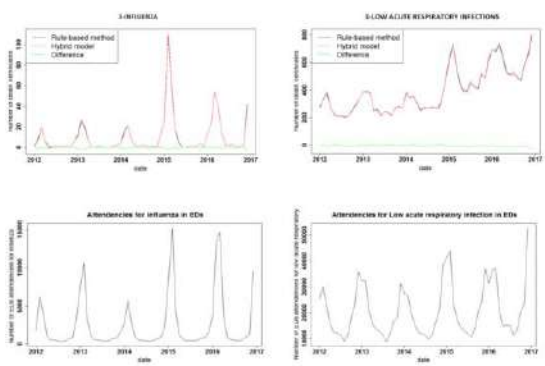


Figure 3: Trends of the monthly numbers of death certificates classified by the rule-based methods and the hybrid model for "influenza" and "low acute respiratory infections"(upper part) and evolution of the monthly numbers of EDs attendances for the same pathologies (lower part); France 2012-2016 data.

Discussion

The current study has shown that more than half of the sixty MSGs achieved both precision and recall over 0.95 for at least one of the two classification methods on the test set (19 MSGs from Group 1, and 6 MSGs from Group 2). When the classification methods were applied to the whole dataset we observed that 13/19 MSGs of Group 1 and 8/12 MSGs of Group 2 as well as the seven MSGs previously evaluated in the prior study had small differences between the monthly numbers of deaths certificates obtained using the two methods, with correlation coefficients close to 1. These results suggest that both the rule-based method and the hybrid model initially implemented, tuned and evaluated to classify medical entities into seven MSGs, are suitable for classification into a larger number of MSGs. This is of particular interest since 10/13 MSGs of Group 1 and 7/8 MSGs of Group 2 have been defined as

needing routine monitoring for the timely detection of outbreaks. Among the other 6/19 MSGs of Group 1 and 4/12 MSGs of Group 2 with performance over 0.95 measured on the test set, 6 had a low visual correlation associated with intermediate or high differences between the monthly numbers of deaths obtained using the two methods. The definitions of those MSGs contained rare and/or a large variety of diseases, like "5-Infectious notifiable diseases", "5-Parasitic infections" or "10-Chronic mental disorders". The training and test sets included an insufficient amount of death certificates containing these medical entities to capture the variety of those rare pathologies.

For two MSGs of the previous study, "Influenza" and "Low acute respiratory infections" with high performance and overlapping trends using both methods, the trends of MSGs were concomitant with those provided by an external data source. These MSGs built using NLP methods correctly reflect the pathologies we want to routinely monitor. This last step should be applied on the remaining MSGs in order to complete this study. The challenge is to find the most appropriate external data sources among specific surveillance systems, notifiable diseases systems, or the Health National Data system.

Group 3, which contained MSGs with heterogeneous precision and recall for the same method, is more complex to analyze. To meet the objective of a reactive detection of events, a high precision is needed to limit false positive cases. To measure the impact of an event, the surveillance system also needs a high recall to avoid underestimation of the impact. Results on the test set showed that recall was mostly higher with the rule-based method than with the hybrid model. However, neither of the two methods provided a consistently higher precision. As both high recall and precision are required to set up an accurate surveillance system, we cannot recommend the use of the studied classification methods to analyze the MSGs of Group 3. A complementary study including an error analysis and a larger training set is required to improve the two methods to classify entities into MSGs from Group 3 and also into MSGs from Groups 1 and 2 that have lower performance.

The current study proposes an approach that goes beyond the traditional evaluation based on a test set. This is interesting, especially when there are many categories to evaluate or when the annotated resources are quite limited and do not capture the large variety of entities to classify. The analysis of the temporal pattern of the number of death certificates for each MSG (including the comparison with an external data source) enabled us to confirm that these indicators (MSGs) accurately capture the targeted health situation we would like to monitor. This approach is also a way to verify whether the misclassifications of one of the two methods are homogeneously distributed over time or are observed in specific periods. Indeed, such a timely mortality surveillance system based on MSGs will be used to provide specific information to health authorities during emergent public health events or during an excess period of deaths (e.g. winter epidemics) to help decision-makers adapt counter measures and prevention messages.

Acknowledgements

The authors thank the IT department of LIMSI for computer support and help in computer setup.

References

[1] Triple S Project, Assessment of syndromic surveillance in Europe, *The Lancet* **378** (2011), 1833-1834.doi: 10.1016/S0140-6736(11)60834-9.

[2] Y. Baghdadi, A. Gallay, C. Caserio-Schönemann, A. Fouillet, Evaluation of the French reactive mortality surveillance system supporting decision making, *Eur J Public Health* (2018), cky251. doi: 10.1093/eurpub/cky251.

[3] World Health Organization, *International statistical classification of diseases and related health problems 10th revision*, 2016.

[4] A. Névéol, R.N. Anderson, K.B. Cohen, C. Grouin, T. Lavergne, G. Rey, A. Robert, C. Rondet, and P. Zweigenbaum, CLEF eHealth 2017 Multilingual Information Extraction task overview: ICD10 coding of death certificates in English and French, in: *CLEF 2017 Evaluation Labs and Workshop: Online Working Notes, CEUR-WS*, 2017, p. 17.

[5] P. Zweigenbaum and T. Lavergne, Multiple methods for multi-class, multi-label ICD-10 coding of multi-granularity, multilingual death certificates, in, CLEF, 2017.

[6] M. Lassalle, C. Caserio-Schönemann, A. Gallay, G. Rey, and A. Fouillet, Pertinence of electronic death certificates for real-time surveillance and alert, France, 2012–2014, *Public Health* **143** (2017), 85-93. doi: 10.1016/j.puhe.2016.10.029

[7] Y. Baghdadi, A. Gallay, C. Caserio-Schönemann, M.-M. Thiam, and A. Fouillet, Towards real-time mortality surveillance by medical causes of death: A strategy of analysis for alert, *Revue d'Épidémiologie et de Santé Publique* **66** (2018), S402. doi: 10.1016/j.respe.2018.05.453.

[8] B. Schölkopf and A.J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*, MIT press, Cambridge, 2001.

[9] P. Zweigenbaum and T. Lavergne, Multiple methods for multi-class, multi-label ICD-10 coding of multi-granularity, multilingual death certificates, in: *CLEF*, 2017.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg, Scikit-learn: Machine learning in Python, *Journal of machine learning research* **12** (2011), 2825-2830.

[11] J.W. Buehler, R.L. Berkelman, D.M. Hartley, and C.J. Peters, Syndromic surveillance and bioterrorism-related epidemics, *Emerging Infectious Diseases* **9** (2003), 1197-1204. doi: 10.3201/eid0910.030231.

[12] W.W. Thompson, D.K. Shay, E. Weintraub, L. Brammer, N. Cox, L.J. Anderson, and K. Fukuda, Mortality associated with influenza and respiratory syncytial virus in the United States, *JAMA* **289** (2003), 179-186. doi: 10.1001/jama.289.2.179.

Address for correspondence
Yasmine Baghdadi: yasmine.baghdadi@santepubliquefrance.fr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190360

A Regularization-Based eXtreme Gradient Boosting Approach in Foodborne Disease Trend Forecasting

Shanen Chen[a], Jian Xu[b], Lili Chen[b*], Xi Zhang[a*], Li Zhang[b], Jinfeng Li[b]

[a] *Department of Industrial Engineering and Management, Peking University, Beijing, China*
[b] *IBM Research - China, Beijing, China*

Abstract

Foodborne disease is a growing public health problem worldwide and imposes a considerable economic burden on hospitals and other healthcare costs. Thus, accurately predicting the propagation of foodborne disease is crucial in preventing foodborne disease outbreaks. Few studies have investigated the dependencies between environmental variables and foodborne disease activity. This study develops a regularization-based eXtreme gradient boosting approach for foodborne disease trend forecasting considering environmental effects to capture dependencies hidden in foodborne disease time series. A real case in Shanghai, China was studied to validate our proposed model along with comparisons to traditional and benchmark algorithms for foodborne disease prediction. Results show that the foodborne disease prediction approach we propose achieves slightly superior performance in terms of one-day-ahead prediction of foodborne disease, and presents more robust prediction for 2-7 days ahead prediction.

Keywords:

Algorithms, foodborne diseases, machine learning

Introduction

Foodborne disease encompasses a wide spectrum of illnesses and is a growing public health concern worldwide, which is the result of eating food contaminated with bacteria, viruses, and toxic or hazardous substances [1]. It is the leading cause of symptoms of nausea, vomiting, abdominal pain, diarrhea, fever, paralysis, rash, and other symptoms across all age groups, especially vulnerable groups such as children, older people and those with impaired immunity [2]. Foodborne diseases are estimated to cause 600 million - almost 1 in 10 people in the world - fall ill and 420,000 die every year [3]. Although recent molecular methods have improved the detection, identification and characterization of foodborne disease in the environment and clinical samples, widespread emergence of the viruses still presents a challenge to the prediction technique. For healthcare authorities, developing emergency preparedness plans to manage the surge in the number of patients increase during a pandemic is important if foodborne disease could be predicted in a timely and accurate manner [4]. The reliable forecast of foodborne disease allows better coordination of mitigation and intervention resources in public health system and reduces the strain on healthcare systems [5].

A vast number of quantitative approaches have been devoted to forecasting foodborne disease [6,7,8,9]. Cawley et al. proposed a sparse Bayesian kernel survival analysis model to predict the growth of a foodborne microbial pathogen, which achieved better accuracy than models based on traditional survival analysis techniques [6]. Rasam et al. applied geographical information system (GIS) and satellite remote sensing technologies in interactive mapping of foodborne disease and analyzing dynamic changes for foodborne transmission risk factor [7]. Mohammed et al. proposed an adaptive neuro-fuzzy inference system (ANFIS) to forecast the total cases of norovirus, and the algorithm showed high accuracy and remarkable prediction of norovirus cases in drinking water [8]. Hill constructed an agent-based model to provide a reasonable representation of how the norovirus spread between students among classrooms in a school [9]. Several studies have contributed to the growing body of evidence that points to the role of environmental variables in the transmission of the foodborne disease [10,11], where linkages between certain environmental factors and specific foodborne disease type have been investigated. However, within current research there has been limited assessment of the global impact of the environment on foodborne diseases, and specifically the temporal transmission of foodborne disease. Incorporating considerations for environmental factors in foodborne disease spread may address uncertainties and non-stationarity existing in foodborne disease observations.

To address this gap, we apply a regularization-based eXtreme gradient boosting approach in the creation of a prediction model that can reveal hidden patterns in foodborne disease data related to environmental factors, particularly meteorology and air quality. The contributions of the study can be divided into three folds. *First*, environmental information is fused into the model to assess their impact on foodborne disease spread. Prediction results are compared to explain the significance in including environmental variables. *Second*, L_1-regularization (least absolute shrinkage and selection operator, LASSO) [13] is employed to identify the most explainable variables and reduce the sparse effect for multiple environmental factors in the prediction model. The LASSO has two advantages over other feature selection techniques in terms of stability and computational efficiency, which are preferable in our case to select most revelent and non-redundant environmental variables for foodborne disease prediction. *Finally*, a systematic comparison between several machine learning models and our approach is conducted through a real dataset from the Shanghai Municipal Center for Disease Control and Prevention (SHCDC) to demonstrate the performance of our proposed model.

The rest of this paper is organized as follows. Methods presents the model details of the proposed regularization-based eXtreme gradient boosting approach for foodborne disease trend forecasting. Results presents a real case study, where foodborne disease time series obtained from Shanghai is applied to demonstrate the performance. Finally, the discussion provides the summary and concluding remarks.

Methods

Data Description

A national foodborne disease monitoring and surveillance system was established in 2010, covering all 31 provinces, major municipalities and autonomous regions in Mainland China. Among those covered regions, Shanghai municipality has been confronted with the challenge of reducing the burden of foodbore disease. To determine a diagnosis of foodborne illness, clinicians review history and and underlying condition for patients who report gastrointestinal symptoms, vomiting, fever, or drowsiness. In collaboration with SHCDC, the daily number of total foodborne disease cases, along with five other channels of foodborne disease such as the daily numbers of positive cases and foodborne cases conducting laboratory test, was collected from 1st-Jan-2014 to 25th-Oct-2017. During this period of 1394 days a total of 52163 cases were reported by selected survillance hospitals. The meteorology data including daily reports of sunshine, humidity, ground temperature, pressure, rainfall, air temperature, wind speed, max wind speed and wind direction from 2014 to 2017 were collected from Shanghai Meteorological Bureau. The daily concentrations of ambient air pollutants from 2014 to 2017 were also obtained from Shanghai Environmental Protection Bureau website, where average values of sulfur dioxide (SO_2), nitrogen dioxide (NO_2), ozone (O_3), particular matter less than 2.5 μm in diameter ($PM_{2.5}$), particulate matter less than 10 μm in diameter (PM_{10}), carbon monoxide (CO) and air quality index (AQI) from different monitoring stations were calculated. The missing values of each variable were imputed with the mean computed from the remaining records of that variable. After preprocessing, these three sources of databases were aggregated according to the recorded dates. With consideration of periodicity such as date, week, month, year, day of week, day of month and day of year, the dataset was composed of a total of 29 variables. For our study, the daily number of total foodborne disease cases was the predictive variable of all models, and meteorology data and ambient air pollutants data were the input variable.

KPSS and Kruskal-Wallis test

The stationarity and linearity analysis of the foodborne disease series were performed. *First*, the non-stationary characteristic of the time series was detected by means of KPSS (Kwiatkowski, Phillips, Schmidt, and Shin) test [14]. Then, Kruskal-Wallis test [15] was applied to confirm the existence of non-linearities, which was performed by splitting the time series into five packages of data and detecting the equality of these five distribution functions to learn whether those samples were from the same distribution.

L_1-regularization for Important Environmental Factor Selection

The prediction task can be challenging given limited data and multiple environmental factors due to overfitting and strong correlations among different environmental factors. Variable selection methods are commonly used to in such scenarios. LASSO was applied in our study to identify an appropriate variable set of environmental factors for improving the prediction accuracy of foodborne disease.

Proposed by Tibshirani, LASSO is widely employed to estimate parameters and select variables in regression analysis [12]. The LASSO estimator is a particular case of the penalized least squares regression using L_1-penalty to shrink regression coefficients toward zero. Thus the most explainable variables whose coefficients are not equal to zero are selected. Typically, we assume that $X \in N \times p$ is the input data matrix, for which the first column is time index, and the remaining $(p-1)$ columns are predictors including environmental components and other channels of foodborne diseases for continuous N days. $Y \in N \times 1$ represents daily number of total foodborne disease cases during the same continuous N days. Given the standardized predictors x_{ij} for $i = 1,2,\dots,N$ and $j = 1,2,\dots,p$, and the response values y_i, the LASSO estimator is described as

$$x_{ij}\hat{\beta}_{LASSO} = arg\,min\{\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p}|\beta_j|\} \quad (1)$$

LASSO is implemented with training data to select the most robust predictors using least angle regression algorithm by employing the LARS package in the R (Version 3.3.2). The following methods are also applied with R based on the corresponding R packages, such as xgboost, stats, adabag, gbm, etc.

eXtreme Gradient Boosting Approach

With the predictor subset selected by LASSO, machine learning models are applied to forecast foodborne disease cases. The gradient boosting model proposed by Friedman has been empirically proven to have high efficiency [13]. As one of popular gradient boosting models, tree boosting is a highly effective and widely used machine learning method. In this paper, a scalable end-to-end tree boosting approach, namely eXtreme gradient boosting (XGBoost) has been applied to perform foodborne disease forecasting. XGBoost is widely used by data scientists to achieve state-of-art results on difficult machine learning challenges [14].

For a given data set with N samples and p features $\mathcal{D} = \{(\mathrm{x}_i, y_i)\}$ $(|\mathcal{D}| = N, x_i \in \mathbb{R}^p, y_i \in \mathbb{R})$, a tree ensemble model uses K additive functions (K trees) to predict the output.

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F}, \quad (2)$$

where $\mathcal{F} = \{f(x) = \omega_{q(x)}\ (q: \mathbb{R}^p \to T, \omega \in \mathbb{R}^T)$ stands for the space of regression trees (also known as CART). Here q represents the structure of each tree that maps an instance to the corresponding leaf index. T is the number of leaves in the tree. Each f_k corresponds to an independent tree structure q and leaf weights ω. For each instance, the decision rules in the trees (given by q) are applied to classify it into the leaves and calculate the final prediction by summing up the weight in the corresponding leaves (given by ω). To learn the set of functions used in the model, we minimize the following objective in equation (3) .

$$L(\emptyset) = \sum_{i=1}^{N} loss(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k), \quad (3)$$

where the first part in equation (3) is the differentiable convex training loss that measures the difference between the prediction $\hat{y}_i$ and the target y_i. The second term Ω is the complexity of the trees. The model complexity $\Omega(f_k)$ is defined by the following equation,

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \left|\left|\omega_j\right|\right|^2, \quad (4)$$

where T is the number of leaves in the tree, λ represents for the regularization weight, γ is the minimum loss reduction and $||\omega_j||$ stands for the score for corresponding leaves. The tree $f_k(x)$ is defined in the following equation (5),

$$f_k(x) = \omega_q(x), \omega \in \mathbb{R}^T, q: \mathbb{R}^d \to 1,2,\dots,T. \quad (5)$$

The tree ensemble model in equation (3) includes function as parameters and cannot be optimized using traditional optimization methods in Euclidean space. Instead, the model is trained in an additive manner. Formally, let $\hat{y}_i^{(t)}$ represents

the prediction of i-th training sample at the t-th iteration, we will need to add f_t to minimize the following objective.:

$$L^{(t)} = \sum_{i=1}^{N} loss\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \quad (6)$$

The f_t that most improves the models is greedily added to the model structure. By embedding the regularization term with equation (4) and (6) it can be further approximated based on the Taylor's expansion as follows:

$$\tilde{L}^{(t)} \approx \sum_{i=1}^{N} \left[loss\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\right] + \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} ||w_j||^2, \quad (7)$$

where $g_i = \partial loss(y_i, \hat{y}_i^{(t-1)}) / \partial \hat{y}^{(t-1)}$, $h_i = \partial^2 loss(y_i, \hat{y}_i^{(t-1)}) / \partial \hat{y}^{(t-1)}$ are the first and second order gradient statistics on the loss function .

Define $I_j = \{i | q(x_i = j)\}$ as the instance set in leaf j. We can rewrite equation (7) as follows.

$$\tilde{L}^{(t)} \approx \sum_{j=1}^{T} [\left(\sum_{i \in I_j} g_i\right) \omega_j + \frac{1}{2}\left(\sum_{i \in I_j} h_i + \lambda\right) \omega_j^2] + \gamma T. \quad (8)$$

Given a fixed decision rule $q(x)$, the optimal weight w_j^* of leaf j can be computedby setting the first-order derivatives of $\tilde{L}^{(t)}$ equal to zero, obtaining the following expression:

$$w_j^* = -\frac{G_j}{H_j + \lambda}, \quad (9)$$

where $G_j = \sum_{i \in I_j} g_i$ and $H_j = \sum_{i \in I_j} h_i$. The corresponding optimal value is calculated as

$$\tilde{L}^{(t)}(q) = -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (10)$$

The optimal XGBoost classifier is obtained with the sequential optimization at each scenario.

Benchmark Algorithms for Comparative Analysis

We implemented benchmark algorithms with their standard implementations and compared them to the XGBoost.

- Linear Regression (LR): The dependent variable is considered as categorical in nature while classification is done by calculating the error [16].
- Support Vector Regression with Linear Kernel (SVR): Attempts to minimize the generalization error bound so as to achieve generalized performance. SVR is based on the computation of a linear regression function in a high dimensional feature space where the input data are mapped via a nonlinear kernel [17].
- Bagging (Treebag): A method for generating multiple versions of a predictor to get an aggregated predictor. The aggregation averages over the versions when predicting a numerical outcome [18].
- Gradient Boosting Model (GBM): Combines the results of different prediction models by iteratively learning from the losses which occurr at the previous step through calculation of negative gradient [13].

Evaluation

The two indices, root mean square error (RMSE) and mean absolute percentage error (MAPE), are selected for evaluation of the errors. The RMSE is a frequently used measure of the differences between values predicted by a model and the values observed, which is the square root of the average of squared errors and scale-dependent. The MAPE is also a popular measure of prediction accuracy of a prediction method in statistics. The absolute value in the calculation is summed for every forecasted point in time and divided by the number of fitted points. Compared to the RMSE, the MAPE is scale-independent. The formulas for calculation are defined as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}, \quad (11)$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{y_i}, \quad (12)$$

where N is the number of data samples, $y_i (i = 1,2, \dots, N)$ is the observed number of foodborne cases and $\hat{y}_i (i = 1,2, \dots, N)$ is the predicted daily number of total foodborne disease cases.

Results

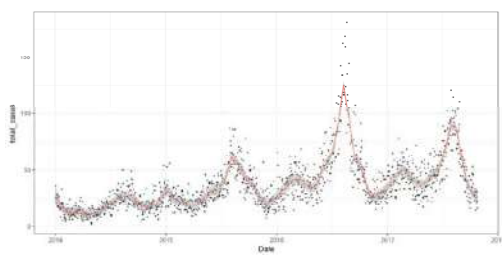
The descriptive statistics summary of the daily number of foodborne disease cases and environment factors are given in Table 1.

Table 1– Decriptive Statistics of the Data (Daily Based)

Variable	Mean	SD	Min	Max
Total foodborne cases	37.30	24.07	2.00	181.00
Sunshine(h)	4.49	4.00	0.00	12.70
Relative humidity(%)	73.17	12.56	35.00	98.00
Ground temperature(℃)	18.66	9.25	-2.70	40.20
Pressure (Pa)	1015.90	8.88	992.80	1039.70
Rainfall (mm)	3.98	12.71	0.00	155.40
Air temperature(℃)	17.71	8.55	-6.10	34.80
Wind speed(m/s)	2.55	0.92	0.50	7.40
Max wind speed(m/s)	4.97	1.22	2.50	10.40
$PM_{2.5}(\mu g/m^3)$	47.31	30.68	5.00	216.00
$PM_{10}(\mu g/m^3)$	66.39	37.06	8.00	256.00
$SO_2(\mu g/m^3)$	15.35	8.55	6.00	75.00
$CO(\mu g/m^3)$	0.80	0.27	0.40	2.20
$NO_2(\mu g/m^3)$	43.38	19.38	5.00	143.00
$O_3(\mu g/m^3)$	106.50	44.54	11.00	286.00
AQI	83.14	37.89	23.00	266.00

In time series plot of the daily number of total foodborne disease cases (Figure 1), a moderate peak is observed in 2015, with two sharp peaks followed in 2016 and 2017 respectively. Before 2015, steady growth in the daily number of total foodborne cases was recognized, which demonstrated the nonlinearity existing in foodborne disease occurrence.

Figure 1– Time Series Plot of Daily Foodborne Disease Cases

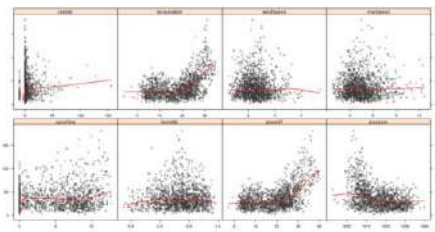


Correlation, Stationarity and Linearity Analysis Between Disease and Environment

The Pearson correlation analysis was implemented to explore the relationship between external factors and foodborne disease cases. Daily numbers of the positive cases, the laboratory testing cases and the norovirus testing cases, as well as the daily percentage of total positive cases were found to be strongly correlated with the daily number of total foodborne disease cases, with positive correlation values of 0.71, 0.65, 0.65 and 0.44 respectively. The qualitative relationships between daily foodborne disease cases and weather conditions were exploited with paired scatter plots as shown in Figure 2. For each subfigure in Figure 2, the x axis denotes the corresponding environmental factor, while the y axis represents the daily number of total foodborne disease

cases. As seen from the figure, weather conditions such as ground temperature, air temperature and pressure were highly

Figure 2- Scatter Plot Matrix Between Foodborne Disease and Environmental Factors



correlated with the daily number of total foodborne disease cases. These findings are further verified through quantitative analysis of Pearson correlation calculation, where the correlations between ground temperature, air temperature, pressure and daily number of total foodborne disease cases reached 0.48, 0.47 and -0.38 respectively. Weak correlations were also observed between air quality conditions and daily number of total foodborne disease cases. Except for O_3, SO_2, NO_2, $PM_{2.5}$, PM_{10}, CO and AQI were negatively correlated with the daily number of total foodborne disease cases, with a maximum absolute value of 0.29 found at NO_2.

The null hypothesis that the foodborne disease series was level (or trend) stationary was rejected by KPSS test results, with p value smaller than 0.01 (p_{trend}<0.01, p_{level}<0.01). The Kruskal-Wallis test also depicited that nonlinearity existed in foodborne disease occurrence with an extreme small p value (p=9.43 x 10^{-144}).

Trend Prediction with Regularization Based XGBoost

The collected foodborne disease cases and environmental datasets were employed to validate the effectiveness of regularization based XGBoost methods. The entire dataset was divided into two partitions, with observations from 1st , Jan, 2014 to 19, Jan, 2017 (1116 samples) for training and observations from 20, Jan, 2017 to 25, Oct, 2017 (279 samples) for testing. The seasonal ARIMA (SARIMA) models were first implemented to encompass the potential critical predictors for instance and the historical daily number of total foodborne disease cases, and then applied as a benchmark model without considering any environmental factors. The parameters in $SARIMA(p,d,q)(P,D,Q)_s$, which represented seasonal factors (s), autoregressive (AR) order (p), degree of differentiating (d), moving average (MA) order (q), seasonal AR order (P), seasonal differential order (D) and seasonal MA order (Q), were selected using both partial autocorrelation and iterative searching approach. Through the comparisons of RMSE and MAPE on the validation dataset (the last 20% of training samples in training dataset), the prediction results showed that $SARIMA(6,0,0)(3,0,0)_7$ achieved best generalization ability with a minimum MAPE of 18.04% on validation dataset compared with other parameter combinations. The selected model ($SARIMA(6,0,0)(3,0,0)_7$) was employed to predict the daily number of total foodborne disease occurrence in 1-7 days' advance with test dataset. The predictive performance steadily decreased when the forecasting horizon increased due to uncertainty and complexity in future foodborne disease occurrence. The best prediction was achieved at 1-day ahead prediction, with a minimum RMSE and MAPE of 11.72 and 21.34% respectively. To illustrate the effects of environmental factors and other monitoring channels of foodborne disease in foodborne disease forecasting, XGBoost method was applied to predict the cases. Before model building, the LASSO was firstly adopted to derivative of 27 external variables for selecting the most explainable features. For 1-day ahead prediction, 12 of 27 variables including daily numbers of other foodborne diseases (positive cases, laboratory testing cases, norovirus positive cases), daily percentage of total positive cases, air temperature, sunshine, humidity, wind speed, maximum wind speed, CO, year, and day of week were found to be critical in improving the performance of XGBoost. With proposed regularization based XGBoost method, we were able to obtain a RMSE of 11.71 and a MAPE of 21.15% on the same test dataset with SARIMA model, slightly improving the predictive performance.

Discussion

We further compared the daily foodborne disease forecasting performances obtained from other machine learning models including linear regression (LR), support vector regression with linear kernel (SVR), bagging (treebag), gradient boosting model (gbm), and XGBoost model. Table 2 shows the forecasting performances obtained on test dataset.

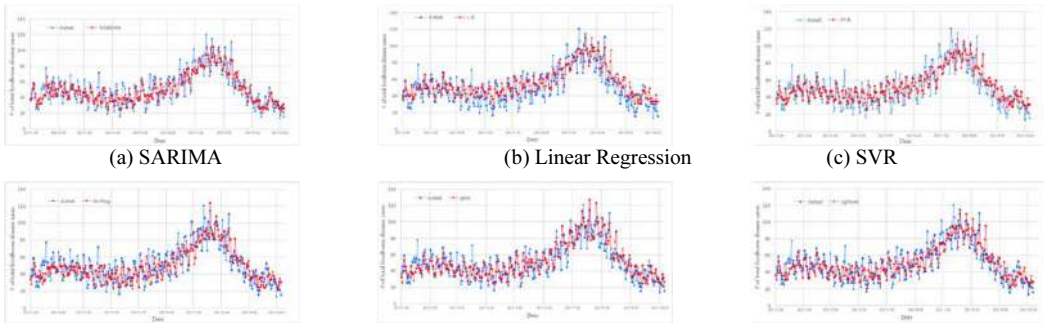
Table 2– Performance of Mutiple Models on Test Dataset

Model	Horizon	RMSE	MAPE (%)	Model	Horizon	RMSE	MAPE (%)
SARIMA	1	11.72	21.34	treebag	1	13.07	23.32
	2	12.49	22.44		2	14.72	25.02
	3	13.37	24.06		3	14.18	24.44
	4	13.98	24.42		4	14.79	25.40
	5	14.89	24.92		5	15.30	27.03
	6	15.79	26.24		6	14.95	26.21
	7	16.88	27.70		7	15.57	26.41
LR	1	12.29	25.43	gbm	1	11.85	21.09
	2	14.06	29.47		2	13.47	24.23
	3	13.61	28.16		3	12.98	23.83
	4	15.49	31.54		4	13.84	26.04
	5	14.87	30.53		5	15.45	27.44
	6	15.35	31.60		6	15.89	28.67
	7	15.78	33.80		7	16.17	27.63
SVR	1	11.65	22.78	XGBoost	1	11.71	21.15
	2	13.21	26.58		2	13.21	24.14
	3	12.88	25.89		3	12.85	25.30
	4	13.98	27.99		4	13.98	27.14
	5	14.16	28.41		5	14.21	26.60
	6	14.31	29.17		6	14.14	26.25
	7	14.51	30.77		7	13.59	25.51

Varying with the forecasting horizon, both SARIMA and XGBoost achieved a stable and robust predictive precision in terms of RMSE and MAPE. However, when the forecasting horizon increased, the predictive errors of XGBoost remained steady, but had a rising trend with SARIMA model. The performances of 1-day ahead prediction on test dataset are further provided in Figure 3 as an example, of which x axis denotes date and y axis is the daily number of foodborne disease cases. From these performances we suggest when the forecasting horizon is within 4 days, SARIMA model can be deployed for real application. When the forecasting horizon is greater than 4, regularization based XGBoost model is recommended.

Conclusion

This paper presents a regularization based model based on eXtreme gradient boosting approach for foodborne disease trend forecasting. Environmental factors including weather conditions and air quality conditions were incorporated into the prediction model and the LASSO technique was applied to shrink the predictors and identify the critical factors that would affect foodborne disease progagation. The real case demonstrated a satisfactory and robust performance in terms of 5-7 day ahead prediction, with a lowest MAPE of 25.52% for 7-day ahead forecasting. However, the model does not significantly improve the predictive performance within 4 days of environmental factors and foodborne disease.



(a) SARIMA (b) Linear Regression (c) SVR

(d) Treebag (e) GBM (f) XGBoost method

Figure 3– Foodborne Disease Prediction with Different Machine Learning Techniques

Future works still need to be done to investigate relationship between specific type of foodborne disease (e.g norovirus foodborne disease) and environmental factors. Moreover, spatial factor should also be modeled for accelerating the investigation of foodborne disease outbreak.

Acknowledgements

The authors would like to thank National Meteorological Information Center and who provided open, online access to weather and meteorological data (https://www.aqistudy.cn/)

References

[1] R.V. Tauxe, M.P. Doyle, T. Kuchenmüller, J. Schlundt, and C.E. Stein, Evolving public health approaches to the global challenge of foodborne infections, *Int J Food Microbiol* **139** (2010), S16-S28.

[2] National Institute of Diabetes and Digestive and Kidney Diseases, Foodborne Illnesses, 2014, Retrieved from https://www.niddk.nih.gov/health-information/digestive-diseases/foodborne-illnesses on 30 October 2018.

[3] WHO News, Food Safety, 2017, Retrieved from http://www.who.int/news-room/fact-sheets/detail/food-safety on 30 October 2018.

[4] D.G. Newell, M. Koopmans, L. Verhoef, E. Duizer, A. Aidara-Kane, H. Sprong, M. Opsteegh, M. Langelaar, J. Threfall, F. Scheutz, J.v. der Giessen, and H. Kruse, Food-borne diseases - The challenges of 20 years ago still persist while new ones continue to emerge, *Int J Food Microbiol* **139** (2010), S3-S15.

[5] M. Thakur, S. Olafsson, J.-S. Lee, and C.R. Hurburgh, Data mining for recognizing patterns in foodborne disease outbreaks, *J Food Eng* **97** (2010), 213-227.

[6 G.C. Cawley, N.L.C. Talbot, G.J. Janacek, and M.W. Peck, Sparse bayesian kernel survival analysis for modeling the growth domain of microbial pathogens, *IEEE Trans Neural Netw* **17** (2006), 471-481.

[7] A.R. Abdul Rasam and A.M. Mohd Noor, Contribution of GIS and remote sensing technologies for managing foodborne diseases in Malaysia, *Control and System Graduate Research Colloquium. IEEE* (2012), 258-261.

[8] H. Mohammed, I.A. Hameed, and R. Seidu, Adaptive neuro-fuzzy inference system for predicting norovirus in drinking water supply, *International Conference on Informatics, Health & Technoogy. IEEE* (*2017)*, 1-6.

[9] A.L. Hill, Norovirus outbreaks: Using agent-based modeling to evaluate school policies, *Winter Simulation Conference. IEEE* (2016),1265-1276.

[10] M.S. Park, K.H. Park, and G.J. Bahk, Combined influence of multiple climatic factors on the incidence of bacterial foodborne diseases, *Sci Total Environ* **610** (2017), 10-16.

[11] R.S. Kovats, S.J. Edwards, S. Hajat, B.G. Armstrong, K.L. Ebi, and B. Menne, The effect of temperature on food poisoning: a time-series analysis of salmonellosis in ten European countries, *Epidemiol Infect* **132** (2004), 443-453.

[12] Tibshirani, Robert, Regression shrinkage and selection via the lasso: A retrospective, *J R Stat Soc Ser B-Stat Methodol* **73** (2011), 273-282.

[13] J.H. Friedman, Greedy function approximation: A gradient boosting machine, *Ann Stat* **29** (2001),1189-1232.

[14] D. Kwiatkowski, P.C.B. Phillips, P. Schmidt, and Y. Shin, Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root, *J Econometrics* **54** (1992), 159-178.

[15] W.H. Kruskal and W.A. Wallis, Use of ranks in one-criterion variance analysis, *Journal of the American Statistical Association* **47** (1952), 583-621.

[16] K. Farid, Applied Linear Regression Models, *Technometrics*,**32**(2004), 352-353.

[17] A.J. Smola and B. Schölkopf, A tutorial on support vector regression, *Stat comput* **14** (2004), 199-222.

[18] L. Breiman, Bagging predictors, *Mach Learn* **24** (1996), 123-140.

Address for correspondence

Lili Chen, chenlili1002@126.com;
Xi Zhang, xi.zhang@coe.pku.edu.cn.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190361

Towards Personalized Lifetime Health: A Platform for Early Multimorbid Chronic Disease Risk Assessment and Mitigation

Ali Daowd[a], Syed Faizan[a], Samina Abidi[b], Ashraf Abusharekh[a], Aaqib Shehzad[a], Syed Sibte Raza Abidi[a]

[a] NICHE Research Group, Faculty of Computer Science, Dalhousie University, Halifax, Canada,
[b] Department of Community Health & Epidemiology, Dalhousie University, Halifax, Canada

Abstract

Chronic diseases are the leading cause of morbidity, disability and mortality worldwide. It is well established that the majority of chronic diseases can be prevented by targeting modifiable lifestyle-related risk factors. Thus, early risk assessment and mitigation at the individual level can significantly reduce the health and economic burden of chronic diseases. Lifetime health has emerged as a potential paradigm to assist individuals to avoid harmful lifestyle-related habits to reduce the risk of chronic morbidity. In this paper, we leverage eHealth and Quantified Self technologies, novel health data visualizations, and artificial intelligence methods to develop a digital-based lifetime health platform (PRISM) to empower individuals to self-assess, self-monitor, and self-manage their risks for multiple chronic diseases and associated morbidities.

Keywords:
Chronic Disease, Precision Medicine, Risk Assessment

Introduction

Chronic Non-Communicable Diseases (NCD) are long term medical conditions that affect the health and quality of life of millions of citizens worldwide. According to the World Health Organization, chronic diseases were responsible for 71% of all deaths worldwide in 2016 [6] . Four major chronic conditions–cardiovascular disease, diabetes, cancers, and chronic respiratory disease–are noted as the leading causes of chronic disease morbidity and mortality in almost all countries [6]. While the onset of chronic diseases manifests in adulthood [5], recent global trends reveal that younger and younger individuals are at higher risk of developing chronic diseases than all other medical conditions combined (i.e. communicable, nutritional, perinatal, and maternal conditions) [6]. The probability of premature death—i.e. between 30 and 70 years of age—from major chronic diseases ranges from 9-12% and 6-12% for men and women respectively in high income countries [6].

In addition to the enormous impact on one's health and quality of life, chronic diseases create adverse economic effects on health systems, communities, and families. It is estimated that chronic diseases will cost the global economy more than $30 trillion over the next few years [7]. As such, strategies to mitigate and prevent the onset of chronic conditions are of national and international concern, with the NCD Countdown alliance aiming for a one-third reduction in the risk of premature death by 2030 [6].

Chronic disease prevention, however, is multi-faceted due to a complex network of underlying factors that influence the risk of a disease [5]. Generally, chronic diseases stem from a wide range of modifiable lifestyle-related risk factors (diet, physical activity, smoking), and non-modifiable risk factors (age, sex, and family history) [5]. Highly prevalent chronic conditions share a range of underlying modifiable and non-modifiable risk factors; therefore, it is not uncommon to find individuals at-risk of multimorbidity. At the individual level, effective chronic disease prevention entails the early identification of risk, and then proactively mitigating these risks throughout the individual's lifetime by addressing modifiable risk factors and increasing adherence to protective lifestyle-factors (i.e. factors that protect against the risk of diseases) [5]. Another vital aspect of chronic disease prevention is continuous tracking of key health metrics, such as blood pressure, and weight.

Current preventive interventions are mostly ineffective in tackling the chronic disease pandemic; traditional health systems are reactive as opposed to being proactive, and episodic as opposed to being lifetime-based [1]. Thus, attention has turned to alternative strategies to deliver personalized and cost-effective chronic disease preventive interventions. *Lifetime Health* has emerged as a potential paradigm to empower and assist citizens to achieve long-term health goals, avoid harmful lifestyle-related habits, and thus prevent the onset of chronic diseases [8]. A central aspect of the lifetime health paradigm is the early identification and assessment of risks at the individual-level, and then proactively targeting those risks.

The objective of this research is to provide citizens with the necessary tools and resources to empower them to (a) self-assess their cumulative chronic disease risk, and (b) provide them evidence-based and personalized self-management recommendations to modify lifestyle related risk factors. We believe that by presenting personalized risk information in an intuitive and informative manner, it is possible to incentivize individuals to pursue behavior modification strategies to mitigate the onset of chronic diseases.

In this paper, we present the digital health-based *Personalized Risk Investigation, Stratification and Mitigation* (PRISM) platform that leverages Quantified Self wearable technologies, artificial intelligence methods, and health data visualizations to empower citizens to self-assess, self-monitor and self-manage their risks for chronic diseases. PRISM currently provides risk assessments for five chronic conditions (Cardiovascular disease, Diabetes, Hypertension, Stroke, COPD), and eight cancers. PRISM aims to help citizens to maintain a healthy status, and in turn help the healthcare system by reducing the cost of managing chronic diseases.

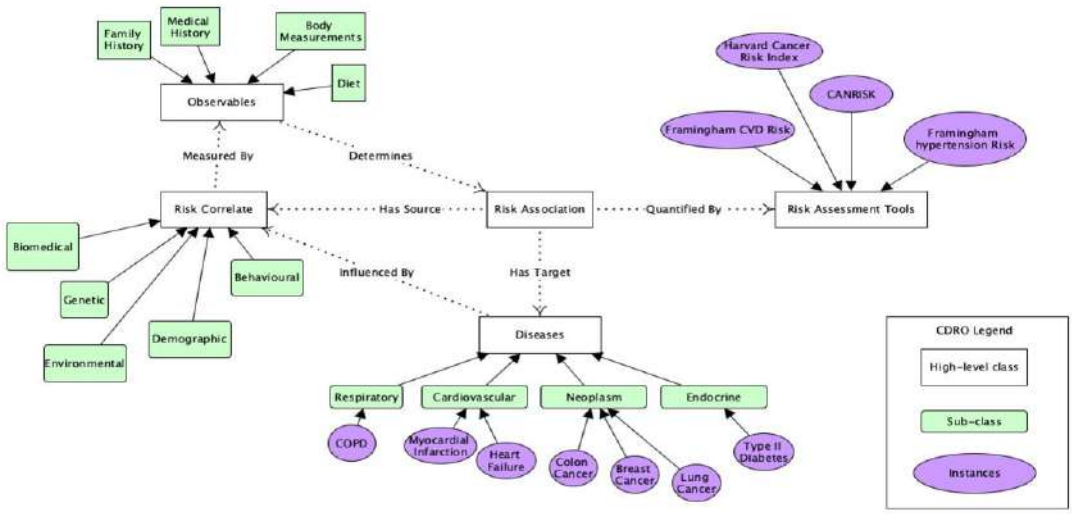


Figure 1 – Chronic Disease Risk Ontology (CDRO)

Related Work

Most existing eHealth-based tools focus on self-assessment, self-monitoring, or self-management of health or chronic disease-related risks separately with varying approaches. Web-based chronic disease risk assessment tools such as Your Disease Risk [3], and My CancerIQ [2] utilize computerized risk assessment algorithms to calculate personalized risks for multiple chronic diseases based on information gathered via short health surveys. These tools employ basic visualization techniques to communicate personalized disease risk scores and provide a set of canned self-management recommendation messages following the completion of segregated disease risk assessments. However, the abovementioned web-based tools do not store any health data and, thus, do not provide any form of self-monitoring features or lifetime health services. Deng et al. [9] on the other hand, presents a web-based health interface to facilitate collection and monitoring of health data, and assessment of a limited number of major chronic diseases. The web-based health platform, called MyHealthAvatar, takes a lifestyle and fitness focused approach and, thus, does not explicitly focus on preventing chronic disease-related risks. The CARRE project [10] is another eHealth-based tool that aims to provide personalized empowerment health services targeting cardiorenal diseases and associated comorbidities. The tool employs a knowledge repository that stores information related risk factors and risk associations specific to cardiorenal diseases. Thus, the tool is disease-specific and does not address a wide range of chronic diseases.

PRISM Design Approach

We have taken a *personalized lifetime health* approach that targets chronic disease prevention at the individual-level. The design and functional scope of PRISM is guided by the lifetime health approach [8]. In functional terms, PRISM is designed to empower citizens to (i) assess their risks for multiple chronic diseases based on personal health data integrated with socio-economic and environmental data to further personalize risk assessments, (ii) receive a cumulative health and multimorbid risk scores based on the aggregation of multiple chronic disease risk factors, (iii) monitor and visualize their personalized health risks and disease-specific risks, (iv) learn about the influence of individual risk factors towards the potential incidence of different chronic diseases, (v) modulate the influence of existing risk factors, to observe the net changes in the risk assessment scores, and (vi) plan personalized risk mitigation and behavior change strategies.

Methods

To achieve the functional scope of PRISM, we take a health knowledge management approach [4] to (i) logically model and map the mutual interactions between risk factors, chronic diseases, and associated multimorbidities; (ii) model and computerize validated chronic disease risk assessment tools; and (iii) utilize interactive health data visualizations to quantify and present personalized risk and health status information.

Chronic Disease Risk Assessment Knowledge Model

The central elements of PRISM are an evidence-based knowledge resource that represents the complex relationships between health determinants—i.e., genetics (family history and health history) and phenotypic information (physiological, behavior and psychosocial parameters), biomedical (conditions and pathology), risk factors and chronic diseases. These relationships are the basis for assessing an individual's comorbid chronic disease risk and for determining the impact of risk modification on the onset of chronic conditions.

We have developed an agile ontology-based knowledge model—referred Chronic Disease Risk Ontology (CDRO)—that represents knowledge about chronic diseases and their risks factors from several validated knowledge sources. The CDRO logically outlines the synergistic associations between risk factors (modifiable and non-modifiable), chronic diseases and multimorbidity (a subsection of CDRO is shown in Fig 1.) The initial design of CDRO is inspired by the work of Third et al. [14]. In our work, we have extended this ontology by introducing formal representation of multimorbid diseases, while focusing on a holistic and health-centered approach to the assessment and prevention of chronic disease risk. CDRO represents the following concepts as high-level classes:

- *Risk Correlates*: The underlying factors that positively or negatively influences the risk of major

chronic and multimorbid diseases. Risk correlates are organized by type (i.e. modifiable vs. non-modifiable) and category (i.e. demographic, environmental, genetic, biomedical, and behavioral). Typically, these risk correlates are surrogates for deeper biomedical mechanisms and pathologies that have a direct association with chronic diseases.

- *Diseases*: All chronic and multimorbid diseases.
- *Observables*: An individual's measurable health-related attributes, which are used to quantify exposure to risk correlates/determinants.
- *Risk Association*: Mutual association between the source of risk (i.e. risk correlate) and target of risk (i.e. chronic disease). A risk association between source risk and target risk is determined by specified observables that provide measures of the risk exposure. A risk association is quantified by the traditional probability measures used in medical literature, including Relative Risk (RR) and Odds Ratio (OR). Risk association is a triple, which consists of a risk source (i.e. risk correlate), risk target (i.e. chronic disease), and observable which is used to determine the threshold by which the risk association holds true.
- *Risk Assessment Tools*: Validated chronic disease risk assessment tools that are represented a set of executable rules encoded in a formal logic expression language. These rules quantify the risk association between a risk or protective factor and chronic disease, and between a chronic disease and associated comorbidities.

In addition to the abovementioned concepts, CDRO contains sub-concepts, ontological relations (object property assertions, data property assertions), and axioms. CDRO is flexible and can be extended to accommodate additional diseases and risk or protective factors.

CDRO serves as the evidence-based knowledge resource for PRISM to (a) assess multiple chronic disease risks; (b) integrate these risks with the individual's health determinants to calculate a co-morbid risk; (c) visualize the influence of multiple risk factors on chronic disease risks; and (d) design behavior modification plans based on personal health profile.

Computerization of Validated Risk Assessment Tools

We computerized validated risk assessment and prevention protocols for: (a) Five chronic conditions: Cardiovascular disease, Diabetes, Hypertension, Stroke, and COPD; and (b) eight cancers: Breast cancer, Colon cancer, Kidney cancer, Ovarian cancer, Melanoma, Pancreatic cancer, Uterine cancer, Lung cancer. These chronic diseases were selected as they are the most prevalent and highly interrelated. This provides us the opportunity to investigate the influence of one disease towards the onset or complication of another disease—i.e. the underlying influence of risk factors leading to co-morbidities in individuals. We used validated risk assessment algorithms such as the Framingham Cardiovascular Risk Score [13] and CANRISK for diabetes [12].

Computerization of the assessment tools involved identifying its constituent risk assessment variables—i.e. individual demographic information, lifestyle information, biomedical (and their ranges to determine acuity levels) and environmental variables. An RDF based risk assessment data model was developed to capture the patient-specific risk assessment variables and calculated risk assessment scores. It may be noted that across multiple risk assessment tools, there is a significant overlap between the risk assessment variables. We capture a variable once and use it across multiple risk assessment tools. We represented the paper-based risk assessment tools as an executable decision tree that takes as input values of its risk assessment variables and outputs a risk assessment score. Each chronic disease risk assessment tool was implemented as a unique decision tree, where these multiple decision trees were integrated along common risk assessment variables to form a risk assessment graph. We have employed a reactive approach for the execution of the computerized risk assessment tools, whereby as new health data is received it is exposed to all the computerized risk assessment tools to generate a risk assessment. A risk assessment decision tree is activated if the received data corresponds to its risk assessment data list. In this way, we ensure that risk assessment is not done selectively rather a holistic risk assessment is performed in response to the availability of current and relevant risk assessment data.

PRISM's web-interface allows users to input their own health data. Additionally, we have developed data input interfaces to capture health data from personal health devices, such as wearable devices and smart watches.

Development of PRISM Lifetime Health Platform

The PRISM platform is developed as web- and mobile- app. The PRISM architecture consists of 5 functional layers:

1. *Data layer*: responsible for capturing and integrating citizens' health data from heterogenous data sources to generate holistic and multi-dimensional personalized health profiles.
2. *Knowledge layer:* encapsulates the CDRO, together with a range of validated and computerized chronic diseases risk assessment tools. This layer uses captured personal health data to generate personalized multimorbid risk trajectories.
3. *Analytics layer:* performs a range of data analytics to analyze end-users' health profiles and generate stratified individual and multimorbid disease risk scores.
4. *Information layer:* uses advance visual analytics to present an interactive dashboard for users to visualize their personalized health status and risk for multiple chronic diseases. We use the D3.JS visual library to generate intuitive and interactive visualizations that display personalized risk information to lay individuals.
5. *Application layer* serves as the digital health interface for end-users to interact with PRISM. This layer was developed by utilizing a user-centered design approach, and consists of (i) the PRISM dashboard as a front page, which utilizes interactive visualizations to provide a summarized overview of an individual's health status and risk for multiple chronic diseases; (ii) a self-monitoring tool, to facilitate exploration of historical trends and patterns of personalized disease risks and health metrics, and to uncover correlations between various metrics by comparing data curves; (iii) a health profile logbook where end-users can input and update their personal health and lifestyle-related information.

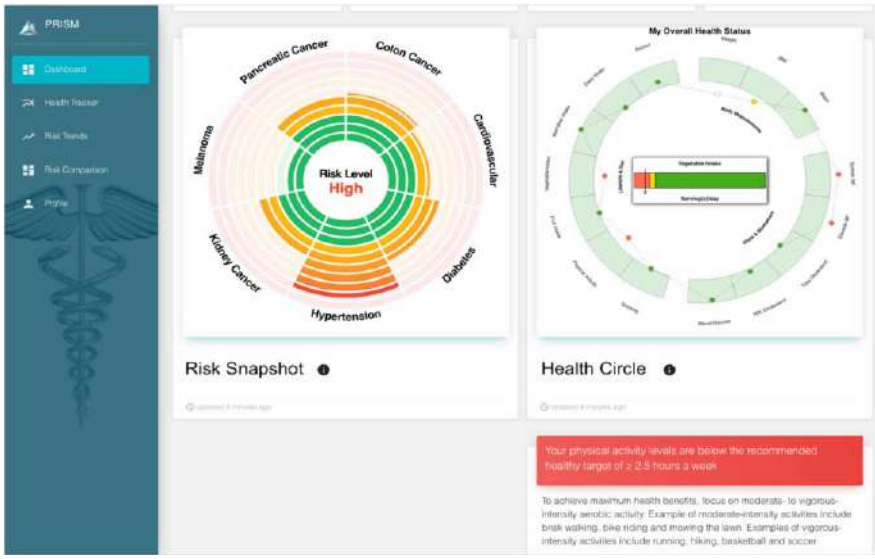


Figure 2 – PRISM Dashboard

Results

The outcome of this research is a web-based citizen empowerment eco-system that aims to provide personalized lifetime health services. Figure 2 shows a screenshot of the PRISM dashboard, with two central visualizations.

Risk Snapshot presents a stratified level of risk (i.e. low vs. moderate vs. high risk) for multiple chronic disease. The visualization is tailored to individuals' personalized health profiles, such that only diseases that exceed a pre-specified risk threshold will be revealed. The disease risk scores are standardized by PRISM's back-end system, such that all scores have the same risk measure unit and range. At the center of the visualization is a stratified level of multimorbid risk, calculated by integrating the singular disease risk levels.

Health Circle presents an overview of an individual's modifiable health metrics. The set of health metrics are organized into different categories, such as body measurements, vitals and biomarkers, and lifestyle factors. The surrounding circular green band represents the recommended healthy range. Therefore, health metrics that are outside the borders of the green band are above or below the healthy range. Furthermore, individual health metrics are color-coded to reflect their impact on an individual's health status. At the center of the circle is a horizontal bullet chart, which conveys the healthy, moderate risk, and high-risk ranges for each metric.

The designed visualizations are dynamic and interactive to encourage active exploration of personalized health data. For example, clicking on a specific disease on the Risk Snapshot will display the disease's associated modifiable risk factors on the Health Circle, while concealing other non-associated factors. Furthermore, clicking on a specific disease on the Risk Snapshot will reveal the disease's risk in numeric format (i.e. 10-year risk probabilities). This helps in providing context to the stratified risk levels.

Another example of PRISM's dashboard interactivity is when the user hovers over a specific risk factor on the Health Circle, the horizontal bullet chart appears at the center of the visualization. Such interactivity enables users to reflect on their health status. Additionally, each health metric is associated with a set of health recommendation messages tailored to individual health profiles. These messages appear below the Health Circle when the users click on a specific metric, and they are color-coded according to the risk level.

Evaluation Framework

A crucial aspect in the development of eHealth-based interventions is to determine the extent to which the target population will use it (i.e. individuals at-risk of chronic diseases). Therefore, the PRISM lifetime health platform will undergo a mixed-methods evaluation study to determine usability of the platform. Participants will be recruited to interact with PRISM using pre-defined goal-based scenarios. While participating in this exercise, participants will be encouraged to think-aloud (i.e. verbally express their thoughts about the platform). Furthermore, interactions with the platform will be recorded to understand what features participants are interested in. Quantitative and qualitative data collected will be used to determine usability of the platform and identify the factors that impact usability from end-user's perspective. Evaluation of usability and functionality of PRISM will provide valuable outcomes to guide the development of subsequent stages of the platform.

Discussion and Concluding Remarks

Personalized and proactive chronic disease preventive strategies represent the most effective approach towards reducing the health and economic burden of chronic diseases at the individual- and population-levels. In this research we have taken a novel approach to demonstrate the feasibility of delivering personalized and preventive lifetime health services to empower citizens to self-asses, self-monitor, and self-manage their health risks via an eHealth-based intervention. PRISM utilizes a health-centered, as opposed to disease-centered, approach to prevent the onset of chronic diseases and associated morbidities. The majority of existing eHealth-based chronic disease interventions are either directed towards chronic disease patients exclusively or individuals at-risk of developing a chronic condition. However, we take a population-wide approach, such that PRISM is intended to target individuals at *any* level of risk. Such interventions are underused in practice

yet represent an effective approach towards chronic disease prevention. Focusing preventive interventions on a subset of the population (i.e. high-risk individuals) is often suggested as a reason for the inefficient current preventive care practices. Therefore, in addition to presenting a novel solution to chronic disease prevention, our research aims to overcome the shortcomings that exist in current preventive health practices due to lack of population-wide intervention strategies.

A contribution of this research is the development of a scalable knowledge model that illustrates the impact of specified risk factors on multiple chronic diseases. Our knowledge model extends the work done by Third et al. by further analyzing the mutual relations between risk/protective factors, chronic diseases, and associated comorbidities. The knowledge model encapsulates a set of validated and evidence-based rules to quantify these complex and interrelated associations in terms of conventional risk measures, such as RRs and ORs.

From a public health perspective, PRISM can be described as an integrated health platform that has the capability of disseminating evidence-based knowledge to the public. Knowledge dissemination in public health is a formal and structured process that aims to spread health-related knowledge to improve population health outcomes. The literature indicates an urgent need for innovative programs that aim to disseminate knowledge regarding chronic disease prevention and management [11]. We believe that PRISM can help in closing this gap by: (i) increasing access of knowledge to various of target groups, including individuals at low, intermediate and high risk of developing chronic diseases; (ii) utilizing the interactive visualizations incorporated to translate preventive care knowledge to lay citizens; and (iii) displaying tailored recommendations by linking preventive care knowledge to citizens' personalized risk profiles.

Future work will include a re-evaluation of features, functionalities and contents of PRISM based on the results of the evaluation study and participants' feedback.

In conclusion, PRISM is an innovative digital health platform that is designed to empower citizens to avoid chronic morbidity and maintain a healthy, disease-free life. Additionally, PRISM offers the potential to help shift outdated reactive health services to personalized and predictive.

References

[1] Transforming health: Shifting from reactive to proactive and predictive care, in: *Transforming Health Market Insights*, https://www.marsdd.com/news/transforming-health-shifting-from-reactive-to-proactive-and-predictive-care/.

[2] My Cancer IQ, in, https://www.mycanceriq.ca 2018.

[3] Your Disease Risk - Siteman Cancer Center, in, https://siteman.wustl.edu/prevention/ydr/ 2018.

[4] S.S.R. Abidi, Healthcare knowledge management: the art of the possible, in: *AIME'07 Proceedings of the 2007 conference on Knowledge management for health care procedures*, Springer, Berlin, Heidelberg, Amsterdam, The Netherlands, 2007, pp. 1-20.

[5] J. Baird, C. Jacob, M. Barker, C.H.D. Fall, M. Hanson, N.C. Harvey, H.M. Inskip, K. Kumaran, and C. Cooper, Developmental Origins of Health and Disease: A Lifecourse Approach to the Prevention of Non-Communicable Diseases, *Healthcare* **5** (2017).

[6] J.E. Bennett, G.A. Stevens, C.D. Mathers, R. Bonita, J. Rehm, M.E. Kruk, L.M. Riley, K. Dain, A.P. Kengne, K. Chalkidou, J. Beagley, S.P. Kishore, W. Chen, S. Saxena, D.W. Bettcher, J.T. Grove, R. Beaglehole, and M. Ezzati, NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards Sustainable Development Goal target 3.4, *The Lancet* **392** (2018), 1072-1088.

[7] D. Bloom, E. Cafiero, E. Jané-Llopis, S. Abrahams-Gessel, L. Bloom, S. Fathima, A. Feigl, T. Gaziano, A. Hamandi, M. Mowafi, D. O'Farrell, E. Ozaltin, A. Pandya, K. Prettner, L. Rosenberg, B. Seligman, A. Stein, C. Weinstein, and J. Weiss, The Global Economic Burden of Noncommunicable Diseases, in: *Program on the Global Demography of Aging.*, 2012.

[8] A. Daowd, S.R. Abidi, A. Abusharekh, and S.S.R. Abidi, A Personalized Risk Stratification Platform for Population Lifetime Healthcare, *Studies in health technology and informatics* **247** (2018), 920-924.

[9] Z. Deng, Y. Zhao, F. Parvinzamir, X. Zhao, H. Wei, M. Liu, X. Zhang, F. Dong, and E. Liu, MyHealthAvatar: A Lifetime Visual Analytics Companion for Citizen Well-being, in: *International Conference on Technologies for E-Learning and Digital Entertainment, Edutainment 2016: E-Learning and Games*, Springer, 2016, pp. 345-356.

[10] E. Kaldoudi, A. Third, G. Gotsis, S. Roumeliotis, N. Karvelyte, N.L. Rimsevicius, K. Pafili, D. Papazoglou, G. Juozalenaite, E. Semertzidou, Z. Visockiene, and K. Zigeridou, D.2.2: Functional Requirements & CARRE Information Model, in: *FP7-ICT-611140 CARRE*, European Commission under the Information and Communication Technologies (ICT) 7th Framework Programme, https://www.carre-project.eu/project-info/deliverables-2/, 2014.

[11] I. Poureslami, L. Nimmon, I. Rootman, and M.J. Fitzgerald, Health literacy and chronic disease management: drawing from expert knowledge to set an agenda, *Health Promotion International* **32** (2017), 743-754.

[12] C.A. Robinson, G. Agarwal, and K.A. Nerenberg, Validating the CANRISK prognostic model for assessing diabetes risk in Canada's multi-ethnic population, *Chronic diseases and injuries in Canada* **32** (2011), 19-31.

[13] R.B.D.A. Sr., R.S. Vasan, M.J. Pencina, P.A. Wolf, M. Cobain, J.M. Massaro, and W.B. Kannel, General Cardiovascular Risk Profile for Use in Primary Care: The Framingham Heart Study, *Circulation* **117** (2008), 743-753.

[14] A. Third, E. Kaldoudi, G. Gkotsis, S. Roumeliotis, K. Pafili, and J. Domingue, Capturing Scientific Knowledge on Medical Risk Factors, in: *8th International Conference on Knowledge Capture (K-CAP 2015)*, Palisades, New York, 2015.

Address for correspondence

Ali Daowd and Syed Sibte Raza Abidi, Faculty of Computer Science, Dalhousie University, E-mail: ali.daowd@dal.ca & ssrabidi@dal.ca

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190362

Integration of FHIR to Facilitate Electronic Case Reporting: Results from a Pilot Study

Brian E Dixon[a,b,c], David E Taylor[b], Myung Choi[d], Michael Riley[d], Trey Schneider[d], Jon Duke[d]

[a] *Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, Indiana, USA,*
[b] *Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, Indiana, USA,*
[c] *Center for Health Information and Communication, Department of Veterans Affairs, HSR&D Service, Indianapolis, IN, USA,*
[d] *Center for Health Analytics and Informatics, Georgia Tech Research Institute, Atlanta, GA, USA*

Abstract

Current approaches to gathering sexually transmitted infection (STI) case information for surveillance efforts are inefficient and lead to underreporting of disease burden. Electronic health information systems offer an opportunity to improve how STI case information can be gathered and reported to public health authorities. To test the feasibility of a standards-based application designed to automate STI case information collection and reporting, we conducted a pilot study where electronic laboratory messages triggered a FHIR-based application to query a patient's electronic health record for details needed for an electronic case report (eCR). Out of 214 cases observed during a one week period, 181 (84.6%) could be successfully confirmed automatically using the FHIR-based application. Data quality and information representation challenges were identified that will require collaborative efforts to improve the structure of electronic clinical messages as well as the robustness of the FHIR application.

Keywords:

Health Information Exchange, Disease Notification, Sexually Transmitted Diseases

Introduction

Sexually Transmitted Infections (STIs)

Undiagnosed and untreated sexually transmitted infection (STI) is associated with adverse outcomes such as infertility, increased HIV transmission and acquisition, and adverse pregnancy outcomes. Several STI health services are recommended by the U.S. Centers for Disease Control and Prevention (CDC) to protect the reproductive and sexual health of young men and women, including annual chlamydia and gonorrhea screening of sexually active women ≤ 24 years of age, pregnant women, and older at-risk women; chlamydia and gonorrhea screening of anatomic sites of exposure (urethral, rectal, or pharyngeal) of men who have sex with men (MSM); retesting of all infected persons after treatment for chlamydia or gonorrhea; and syphilis testing of pregnant women as well as sexually active MSM [6].

Surveillance of STIs and STI Services

Surveillance, a cornerstone of public health [23; 29], is the routine assessment of disease prevalence and burden as well as the utilization of health care services. Ministries of health seek to perform surveillance on a range of diseases including STIs. For example, most ministries seek to monitor the quality of STI health services received by at-risk groups such as adherence to recommendations for chlamydia and gonorrhea testing and retesting, syphilis testing, test results, patient and partner treatment, and the incidence of adverse STI outcomes.

Surveillance of STIs relies upon physicians and laboratories to manually, spontaneously report STI cases to public health authorities [2]. However, passive approaches are known to be burdensome for reporters, producing incomplete and delayed reports which can hinder the assessment of disease in the community and potentially delay the recognition of patterns and outbreaks [17; 20; 28]. For example, in a recent analysis of STI cases laboratories reported between 63.1% and 71.7%, and physicians reported between 6.3% and 44.4%, of syphilis, gonorrohea, and chlamydia cases [15].

Electronic Reporting of STIs

While most U.S. health agencies continue to publish offiical paper-based forms for STI case reporting [10; 16], surveillance practice is evolving towards electronic methods for data capture. The adoption of electronic health record (EHR) systems and health information exchange (HIE) among clinical organizations and systems [3; 4], driven by policies like the 'meaningful use' program in the United States [7], is creating an information infrastructure that public health organizations can leverage for improving surveillance practice [9].

To date, the focus of modernizing STI reporting has been on the implementation of electronic laboratory reporting (ELR). ELR messages utilize HL7 (Health Level 7) Version 2 standards to encode information about tests ordered and test results pertaining individual patients. The rapid adoption of ELR over the past decade now enables over two-thirds of health departments in the U.S. to improve the surveillance of STIs and other conditions [22]. Yet there are key data missing from ELR messages that public health agencies need to investigate STI cases. For example, at the time the lab result is electronically delivered to the physician, the ELR message does not contain the treatment to be prescribed by the physician. Therefore public health authorities need case information from providers beyond what is available in the initial ELR message.

To access complete information on STI cases, public health authorities seek to implement electronic case reporting (eCR) where case reports from providers are generated or submitted electronically. The goal is to leverage EHR systems and HIE networks to facilitate eCR. Although desired, there exist few standardized methods to support eCR within commercial EHR systems and few existing implementations of eCR.

Research Objective

Given the need for better community-level surveillance of STIs and limited experiences with eCR, we sought to develop and test a standards-based eCR service within the context of

an existing HIE network. The goal was to establish the feasibility of such an approach to support public health work.

Methods

To examine whether eCR processes could be automated, we implemented and tested a standards-based application within an existing HIE network. The application received ELR messages indicating a positive lab result for chlamydia or gonorrhea and returned a completed eCR report with case information extracted from the patient's EHR. The completed eCR reports were stored in a local database to enable analysis for the study, but this repository could be used to transmit completed reports to a public health authority. Our work received approval from the Institutional Review Board (IRB) at Indiana University.

Geography and Population Information

The State of Indiana ranks 15th among U.S. states by population with just under 6.5 million residents, according to the 2010 census. Consistent with national data, minority race and ethnicity are over-represented in STIs. For example, the 2015 rate of gonorrhea among black (African-American) individuals was 836/100,000 people compared to the rate among whites (Caucasian) of 87.7 and for Hispanic individuals of 85.0. The rates for Chlamydia were 2234 for black, 319 for white, and 545 for Hispanic, and the rates for primary and secondary syphilis were 26.8, 6.6, and 16.6, respectively.

The Indiana State Department of Health (ISDH) STD Control Program divides the state's 92 counties into ten districts for morbidity reporting and disease intervention purposes. These district offices are the recipients of contracts with the STD Program for the state's approximately 30 disease intervention specialists. The Marion County Public Health Department (MCPHD) STD Control Program has responsibility for STD reporting in District 5, which includes Marion County (Indianapolis) and the seven surrounding counties: Boone, Hamilton, Hancock, Hendricks, Johnson, Morgan, and Shelby. This district makes up the majority of the Indianapolis MSA. District 5 (population of 1.7 million), and Marion County (population of 903,393) always account for the largest share of Indiana's STI morbidity. In 2015, District 5 accounted for 39% of the state's chlamydia and 47% of the state's gonorrhea morbidity. This reflects, in part, racial health disparities in the district which is substantially more diverse than the state.

According to the CDC's 2015 STD Surveillance Report, Indiana reported a total of 28,886 cases of Chlamydia and ranked 27th among states in rate (437.9/100,000) while Marion County ranked 25th among U.S. counties and independent cities at 949.3 cases/100,000 people. Indiana is ranked 23rd among states for gonorrhea with a case rate of 118.9/100,000 people, while Marion County is ranked 16th among U.S. counties and independent cities in the rate of gonorrhea cases with 344.1 cases/100,000 people.

Indiana Network for Patient Care

The Indiana Network for Patient Care (INPC) is one of the largest community-based HIE networks in the United States [24]. The INPC connects 117 hospitals representing 38 health systems with physician practices, long-term post-acute care facilities, laboratories, and radiology centers. The INPC maintains nearly six billion structured observations for over 12 million individuals. Nearly two million electronic health care transactions are processed every day.

Since 2000, the INPC has leveraged electronic laboratory messages sent from hospitals to automate the reporting of notifiable disease information to public health authorities. Using a technology dubbed the 'Notifiable Condition Detector' or 'NCD,' developed by the Regenstrief Institute, the INPC examines each incoming electronic lab message to determine if the results should be reported to public health authorities. In other words, the NCD is how the INPC facilitates ELR. In prior studies, the NCD was shown to have good sensitivity and specificity as well as improve the completeness and timeliness of public health reporting processes [11; 18]. This study leveraged the NCD to identify positive lab tests for chlamydia and gonorrhea sent during the study period. Specifically, the NCD identified tests from a value set defined in CDC case definitions and published by the Public Health Informatics Institute [26].

A FHIR-based Service for eCR

In partnership with the Georgia Tech Research Institute, the Regenstrief Institute implemented a FHIR-based application within the INPC. FHIR (Fast Healthcare Internet Resources) is an emerging HL7 standard that seeks to expose discrete health data through web services [19]. Using a FHIR-compliant server, organizations can expose health data as FHIR resources to external applications that can use requested resources to perform various functions. FHIR services have been integrated into existing EHR platforms like OpenMRS [21], i2b2 [25], and OMOP [1; 8].

For this study, the Regenstrief Institute installed a FHIR-based eCR application developed by Georgia Tech, entitled the Public Health Case Reporting (PHCR) platform. The application receives as input HL7-compliant ELR messages (Verion 2.5.1 Observation Result messages) from the NCD. These messages represent positive lab results for individuals tested for chlamydia and/or gonorrhea. The positive lab test messages trigger the PHCR application to query a previously-implemented FHIR-compliant server that exposes INPC data as resources for additional details about the disease case. The FHIR service running on top of the INPC provides the requested resources which are used by the PHCR application to populate an eCR along with the data from the original ELR message. These data are stored in a local database that permit the eCR to be submitted to a public health authority. Georgia Tech further developed a Web-based dashboard that enables the eCR data to be visualized from the database, which is useful for testing purposes as well as quality control. Source code for the project is available via GitHub in two distinct repositories: https://github.com/gt-health/ecr_manager and https://github.com/gt-health/PACER

The architecture implemented for the pilot study is depicted in **Figure 1**. The application developed by Georgia Tech is labeled as the PHCR Controller. The FHIR-based service that interacts with the INPC is labelled as the FHIR Controller. Messages from the NCD are fed into the PHCR Controller using a HL7 Version 2.5.1 Receiver. The Dashboard is a web application that displays eCR records stored in the local database connected to the PHCR Controller.

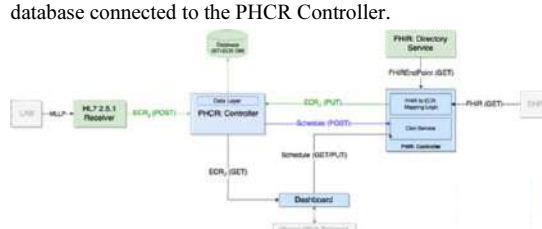


Figure 1– Architecture.

Once the FHIR-based service in the INPC returns a completed eCR, the report data are stored in a local database. Using a web application developed by Georgia Tech, the completed eCR reports can be viewed in a web browser. Figure 2 depicts part of a completed eCR report in a web browser for a test patient. The eCR contains details on the patient, guardian (if under 18 years of age), diagnosis, medications, and lab results. Also available are data on symptoms, health care facility, provider, clinic visits, travel history, and immunization history. These are the data elements important to disease investigators at public health authorities.

Figure 2 – Screenshot of electronic case report viewer application showing test patient information, including demographics, diagnosis, and laboratory results.

Data Collection and Analysis

Once implemented within the INPC, the FHIR-based eCR service was tested for one week (November 30, 2017 to December 6, 2018). Data were collected from the incoming ELR messages received by the INPC as well as the eCR reports generated by the FHIR-based service. Patient details, confirmatory lab test details, and corresponding ICD diagnoses from the eCR were collected to ensure that the correct linkages were made between initial ELR and final eCR for a given patient. Error logs were captured to identify issues with the service as well as potential mismatches between patients identified in the ELR messages and known patients in the INPC.

A descriptive analysis was performed to summarize the results of the pilot test. The throughput of the service was calculated along with general descriptions of the population with a positive STI observed during the pilot period. R (version 3.4.3) was used to calculate descriptive statistics and ggplot was used to create histograms and bar charts.

Results

A total of 214 ELR messages were received by the INPC from 16 health systems during the pilot test period. All (100%) ELRs were correctly matched to a patient's longitudinal medical record in the INPC via the FHIR service.

A date of disease onset was confirmed in the patient's medical record using ICD diagnosis codes for only 181 (84.6%) patients, enabling the FHIR service to return a completed eCR to the public health agency. Additional errors included:

- 5 (2.3%) ELR messages were missing test dates; and
- 4 (1.9%) ELR messages had phone numbers in an invalid format.

The distribution of patient age, stratified by gender, is depicted in Figure 3. Overall there were more females (N=157) diagnosed with an STI than males (N=57). However, the median age for both groups was similar (22 years for females and 23 years for males).

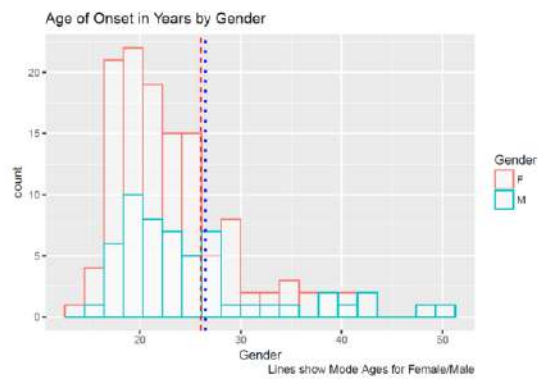


Figure 3 – Distribution of age by gender for those with a positive chlamydia or gonorrhea test.

The most prevalent laboratory test codes are summarized in **Table 1**. There were a total of 680 test results observed as several ELR messages contained multiple test results. Of the top seven lab tests observed, five were identified in the ELR message using the LOINC coding standard and two were identified using local lab codes (MIDAM is a regional lab located in Indianapolis, Indiana).

Table 1 – Prevalent laboratory test codes observed in electronic lab messages exchanged using the FHIR service.

Laboratory Test Code	Laboratory Test Code System	Laboratory Test Description	Count
21613-5	LOINC	*Chlamydia trachomatis* DNA	117
31208-2	LOINC	Specimen source identified	94
970000571	MIDAM	*Chlamydia trachomatis+ Neisseria gonorrhoeae* rRNA	93
4993-2	LOINC	*Chlamydia trachomatis* rRNA	89
5028-6	LOINC	*Neisseria gonorrhoeae* rRNA	89
10001637	MIDAM	*Chlamydia trachomatis* rRNA	52
24111-7	LOINC	*Neisseria gonorrhoeae* DNA	49

Discussion

In a pilot study to establish the feasibility of a standards-based approach to automate the collection of information in support of electronic case reporting for public health surveillance, we implemented a FHIR-based application that could query an HIE network for data necessary for eCR work processes. Real-world ELR messages for positive cases of chlamydia and gonorrhea were transmitted to the application over the course of one week. The application successfully queried the FHIR-based service at the HIE for 100% of lab positive results. For a high proportion (85%) of cases, the application could automate completion of the eCR for transmission to a public health authority. Therefore the pilot project established strong feasibility for automating eCR information flows, which has the potential to save time and cost for the health system as most eCR processes currently rely on clinical and public health personnel to call, fax, or manually enter information to and from organizations. Although feasible, we recognize that additional testing, refinement and study of FHIR-based approaches will be necessary to implement and scale eCR applications to automate information capture.

Although the pilot was considered successful, some errors challenged the application. ELR messages, the input that triggered the eCR process, were missing test dates in a small proportion (2.3%) of cases. In other cases (1.9%), the patient's phone number was improperly formatted. Data quality issues such as completeness and improper data representation are common in health care as documented in prior studies [12; 31]. Similarly, one quarter (25%) of the most common lab results were encoded using a local laboratory information system terminology as opposed to the internationally recognized standard LOINC. This challenge has also been observed in prior examinations of routine ELR messages sent to public health organizations [13]. These data quality and standardization challenges require work to make solutions like the PHCR application more reliable across the wide variation of ELR data feeds found in the health care system.

While a high proportion (85%) of cases were confirmed using EHR data returned from the INPC, several patient records were missing an ICD-based diagnosis that the eCR application requires to confirm a positive case of disease. Since each patient had a positive, confirmatory laboratory result for one of the two target diseases, these patients should have the respective disease documented in their EHR. The most likely reason why this diagnosis was missing from the EHR is clinic workflow as the lab result was reported to clinicians after the patient was no longer in the clinic or the emergency department and therefore clinic staff did not go into the EHR to update the record.

Tackling the data quality and standardization challenges will enable applications like the PHCR to better automate public health reporting processes. Health care organizations, information system vendors, and HIE networks can and should work to ensure that data are complete and properly represented in electronic messages using available health information standards. Solutions like terminology mapping exist to support efforts at improving data standardization [5]. Efforts also exist to support data quality improvements [14; 30].

Furthermore, applications like the PHCR need to be flexible and adapt to information feeds that may not perfectly provide all of the data necessary to trigger a case report for public health. This may require public health organizations to relax the rules for confirming a case, or application developers may need to configure software to enable eCR information moving forward even if the report is not complete. Thus we all have work to do in order to make applications like PHCR robust.

Currently the CDC, with support from the Robert Wood Johnson Foundation, is conducting a pilot program to test a 'digital bridge' between clinical and public health organizations for notifiable disease reporting in conjunction with the meaningful use program [27]. This project could provide a method for scaling automated eCR approaches beyond what we tested in this study. However, this project has not yet published early findings or preliminary results. More implementation and evaluation of these efforts will be required to achieve adoption rates as high as ELR. Furthermore, public health organizations should investigate policy drivers that may encourage eCR application adoption by health systems.

Conclusions

A pilot study to examine the implementation of a standards-based approach to support electronic case reporting for public health demonstrated feasibility. While successful, the pilot study identified errors and challenges that need to be addressed before a FHIR-based approach to electronic case reporting can be implemented and scaled across the health system. Technical and workflow improvements will be required to facilitate broad adoption of standards-based eCR in support of public health.

Acknowledgements

The research reported in this publication was supported by the Centers for Disease Control and Prevention (CDC) and U.S. Department of Health and Human Services (HHS). The Department of Veterans Affairs (VA), Veterans Health Administration, Health Services Research and Development Service (CIN 13-416) further employs Dr. Dixon as a Health Research Scientist at the Richard L. Roudebush VA Medical Center in Indianapolis, Indiana, USA. The content is solely the responsibility of the authors and does not necessarily represent the official views of CDC, HHS, or VA. Gratitude to Dr. Ninad Mishra and Greg Pierce for their guidance in this work.

References

[1] GT-FHIR2, the OMOP on FHIR Project, (n.d.). http://omoponfhir.org/ (accessed March 27, 2019).

[2] Centers for Disease Control and Prevention, Progress in improving state and local disease surveillance--United States, 2000-2005, *MMWR Morb Mortal Wkly Rep* **54** (2005) 822-825.

[3] J. Adler-Milstein, D.W. Bates, and A.K. Jha, Operational health information exchanges show substantial growth, but long-term funding remains a concern, *Health Aff (Millwood),* **32** (2013) 1486-1492. doi: 10.1377/hlthaff.2013.0124.

[4] J. Adler-Milstein, C.M. DesRoches, P. Kralovec, G. Foster, C. Worzala, D. Charles, T. Searcy, and A.K. Jha, Electronic Health Record Adoption In US Hospitals: Progress Continues, But Challenges Persist, *Health Aff (Millwood),* **34** (2015) 2174-2180. doi: 10.1377/hlthaff.2015.0992.

[5] J.M. Alyea, B.E. Dixon, J. Bowie, and A.S. Kanter, Standardizing Health-Care Data Across an Enterprise, in: *Health Information Exchange: Navigating and Managing a Network of Health Information Systems,* B.E. Dixon, ed., Academic Press, Waltham, MA, 2016, pp. 137-148.

[6] Centers for Disease Control and Prevention, 2015 Sexually Transmitted Diseases Treatment Guidelines, (2015). https://www.cdc.gov/std/tg2015/default.htm (accessed March 27, 2019).
[7] Centers for Medicare and Medicaid Services, Meaningful Use, in: Centers for Medicare & Medicaid Services, Baltimore, MD, 2013.
[8] M. Choi, R. Starr, M. Braunstein, and J. Duke, OHDSI on FHIR Platform Development with OMOP CDM mapping to FHIR Resources, in: *OHDSI Symposium*, Observational Health Data Sciences and Informatics, Washington, DC, 2016.
[9] B. Dixon and S. Grannis, Public Health Informatics Infrastructure, in: *Public Health Informatics and Information Systems,* J.A. Magnuson and J.P.C. Fu, eds., Springer, London, 2014, pp. 69-88.
[10] B.E. Dixon, J.F. Jones, and S.J. Grannis, Infection preventionists' awareness of and engagement in health information exchange to improve public health surveillance, *Am J Infect Control,* **41** (2013) 787-792. doi: 10.1016/j.ajic.2012.10.022.
[11] B.E. Dixon, J.J. McGowan, and S.J. Grannis, Electronic laboratory data quality and the value of a health information exchange to support public health reporting processes, *AMIA Annu Symp Proc,* **2011** (2011) 322-330.
[12] B.E. Dixon, M. Rosenman, Y. Xia, and S.J. Grannis, A vision for the systematic monitoring and improvement of the quality of electronic health data, *Stud Health Technol Inform,* **192** (2013) 884-888.
[13] B.E. Dixon, D.J. Vreeman, and S.J. Grannis, The long road to semantic interoperability in support of public health: experiences from two states, *J Biomed Inform,* **49** (2014) 3-8. doi: 10.1016/j.jbi.2014.03.011.
[14] B.E. Dixon, C. Wen, T. French, J. Williams, and S.J. Grannis, Advanced Visualization and Analysis of Data Quality for Syndromic Surveillance Systems, *Online J Public Health Inform,* **10** (2018) e28. doi: 10.5210/ojphi.v10i1.8351
[15] B.E. Dixon, Z. Zhang, P.T.S. Lai, U. Kirbiyik, J. Williams, R. Hills, D. Revere, P.J. Gibson, and S.J. Grannis, Completeness and timeliness of notifiable disease reporting: a comparison of laboratory and provider reports submitted to a large county health department, *BMC Medical Informatics and Decision Making,* **17** (2017) 87. doi: 10.1186/s12911-017-0491-8.
[16] S.M. Downs, V. Anand, M. Sheley, and S.J. Grannis, The last mile: Using fax machines to exchange data between clinicians and public health, *Online Journal of Public Health Informatics,* **3** (2011) 1-14. doi: 10.5210/ojphi.v3i3.3892.
[17] T.J. Doyle, M.K. Glynn, and S.L. Groseclose, Completeness of notifiable infectious disease reporting in the United States: an analytical literature review, *Am J Epidemiol,* **155** (2002) 866-874.
[18] M. Fidahussein, J. Friedlin, and S. Grannis, Practical Challenges in the Secondary Use of Real-World Data: The Notifiable Condition Detector, *AMIA Annu Symp Proc,* **2011** (2011) 402-408.
[19] Health Level 7, FHIR Overview, in, HL7.org, 2017.
[20] R.A. Jajosky and S.L. Groseclose, Evaluation of reporting timeliness of public health surveillance systems for infectious diseases, *BMC Public Health,* **4** (2004) 29.
[21] S.N. Kasthurirathne, B. Mamlin, G. Grieve, and P. Biondich, Towards Standardized Patient Data Exchange: Integrating a FHIR Based API for the Open Medical Record System, *Stud Health Technol Inform,* **216** (2015) 932.
[22] E. Lamb, J. Satre, G. Hurd-Kundeti, B. Liscek, C.J. Hall, R.W. Pinner, L. Conn, J. Zajac, M. Smallwood, and K. Smith, Update on progress in electronic reporting of laboratory results to public health agencies - United States, 2014, *MMWR Morb Mortal Wkly Rep,* **64** (2015) 328-330.
[23] L.M. Lee and S.B. Thacker, The cornerstone of public health practice: public health surveillance, 1961--2011, *MMWR. Surveillance summaries : Morbidity and mortality weekly report. Surveillance summaries / CDC,* **60 Suppl 4** (2011) 15-21.
[24] J.M. Overhage, The Indiana Health Information Exchange, in: *Health Information Exchange: Navigating and Managing a Network of Health Information Systems,* B.E. Dixon, ed., Academic Press, Waltham, MA, 2016, pp. 267-279.
[25] N. Paris, M. Mendis, C. Daniel, S. Murphy, X. Tannier, and P. Zweigenbaum, i2b2 implemented over SMART-on-FHIR, *AMIA Jt Summits Transl Sci Proc,* **2017** (2018) 369-378.
[26] Public Health Informatics Institute, Advancing Electronic Case Reporting (eCR) of Sexually Transmitted Infections (STIs), Decatur, GA, 2016.
[27] Public Health Informatics Institute, Digital Bridge, (2017). https://digitalbridge.us/infoex/ (accessed March 27, 2019).
[28] B.J. Silk and R.L. Berkelman, A review of strategies for enhancing the completeness of notifiable disease reporting, *J Public Health Manag Pract,* **11** (2005) 191-200.
[29] S.B. Thacker, J.R. Qualters, and L.M. Lee, Public health surveillance in the United States: evolution and challenges, *MMWR. Surveillance summaries : Morbidity and mortality weekly report. Surveillance summaries / CDC,* **61** (2012) 3-9.
[30] N.G. Weiskopf, S. Bakken, G. Hripcsak, and C. Weng, A Data Quality Assessment Guideline for Electronic Health Record Data Reuse, *EGEMS (Wash DC),* **5** (2017) 14. doi: 10.5334/egems.218.
[31] A. Wright, A.B. McCoy, T.T. Hickman, D.S. Hilaire, D. Borbolla, W.A. Bowes, 3rd, W.G. Dixon, D.A. Dorr, M. Krall, S. Malholtra, D.W. Bates, and D.F. Sittig, Problem list completeness in electronic health records: A multi-site study and assessment of success factors, *Int J Med Inform,* **84** (2015) 784-790. doi: 10.1016/j.ijmedinf.2015.06.011.

Address for correspondence

Brian E. Dixon, PhD
Center for Biomedical Informatics
Regenstrief Institute, Inc.
1101 W. 10th St., RF 336
Indianapolis, IN 46202 USA
+1 (317) 278-3072

bedixon@regenstrief.org

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190363

Behavioural Phenotyping of Daily Activities Relevant to Social Functioning Based on Smartphone-Collected Geolocation Data

Paolo Fraccaro[b], Stuart Lavery-Blackie[a], Sabine N Van der Veer[a], Niels Peek[a,c]

[a] *Centre for Health Informatics, Division of Informatics, Imaging and Data Sciences, Manchester Academic Health Science Centre, The University of Manchester, Manchester, UK*
[b] *IBM Research UK, Sci-Tech Daresbury Campus, Daresbury, UK*
[c] *NIHR Manchester Biomedical Research Centre, Manchester Academic Health Science Centre, The University of Manchester, Manchester, UK*

Abstract

Smartphones offer new opportunities to monitor health-related behaviours in the real world. This allows researchers to go beyond traditional data collection methods, such as interviews and questionnaires that suffer from recall bias and low spatio-temporal resolution. In this study, we present an experiment that uses advanced analytical methods to identify daily activities relevant to assess social functioning, from geolocation data. Twenty-one healthy volunteers used a smartphone to continuously record their GPS location for up to 10 days. Participants also completed a diary to record their daily activities that was used as ground truth. Using clustering algorithms and semantic enrichment methods we were able to predict these activities from the GPS data with a precision of 0.75 (standard deviation [SD] 0.13) and a recall of 0.60 (SD 0.11). Although performed on a limited sample, our study shows potential for continuous, and passive geolocation-based monitoring of patient behaviour in mental health.

Keywords:

Schizophrenia [MeshTerm]; Smartphone [MeshTerm]; Geographical Positioning System.

Introduction

The use of personal digital devices with geolocation capabilities continues to grow, with an estimated three billion people using smartphones around the world [1]. This offers new opportunities for gaining insight into human behaviours based on geolocation data, at cheaper costs and higher spatio-temporal resolution than traditional methods (e.g. questionnaires, activity diaries or interviews) [2].

These new opportunities are particularly relevant to domains such as mental illness, where resources are incresingly lacking and changes in behaviours are a key factor to monitor [3,4]. For example, Social functioning is one of the mainconcepts used to monitor the mental health status of people living with psychotic disorders [5]. Social functioning considers aspects such as withdrawal, employment, independence, interpersonal functioning, social activities and recreational activities [5]. These are currently assessed via self-reported questionnaires, or by observations of relatives or clinical staff. Though the validity and reliability of these instruments have been confirmed in previous studies, they have a high risk of recall and selection bias in reported information because they rely on patients' and relatives' memory, as well as a willingness and ability to record past behaviours [6–8]. Furthermore, such questionnaires are recorded infrequently, and may fail to capture the dynamic phenotypic expression of psychotic disorders characterised by interacting and fluctuating symptoms [9].

Mobility metrics derived from geolocation data have been shown to be correlated with disease status in different mental health conditions such as depression and anxiety [10,11], dementia [12,13], and bipolar disorders [14]. However, often such geolocation-derived metrics (e.g. distance travelled, number of places visited, time spent out from home) are rather crude; much more details about patients behaviours could be retrieved from geolocation data with less information loss.

In this paper, we present a method that uses clustering and semantic enrichment via the OpenStreetMaps (OSM) API [15] to identify daily activities relevant to assessing social functioning from passively and routinely collected geolocation data. We tested our method in a population of 21 healthy volunteers, from whom we collected both geolocation and daily activity data for a period of ten days.

Methods

Study population

We conducted the study among students and staff at the faculty of Biology, Medicine and Health at University of Manchester in spring 2017. Participants were recruited via University and student mailing lists. Each participant received a £20 at the end of the study, as a compensation for their time and effort. The study was approved by the University Research Ethics Committee 2 at the University of Manchester (ID number: 2017-0886-2473).

Data collection

Each participant was asked to collect data for ten randomly chosen days during a period of four weeks after entering the study. Participants received a schedule indicating randomly selected recording days and email reminders sent by a member of the research team (SLB). Participants could ask for data recording re-scheduling, if needed.

On data recording days, participants were instructed to collect geolocation data using an application on their smartphone, called GPSLogger for Android [16]. Participants who did not own an Android phone were provided with one (i.e. Samsung

Galaxy S6) for the duration of the study. GPSLogger for Android collects raw geolocation data (i.e., geolocation timestamp, latitude, longitude, altitude, speed and direction) without profiling or analysing it. For this study, the sample frequency was set to one sample every 10 seconds, and to limit noise we set GPSLogger for Android to record only data points with an accuracy of at least 40 meters. As prescribed by Android in their developer manual [17], this means that each data point had a 68% probability to be within a circle having 40 meters radius and centre on the recorded location. To avoid missing data due to an empty battery, we instructed participants to fully charge their phone at least once a day. Participants were asked to start recording data in the morning, perform everyday activities as normal, and stop recording when going to bed. A data recording day was considered as valid if the participants recorded for more than one hour during the day. After each data recording day, the data were transferred automatically to a secure server at a University of Manchester via a secure connection. As a gold standard for evaluating the accuracy of the algorithms to identify daily activities from geolocation data, participants were also asked to complete an activity diary for each data recording day. This was structured as a diagram containing the places they visited and activities they undertook, including time stamps. Days recorded in the diary were considered as valid if they reported a clear sequence of all the activities performed during the day.

Geolocation data processing

In this study, we built on our previous work [18], where we developed a three-step analytical pipeline to predict activities from raw geolocation data (see Figure 1).

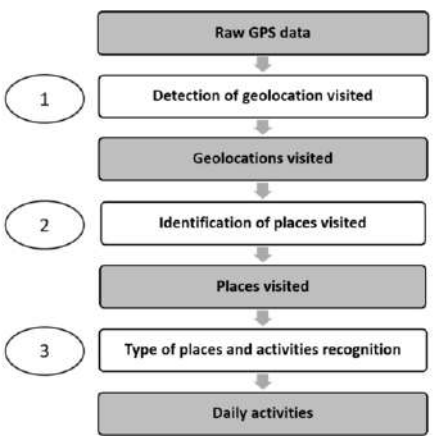


Figure 1– Pipeline to move from raw geolocation data to daily activities [18].

Geolocation visited detection

Time- and density-based methods can be used to identify the geolocations visited from raw geolocation data points. The time-based method was developed by Ashbrook & Starner [19]. It assumes that, when the geolocation signal of a person is lost for a certain time period (defined by a time threshold), it is reasonable to think that the individual has entered a building and performed an activity during this time. However, this method can lead to false recognition since the signal can also be lost in other scenarios (e.g. when a person is passing through a tunnel). Therefore, we modified the algorithm by introducing a distance threshold of 50 meters that has also to be met. Thus, only when the time interval between two consecutive geolocation data points was greater than the time threshold and they were not more than 50 meters apart, it was assumed that the person had stopped to perform an activity in a building.

The density-based method was developed by Fehér et al. [20]. It assumes that when a person has stopped at a geolocation to perform an activity, their geolocation data points will move slightly within a certain range (i.e. a circle with a radius of 50 metres) for a certain period (i.e. 10 minutes). The centroid of this circle is classified as a visited geolocation.

Because the two methods use complementary strategies to identify geolocations visited, we combined the results from the two methods. For each group of data points identified as a visited geolocation by either method, we retained the centroid and the number of points composing the group. For both methods, we applied various time thresholds (5, 6, 7, 8, 9 and 10 minutes).

Places visited

A place visited was defined as a unique location with which a group of geolocations visited close to each other could be identified (see Figure 2). We found places visited by using a modified k-means clustering algorithm [19], which was applied to the geolocations visited found in the previous step. The algorithm starts from the first identified geolocation visited and clusters it with all the geolocations visited within 50 meters. The coordinates of the place visited are subsequently calculated as a weighted mean, where the weights are the number of points of which each geolocation visited is composed. This accounts for the fact that geolocation visited from the density-based method are actually representative of several data points, which have been grouped together in the previous step. In the next iteration, the geolocations visited within 50 meters of the centroid are assigned to the cluster, with the centroid of the cluster that is updated with the newly-added geolocations visited. This process continues until the centroid coordinates no longer move. At this stage all the geolocations visited within the cluster are assigned to the same place visited, which is represented by the centroid coordinates. The algorithm restarts with the first geolocation visited that was not assigned to the current cluster. Ultimately, this step produces a list of cluster centroids that represent the places visited by the person.

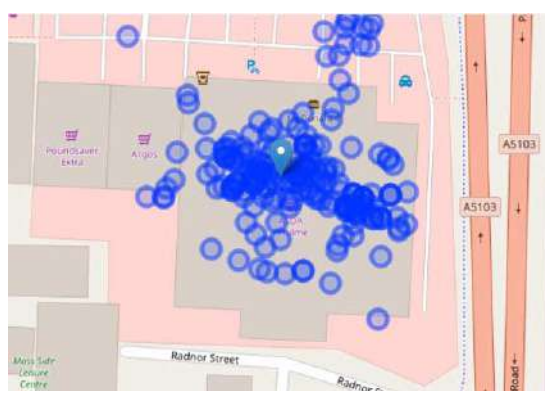


Figure 2–Place visited (map marker) identified from a group of geolocations visited (blue circles).

Semantic enrichment

We used the OpenStreetMap (OSM) platform for semantic enrichment of the places visited. OSM is a collaborative, free-editable and reusable map database, which offers information about places all over the world [21]. OSM was accessed via the

package osmdata [22] in the R statistical computing software, which allows queries to the OSM API to retrieve information about the surrounding Places of Interest (POIs) of specific geolocation coordinates. Such POIs have a geographical and a semantic component. The former represents the geolocation coordinates of the POI shape on the map (e.g. buildings, roads or wider areas such as a university campus). The latter describes the type of POI (e.g. university, restaurant, and bar) and other additional features (e.g. address, opening times, and others), and thus provides semantic enrichment.

To identify the most likely POI for a visited place we implemented the following strategy. First, we checked whether a place visited was within a building. If that was the case, we controlled whether there was a POI within that building with a relevant key-value pair. If that was true, the place visited was associated to that POI (in case of multiple POIs it was associated with the closest). If there was no POI inside the building, we checked whether the building had a relevant key-value pair and associated the place with that. If that was not the case, we looked at whether the building was part of a wider area or group of buildings (e.g. university campus, hospital, residential area). In that case, the place was associated to the smallest area including the building. Finally, if the place was not within a building, the place was associated with the closest POI on the map within a maximum of 50 meters. If none of the above conditions were met, the place visited remained unclassified.

After identifying the most likely POI for a visited place, we assigned an activity to that place using a bespoke ontology that we manually developed for this purpose. In the ontology we used all possible semantic tags that OSM uses, which is formed by a combination of a key and a value (e.g. building = university). OSM already divides such key-value pairs in relevant categories on the basis of the type of place. We used these categories to associate activities relevant to social functioning to the identified POIs (Table 1). For example, the category "sustenance", which contains POIs where people can eat (i.e. pubs, restaurants and bars), was associated to social activities such as "eating out" or "going to the pub"; or the OSM "education" category, which includes places such as schools and college, was associated to "attending classes/lectures". Some OSM categories did not have direct link to activity categories in our study. Such places were then assigned as "other" by two authors in the most relevant activity category from table 1.

Table 1– Activity categories and related activity examples

Area of social functioning	Activity examples
Employment	Working, attending lectures/classes
Shopping	Groceries, clothes, other
Sports	Swimming, gym, team sports, other
Social activities	Visiting friends/relatives, going to a pub/night club/, eating out, other
Recreational activities	Going to the cinema, going to a concert, going to the theatre, other
Other	Religion related, financial related

We also identified each participant's home and work place. The home was identified as the place that most often was either the first or last place visited across all recorded data among the places with a "residential" tag in OSM. As all our participants were working or studying in Manchester, the algorithm would look for eligible places in Manchester first, then in the county (Greater Manchester) and finally in the region (North West England). Once the algorithm found the participant's home, all other POIs with a "residential" tag were classified as "visiting family and friends" under the "Social activity" category. As our participants were studying and working in different locations (e.g. attending lectures or meetings in different university buildings, or training in different hospitals), we decided to use the information about their profession (e.g. university staff and medical students) to identify their work place. Particularly, we labelled as work all places with a tag related to university or hospital.

Figure 3 – Results from the OSM API that highlight all the buildings with a key-value pair that matched our ontology around the POI under study (red star).

Outcome measures

The main outcome measure of this study was the accuracy of our algorithm in predicting the daily activities that were reported in the participant's activity diary. Our secondary outcome was to estimate the accuracy with which our algorithm would estimate the number of unique activities performed during the day (e.g. often called "Number of places visited" or "Number of significant locations" in the literature), as this has been shown to be an important indicator of disease status in mental health conditions [6,11,23,24].

We compared the activities predicted by our algorithm to what was recorded by participants in their activity diaries. We considered activities as predicted correctly if there was a match between the activity category identified by the algorithm, such as the areas of social functioning identified in Table 1, and what was reported in the activity diary. To calculate performance, we also accounted for the number of times an activity was performed. For example, if someone had reported to have gone shopping two times during the day, and our algorithm detected three shopping sessions, two of the detected sessions would be considered as correctly classified (e.g. true positive) while the third would be considered as incorrectly classified (e.g. false positive). We assessed performance with recall, precision and F1 scores. Recall is defined as the number of correctly classified activities out of those recorded in the social functioning diary. Precision is defined as the number of correctly classified activities out of the total number of activities identified by the algorithm. The F1 score is defined as the harmonic mean of precision and recall. To have an idea of the extent to which a completely passive algorithm could associate places visited to activities on OSM, we also calculated the proportion of places visited classified, out of all the ones identified for each participant. Finally, we calculated the root mean squared error (RMSE) for the number of activities

inferred per day by our algorithm, compared to what reported by participants in their activity diaries.

Results

Participants' characteristics

We recruited 21 participants among students and staff at the University of Manchester. Overall, fourteen participants (56%) were female, and the mean age was 25.4 years (Standard Deviation [SD], 6.7 years). The majority were students (seventeen), with only four being staff at the university. Five participants used their own smartphone for the study, all others used a smartphone that we handed out to them.

Collected data

Out of the 210 days (i.e. 10 days times 21 participants) that potentially could have been collected, 157 were analysable (i.e. having both valid geolocation and activity diary data). Of these, we had to exclude four days, as the objective was to analyse participants' daily routines but two participants went back home to their families outside the UK during the Easter Holidays. Overall, nineteen participants provided at least one analysable day and were included in the analysis. For these, the mean number of analysable days provided was 8.3 (SD 1.7). In the activity diary, each participant reported a mean of 7 unique activity categories (SD 1), for a mean total of 46.4 (SD 10.8) activities recorded throughout the study.

Evaluation

Table 2 shows the performance of our method in inferring daily activities based on geolocation data. Performance was similar across the different time thresholds. Recall was 0.58 (SD 0.11) for the 10 minutes threshold and 0.60 (SD 0.11 and 0.12) for the 6 minutes and 8 minutes thresholds. Precision was 0.72 (SD 0.13) for the 5 minutes threshold and 0.76 (SD 0.13) for the 10 minutes threshold. The F1 score was around 0.65 (SD 0.11) for all the time thresholds we tried. The mean proportion of uniquely identified places classified was always above 0.90, with the exception of the five minutes threshold. Finally, the RMSE of unique places identified was 1.1 (SD 0.3) across all thresholds (results not shown in the table).

Table 2 – Overall mean performance across all participants.

Time threshold (min)	Mean Recall (SD)	Mean Precision (SD)	Mean F1 score (SD)	Mean proportion of labelled places (SD)
5	0.59 (0.14)	0.72 (0.13)	0.64 (0.11)	0.89 (0.11)
6	0.60 (0.12)	0.75 (0.13)	0.65 (0.1)	0.92 (0.06)
7	0.59 (0.11)	0.74 (0.13)	0.65 (0.1)	0.92 (0.07)
8	0.60 (0.11)	0.75 (0.13)	0.65 (0.1)	0.93 (0.06)
9	0.58 (0.11)	0.76 (0.14)	0.65 (0.1)	0.92 (0.06)
10	0.58 (0.11)	0.76 (0.13)	0.64 (0.1)	0.91 (0.06)

Discussion

We conducted an experiment that uses clustering and semantic enrichment to derive people's daily activities from geolocation data in a study with 21 healthy volunteers. Performance was similar across the time thresholds we tested, with an F1 score of 0.65. The RMSE of unique places identified was 1.1 (SD 0.3) across all thresholds.

Our study is the first to combine different methods (time-based and density-based) to identify geolocations visited in one analytical pipeline, as well as performing semantic enrichment using geographical information from OSM. To the best of our knowledge, Saeb et al. [10] is the only previous study that used geolocation data to infer actual daily activities relevant to mental health. They used the Foursquare API [25] that can be used to identify places for travel & transport, nightlife, spiritual activities, outdoors & recreation, arts & entertainment, medical offices, food, home, and shopping. In their study, they tested two different strategies. The first one was totally passive with POIs indentified only by using the Foursquare API based on geolocation data. However, this strategy disregards activities such as visiting friends and relatives or work, which are important indicators of social functioning. For the second strategy, they combined the information provided by Foursquare API with other phone and sensor data (e.g. screen utilisation, sound, and accelerometers among the others) to train a predictive model on the daily activities recorded by participants. Although this method achieved a mean AUC of 0.88, it requires a high involvement by participants, who have to label the data to train the model.

Our results show potential for geolocation data to be used for monitoring social functioning in mental health. However, before this type of methods can regularly be used in practice, ethical and privacy concerns of using geolocation data to monitor people whereabouts need to be overcome [26]. This could be done with two main strategies. First, developers should always give people the option to stop recording their location at any point. Furthermore, as much computation as possible should be done on the person's device in order to transmit limited amount of sensitive information. Second, the purpose and advantages of using this type of remote monitoring strategy should be clearly explained to patients and families. In fact, people are willing to share their geolocation data in other fields, if they clearly understand the purpose and perceive a benefit out of it (e.g. Google Maps [27]). However, as behavioural phenotyping based on geolocation data is still in its infancy, providing evidence of the advantages of this approach is not straightforward.

For the future, we envisage different ways in which our method could be improved. First, locations where people actually stopped could be better identified by integrating the speed measurements within the first step of our analytical pipeline. Second, additional information about POIs (e.g., opening and closing times) could be used to derive more sophisticated rules for identifying relevant POIs during the semantic enrichment. Finally, active learning could be explored to train our algorithm by asking participants to label, in real time, locations for which there is a high level of uncertainty.

Our analysis has several limitations. First, our study was performed on a small study sample who have been followed for a limited amount of time. Second, our ground truth was based on participants willingness and ability to recollect their activities and whereabouts. Therefore, we acknowledge that there might have been instances in which participants erroneously reported, or forgot to report, some activities. However, without actively following participants during the study this is the only practical and acceptable strategy to collect the ground truth in these type of studies. Finally, since OSM is an open source project where users can add and edit information, some records in this database might be inaccurate or out-dated.

Conclusions

We predicted daily activities of 21 healthy volunteers from geolocation data that was passively recorded on their smartphones. Using clustering algorithms and semantic enrichment methods we obtained a precision of 0.75 (SD 0.13) and a recall of 0.60 (SD 0.11). The study demonstrates the feasibility of monitoring daily activities relevant to social functioning, and has potentially valuable applications in mental health.

Acknowledgements

This project was funded by the Engineering and Physical Sciences Research Council (grant EP/P010148/1; The Wearable Clinic: Connecting Health, Self and Care). Map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org.

References

[1] Interim Ericsson Mobility Report, 2018. https://www.ericsson.com/assets/local/mobility-report/documents/2018/emr-interim-feb-2018.pdf.

[2] K. Stanley, E.-H. Yoo, T. Paul, et al., How many days are enough?: capturing routine human mobility, *Int. J. Geogr. Inf. Sci.* **32** (2018) 1485–1504. doi:10.1080/13658816.2018.1434888.

[3] M. Olfson, Building The Mental Health Workforce Capacity Needed To Treat Adults With Serious Mental Illnesses, *Health Aff.* **35** (2016) 983–990. doi:10.1377/hlthaff.2015.1619.

[4] C. Hollis, R. Morriss, J. Martin, et al., Technological innovations in mental healthcare: harnessing the digital revolution, *Br. J. Psychiatry*. **206** (2015) 263 LP-265.

[5] M. Birchwood, J. Smith, R. Cochrane, et al., The Social Functioning Scale. The development and validation of a new scale of social adjustment for use in family intervention programmes with schizophrenic patients., *Br. J. Psychiatry*. **157** (1990) 853–9. doi:10.1192/bjp.157.6.853.

[6] S. Abdullah, M. Matthews, E. Frank, et al., Automatic detection of social rhythms in bipolar disorder., *J. Am. Med. Inform. Assoc.* **23** (2016) 538–543.

[7] A. Doryab, M. Frost, M. Faurholt-Jepsen, et al., Impact factor analysis: combining prediction with parameter ranking to reveal the impact of behavior on health outcome, *Pers. Ubiquitous Comput.* **19** (2015) 355–365. doi:10.1007/s00779-014-0826-8.

[8] J.-P. Onnela, and S.L. Rauch, Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health, *Neuropsychopharmacology*. **41** (2016) 1691–1696. http://dx.doi.org/10.1038/npp.2016.7.

[9] M.J. Owen, A. Sawa, and P.B. Mortensen, Schizophrenia, *Lancet*. **388** (2016) 86–97. doi:http://dx.doi.org/10.1016/S0140-6736(15)01121-6.

[10] S. Saeb, G.E. Lattie, P.K. Kording, et al., Mobile Phone Detection of Semantic Location and Its Relationship to Depression and Anxiety, *JMIR Mhealth Uhealth*. **5** (2017) e112. doi:10.2196/mhealth.7297.

[11] S. Saeb, M. Zhang, C.J. Karr, et al., Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study, *J. Med. Internet Res.* **17** (2015) e175. doi:10.2196/jmir.4273.

[12] M. Wettstein, H.-W. Wahl, N. Shoval, et al., Out-of-Home Behavior and Cognitive Impairment in Older Adults, *J. Appl. Gerontol.* **34** (2012) 3–25. doi:10.1177/0733464812459373.

[13] M. Wettstein, H.-W. Wahl, N. Shoval, et al., Cognitive status moderates the relationship between out-of-home behavior (OOHB), environmental mastery and affect., *Arch. Gerontol. Geriatr.* **59** (2014) 113–21. doi:10.1016/j.archger.2014.03.015.

[14] M. Faurholt-Jepsen, M. Bauer, and L.V. Kessing, Smartphone-based objective monitoring in bipolar disorder: status and considerations, *Int. J. Bipolar Disord.* **6** (2018) 6. doi:10.1186/s40345-017-0110-8.

[15] OpenStreetMap contributors, Planet dump retrieved from https://planet.osm.org , (2017).

[16] GPSLogger for Android, (n.d.). https://gpslogger.app/ (accessed June 4, 2018).

[17] Android, Android developer manual - Location, (n.d.). https://developer.android.com/reference/android/location/Location.html#getAccuracy().

[18] S. Difrancesco, P. Fraccaro, S.N. van der Veer, et al., Out-of-Home Activity Recognition from GPS Data in Schizophrenic Patients, *2016 IEEE 29th Int. Symp. Comput. Med. Syst.* (2016) 324–328. doi:10.1109/CBMS.2016.54.

[19] D. Ashbrook, and T. Starner, Using GPS to learn significant locations and predict movement across multiple users, *Pers. Ubiquitous Comput.* **7** (2003) 275–286. doi:10.1007/s00779-003-0240-0.

[20] M. Fehér, and B. Forstner, Identifying and utilizing routines of human movement, in: Second East. Eur. Reg. Conf. Eng. Comput. Based Syst. Identifying, 2011. doi:10.1109/ECBS-EERC.2011.28.

[21] OpenStreetMap API, (n.d.). https://wiki.openstreetmap.org/wiki/API.

[22] M. Padgham, osmdata: Import "OpenStreetMap" Data as Simple Features or Spatial Objects, (n.d.). https://cran.r-project.org/web/packages/osmdata/index.html.

[23] L. Canzian, and M. Musolesi, Trajectories of Depression: Unobtrusive Monitoring of Depressive States by Means of Smartphone Mobility Traces Analysis, in: Proc. 2015 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput., ACM, New York, NY, USA, 2015: pp. 1293–1304. doi:10.1145/2750858.2805845.

[24] L.J. Faherty, L. Hantsoo, D. Appleby, et al., Movement patterns in women at risk for perinatal depression: use of a mood-monitoring mobile application in pregnancy, *J. Am. Med. Informatics Assoc.* **24** (2017) 746–753. http://dx.doi.org/10.1093/jamia/ocx005.

[25] Foursquare, (n.d.). https://developer.foursquare.com/ (accessed June 4, 2018).

[26] D.M. Goldenholz, S.R. Goldenholz, K.B. Krishnamurthy, et al., Using mobile location data in biomedical research while preserving privacy, *J. Am. Med. Informatics Assoc.* **25** (2018) 1402–1406. doi:10.1093/jamia/ocy071.

[27] Google, Google Maps API, (n.d.). http://www.webcitation.org/6ubEADyQl.

Address for correspondence

Professor Niels Peek

Email: niels.peek@manchester.ac.uk

Phone: +44 (0) 161 306 5058

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190364

Finding Options Beyond Standard of Care in Oncology: A Proposal for Workflows Utilizing Knowledge Databases

Katrin Glocker[a], Janko Ahlbrandt[a], Alexander Knurr[a], Peter Horak[b], Christoph Heining[c] and Frank Ückert[a]

[a] Division of Medical Informatics for Translational Oncology, German Cancer Research Center (DKFZ), Heidelberg, Baden Württemberg, Germany,
[b] National Center for Tumor Diseases (NCT), Heidelberg, Baden Württemberg, Germany,
[c] National Center for Tumor Diseases (NCT) Partner Site Dresden University Cancer Center UCC, Dresden, Saxony, Germany

Abstract

With the novel approach of molecularly stratified therapies based on genetic characteristics of individual tumors, the need for databases providing information on molecular alterations and targeted treatment options is increasing rapidly. In Molecular Tumor Boards (MTB) professionals discuss molecular alterations and provide biological context for therapeutic options using external knowledge databases. The identification of informative databases and the information on their specific contents can greatly facilitate and standardize the functioning of a MTB. In this work we present a list of databases which have been deemed useful and relevant for MTB in a clinical setting. We describe workflows to recommend the use of specific databases at different steps in the clinical curation process. Information obtained from these databases is a necessary prerequisite to evaluate molecular alterations and devise rational targeted therapies in MTB.

Keywords:

Precision Medicine; Databases, Genetic; Databases, Pharmaceutical

Introduction

According to the American Cancer Society, cancer was the cause of approximately 600.000 deaths in the US alone in 2017, which makes it the second leading cause of death. Furthermore, 1,5 million new cancer cases were diagnosed in the same year [1].

The treatment of this widespread disease poses many challenges. There is a widespread molecular variation of cancer types among each other. In addition, cancers that are considered the same type show also large intra- and intertumoral heterogeneity. This results in therapies which are used to treat a specific cancer type that can have extremely different outcomes in different patients. Kwak et al. could show that patients with ALK fusion genes showed limited response to first line chemotherapy in comparison with other patients with the same cancer type (lung adenocarcinoma). A treatment with Crizotinib, however, improved the outcome in these patients significantly [2]. The variance of responses to therapy in oncology creates the need for a more personalised way of treating patients based on their specific genetic, biomarker, phenotypic or psychosocial characteristics. This new paradigm is summarized in the concept of precision medicine [3].

Naturally this new approach also provokes new challenges. Oncologists have to get a deep insight into various variables which until recently represented only a side note or were not considered at all. Establishment of Molecular Tumor Boards (MTB) provides the opportunity to discuss molecularly characterized cases among specialists like molecular pathologists, clinical oncologists, biologists and bioinformaticians. Cases discussed in MTB are normally limited to those patients who did not benefit from or respond to standard therapies. At the National Center for Tumor Diseases in Heidelberg (NCT) an MTB was established in 2013 and meetings are currently held twice a week. In preparation for an MTB the patients' tissue samples are sequenced at the German Cancer Research Center (DKFZ). Bioinformaticians prefilter the data for mutations which are relevant for clinical evaluation. The molecular oncologist receives this prefiltered list and clinical data for each patient. With these parameters the oncologists first investigate individual molecular drivers and recommend treatment options which are subsequently discussed in the MTB [4,5].

Preparing a MTB is a time-consuming task. The need for information about genes, variants, pathways or studies is served by many different publicly available databases. As these databases are highly specialized, a deep and detailed understanding is required to connect and combine information from multiple sources. The goal of our study is to identify required steps to reveal known or inferred associations between mutations and possible therapies. For each step, we propose a category of knowledge databases, that can facilitate the process. Thereby we aim to support oncologists in finding therapy options for specific patients based on their individual genetic characteristics.

Methods

In order to identify databases which are required to gather essential information needed to give substantiated therapy suggestions we interviewed experienced oncologists at the NCT in Heidelberg. These oncologists routinely prepare an MTB twice a week and impose strict requirements for the databases they use to gather and prepare evidence for suggested treatments. A strong emphasis is set on publicly available web applications, which are easily accessible, do not require an installation and thus are immediately utilized. The list of databases which was provided by the oncologists was complemented in close consultation by additional databases we found by using scientific and non-scientific search engines.

After a general assessment of the content, the databases were grouped by their specialization into five subgroups (Table 1).

Table 1 – List of Selected Scientific Databases Used to Prepare a MTB. Based on Glocker et al. 2018 [6].

Specialization	Database	Content
Genetics (Variants/ Proteins) (G)	COSMIC	Genes, diagnoses
	CIViC	Genes, variants
	OMIM	Genes, proteins, pathways, diseases, expression, drugs, function, localisation, publications
	OncoKB	Genes, genetic variation, cellular coactors / processes
	ClinGen	Genes, variants, tumor types, drugs with evidence levels
	cBioPortal	Genes, variants, interventions, diseases
	Clinvar	Analytics tool, genetic variations, clinical evidence levels
	UniProt	Proteins, regulations, domains, functions, diagnoses, Protein variants
	Expression Atlas	Organism, localisation of proteins in body and brain, cell lines, diseases
	Varsome	Genes, variants, region browser, transcripts
	CRAVAT	High-throughput assessment and priritization of alterations, predictive scores
Pathways (P)	Reactome	Genes, pathways, cellular processes, reactions, protein complexes
	Kegg	Pathways, hierarchies, genes and proteins, similarity between gene sequences, small molecules, networks, gene variations, diseases, drugs, substances
Medication (M)	PharmGKB	Genes, chemicals, diagnoses, pathways, dosing guidelines, drug labels
	Guide to Pharmacology	Diseases, targets, inhibitors
	DGIdb	Genes, inhibitors, gene-drug interactions
Literature (L)	Pubmed	Literature
Studies (S)	ClinicalTrials	Studies
	EudraCT	European studies

In collaboration with oncologists, we identified examples for commonly used workflows in clinical routine using molecular findings. Drawing from these examples, we engineered exemplary (prototypical) workflows of steps required to gain hints for therapy suggestions for different patients or cases.

Results

Associating genetic and clinical information to give therapy suggestions is a task that often requires external knowledge. This knowledge is dispersed in various publicly available databases and has to be interlinked by a professional. In order to support physicians in the search for relevant information, typical clinical workflows were formalized. In these workflows, clinical problems were constituted and databases for the solution of the problem were proposed.

In clinical routine oncologists are confronted with difficult cases. A common task is to search for new treatments for patients who progressed on guideline-recommended therapies. Steps involved in a search based on clinical parameters are shown in the first workflow (Figure 1). The clinician starts with a search for studies a patient may benefit from in databases with the specialization on studies. Searching available medical literature in Pubmed also can give an indication to case reports, clinical trial reports or preclinical evidence for treating the specific diagnosis (Figure 1).

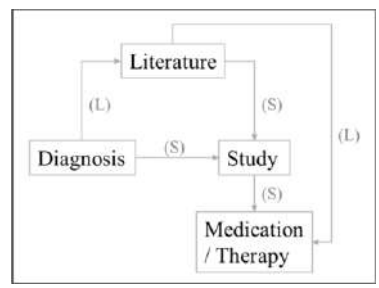


Figure 1 – Typical Clinical Workflow to Find Therapies for Patients Who Do Not Respond to Guideline Conform Therapies. To gain deeper insights with the help of information from external knowledge, databases are recommended for the steps of the workflow. (L): Literature, (S): Studies.

For a workflow of precision medicine, molecular diagnostics is performed upfront and has to be considered. This leads to lists of molecular alterations, which are reviewed and curated by the physician. In this case the diagnosis is often of secondary importance and genetic alterations are the mainpoint of focus. One way to find a treatment for a patient with a specific mutation is to research the respective gene or protein. This information is located in the databases listed with the specialization on genetics in Table 1. With this information, it is possible in some cases to find a drug or a study targeting this protein in medication databases or in studies databases. If no evidence of a direct link between a specific alteration and a drug is available the possibility to interrupt the pathway can be investigated in Kegg or Reactome. With the knowledge about the pathway which is impeded by the mutated protein the possibility to treat a patient with targeted drugs increases as proteins up- or downstream within the pathway may function as targets for available therapies. This newly identified targets can be researched with information from medication databases or studies databases.. For patients with cancers of unkown primary (CUP), the mutation of genes and respective proteins can shed light on the origin of the disease. This may open up

new options for medication or other therapies within established clinical pathways (Figure 2).

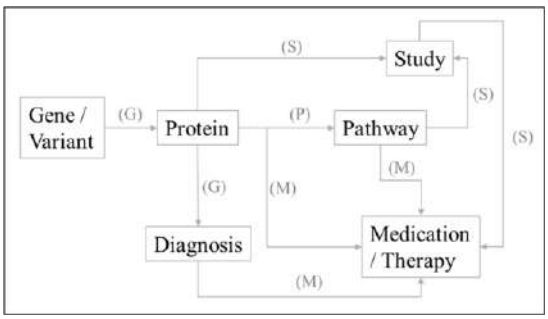


Figure 2 – Typical Clinical Workflow in Precision Medicine. To gather important information databases are recommended for the steps of the workflow. (G): Genetics (Variants/Proteins), (M): Medication, (P): Pathways, (S): Studies.

For example, with the information of a patient with the diagnosis CUP and a specific mutation in the IDH2 gene (IDH2 p.R172W) the physician can query COSMIC for the mutated gene. The input of IDH2 in the search field reveals general information about the gene and its associated protein, as well as a list of known variants. The search for p.R172W results in two hits on the same position. More detailed information can be obtained by following on one of the hits. The tissue distribution of the specific variant of the protein together with the literature used for references suggest a disease in the biliary tract (cholangiocarcinoma), chondrosarcoma or acute lymphatic leukemia. In a second step an inhibitor targeting IDH2 can be searched in the Guide to Pharmacology where Enasidenib is proposed as an FDA approved IDH2 inhibitor for AML (acute myeloid leukemia). Even if there are no direct inhibitors for a specific mutated protein, the oncologist has the chance to find a fitting clinical study for the mutated protein or even a study for cholangiocarcinoma patients.

Discussion

With the identification of workflows, medical and oncological questions can be solved more efficiently with the help of relevant databases. The information gathered from external databases can offer a new view on the situation of the patient. Starting with little information, like the mutated gene, the oncologist can get a notion of the specific diagnosis with which new diagnostics can be performed to verify this tentative diagnosis and new therapy options may arise. Furthermore the oncologist can achieve a better overview on approved inhibitors and available clinical studies. This information can help finding therapy options for patients with therapy-refractory advanced cancers.

As the essential information is spread in various databases a workflow can be useful to grasp the important facts. With the classification of the databases a first step for choosing the right databases for a clinical evaluation is made. Figure 1 and Figure 2 offer algorithms for choosing specific knowledge databases which can be used in various clinical problems arising in precision oncology. Naturally, this list shows only a mere fraction of knowledge bases found in the web but presents a cross section of databases used routinely by trained molecular oncologists in the MTB at NCT Heidelberg.

Furthermore, the recommendation of databases is only the first step in a long way to find therapy options for a patient. Professionals have to query the databases one after the other in order to ensure the completeness of the information gathered. This is still a time consuming task even with a set of recommended databases. In order to facilitate and accelerate the search for relevant information the next step will be an automation of the process. This would involve the interpretation of all databases in a structured manner via API integration. With the automated query of various databases we want to present relevant facts in one view to enable oncologists to grasp all information at one glance.

Conclusions

In order to facilitate the overwhleming task of research therapy options for patients with complex molecular alterations, public databases were categorized. Categorized databases were then recommended for different steps in typical clinical workflows. With this effort we want to support clinicians who face challenges of bringing molecular alterations into a therapeutically relevant context. This work is a starting point for the preparation of an MTB in hospitals which want to provide patients with novel therapeutic options and perspectives by considering specific molecular characteristics of their tumors.

Acknowledgements

This work was conducted within the frame of the DataThereHouse (DTH) project in cooperation with the National Center for Tumor Diseases (NCT) Heidelberg, supported by the Federal Ministry of Education and Research (BMBF). The project is concerned with the integration of data from different sources including the Heidelberg University Hospital as well as the German Cancer Research Center (DKFZ) Heidelberg for all oncological entities.

References

[1] R.L. Siegel, K.D. Miller, and A. Jemal, Cancer statistics, 2017, *CA: A Cancer Journal for Clinicians*. **67** (2017) 7–30. doi:10.3322/caac.21387.

[2] E.L. Kwak, Y.-J. Bang, D.R. Camidge, A.T. Shaw, B. Solomon, R.G. Maki, S.-H.I. Ou, B.J. Dezube, P.A. Jänne, D.B. Costa, M. Varella-Garcia, W.-H. Kim, T.J. Lynch, P. Fidias, H. Stubbs, J.A. Engelman, L.V. Sequist, W. Tan, L. Gandhi, M. Mino-Kenudson, G.C. Wei, S.M. Shreeve, M.J. Ratain, J. Settleman, J.G. Christensen, D.A. Haber, K. Wilner, R. Salgia, G.I. Shapiro, J.W. Clark, and A.J. Iafrate, Anaplastic Lymphoma Kinase Inhibition in Non–Small-Cell Lung Cancer, *New England Journal of Medicine*. **363** (2010) 1693–1703. doi:10.1056/NEJMoa1006448.

[3] J.L. Jameson, and D.L. Longo, Precision Medicine — Personalized, Problematic, and Promising, *New England Journal of Medicine*. **372** (2015) 2229–2234. doi:10.1056/NEJMsb1503104.

[4] J. Perera-Bel, B. Hutter, C. Heining, A. Bleckmann, M. Fröhlich, S. Fröhling, H. Glimm, B. Brors, and T. Beißbarth, From somatic variants towards precision oncology: Evidence-driven reporting of treatment options in molecular tumor boards, *Genome Medicine*. **10** (2018). doi:10.1186/s13073-018-0529-2.

[5] P. Horak, B. Klink, C. Heining, S. Gröschel, B. Hutter, M. Fröhlich, S. Uhrig, D. Hübschmann, M. Schlesner, R. Eils, D. Richter, K. Pfütze, C. Geörg, B. Meißburger, S. Wolf, A. Schulz, R. Penzel, E. Herpel, M. Kirchner, A. Lier, V. Endris, S. Singer, P. Schirmacher, W. Weichert, A. Stenzinger, R.F. Schlenk, E. Schröck, B. Brors, C. von Kalle, H. Glimm, and S. Fröhling, Precision oncology based on omics data: The NCT Heidelberg experience: Horak et al., *International Journal of Cancer*. **141** (2017) 877–886. doi:10.1002/ijc.30828.

[6] K. Glocker, J. Ahlbrandt, A. Knurr, P. Horak, S. Kreutzfeldt, C. Heining, and F. Ückert, Real-world IT toolkits for modern oncology – a market overview of molecular knowledge-based systems, *German Medical Science GMS Publishing House*. (2018).

Address for correspondence

Katrin Glocker
Research Associate
Division of Medical Informatics for Translational Oncology

German Cancer Research Center (DKFZ)
Foundation under Public Law
Im Neuenheimer Feld 280
69120 Heidelberg
Germany
phone: +49 6221 42-5130
fax: +49 6221 42-5109

k.glocker@dkfz-heidelberg.de
https://www.dkfz.de/en/mitro/

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190365

Towards a Framework for National eHealth Evaluation and Monitoring: A Combined Top-Down and Bottom-Up Approach Using Sweden as Example

Chen Hsi Tsai, Sabine Koch

Department of Learning, Informatics, Management and Ethics, Health Informatics Centre, Karolinska Institutet, Stockholm, Sweden

Abstract

Conducting evaluation and monitoring of national eHealth implementation aligned with national eHealth strategies should be prioritized concerning the insufficient scientific evidence and the current challenges in measuring eHealth effects and impacts where most approaches aim to measure at the micro level. This study aimed to build an evaluation and monitoring framework for national eHealth, using the Swedish national eHealth strategy as example, and to develop the process in doing so. Combining both top-down and bottom-up approaches, the WHO-ITU national eHealth strategy toolkit, as a systematic guidance, and two Swedish reports were used for development of the framework. Experts' opinions on the framework were collected and converged by the Delphi technique. The final draft suggested a framework containing 19 eHealth outcomes, 13 eHealth outputs, and 107 eHealth outcome and output indicators for 4 prioritized stakeholders, which can support comprehensive measurements to follow up the current advancement of eHealth in Sweden.

Keywords:

eHealth, Health impact assessment, Health policy

Introduction

Following the global adoption of eHealth, many have come to acknowledge the value of a national or regional eHealth strategy. A national eHealth strategy is a comprehensive nationwide approach of pinpointing and implementing eHealth solutions with appropriateness [1]. To ensure that national entities invest wisely in eHealth with scarce resources and to facilitate a more rapid dissemination of sustainable national eHealth solutions are a few benefits that can be realized once a national eHealth strategy is in place [1].

The Swedish national eHealth strategy was issued and updated in 2006, 2011, and 2016 by the Swedish Ministry of Health and Social Affairs and the Swedish Association of Local Authorities and Regions (SALAR) [2-4]. In the latest document, namely, Vision for eHealth 2025, Sweden wished to "be the best in the world at using the opportunities offered by digitization and eHealth to make it easier for people to achieve good and equal health and welfare…[4]" The three action areas: the work on regulatory frameworks in eHealth, the consistent use of information structure/terms, and the establishment of technical standards, were the main focus for future work towards the national vision of 2025.

Efforts in Swedish national eHealth evaluation and monitoring was first seen in the report, "eHealth and IT in Swedish County Councils" [5] (referred to as the SLIT report), published by the SLIT group of Sweden (IT Strategists/IT Managers/CIO's in the county councils of Sweden) since 2003. The SLIT report aimed to provide a picture of the extent of eHealth solutions and IT introduced and used in health care annually. Even though many existing indicators and measurements were presented, the ambition of the report was not to collect and present data on the impacts or effects of national eHealth but on existence, use, expenditure, and future local strategies.

A nationally coordinated effort was not carried out until 2018 when the Follow-up Model for eHealth and the First Pilot Measurement [6] (referred to as the eHealth 2025 model), the first Swedish follow-up model for eHealth according to the Swedish Vision for eHealth 2025, was published by the steering group of eHealth 2025, a collaboration of the Swedish Government and SALAR. The model, which was based on existing data/indicators from Swedish authorities and national or international organizations, consisted of a list of values that the national eHealth should contribute to realize and three follow-up areas. The purpose of the model was to measure the progress of Swedish national eHealth against the national strategy and to stimulate discussion of how eHealth should be developed and prioritized in order to achieve the Vision for eHealth 2025.

Aiming to deploy guidance for the development of national eHealth vision, action plan, and evaluation and monitoring framework for all governments, the WHO-ITU National eHealth Strategy Toolkit [7] (referred to as the WHO-ITU toolkit) was published by the World Health Organization (WHO) in collaboration with the International Telecommunication Union (ITU) in 2012. Although monitoring and evaluation have often been overlooked and under-valued in the process of adopting national eHealth [8], the third part of the WHO-ITU toolkit was dedicated to a systematic top-down approach for developing an evaluation/monitoring framework as part of the national eHealth strategy, which should be crucial in order to indicate the progress of national eHealth and assess the impacts followed by the efforts [7].

The challenging nature of planning and conducting evaluation and monitoring in eHealth results in the lack of scientific evidence in procedures used to assess eHealth solutions [9]. Moreover, a challenge exists where most monitoring and evaluations only measure the adoption and usage of eHealth at a micro level. In respect of the challenges, it is urgent to generate scientific guidance, which enables a contextualized evaluation and monitoring on effect and impact of eHealth at a national level aligned with national eHealth strategy.

Using the Swedish context as example, the aims of the study were:

1. to build a draft of a framework for national eHealth monitoring in relation to the Swedish national eHealth strategy.
2. to develop and demonstrate the method of drafting the framework.

Methods

Study Design

The study was designed into four phases, generating a preliminary framework, revising based on Delphi, proposing indicators, and establishing the final framework.

Phase 1: Generating a Preliminary Framework

In the first phase of the study, the three documents of the Swedish national eHealth strategy [2-4] and the WHO-ITU toolkit [7] were chosen to be reviewed carefully.

To align the Swedish national eHealth strategy with the systematic guidance of the WHO-ITU toolkit, a top-down approach was taken at this phase. According to the WHO-ITU toolkit, major stages for developing a national evaluation and monitoring framework included defining output and outcomes indicators for identified stakeholders, defining baseline and target measures for indicators, and defining supporting governance and processes as shown in Figure 1. This study was delimited to the first stage, define output and outcomes indicators for identified stakeholders, given the sequential process and the requirement of a more realistic implementation setting for the latter two stages.

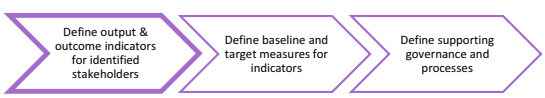


Figure 1– Stages for Developing a National eHealth Evaluation and Monitoring Framework [7]

The backbone for the first stage of developing a framework was extracted from the WHO-ITU toolkit (as shown in Figure 2) for further alignment with the Swedish national eHealth strategy. eHealth outputs were defined as deliverables related to the adoption/take-up of eHealth and shall be delivered through the national action plan; whereas eHealth outcomes were related to the results of the adoption/take-up of eHealth solutions and shall be based on the national eHealth vision [7].

By rigorously analyzing the three documents of the Swedish national eHealth strategy, the national visions and the corresponding prioritized stakeholders were identified. Thereafter, the Swedish national eHealth visions were rephrased in categorizations of the identified stakeholders to fill in the eHealth outcomes of the backbone. The national eHealth action areas mentioned in the Swedish national eHealth strategy were recreated in more general forms and linked to the eHealth outcomes. As a result, the preliminary framework of the study was generated by the two authors of the study.

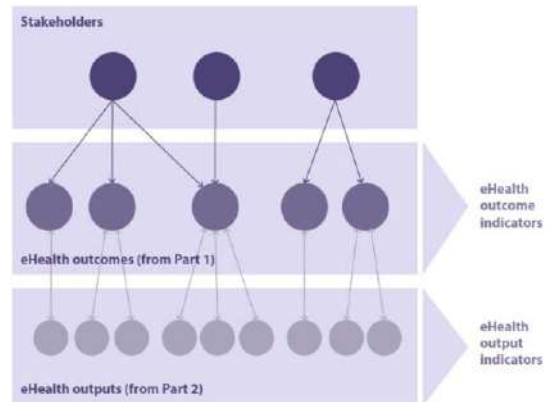


Figure 2– Relationship Between Stakeholders, eHealth Outcomes and Outputs, and Indicators [7]

Phase 2: Revising Based on Delphi

In consideration of the fractional knowledge of eHealth evaluation and monitoring for the Swedish national context, the Delphi technique was chosen in order to seek consensus of opinions on the developed framework with structured group communication [10].

The second phase focused on obtaining feedback from the Delphi panel of experts through two rounds of questionnaires. The foreword, study description, and background knowledge were presented to the respondents in the initial round of the Delphi questionnaire. Provided with a simplified version of the preliminary framework, the respondents were asked if the items in the framework were clear, unambiguous and complete and if anything should be removed, modified, or added. Validation of the questionnaire was done by conducting a pilot round with 2 experts who were not invited to join the Delphi panel. Answers and feedback were analyzed and utilized for revising the preliminary framework.

A recapitulation of feedbacks and the revised framework were provided in the second round of the questionnaire, where the respondents were asked to rate the importance of the listed eHealth outcomes with a five-point Likert scale of which the level of one indicated not important at all; and the level of five indicated absolutely essential. Ratings were analyzed with descriptive statistics to present the aggregated judgments and convergence of opinions from the Delphi panel.

Phase 3: Proposing Indicators

After the revision and the importance rating, the outcome and output indicators were further defined during the third phase. The eHealth 2025 model [6] and the SLIT report from 2017 [5] were chosen as references to be reviewed. A bottom-up approach was taken to fill in the revised framework with existing suitable indicators that were used for measurements in Sweden shown in the reports as well as to propose new indicators created by the authors.

Phase 4: Establishing the Final Framework

Eventually in the last phase, with the proposed eHealth outcome and output indicators, the revised framework was updated as the final draft of the eHealth evaluation and monitoring framework for Sweden (referred to as the final framework).

Table 1– Examples of the Revision of Modified and Additional eHealth Outcomes and Outputs (revision marked as blue)

eHealth outcome	eHealth output
C4. Improve the ability of citizens and patients • to gain trouble-free access to information on patients' rights or duties in relation to health and social care	
C5. Improve the ability of citizens and patients • to gain trouble-free access to information on available health and social care options, their quality, accessibility, staff friendliness and results; information on choice in care, the care guarantee and its significance in a user-friendly way.	Quality assessment guidelines for health and social care service established National registers of available health and social care options and providers established

Study Sample

Despite the controversial issues on what makes an expert in the Delphi panel [11], an expert for this study was defined as an individual involved in the Swedish national eHealth strategy work for the last ten years. Experts from national reference groups who have been involved in national eHealth strategy work from the beginning were considered as ideal study participants. Purposive sampling was used when selecting study participants. Verbal invitations and emails were sent to 17 experts to ask for willingness to join the Delphi panel. A positive response via email or verbally was seen as successful recruitment. 6 experts expressed willingness to participate. However, only 4 experts answered the initial round of the Delphi questionnaire and 5 experts in the second round.

As the study was primarily based on document review except the collection of expert opinions through the Delphi technique, obtaining consent was considered unnecessary. Nevertheless, there was no sensitive personal data acquired from the respondents and their identities were not disclosed.

Data Collection and Analysis

Qualitative data collected through the initial round of the questionnaires was compared with the preliminary framework to observe the respondents' feedback on outcomes/outputs that were added, removed, and modified. The comparison was used as the basis for revision of the preliminary framework.

Statistical analysis was carried out to present the central tendencies, mean and median, and levels of dispersion, standard deviation and quartile deviation, of the rated importance of the eHealth outcomes in the second round. Due to the single round of importance rating, it was difficult to draw conclusion and claim that the items have reached consensus. Therefore, quartile deviation was calculated as an indication of the consensus level for each item rated [12]. An item with a quartile deviation that was less than 0.60 was considered to have achieved a level of high consensus, whereas, a quartile deviation that was greater than 0.60 and less than 1.00 was seen as achieving a level of moderate consensus [12]. In other words, merely the consensus levels were reported without omitting any item from the framework after the second round of the questionnaire.

Results

In the preliminary framework, a total of 4 prioritized stakeholders, namely citizens and patients, health and social care professionals, researchers, and decision-makers, were identified. 15 unique national eHealth visions were extracted and transformed into 16 unique eHealth outcomes. 12 repetitive national eHealth action areas were linked with the 17 eHealth outcomes and rephrased as 7 repetitive eHealth outputs.

4 respondents in the panel answered the initial round of the questionnaire. They possessed different backgrounds and have worked in public health care organizations, research institutes, industries and national entities over the years. The comparison between the preliminary framework and the respondents' feedback suggested revision on the preliminary framework. The revision contained six modified eHealth outcomes, two additional eHealth outcomes, two modified eHealth outputs, and six unique additional eHealth outputs. A modified outcome and two additional outputs with the linking outcome were extracted from the stakeholder, citizens and patients, to show as examples in Table 1.

5 respondents answered the second round of the questionnaire. Four of the eHealth outcomes containing two sub outcomes and one containing three sub outcomes were divided into separate items resulting in 25 eHealth outcomes to be rated. The 25 eHealth outcomes were labeled according to stakeholders, for instances, the first outcome of citizens and patients was C1. 18 eHealth outcomes achieved the level of high consensus (quartile deviation ranging from 0.25 to 0.50), whereas, 7 eHealth outcomes achieved the level of moderate consensus (quartile deviation ranging from 0.75 to 1.00). The descriptive statistics of the 7 eHealth outcome with level of moderate consensus were presented in Table 2 on the next page.

The indicators used in the two reports [5, 6] were selected and proposed as the eHealth outcome and output indicators, which resulted in the establishment of the final framework containing 107 indicators. An eHealth outcome and its related indicators was taken from the final framework as example shown in Table 3. The first indicator in black color was extracted from the eHealth 2025 model [6]; the second in dark blue was from the SLIT report 2017 [5]; and the third in light blue was developed by the authors of the study.

Table 3– An Example of an eHealth Outcome and its Related Indicators Extracted from the Final Framework

eHealth outcome	eHealth outcome indicator
H6-8. Improve the ability of health and social care professionals	Proportion of prescribers who received at least one report from My Prescription Service during the year
• to make use of ICT-based decision support systems and	Number of usage sessions of the Swedish Information Services of Medication (Svenska informationstjänster för Läkemedel)
• to gain access to research and guidelines whenever in care setting	Number of usage sessions of each type of decision support (e.g. digital warnings on specific patient risks, reminders of actions that need to be taken, etc.)
• to have access to support of standardized care processes	

Table 2– Rating of Importance and Consensus on the 7 eHealth Outcomes with Level of Moderate Consensus

eHealth outcome	Mean	Median	Standard deviation	Quartile deviation
C2. Enable citizens and patients to track health and social care personnel who have had access to their records	4.20	4.00	0.84	0.75
C6.Enable citizens and patients to write comments addressed to authorized care professionals	4.20	5.00	1.10	1.00
C8. Enable citizens and patients to obtain advice about care and health based on the information they documented	4.40	5.00	0.89	0.75
C9. Improve the ability of citizens and patients to find and contact health and social care units of different types (for problem discussion)	4.00	4.00	1.00	1.00
H3. Ensure information registered in one place in the health and social care system is accessible at other points such as state entities, county councils, municipalities and private/third sector professionals	4.00	4.00	1.22	1.00
H7. Improve the ability of health and social care professionals to gain access to research and guidelines whenever in a care setting	4.40	5.00	0.89	0.75
D3. Improve the ability of decision-makers to gain access to high-quality data with respect for citizen/patient integrity through automatic and secure transfer of registered information in the health and social care systems to health databases and national quality registers for management and follow-up purposes	4.00	4.00	1.00	1.00

Discussion

Analysis of Results

According to the results, respondents' opinions varied on the importance of 7 eHealth outcomes as they achieved a level of moderate consensus. It was difficult to draw any conclusions on whether some eHealth outcomes were more important than others. However, since all eHealth outcomes received high rating of importance with mean values greater or equal to 4.00, the outcomes of the developed framework seemed to be relevant.

With the two reports [5, 6] as references, some of the eHealth outcomes and outputs, not all of them, could be filled in with indicators. Yet we could not conclude in how far the outcomes and outputs were met during or after implementation of the Swedish national eHealth strategy since a benchmark should be conducted beforehand to realize the progress and long-term impacts [13, 14].

Regarding the method used for building the framework, several points can be observed and suggested from the experience of this study. When comparing the eHealth 2025 model and the developed framework of this study, we saw the difference between development approaches. A bottom-up approach was mainly used for the eHealth 2025 model, where the structure of the model was put together by collecting pieces of results from other similar work. This might lead to a model/framework considered fragmented. This study, on the other hand, utilized the WHO-ITU toolkit as an important reference which provided a systematic way of developing the backbone of the framework with a top-down approach. In this case, the framework could be better aligned with the visions and actions of the national eHealth strategy. Thereafter, a bottom-up approach was taken to fill in indicators for measurements against the national eHealth strategy. Also, it would be essential to take in all major documents and updates of the national eHealth strategy while working on the establishment of such a model/framework considering the comprehensiveness and evolvement of the strategy. The Delphi technique was used as part of the process for seeking feedback from experts on the backbone of the developed framework. However, other methods such as nominated group, which would require a physical assembly, could be used depending on the practicalities.

Limitations

Having only 6 experts accept the invitation suggested a much smaller Delphi panel with fewer experts as compared to the recommended number, 15 to 20 experts [15]. Therefore, the result of the panel formation was compromised even though a bigger and more diverse panel with experts possessing different backgrounds was originally planned. The response rates were, in turn, low for both rounds of the Delphi questionnaire, with 4 and 5 respondents respectively. Although some suggested that two iterations of Delphi questionnaires were sufficient [11], the convergence and level of consensus on the eHealth outcomes generated by this study might be critically scrutinized as most studies suggested three to four rounds were required [10, 15]. Nevertheless, concerning the constraints of time, finance, and practical matters, the Delphi technique was considered suitable for this study.

Comparison with Other Similar Work and Suggestion for Future Research

There have been international programs seeking to evaluate and monitor the impacts of national eHealth. Unfortunately, most of them focused on the impacts of a single national eHealth service/solution [16, 17], such as the electronic health record systems. Interestingly, the Australian National E-Health Transition Authority has published a report [18], which aimed to summarize the achievements of national eHealth implementation according to Australian national eHealth policies and programs. The indicators for measurements presented in the Australian report and the eHealth outcome and output indicators of the framework developed by this study shared similar aims as to assess the adoption and impacts of national eHealth and the implementation of a national eHealth strategy. However, most indicators from both sources captured data regarding the adoption rather than the impacts.

A benchmark of comprehensive measurements on the current status of national eHealth in Sweden and the implementation of the national eHealth strategy using the framework developed by the study can be a question for further research so that the progress of achievement on the national eHealth outcomes and outputs according to the national eHealth strategy can be compared with baseline data and presented in the future.

Regarding the obstacles of measuring the impacts of national eHealth, the Nordic eHealth Research Network as part of the Nordic Council of Ministers have revisited the issue in their reports [13, 14, 19]. As suggested, parallel development of policy/strategy and the identification of appropriate corresponding indicators [13] can be a better solution in the future in order to smooth out the difficulties in following up on the progress of an eHealth policy/strategy and its impact. Close collaborations between researchers and policy makers can be beneficial [13] as many of the eHealth outcomes that we aimed to achieve are currently stated on a high level since the policy makers have set the national eHealth visions in such manner with their expertise. Researchers, with their skill set, can collaborate with policy makers and help break up the high-level eHealth visions into detailed measurable items and identify adequate indicators in order to successfully monitor the national eHealth advancement. After all, eHealth impacts and outcomes monitoring requires not only the identification of the types of eHealth solutions and their impact mechanisms aligned with the eHealth strategy but also the identification of related care quality indicators and data [14].

Despite the obstacles and current challenges, we hope the experience of this study can contribute to the knowledge of developing a more comprehensive evaluation and monitoring framework of national eHealth with a systematic approach provided by the WHO-ITU toolkit for countries who wish to do so.

Conclusions

In conclusion, the study initiated a discussion in national eHealth evaluation and monitoring by presenting a method of drafting a framework for national eHealth evaluation and monitoring using a draft of the Swedish framework as an example [20]. Developed with a systematic approach and aligned with the Swedish national eHealth strategy, the draft of the framework, which contains 19 eHealth outcomes, 13 eHealth outputs, and 107 indicators for 4 prioritized stakeholders, this can be used for comprehensive national-scale measurements to follow up on the current advancement of Swedish national eHealth.

Acknowledgements

This research was funded by the Swedish Research Council for Health, Working Life and Welfare – FORTE (Dnr 2016-07324).

References

[1] R.E. Scott and M. Mars, Principles and framework for eHealth strategy development, *J Med Internet Res* **15** (2013), e155.
[2] Swedish Ministry of Health and Social Affairs. *National Strategy for eHealth Sweden*, (S2006:019), Stockholm, 2006.
[3] Swedish Ministry of Health and Social Affairs. *National eHealth – the strategy for accessible and secure information in health and social care*, (S2011:023), Stockholm, 2011.
[4] Swedish Ministry of Health and Social Affairs. *Vision e-hälsa 2025 - gemensamma utgångspunkter för digitalisering i socialtjänst och hälso- och sjukvård*, (S2016:004), Stockholm, 2016.
[5] L. Jerlvall and T. Pehrsson, *eHälsa och IT i landstingen*, Inera och SLIT Gruppen, Stockholm, 2017.
[6] eHälsa 2025 - the Swedish government and the Swedish association of local authorities and regions. *Uppföljningsmodell för e-hälsa och en första testmätning – Ett diskussionsunderlag*, Stockholm, 2018.
[7] World Health Organization and International Telecommunication Union. *National eHealth Strategy Toolkit*, Geneva, 2012.
[8] C. Hamilton, The WHO-ITU national eHealth strategy toolkit as an effective approach to national strategy development and implementation, *Stud Health Technol Inform* **192** (2013), 913-916.
[9] E. Ammenwerth, Evidence-based Health Informatics: How Do We Know What We Know?, *Methods Inf Med* **54** (2015), 298-307.
[10] C. Powell, The Delphi technique: myths and realities, *J Adv Nurs* **41** (2003), 376-382.
[11] F. Hasson, S. Keeney, and H. McKenna, Research guidelines for the Delphi survey technique, *J Adv Nurs* **32** (2000), 1008-1015.
[12] M.C. Holden and J.F. Wedman, Future-Issues of Computer-Mediated Communication - the Results of a Delphi Study, *Etr&D-Educ Tech Res* **41** (1993), 5-24.
[13] Nordic Council of Ministers. *Nordic eHealth Indicators : Organisation of research, first results and the plan for the future*, TemaNord NV - 2013:522, Copenhagen, 2013.
[14] Nordic Council of Ministers. *Nordic eHealth Benchmarking*, TemaNord NV - 2015:539, Copenhagen, 2015.
[15] C.C. Hsu and B. Sandford, The delphi technique: making sense of consensus, *Practical Assessment, Research & Evaluation* **12** (2007), 1-8.
[16] Australian Digital Health Agency. *MyEHR to National eHealth Record Transition Impact Evaluation - Phase 1 Report*, Sydney, 2015, https://www.digitalhealth.gov.au/about-the-agency/publications/reports/benefit-and-evaluation-reports/myehr-to-national-ehealth-record-transition-impact-evaluation-phase-1-report, last access 15 November 2018.
[17] Centers for Disease Control and Prevention. *Meaningful Use*, Atlanta, 2017, https://www.cdc.gov/ehrmeaningfuluse/introduction.html, last access 15 November 2018.
[18] Australian Digital Health Agency. *Evolution of eHealth in Australia - Achievements, lessons, and opportunities*, Sydney, 2016, https://www.digitalhealth.gov.au/about-the-agency/publications/reports/benefit-and-evaluation-reports/evolution-of-ehealth-in-australia-achievements-lessons-and-opportunities, last access 15 November 2018.
[19] Nordic Council of Ministers. *Nordic eHealth benchmarking : From piloting towards established practice*, TemaNord NV - 2017:528, Copenhagen, 2017.
[20] C.H. Tsai, Towards a framework for eHealth evaluation and monitoring of the implementation of the Swedish national eHealth strategy [master's thesis], Stockholm, Karolinska Institutet, 2018.

Address for correspondence
Chen Hsi Tsai, Department of Learning, Informatics, Management and Ethics (LIME), Health Informatics Centre, Karolinska Institutet, Tomtebodavägen 18b, 171 77 Stockholm, Sweden, **email:** chen.tsai@ki.se

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190366

Mining Social Media for Perceptions and Trends on HIV Pre-Exposure Prophylaxis

Christine Kakalou[a], Jeffrey V Lazarus[b], Vassilis Koutkias[a]

[a] Institute of Applied Biosciences, Centre for Research & Technology Hellas, Thermi, Thessaloniki, Greece,
[b] Barcelona Institute for Global Health (ISGlobal), Hospital Clínic, University of Barcelona, Spain

Abstract

Pre-Exposure Prophylaxis (PrEP) is an approach for preventing the human immunodeficiency virus (HIV), which entails the administration of antiretroviral medication to high-risk seronegative persons. If taken correctly, PrEP can reduce HIV infection risk by more than 90%. The aim of this study was to identify and examine PrEP-related perceptions and trends discussed on Twitter. Using open-source technologies, text-mining and interactive visualisation techniques, a comprehensive data gathering and analytics Web-based platform was developed to facilitate the study objectives. Our results demonstrate that monitoring of PrEP-related discussions on Twitter can be detected over time and valuable insights can be obtained concerning issues of PrEP awareness, expressed opinions, perceived barriers and key discussion points on its adoption. The proposed platform could support public-health professionals and policy makers in PrEP monitoring, facilitating informed decision making and strategy planning for efficient HIV combination prevention.

Keywords:
HIV; Data Mining; Social Media.

Introduction

Pre-exposure prophylaxis (PrEP) is the administration of antiretroviral therapy to prevent seronegative persons from getting infected with human immunodeficiency virus (HIV). It is most commonly taken orally as a pill either daily or on demand, but other forms are also being tested in clinical trials (e.g. vaginal gels, injectables, and implants). The World Health Organization (WHO) recommendations state that individuals at high-risk should be offered PrEP as an additional prevention choice, as part of a broad HIV prevention strategy [1], outlining five key populations at-risk for contracting HIV [2], i.e. (a) men who have sex with men (MSM), (b) people in prisons and other closed settings, (c) people who inject drugs, (d) sex workers, and (e) transgender persons.

The rapid spread of social media has changed the way that people share their personal health experiences with others as well as the manner of seeking disease information and treatment alternatives [3]. Through social media, huge volumes of data are constantly being generated, providing a significant data source for research. In particular, patient-generated data referring to personal experiences in a way that is unsolicited, spontaneous and up-to-date, in large volumes and with a high degree of variance, is of great interest for research on public health surveillance [3]. Social networks focusing on health-related topics (e.g. medication and treatment options, side-effects, etc.) are particularly useful, since they are providing a wealth of information that includes strong semantic associations. The exploitation of social media facilitates the need to "hear the patient's voice", providing a unique opportunity to obtain input that cannot be acquired through official healthcare channels [4], either because certain groups don't have access to them, or because people are reluctant to provide sensitive information.

However, mining social media presents considerable challenges, which are linked with the inherent characteristics of user-generated data, e.g. vastness and no structure, recency, uniqueness, frequency, and salience [5]. The analysis techniques, relying typically on text mining/Natural Language Processing (NLP), need to be adjusted to the intricacies of the short, informal and colloquial text that people use to express themselves when posting online [6]. Specifically, in health-related use cases, the use of layman, descriptive, subjective and often misspelled terms complicate NLP tasks such as syntactical analysis, tokenizing and direct matching with medical lexicons. Supervised learning algorithms are rendered more difficult by the high amount of noise in the frequently vast collection of data, which demands more computational power and poses significant problems concerning performance.

Twitter is a very popular microblogging and social networking platform, through which users can post "tweets" i.e. short messages consisting of 280 characters maximum. Twitter is a common data source for opinion and knowledge mining on public health topics [5], such as pharmacovigilance [7], [4], prescription drug abuse [8], epidemiology surveillance [9], chronic diseases [10], mental health [11], and quality of care in hospitals [12] to name a few. One of the key barriers in sexual health communication is the patients' reluctance to seek information through their close personal networks [13]. Thus, the perceived anonymity offered by social media allows users to discuss sexual health issues and share their experiences within their online social circles.

This work aimed to examine whether PrEP-related trends and useful information can be identified in Twitter discussions by using text-mining techniques. While Twitter data have been exploited for PrEP-related analysis [13]-[15], the current study elaborated on systematizing and reinforcing such analyses through a comprehensive Web-based platform integrating various text-mining and visual analytics methods. In particular, an Exploratory Data Analysis (EDA) was employed, in order to detect the main discussion topics, the most frequent terms and popular hashtags along with possible side-effects regarding PrEP. In addition, sentiment analysis was used to investigate the public awareness on PrEP, the views and emotions towards it, as well as potential barriers for its adoption. The geographic distribution of the collected tweets along with the sentiment analysis scores were assessed, in order to discover potential insights on regions with low PrEP awareness or where PrEP meets obstacles for its acceptance.

In the following, we present the employed methods for the development of the platform, the obtained results, and discuss our findings along with future work directions.

Methods

Data Acquisition

Data acquisition relied on the SEST platform, an in-house developed software application [7]. SEST focuses on the systematic data aggregation and exploitation from various sources of unstructured text through NLP techniques, aiming to support public health surveillance applications. Built upon the microservices architecture, it enables the repurposing of its modules to accommodate alternative workflows/scenarios, interoperability with other systems through standards-based data exchange and semantic annotation of NLP results for further analysis.

Using SEST, which exploits the Twitter Streaming API [16], tweets were collected from September 2017 to December 2017 through two different collection streams:

1. "PrEP Datastream": tweets gathered with keywords: (truvada) OR (hiv AND PrEP) OR (preexposure AND prophylaxis AND hiv)
2. "Pill Datastream": tweets gathered with keywords: (pill AND prevent AND hiv) OR (pill AND protect AND hiv) OR (pill AND protect AND AIDS)

The "Pill Datastream" was set up to obtain tweets related to PrEP from users that might be unaware of the terminology of PrEP and/or its commercial name, i.e. truvada. The collected data were stored in MongoDB [17], an open-source, flexible and scalable NoSQL database that stores data in JSON format, while the RESTHeart API [18] was used for the create, read, update and delete (CRUD) operations in the SEST platform. For each gathered tweet, its content was stored in JSON format along with its unique id, the creator name, and the creation timestamp. For geolocation, only the location stated in the creator's bio was stored and not the geographical coordinates for each tweet (if available), aiming to preserve the creator's privacy. In addition, only tweets written in English were collected.

Data Analytics

Various types of analytics were applied in the collected tweets based on text-mining and NLP techniques implemented in R (Figure 1). In the preprocessing stage, punctuation and stopwords were removed from the text and special points of interest, namely, hashtags (i.e. #PrEPworks), mentions (i.e. @UNAIDS) and URLs, were detected using regular expressions. Through the Term Frequency - Indirect Document Frequency (TF-IDF) normalisation algorithm, term frequencies of the corpus were calculated and word associations with a correlation threshold of 0.2 were presented. The main discussion topics and the 7 most prevalent terms per topic were detected using the Latent Dirichlet Allocation (LDA).

Sentiment analysis was performed to discover the polarity of expressed opinions in the corpus overall, but also per tweet and per creator. For the polarity calculation, the AFINN lexicon was used that has been especially developed for sentiment mining of microblogging text, featuring also slang words [19]. The emotions expressed within the corpus were mapped to the 8 basic human emotions (namely, anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) via the NRC Word-Emotion Association Lexicon [20]. Both lexicons had to be adjusted to exclude terms that skewed the polarity, such as "negative" and "positive", which in this specific context refer to the HIV status of an individual, and to include colloquial terms and abbreviations often mentioned in the collected tweets (i.e. "ftw", which means "for the win", and "dope" meaning "great").

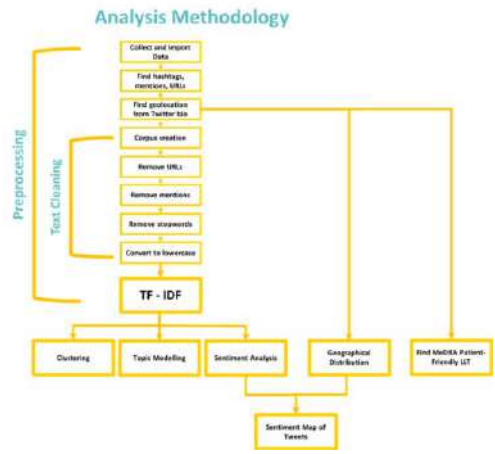


Figure 1 - Analysis Methodology

Since more than 23,000 tweets contained location information in the users' bios, they were extracted and converted to coordinates via the Google Geolocation API [21]. Locations that returned more than one set of coordinates or that did not correspond to real locations were rejected.

Possible mentions of side effects were identified by matching terms in the corpus of the recently released MedDRA® patient-friendly Low-Level Term (LLT) lexicon [22], which contains 1,440 terms that patients use to describe Adverse Drug Reactions (ADRs) and maps them to MedDRA® terms.

EDA was then applied to enhance the understanding of the collected data, uncover underlying trends, outline important variables, and detect outliers/unexpected behaviours that might signify important insights regarding the research questions. Visualisation plays a pivotal role in EDA, as it enables intuitive pattern recognition, when studying public health trends over time.

The Shiny package in R was used for the visualisation of the analysis; an interactive Web application (Figure 2) was deployed online, providing an overview of the acquired data along with suitable visualisation techniques for each aspect of the analysis and the ability to select text-mining parameters such as frequency and correlation thresholds where applicable.

Figure 2 - Homepage of the PrEP trends monitoring platform

Results

Over the course of 4 months, we identified more than 32,000 PrEP-related tweets (31,854 from the "PrEP Datastream" and 804 from the "Pill Datastream"), which constitute only a small fraction of the total Twitter datastream at a given time. 60.5% and 69.3% of the tweets were retweets in the "PrEP Datastream" and the "Pill Datastream", respectively. We expected that the "Pill Datastream" would contain original content from people seeking information about PrEP, but the bulk of

tweets came from institutional or NGO campaigns that propagated messages like "PrEP is a pill that you can take once a day to prevent HIV".

In the following, we present the results of each analysis step.

Frequent and Associated Terms

Using TF-IDF normalisation the most frequent terms address prevention, the key at-risk populations (i.e. "msm", "gay", "trans"), and countries or cities where PrEP becomes available such as Canada, Ireland, and the United Kingdom. Generic PrEP as well as free access to it or insurance coverage was similarly repeatedly discussed. Also notable was the high number of tweets discussing clinical trials and, more specifically, the inclusion of a more diverse sample, such as transgender people and women from different countries. The PrEP discourse within Black African communities seems to diverge, ranging from strong acceptance as a tool of empowerment among black women to a high level of doubt and stigmatisation within the black MSM community.

Word associations display the most similar terms to a selected word of interest above a user-defined threshold. The terms more closely associated with "buy" reveal that people seek to buy truvada "otc" meaning "over the counter" and find the "cheapest" options such as "viread", another antiretroviral drug.

Topic Modelling

Using the LDA algorithm, we identified the 50 and 30 top discussion topics for the "PrEP Datastream" and the "Pill Datastream", respectively, with 7 terms per topic shown. For the visualisation, the LDAvis [23] method was used, which presents the topics globally and reveals how much they differ from each other. All topics discussed were found relevant to PrEP and tend to cluster together in large numbers, which is expected considering the very similar context in all PrEP-related tweets. Smaller clusters and topics that are more separate from the large cluster represent specific events/tweets that were largely propagated in the Twitter stream, such as topic #21 (Figure 3) that contains the terms "thanks", "uk", "cases", "percent", "men", "plunge", "gay", and clearly refers to the impressive decline in HIV infections amongst MSM in London after PrEP was made available in clinics.

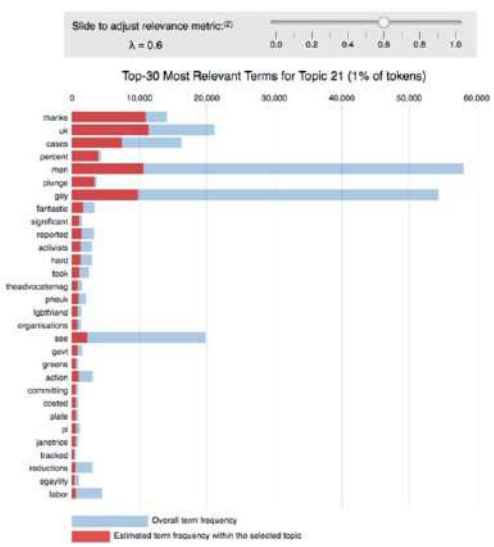


Figure 3 - Most relevant terms for topic #21 using the LDA algorithm

Sentiment Analysis

The overall sentiment was found positive, scoring 0.208 and 0.328 in a scale of -1 (absolutely negative) to +1 (absolutely positive) for the "PrEP" and "Pill" datastreams, respectively. The sentiment score densities calculated for each datastream show a more uniform distribution of sentiment for the "PrEP Datastream", while the "Pill Datastream" features a significant cluster of negative sentiment that, when traced back, reveals tweets expressing doubt and cautiousness regarding the existence of a pill that can prevent HIV infections. The most commonly expressed emotions in the corpus are trust and anticipation, with fear being very prominent in the "Pill Datastream".

Hashtags

Hashtags are a pivotal element of Twitter data analysis; not only do they signal emphasis in the context of a tweet text, they also act as nodes within the network, connecting various users and posts through a shared discussion topic. The use of a specific hashtag for a cause, event or awareness campaign is a common communication strategy adopted by organisations worldwide. The hashtags in the corpus were separately analysed and presented in a streamgraph visualisation. As expected, around the 1st of December (Word AIDS Day) a substantial peak was found for #WorldAIDSDay, #WAD2017, but also for HIV-related terms such as #TasP (Treatment as Prevention), #condoms, #EndTheEpidemic, etc. While this observation is evident, it serves as a validation control for accurately identifying trends over time using data-mining techniques. Observations of smaller peaks offer insights on events that influence the audience's perception and awareness of PrEP, such as the #Ireland hashtag that peaked after November 9th, 2017 when Ireland's first clinic monitoring PrEP was opened (Figure 4).

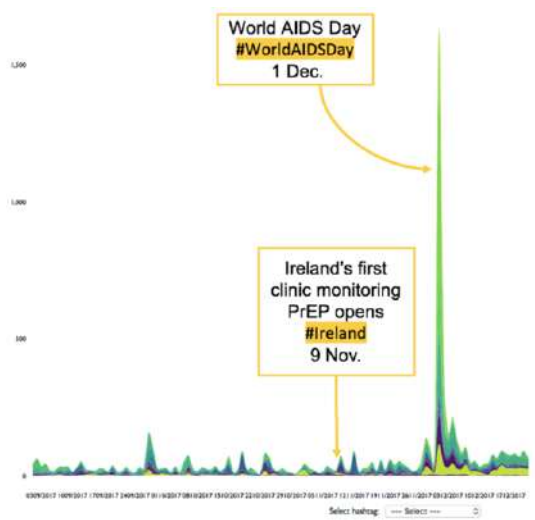


Figure 4 - Hashtags timeline

References to Side-Effects

The identified mentions to side effects that matched terms in the MedDRA® LLT lexicon concerned notably references to headaches, miscellaneous pain, and flu symptoms. A lot of noise was found in this mapping, because some lexicon terms are used in a different context than side-effect reporting, namely, "fatigue" for the phenomenon of "pill fatigue", "infection" as in "HIV infection", "worry" for being anxious about contracting HIV, and not as a result of taking PrEP medication.

It was evident that this type of analysis requires more sophisticated analysis and further research.

Geographic Distribution

The geographic distribution of tweets, in the form of an interactive map, holds special interest, with the largest clusters of PrEP discourse being in the USA, Western Europe, and South Africa. Notably, the distribution shows a similarity with the UNAIDS 2016 map of countries [24] that have either approved PrEP for commercial use, or are in the pilot phase (Figure 5). There is an apparent lack of PrEP-related tweets in Russia, China and Japan that although probably attributed to the language barrier and the existence of alternative social networks to replace the - often blocked entirely - western social media, may also hint at clusters of populations with intense HIV stigmatisation, especially considering the high HIV prevalence and increased rate of infection in those countries.

Figure 5 - Geographic distribution of tweets, and UNAIDS map of PrEP availability

The sentiment score for each tweet was calculated and is displayed on an interactive "sentiment map" (Figure 6). Although significantly different clusters of positive or negative tweets are not evident, there is a slightly more positive concentration of tweets in the UK, Ireland, and the USA.

Figure 6 - Geographic distribution with color-coded sentiment score per tweet

Influencers

Twenty-three (23) Twitter accounts per datastream, the vast majority corresponding to PrEP advocacy groups, LGBT networks and sexual health organisations, were found to post heavily about PrEP, representing more than 9% of the total posts collected. Those accounts constitute "PrEP Influencers", since they contribute to the dissemination of PrEP awareness across their networks and have a large number of followers (>10k mean follower count). The number of tweets per creator and per datastream as well as the sentiment score for the respective tweets is presented with a Treemap visualisation (Figure 7). Some accounts show negative sentiment associated with their posts as they mention issues like stigmatisation or feelings of guilt and isolation that may accompany a PrEP regimen.

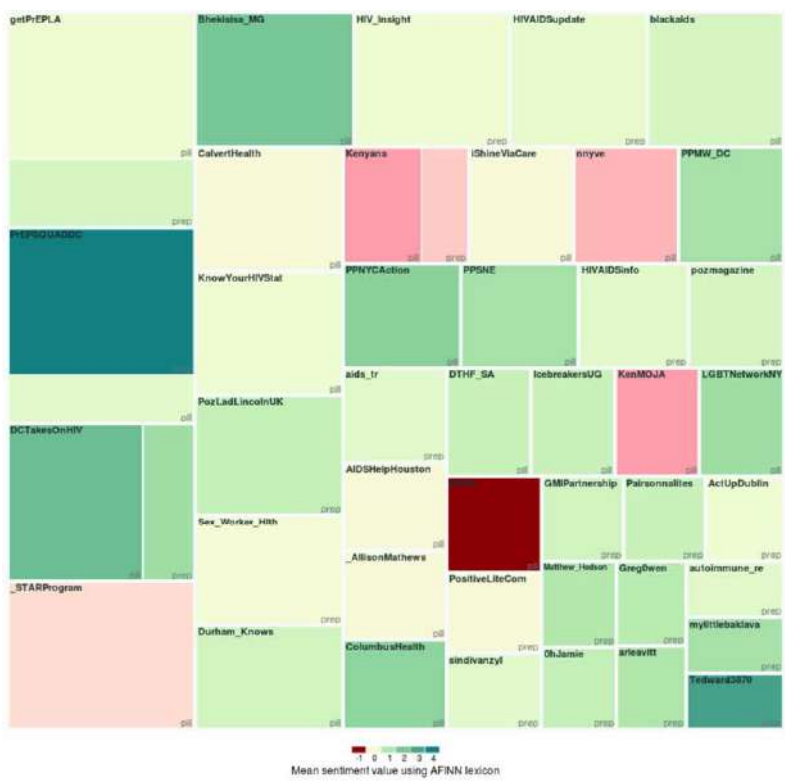


Figure 7 - Treemap of top 23 influencers and their expressed sentiment score for each datastream

Discussion and Conclusion

The collection of tweets during the 4-month period proved to be quite prolific and with very low amounts of noise, indicating that PrEP discussions can be identified successfully on Twitter. PrEP discourse seems to be continuously increasing on Twitter and in a rather fast rate, considering the collected data over time. Users disseminate information about PrEP, engage in sharing ideas, experiences and considerations about it, while activists, HIV support and PrEP advocacy groups appear to drive the demand for free and accessible-to-all PrEP worldwide.

Overall sentiment regarding PrEP is positive and the main concerns refer to stigma, medication cost, insurance coverage as well as vulnerable population groups with limited access to PrEP. The various interactive visualisation techniques that we employed allowed for a more flexible inspection and exploration of the collected data. Detection of critical events that impact the target audience's reaction to PrEP is possible by combining the monitoring of trending hashtags, frequent terms and most discussed topics on the interactive platform provided.

The preliminary, shallow text-mining analysis provided satisfactory results in some context, such as the identification of the most frequent terms and word associations, as well as hashtags monitoring. However, other aspects of the analysis, e.g. side-effect mentions and the sentiment analysis would greatly benefit from more sophisticated NLP tasks, especially considering the unofficial/layman, brief, full of typos and abbreviations language contained in tweets. Besides addressing the above limitations, Machine Learning methods is another necessary future step to enable the accurate detection of personal vs informative tweets, a task which constitutes one of our future work directions. To effectively address this requirement, we are currently exploiting a significantly larger data corpus, which was acquired through our platform.

Overall, our results indicate that the current work provides the foundations for a comprehensive analytics platform that can offer insights to public health professionals and policy makers, so as to assist them in taking informed decisions and developing efficient strategies for improving HIV prevention.

References

[1] WHO | Pre-exposure prophylaxis, *WHO*. (2018). https://www.who.int/hiv/topics/prep/en/ (accessed November 19, 2018).

[2] WHO, Cascade data use manual: to identify gaps in HIV and health services for programme improvement., Geneva, 2018.

[3] R. Harpaz, et al., Text mining for adverse drug events: the promise, challenges, and state of the art, *Drug Saf.* **37** (2014), 777-90.

[4] M. Salath, Digital pharmacovigilance and disease surveillance: combining traditional and big-data systems for better public health, *J Infect Dis.* **214** (2016), S399-403.

[5] A. Sarker, et al., Utilizing social media data for pharmacovigilance: A review, *J Biomed Inform.* **54** (2015), 202-12.

[6] A. Sarker and G. Gonzalez, Portable automatic text classification for adverse drug reaction detection via multi-corpus training, *J Biomed Inform.* **53** (2015), 196-207.

[7] P. Natsiavas, N. Maglaveras, and V. Koutkias, A public health surveillance platform exploiting free-text sources via natural language processing and linked data: application in adverse drug reaction signal detection using PubMed and Twitter, in: D. Riaño, R. Lenz, and M. Reichert (Eds.), *Knowl Represent Heal Care*, Springer International Publishing, Cham, 2017: pp. 51-67.

[8] C.L. Hanson, B. Cannon, S. Burton, and C. Giraud-Carrier, An exploration of social circles and prescription drug abuse through Twitter, *J Med Internet Res.* **15** (2013), e189.

[9] M. Santillana, A.T. Nguyen, M. Dredze, M.J. Paul, E.O. Nsoesie, and J.S. Brownstein, Combining search, social media, and traditional data sources to improve influenza surveillance, *PLoS Comput Biol.* **11** (2015), 1-15.

[10] M. Roccetti, et al., Attitudes of Crohn's disease patients: infodemiology case study and sentiment analysis of Facebook and Twitter posts, *JMIR Public Heal Surveil.* **3** (2017), e51.

[11] P. Burnap, G. Colombo, R. Amery, A. Hodorog, and J. Scourfield, Multi-class machine classification of suicide-related communication on Twitter, *Online Soc Networks Media* **2** (2017), 32-44.

[12] J.B. Hawkins, et al., Measuring patient-perceived quality of care in US hospitals using Twitter, *BMJ Qual Saf.* **25** (2016), 404-13.

[13] M.L. McLaughlin, et al., Propagation of information about Pre-Exposure Prophylaxis (PrEP) for HIV prevention through Twitter, *Health Commun.* **31** (2016), 998-1007.

[14] P. Breen, J. Kelly, T. Heckman, and S. Quinn, Mining Pre-Exposure Prophylaxis trends in social media, in: *Proc IEEE Int Conf Data Sci Adv Anal*., 2016, pp. 214-21.

[15] J. Schwartz and J. Grimm, PrEP on Twitter: information, barriers, and stigma, *Health Commun.* **32** (2017), 509-16.

[16] Twitter Streaming API, (n.d.). https://developer.twitter.com/en/docs/tweets/filter-realtime/guides/connecting.html.

[17] What Is MongoDB? | MongoDB, https://www.mongodb.com/what-is-mongodb (accessed February 18, 2018).

[18] Documentation - RESTHeart - Confluence, (2016). https://softinstigate.atlassian.net/wiki/spaces/RH/pages/9207957/Documentation (accessed March 10, 2018).

[19] AFINN, (n.d.). http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010 (accessed February 24, 2018).

[20] NRC Emotion Lexicon, http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm (accessed February 25, 2018).

[21] The Google Maps Geolocation API | Google Maps Geolocation API | Google Developers, (2018). https://developers.google.com/maps/documentation/geolocation/intro (accessed March 7, 2018).

[22] Patient-Friendly Term List | MedDRA, (2018). https://www.meddra.org/patient-friendly-term-list (accessed February 27, 2018).

[23] C. Sievert, and K.E. Shirley, LDAvis: A method for visualizing and interpreting topics, in: *Proc Work Interact Lang Learn Vis Interfaces*, 2014: pp. 63-70.

[24] UNAIDS, Fact sheet - Latest statistics on the status of the AIDS epidemic | UNAIDS, (n.d.). http://www.unaids.org/en/resources/fact-sheet (accessed May 5, 2017).

Address for correspondence

Christine Kakalou, Institute of Applied Biosciences, Centre for Research & Technology Hellas, 6th Km. Charilaou-Thermi Road, P.O. BOX 60361 GR – 57001, Thermi, Thessaloniki, Greece, email: ckakalou@certh.gr.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190367

Qualitative and Quantitative Analysis of Web Forums for Adverse Events Detection: "Strontium Ranelate" Case Study

Pierre Karapetiantz[a], Bissan Audeh[a], Juliette Faille[a], Agnès Lillo-Le Louët[b], Cédric Bousquet[a]

[a] Sorbonne Université, Inserm, université Paris 13, Laboratoire d'informatique médicale et d'ingénierie des connaissances en e-santé, LIMICS, F-75006 Paris, France
[b] Centre régional de pharmacovigilance HEGP, AP-HP Paris, France

Abstract

Social media are proposed as a complementary data source for detection and characterisation of adverse drug reactions. While signal detection algorithms were implemented for generating signals in pharmacovigilance databases, the implementation of a graphical user interface supporting the selection and display of algorithms' results is not documented in the medical literature. Although collecting information on the chronology and the impact of adverse drug reactions is desirable to enable causality and quality assessment of potential signals detected in patients' posts, no tool has been proposed yet to consider such data. We describe here two approaches, and the corresponding tools we implemented for: (1) quantitative approach based on signal detection algorithms, and (2) qualitative approach based on expert review of patient's posts. Future work will focus on implementing other statistical methods, exploring the complementarity of both approaches on a larger scale, and prioritizing the posts to manually evaluate after applying appropriate signal detection methods.

Keywords:
Drug-Related Side Effects and Adverse Reactions, Social Media, Pharmacovigilance

Introduction

Pharmacovigilance, defined as "the science and activities relating to the detection, assessment, understanding, and prevention of adverse effects or any other drug-related problem" [1] is mainly based on spontaneous reporting. A valid pharmacovigilance case should include four elements: a patient, at least one suspect medicinal product, one suspect adverse reaction and an identifiable reporter [2]. The presence of additional information, such as treatment dates, time-to-onset (TTO), drug indication, patient characteristics and outcome, ensures high quality information for Adverse Drug Reactions (ADR) [3] and allows causality assessment similarly to case reports usually evaluated in usual pharmacovigilance workflows.

However, one of the major limitations of the current pharmacovigilance systems is underreporting. Indeed, the underreporting rate is estimated to be between 78% and over 99% in France [4]. To overcome underreporting, social media was proposed as a potential source of information to complement case reports. Several previous studies have explored the added value of social media for pharmacovigilance [5-12].

Although social media cannot be considered as a "first line signal detection system" [11], research on social media for pharmacovigilance is required, as data extracted from this resource is different from data stored in pharmacovigilance databases [12]. Our work on the Vigi4Med project [12] investigated whether users' posts in social media were informative enough to be considered as pharmacovigilance reports (qualitative evaluation), and whether traditional signal detection methods were adapted to social media (quantitative evaluation).

Several tools facilitating the evaluation of posts have been described in the literature, such as Insight Explorer, MedWatcher social and PREDOSE [13-15]. PREDOSE focuses on drug abuse rather than ADRs surveillance. MedWatcher social does not allow the collection of supplementary data to perform causality assessment, and to measure if online posts can be considered as potential case reports. To our knowledge, Insight Explorer is the only tool that allows collecting of large amounts of data for causality and quality assessment. Indeed, users can enter data such as poster's information (patient age rank and gender, etc.), and product's description (route, length of use, dosage, etc.), but cannot enter information about case seriousness, adverse event's (AE) expectedness, and chronology. We propose that such tools should support additional information such as TTO, dechallenge (outcome when withdrawing a drug), and rechallenge (outcome when reintroducing a drug).

For the quantitative approach, signal detection methods have already been implemented to extract ADRs from social media [16]. To our knowledge, no graphical user interface has previously been described in the medical literature to facilitate the choice and application of signal detection methods when applied to social media data.

In this paper, we present the methods and tools that we developed and used to analyze the utility of social media for pharmacovigilance from qualitative and quantitative points of view. To demonstrate the performance and the usage of our approaches, we chose as a case study "strontium ranelate", a drug used for the treatment of osteoporosis since 2004 with safety concerns: after its marketing authorization in 2014, this drug was restricted in Europe due to its cardiovascular risk. Since March 2017, the marketing authorization holder has withdrawn the drug from the market.

Methods

Our qualitative approach is based on visualization and evaluation of a selection of users' posts from web forums by a pharmacovigilance expert, and on comparing posts' informativeness to French Pharmacovigilance Database

(FPVD)'s cases. This selection of posts is obtained automatically by extracting from web forums, annotating and filtering posts that contain at least one mention of "strontium ranelate" (or one of its commercial names), and one mention of an AE. The quantitative approach focuses on the analysis of the signals generated by statistical methods on social media.

In this section, we describe the steps we achieved to create our datasets, present our qualitative and quantitative analysis approaches and the tools that were developed for each of these approaches.

Dataset creation

FPVD's Dataset

Created in 1985, the French Pharmacovigilance Database (FPVD) centralizes anonymized reported cases of ADRs in France. By June 17, 2015, when we completed the extraction, the FPVD contained 582,193 reports. AEs recorded in the FPVD are coded using the Medical Dictionary for Regulatory Activities (MedDRA) which is an international medical dictionary approved by the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. This dictionary is organized in 27 System Organ Classes (SOC) – the highest hierarchical level – and in 4 other levels including the Preferred Terms (PT) level, traditionally used in pharmacovigilance.

FPVD extraction requests were performed by the ANSM, the French Agency for Drug Safety (*Agence Nationale de Sécurité du Médicament et des Produits de Santé*). The request criteria were:

- Investigated period: same study period as for the forums (from 01/01/2004 to 06/17/2015).
- Drugs: strontium ranelate coded as suspect.
- All cases whatever the seriousness and AE description.

Web forums' dataset

We developed the open source tool Vigi4Med Scraper [17] to extract over 60 million posts from 22 French medical forums. Natural Language Processing (NLP) and machine learning techniques were then applied to identify drugs and AEs in the extracted posts [18].

For the qualitative approach, we created a filtered dataset composed of posts containing at least one mention of "strontium ranelate" (or any of its variants) and at least one AE. This filtered dataset was automatically selected from a total of 6,569,555 posts in which 55,350,564 drug/AE pairs were identified. The period included in our study starts on 01/01/2004, which is the commercialization date of "strontium ranelate" in France, and concludes on 06/17/2015, the date of the last extracted post.

For the quantitative approach, we normalized all the mentions of drugs and AEs detected automatically in the posts by the NLP and machine learning methods to improve precision [19].

Ethical aspects

This French study has to comply with the European General Data Protection Regulation (GDPR). First, according to Article 6, processing is lawful when "(a) the data subject has given consent to the processing", but also applies when "(e) processing is necessary for the performance of a task carried out in the public interest". Second, according to Article 9, "Processing of personal data […] concerning health" shall apply when "(e) processing relates to personal data which are manifestly made public by the data subject". Third, our data protection officer was notified of the data collection, and posts were stored on a secure server. Finally, the GDPR does not apply to "personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable" which explains why we processed only pseudonymized information (see also Article 89 for the question of archiving data).

Qualitative analysis evaluation

Forums

For the qualitative approach, the pharmacovigilance experts were asked to evaluate forum posts from the filtered dataset. For each post, its characteristics and informativeness was evaluated regarding the following criteria:

- *The date of the post is included in the period of the study.*
- *The post contains a reporter, a patient, a drug and at least one AE that is possibly related to the drug and that can be coded by MedDRA.*
- *The post is from a user in France when the information was available.*
- *The post is about a personal experience.*
- *The post is not identified by the expert as a duplicate (the post is not exactly similar to or too close to another one according to the expert).*

The reviews of the posts were performed by a trained pharmacist. For doubtful or complicated posts, a trained pharmacovigilance expert made the final evaluation.

Graphical user interface for visualizing and validating user's posts

To enable pharmacovigilance experts to visualize and validate users' posts, a validation tool was developed using PHP and baseX, an open source XML database engine allowing indexation and interrogation of XML files (http://basex.org/). The input of this tool, in our case, is the XML file that contains users' posts that were automatically annotated by the NLP and machine learning tools.

As depicted in Figure 1, the different variables to assess by the pharmacovigilance specialist via the user interface were:

- Patient information: age, sex, medical history, and pregnancy status;
- Treatment information: drug, indication, dose, time period, and outcome;
- AE information: MedDRA term (as used in the FPVD) coded using PTs and SOCs, start date, and duration.

To evaluate a case, the expert had the possibility to evaluate and provide:

- Compatibility of the TTO, dechallenge and rechallenge;
- Drug causality;
- Case seriousness, based on Pharmacovigilance criteria (death, life threatening, etc);
- Expectedness: ADR labeled in the SPCs.

With this tool, validated users' posts can be exported to an XML file that contains new tags for the manual information registered by the pharmacovigilance specialist.

Comparison

We compared the informativeness (i.e. presence/absence of the information) of forums posts to the informativeness of case reports in the FPVD using the following variables:

- Patient's profile: age/class of age and sex;
- Treatment information: indication, dose and duration;
- ADR's outcome.

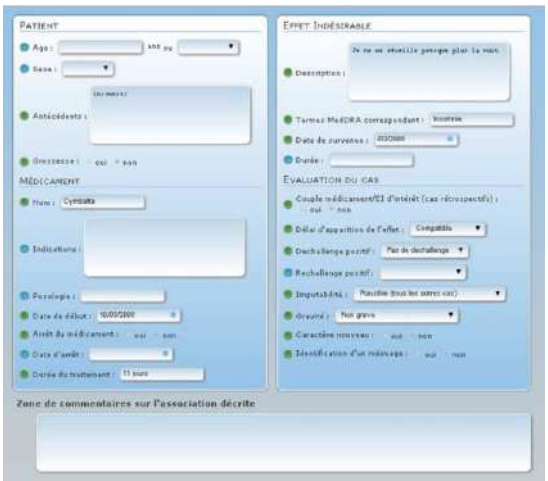


Figure 1– Screenshot of the validation tool

Furthermore, we compared the two data sources using the following variables:

- Patient's profile: age and sex;
- ADR's information: nature (SOC), seriousness and unexpectedness.

Student test was performed to compare the age (comparison of the mean). Chi2 and exact Fisher tests were used to compare qualitative variables. Tests were considered significant when p-value was lower than 5%.

In order to compare the SOC distribution, we conserved the four most represented SOCs in both data sources (with at least a size of five) and grouped the others in one category. Statistical calculations and comparisons were performed using R [21].

Quantitative evaluation

Graphical user interface to parameterize signal detection algorithms

The XML file exported from the validation tool was imported into R for statistical analysis. In order to easily and quickly apply signal detection algorithms to large amount of data extracted from web forums, we developed a web service that proposes traditional pharmacovigilance signal detection methods based on disproportionality. We used the implementation of the PhViD package to apply signal detection algorithms and decision rules as described below. For each of the algorithms, the graphical user interface enabled selection of the period to include, the minimum number of couples needed to generate a signal, and the drug to analyze, as shown in Figure 2. After choosing a method, the user has access to the drug/AE couples considered as a signal, the decision rule parameters value, and the ID of the posts describing the couple.

The decision rules we used to statistically consider a drug-AE couple as a signal [21] were based on the value of the lower bound of the 95% confidence interval ($CI_{95\%}$) of a statistic or of the p-value corrected with the False Discovery Rate (FDR) method:

1. Proportional Reporting Ratio (PRR):
   - $n \geq 3,\ PRR > 2\ and\ CI_{95\%}\left(ln(PRR)\right) > 0$
   - $CI_{95\%}\left(ln(PRR)\right) > 0, with\ FDR < 5\%$
2. Reporting Odds Ratio (ROR):
   - $CI_{95\%}\left(ln(ROR)\right) > 0$
   - $CI_{95\%}\left(ln(ROR)\right) > 0, with\ FDR < 5\%$
3. Reporting Fisher Exact Test (RFET, midRFET):
   - $RFET, with\ FDR < 5\%$
   - $midRFET, with\ FDR < 5\%$
4. Bayesian Confidence Neural Network (BCPNN):
   - $beta\ distribution\ a\ priori\ and$
     - $CI_{95\%}(Information\ Component) > 0$
     - $CI_{95\%}(Information\ Component) > 0\ with\ FDR < 5\%$
5. Gamma Poisson Shrinker (GPS):
   - $CI_{95\%}\left(ln(Empirical\ Bayes\ Geometric\ Mean)\right) > 2$
   - $CI_{95\%}\left(ln(Empirical\ Bayes\ Geometric\ Mean)\right) > 2, with\ FDR < 5\%$

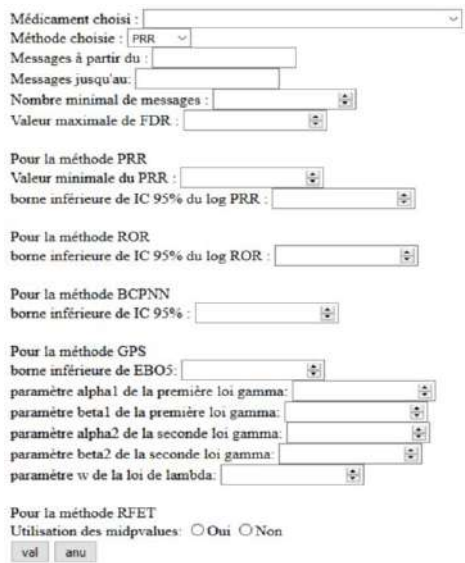


Figure 2– Screenshot of the signal detection tool (the input screen)

Results

Qualitative study

General findings

437 relevant cases were identified in the FPVD, with an average of 1.9 ADRs per case (840 ADRs in total). 387 relevant posts were detected in the web forum dataset. Among these posts, 47 were validated (12%) and 109 ADRs were detected (2.3 per post in average). These posts came from 7 of the 22 selected websites. Most of them came from esante.fr (60%) and 84% of the posts came from only 3 websites (esante.fr, doctissimo.fr and atoute.org).

Comparison of the informativeness

The informativeness of the cases in the FPVD was significantly higher than in the forums concerning patient information (95.7% vs 34% for age, 38.3% for class of age and 99.5% vs 40.4% for sex), the dose (49.9% vs 4.3%) and the outcome (89.5% vs 27.7%). Only the indication of the treatment was significantly more frequently known in forums than in the FPVD (55.3% vs 45.8%). There was no significant difference between the informativeness of the duration of treatment (66% in forums vs 74.1% in the FPVD).

Comparison of patients' profiles

Patients were significantly younger in forums than in the FPVD (mean age of 59.9 vs 72.0, p<0.0001) and they were mostly women in both data sources (94.7% in forums vs 97.2% in the FPVD) as presented in Table 1. No pregnancy was identified in forums, and this information was not available in the FPVD extraction.

Table 1– Comparison of the age and sex of patients in the FPVD and in the forums

Age Mean (sd) *Median (Q1-Q3)*		Sex Women *Men*	
Forums	*FPVD*	*Forums*	*FPVD*
59.9 (7.9) *58.0 (54.8-62.3)*	72.0 (11.1) *73.0 (64.0-80.0)*	18 (94.7%) *1 (53%)*	423 (97.2%) *12 (2.8%)*

Comparison of ADRs' profiles

Only two serious cases were found in forums against 203 in the FPVD (4.3% vs 46.5%, p<0.0001). No death was identified in forums whereas 7 were found in the FPVD.

There was no significant difference between the two sources for the number of cases containing at least one unexpected ADR (25.5% in forums vs 22.1% in the FPVD). Only three PT terms identified in forums were not reported in the FPVD: "Weight increased", "depression" and "hypertension".

There were more SOCs describing the ADRs in the FPVD than in forums (22 vs 11) which can explain the fact that the four first SOCs in the FPVD represented only 60% of the ADRs found (487 among 840) whereas they represented 72% of the ADRs found in the forums (79 among 109). The four SOCs most represented in the FPVD were "Skin and subcutaneous tissue disorders" (31%), "Gastrointestinal disorders" (13%), "Respiratory, thoracic and mediastinal disorders" (10%), and "General disorders and administration site conditions" (8%). In forums, the four most represented SOCs were "Nervous system disorders" (22%), "Gastrointestinal disorders" (21%), "Musculoskeletal and connective tissue disorders" (16%), and the SOCs "General disorders and administration site conditions" and "Skin and subcutaneous tissue disorders" (both representing 14% of the ADRs). Thus, SOC's distribution was significantly different between the FPVD and the forums (p < 0.0001), SOCs' distribution being more diverse in the FPVD.

Quantitative study

After normalization of drugs and AEs, we found 2,582,287 couples corresponding to 194,318 unique couples in 731,043 posts from 07/03/2000 to 07/02/2015.

143 unique drug /AE couples with strontium ranelate (1,021 couples in total) were detected in 297 posts published between 06/28/2005 and 06/09/2014. 21 of the 143 AEs (15%) found with strontium ranelate were either the indication ("Osteoporosis", "Bone loss", "Osteopenia" or "Osteoporosis postmenopausal") or diseases linked to or increasing the risk of osteoporosis. These 21 AEs represented 51% (524) of the 1,021 couples found with strontium ranelate in total.

21 AEs were considered as signals applying all the decision rules. 33 percent of these 21 AEs were from the SOC "Musculoskeletal and connective tissue disorders" and 19% from the "Neoplasms benign, malignant and unspecified (incl cysts and polyps)" SOC.

Among the 21 statistical signals, three (14%) were the direct indication ("Osteoporosis", "Bone loss" and "Osteopenia"), 10 (48%) were diseases linked to or increasing the risk of osteoporosis (among them: "Cystic fibrosis", "Breast cancer", "Psoriasis", "Neoplasm malignant", "Multiple fractures", "Multiple sclerosis", and "Rheumatoid arthritis"), and eight (38%) were AEs among which three were unexpected: "Atrial fibrillation", "Rhabdomyolysis" and "Toxic epidermal necrolysis" representing five, three and three couples respectively. These 11 couples corresponded to eight posts reporting two lists of drugs recommended to avoid, one from a consumer advocacy association (UFC-QUE CHOISIR) in March 2011, and another from a national association for patients suffering from ADR.

Discussion

This retrospective study allowed us to identify 47 pharmacovigilance cases concerning strontium ranelate among 387 posts extracted from French language forums reviewed by pharmacovigilance experts.

Although the average number of ADRs per post/case was the same in the forums and the FPVD (2.3 vs 1.9), forums' posts were significantly less informative than FPVD cases concerning patient's information (age, sex), treatment's dose and duration as well as the AE's evolution. Only the indication of the treatment was significantly more frequently known in forums than in the FPVD. This result was also found in a larger scaled study that we previously achieved for six drugs [12].

Patients from both data sources were mostly women. Based on the 34% of posts containing gender information, patients in forums were younger than those in the FPVD (mean age of 60 *vs* 72). These results are concordant with published studies: women and young patients are more prone to publish posts on social media [6, 12].

The proportion of non-serious cases reported in web forums was significantly higher than those reported in the FPVD (95.7% vs. 53.5%) which confirms the results found in the medical literature [6, 7, 12], but no significant difference was found between both sources for the number of cases containing at least one unexpected ADR.

The distribution of the SOCs was significantly different ($p < 0.0001$) between the two sources and was more diverse in the FPVD than in the forums: the three most frequent SOCs represented 60% of those in the FPVD vs. 72% in the forums. This could be due to the collaborative aspect of web forums: a first post could lead to several others reporting the same type of ADR in the discussion topic.

If tools have already been developed to evaluate posts from social media [13-15], none gave the possibility to evaluate the medical history, pregnancy status, indication, rechallenge, TTO, drug causality, seriousness or expectedness of the AE. Moreover, no graphical user interface has previously been described in the medical literature to facilitate the application of signal detection methods on social media data.

A major bias of the quantitative study is that the NLP tool could not detect causality – thus, three unexpected severe ADRs considered as signals by the signal detection algorithms were actually false positives - while the qualitative study identified 47 cases showing causality relations from 387 posts (12%). Three AEs identified in forums in the qualitative approach and not reported in the FPVD were not detected as signals in the quantitative approach. However, these AEs may bring interesting information complementary to data available in spontaneous reports. Moreover, the quantitative evaluation also showed that indications were captured as AEs in some cases: 51% of the 1,021 AEs found were either the indication, either diseases linked to or increasing the risk of osteoporosis.

Another limitation was that information concerning strontium ranelate and other drugs was probably under-estimated, as we used a perfect match to normalize drug terms. This study illustrates current limitations of using social media to complement spontaneous reports for signal detection.

Conclusions

We developed two tools to facilitate the evaluation of web forums' posts for pharmacovigilance purpose and used them in a case study focusing on one drug. We found that these tools are complementary. Findings of the quantitative study may be biased by false positives. A qualitative study is difficult to perform on a large number of drugs and posts as it requires important time and workforce. This is the reason why a qualitative approach is usually restricted to signals detected by the quantitative approach.

Future work will focus on reducing bias related to the absence of detection of causality at the linguistic level, implementing other statistical methods (logistic regression, change-point analysis, etc.), and exploring the complementarity of the two approaches on a larger scale, prioritizing the posts that should be manually evaluated after having applied appropriate signal detection methods.

Acknowledgements

This work was funded by grant AAP-2013-052 from the ANSM, the French Agency for Drug Safety (*Agence Nationale de Sécurité du Médicament et des Produits de Santé*) through the Vigi4MED research project and Convention n° 2016S076 through the PHARES project.
Damien Leprovost implemented the tool used for the qualitative study. Rim Aboukhamis evaluated the posts in the qualitative study.
We acknowledge ANSM for providing the FPVD extraction, CHU of Rouen, CISMeF team for providing RacinePharma, Vidal for linking it to the ATC, classification and CNRS-LIMSI for detecting mentions of drugs and AEs in posts.

References

[1] World Health Organization (2018). *Pharmacovigilance.* Geneva: World Health Organization. doi: 10.3414/ME12-02-0004

[2] European Medicine Agency (2017). *Guideline on Good Pharmacovigilance Practices (GVP) - Module VI – Management and Reporting of Adverse Reactions to Medicinal Products.* London: EMA.

[3] I. R. Edwards, M. Lindquist, B. E. Wilholm, B, and E. Napke. Quality criteria for early signals of possible adverse drug reactions. *Lancet.* **336** (1990) 156–158.

[4] L. Hazell, SA. Shakir. Under-Reporting of Adverse Drug Reactions: A Systematic Review. *Drug Saf.* **29** (2006) 385–396.

[5] B. W. Chee, R. Berlin, and B. Schatz, Predicting Adverse Drug Events from Personal Health Messages, *AMIA. Annu. Symp. Proc.* **2011** (2011) 217–226.

[6] S. Golder, G. Norman, and YK. Loke, Systematic review on the prevalence, frequency and comparative value of adverse events data in social media. *Br. J. Clin. Pharmacol.* **80** (2015) 878–888. doi: 10.1111/bcp.12746

[7] J. Lardon, R. Abdellaoui, F. Bellet, H. Asfari, J. Souvignet, N. Texier, et al. Adverse drug reaction identification and extraction in social media: a scoping review. *J. Med. Internet Res.* **17** (2015). doi: 10.2196/jmir.4304

[8] A. Sarker, R. Ginn, A. Nikfarjam, K. O'Connor, K. Smith, S. Jayaraman, et al. Utilizing social media data for pharmacovigilance: a review. *J. Biomed. Inform.* **54** (2015) 202–212. doi: 10.1016/j.jbi.2015.02.004

[9] R. Sloane, O. Osanlou, D. Lewis, D. Bollegala, S. Maskell, and M. Pirmohamed. Social media and pharmacovigilance: a review of the opportunities and challenges. *Br. J. Clin. Pharmacol.* **80** (2015) 910–920. doi: 10.1111/bcp.12717

[10] O. Caster, J Dietrich, ML. Kürzinger, M. Lerch, S. Maskell, GN. Norén, S. Tcherny-Lessenot, B. Vroman, A. Wisneiwski, J. van Stekelenborg. Assessment of the Utility of Social Media for Broad-Ranging Statistical Signal Detection in Pharmacovigilance: Results from the WEB-RADR Project. *Drug Saf.* **41** (2018) 1355–1369. doi: 10.1007/s40264-018-0699-2.

[11] C. Bousquet, B. Audeh, F. Bellet, A. Lillo-Le Louët. Comment on "Assessment of the Utility of Social Media for Broad-Ranging Statistical Signal Detection in Pharmacovigilance: Results from the WEB-RADR Project". *Drug Saf.* **41** (2018) 1371–1373. doi: 10.1007/s40264-018-0747-y.

[12] P. Karapetiantz, F. Bellet, B. Audeh, J. Lardon, D. Leprovost, R. Aboukhamis, et al. Descriptions of adverse drug reactions are less informative in forums than in the French pharmacovigilance database but provide more unexpected reactions. *Front Pharmacol*. **9** (2018). doi: 10.3389/fphar.2018.00439.

[13] T. Casperson. et al. Strategies for distributed curation of social media data for safety and pharmacovigilance. *Proceedings of the International Conference on Data Moining (DMIN)*. The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp) (2016) 118–124.

[14] C. Freifeld, "Digital pharmacovigilance: the medwatcher system for monitoring adverse events through automated processing of internet social media and crowdsourcing." (2014). Thesis.

[15] D. Cameron et al., PREDOSE: a semantic web platform for drug abuse epidemiology using social media. *J Biomed Inform*. **46** (2013) 985–97. doi: 10.1016/j.jbi.2013.07.007.

[16] S. Yeleswarapu et al., A pipeline to extract drug-adverse event pairs from multiple data sources. *BMC Med Inform Decis Mak*. **14** (2014) 13. doi: 10.1186/1472-6947-14-13.

[17] B. Audeh, M. Beigbeder, A. Zimmermann, P. Jaillon, C. Bousquet. Vigi4Med Scraper: A Framework for Web Forum Structured Data Extraction and Semantic Representation. *PloS one*. **12** (2017).

[18] C. Grouin, F. Morlane-Hondere, P. Zweigenbaum, Identification of Drug-Related Medical Conditions in Social Media, *LREC*, *International Conference on Language Resources and Evaluation* (2016).

[19] P. Karapetiantz, B. Audeh, A. Lillo-Le Louët, C. Bousquet. Signal Detection for Baclofen in Web Forums: A Preliminary Study. *Stud Health Technol Inform*. **247** (2018) 421-425.

[20] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2016).

[21] I. Ahmed. Détection automatique de signaux en pharmacovigilance : Approche statistique fondée sur les comparaisons multiples (2009). Doctoral dissertation, Paris 11.

Address for correspondence

Pierre Karapetiantz

E-mail: pierre.karapetiantz@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190368

Open Source HMIS Enabled Evaluation of Financial Burden of Disease and Patient Coverage in Three University Hospitals in Great Lakes Africa

Gustave Karara[a,c], Frank Verbeke[c] , Jean-Claude Byiringiro[b], Franck Nziza[c], Ronald Buyl[a], Marc Nyssen[a]

[a] *Department of Public Health, Biostatistics and Medical Informatics, Vrije Universiteit Brussel, Brussels, Belgium*
[b] *Division of Clinical, Education and Research, University Teaching Hospital of Kigali, Kigali, Rwanda*
[c] *Burundi Health Informatics Association, Bujumbura, Burundi*

Abstract

Since the eighties, case mix evaluation methods based on diagnosis-related groups (DRG) were gradually introduced in developed countries. These methods of assessing the costs of diseases to measure the productivity of the hospital have been introduced in management softwares that are not accessible to low-income countries. In this study, the authors applied these methods to an open source hospital management information system (HMIS) implemented in three university hospitals in Great Lakes Africa. A comparative study of the financial burden of five major diseases, monitored as part of a universal health coverage (UHC) analysis, was carried out. The level of coverage of patients in the hospitals was evaluated and the impact of UHC policies demonstrated. Although the financial protection of patients treated in the three hospitals had improved, HIV and tuberculosis treatments that ought to be free, remained a considerable financial burden for the patient.

Keywords

Diagnosis-Related Groups; Hospitals, University; Africa

Introduction

Knowing the financial burden of disease remains a challenge in sub-Saharan countries. In this study we focused on three countries in the Great Lakes region: Rwanda, Burundi and DRC. WHO and the WB [1] evaluated the coverage of the population for the treatment or prevention of tracer health services and Boema et al. [2] estimated coverage for five diseases as a metric for the UHC progress in countries. The evaluation mainly used household surveys from the Demographic and Health Survey (DHS). The indicators of health coverage focused on maternal and newborn care, the traitement of HIV infection, hypertension, tuberculosis (TB) and diabetes. Health coverage diminishes when household income is lower [3; 4]

Tuberculosis treatment was highly covered in Rwanda, Burundi and DRC (between 80 and 90%) thanks to the national TB programs. ARV therapy was accessible to 67%, 40% and 20% of population with advanced HIV infection in Rwanda, Burundi and DRC respectively. But values for diabetes coverage remained particularly low in the three countries (<10%). The evaluation focused on the general framework of health services accessibility and coverage, without zooming in on the individual patient's financial situation in health facilities. Questions such as “What is the patient's financial coverage for the treatment of these diseases monitored in the context of the UHC?”, “What is the patient's ability to pay for the treatment?”, “What role does health insurance play?” remained unanswered.

According to the DHS [5], UHC policies in the three countries in the Great Lakes region are very different. Rwanda opted for a community based health insurance (Mutuelle de santé) which covers more than 90% of the population. Burundi has strengthened its UHC policy with free healthcare for pregnant women and children under five years old, reaching coverage of 23-30% of the population. In the DRC, we observe a low level of population health coverage (2%). In DRC, patients nearly pay for the full costs of health services they consume.

Alternative methods, based on health facility data, have been developed with the same purpose of measuring financial burden of diseases and to promote better quality of care. These methods are at the basis of the case mix concept introduced in the USA around the 1980s [6] This concept is based on DRGs (Diagnosis Related Groups), and was developped by a team of researchers at the Yale University and consequently introduced in the *MEDICARE* prospective payment system [7]. The DRG is defined as the first “health management tool” to group patients in clinically meaningful categories with homogeneous resources consumption [6]. The DRG-based payment systems have been introduced in Europe in the mid-nineties [8]. The first evaluation of the use of the case mix tool in Africa is found in private health facilities in South Africa in the early 2000s [9] . These were pilot experiments introduced in *3M* software which used the ICD-10 classification for diagnostics and CPT for medical procedures in service pricing. The challenge in sub-Saharan health facilities remained to find affordable tools that integrated these international standards, were easy to use and better adapted to the local context.

In 2009, we worked out a set of disease grouping codes in an attempt to enable efficient evaluation of clinical activity in a typical sub-Saharan health facility [10]. This classification was called the KHIRI Pathology Grouping Set (KPGS). KPGS is a bi-classified grouping system, based on ICD-10 and ICPC-2 classification standards. The KPGS classification aimed to be a simplified sub-Saharan implementation of DRG codes, addressing clinical conditions that better matched local sub-Saharan health management reality [11]. We introduced this classification in an open source hospital management information system (HMIS) called *OpenClinic GA* [10; 12]. *OpenClinic GA*-HMIS was developed by a research team of the Vrije Universiteit Brussel (VUB) and has been put in the public domain (https://sourceforge.net/projects/open-clinic/). The first implementation was at the university hospital of Kigali (CHUK), where the KHIRI (Kigali Health Informatics Research Institute) was established [13]. At a later stage, a data warehouse enabling fully automatic extraction of KPGS coded health indicators from *OpenClinic GA* hospital databases worldwide, named the *Global Health Barometer* (GHB, http://www.globalhealthbarometer.net), was added to the project.

This study attempts to use the KPGS classification to provide case mix evaluation extracted from OpenClinic GA databases

for estimating the financial burden of diseases and the level of patient coverage provided by health insurance schemes.

Methods

The study was conducted during a 6-year period from 2010 to 2016 in three university hospitals of Rwanda, Burundi and DRC by collecting data through *the OpenClinic GA* software. Data related to patient identification, ADT (admission, discharge, transfer), patient- and insurance invoicing, diagnostics and other clinical medical records content.

The following hospitals participated in the study:

1. The University Teaching Hospital of Kigali (CHUK) in Rwanda where the first implementations of OpenClinic GA started in 2007.
2. The University Teaching Hospital of Kamenge (CHURK) in Burundi where the OpenClinic GA implementation started in 2014, and
3. The Provincial General Reference Hospital of Bukavu (HPGRB) in DRC where the OpenClinic GA implementation started in 2011.

The process of *OpenClinic GA* implementation included the setup and configuration of the software, staff training, follow-up, quality control, monitoring and evaluation.

In each of the hospitals, we analysed and compared:

1. The main disease groups encountered in the hospitals. Data were extracted from the GHB data warehouse. We focused on the five disease groups monitored within the framework of UHC: Maternal and new-born health, diabetes, hypertension, tuberculosis and HIV.
2. The financial burden was calculated in terms of care costs generated by each of the five disease groups and split up into patient out-of-pocket and health insurance payments. This information was directly extracted from the hospital OpenClinic GA databases. Disease cost factors have been furthermore identified using the CALCO method. This approach takes into account the amounts of health services consumed and, the disability and mortality weight factors of a disease group compared to associated pathologies (comorbidity) recorded during one and the same encounter [11] [14]. Care deliveries and medical procedures had been registered using local coding schemes. Diagnoses were coded using ICD-10 and ICPC-2 and have been mapped onto the KPGS classification.
3. Levels of patient coverage were derived from average amounts paid by patients and reimbursed by insurer schemes. We calculated the patient health service payment rate (PHSP) by dividing the patient out-of-pocket payment (POOP) by the total amount of health services consumed. The PHSP defines the level of the patient share in the health services consumed and should, in the light of UHC, not exceed 25% of the total cost of health services consumed. To avoid ruinous expenses for the patient, the POOP should also not exceed 25% of the GDP per capita in the country.

Results

Main KPGS-disease groups monitored

Table 1 shows the top 10 of disease groups registered in the three university hospitals during the year 2016. These disease groups represent respectively 58%, 66% and 59% of all cases at CHUK, CHURK and HPGRB. At CHUK in Rwanda, the first in-patient disease group is related to the complications of pregnancy, childbirth and the puerperium (1679 cases). This disease group (KPGS code 15B) covered 13% of all cases in 2016 at CHUK and was also found in the top 10 of disease groups at CHURK (10th group, 2.1%) and HPGRB (3rd group, 8.1%). Genitourinary system diseases (KPGS code 14O) constituted another disease group that found its way to the top 10 disease groups recorded in each of the three university teaching hospitals.

Malaria (KPGS code 1V) was the first disease group encountered at CHURK in Burundi. This pathology group did not appear in the top 10 pathologies group encountered either at CHUK or HPGRB.

Figure 1 presents the disease group overlap for the top 10 disease groups (KPGS codes) in the three hospitals.

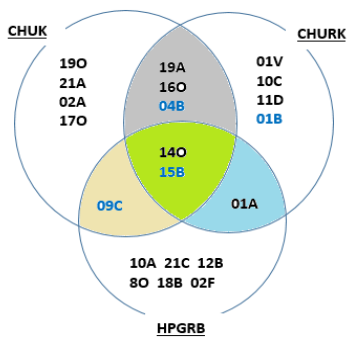


Figure 1- Top 10 KPGS codes overlap in three university hospitals of Rwanda, Burundi and DRC (2016)

If we focus on the five disease groups monitored as indicators of UHC progress, we see that:

1. 15B (Other complications of the pregnancy, childbirth and the puerperium) is found in the top 10 of pathologies in the three hospitals.
2. 04B (Diabetes mellitus) is encountered in the top 10 of diseases at CHUK of Rwanda and CHURK of Burundi.
3. 09C (Hypertensive diseases) is encountered in the top 10 of diseases at CHUK (Rwanda) and HPGRB (DRC).
4. 01B (Tuberculosis) is encountered in the top 10 of pathologies treated at CHURK in Burundi only.

HIV, the fifth treatment intervention monitored by the health coverage at global level was not among the top 10 disease groups in any of the three hospitals in 2016. In 2010, HIV still represented a morbidity rate of 4.5% (1594 in-patient cases, ranked at the third place) at CHUK in Rwanda.

Financial burden per pathology group

Table 2 shows the disease weight scores used from the CALCO method and the number of in-patient cases related to the five disease groups monitored in the framework of health coverage analysis during the period of study.

The data analysis period was different in the three hospitals. For each hospital, we only took into account the period during which financial information and diagnostics have been recorded in the *OpenClinic GA* system.

Figure 2 illustrates a graph of financial burden by disease group. It also shows a table of the POOP and the health insurance share in disease treatment costs. We have expressed the money values in the three countries in international dollars for the sake of consistency and comparability.

Table 1- Top 10 of disease groups in three university hospitals of Rwanda, Burundi and DRC (2016)

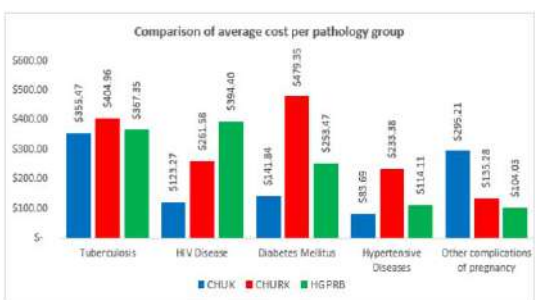
Top 10 of pathologies (CHUK)			
Code	Pathology	# Cases	%
15B	OTHER COMPLICATIONS OF THE PREGNANCY, CHILDBIRTH AND THE PUERPERIUM	1679	13.1%
14O	DISEASES OF THE GENITOURINARY SYSTEM	1515	11.8%
19A	FRACTURES	1429	11.1%
16O	CERTAIN CONDITIONS ORIGINATING IN THE PERINATAL PERIOD	575	4.5%
19O	OTHER INJURY, POISONING AND CERTAIN OTHER CONSEQUENCES OF EXTERNAL CAUSES	554	4.3%
09C	HYPERTENSIVE DISEASES	512	4.0%
21A	PERSONS ENCOUNTERING HEALTH SERVICES FOR EXAMINATION AND INVESTIGATION	335	2.6%
02A	MALIGNANT NEOPLASMS, STATED OR PRESUMED TO BE PRIMARY, OF SPECIFIED SITES, EXCEPT OF LYMPHOID, HAEMATOPOIETIC AND RELATED TISSUE	308	2.4%
04B	DIABETES MELLITUS	302	2.4%
17O	CONGENITAL MALFORMATIONS, DEFORMATIONS AND CHROMOSOMAL ABNORMALITIES	264	2.1%
		7473	58.2%

Top 10 of pathologies (CHURK)			
Code	Pathology	# Cases	%
01V	MALARIA	444	20.7%
19A	FRACTURES	224	10.4%
16O	CERTAIN CONDITIONS ORIGINATING IN THE PERINATAL PERIOD	185	8.6%
14O	DISEASES OF THE GENITOURINARY SYSTEM	110	5.1%
04B	DIABETES MELLITUS	105	4.9%
10C	PNEUMONIA	82	3.8%
11D	HERNIA	74	3.5%
01A	INTESTINAL INFECTIOUS DISEASES	74	3.5%
01B	TUBERCULOSIS	64	3.0%
15B	OTHER COMPLICATIONS OF THE PREGNANCY, CHILDBIRTH AND THE PUERPERIUM	44	2.1%
		1406	65.6%

Top 10 of pathologies (HPGRB)			
Code	Pathology	# Cases	%
14O	DISEASES OF THE GENITOURINARY SYSTEM	499	16.3%
10A	ACUTE UPPER RESPIRATORY INFECTIONS	248	8.1%
15B	OTHER COMPLICATIONS OF THE PREGNANCY, CHILDBIRTH AND THE PUERPERIUM	248	8.1%
21C	PERSONS ENCOUNTERING HEALTH SERVICES IN CIRCUMSTANCES RELATED TO REPRODUCTION	230	7.5%
09C	HYPERTENSIVE DISEASES	147	4.8%
12B	DERMATITIS AND ECZEMA	125	4.1%
01A	INTESTINAL INFECTIOUS DISEASES	88	2.9%
8O	DISORDERS OF THE EARS AND PROCESSUS MASTOIDEUS	84	2.7%
18B	SYMPTOMS AND SIGNS INVOLVING THE DIGESTIVE SYSTEM AND ABDOMEN	77	2.5%
02F	BENIGN NEOPLASMS	76	2.5%
		1822	59.4%

Table 2- Number of in-patient cases and weight score in five disease groups at CHUK, CHURK and HPGRB during the period of study

Code	Pathology group	Weight-Score	CHUK (2010-2016)	CHURK (2014-2016)	HGPRB (2014-2016)
01B	Tuberculosis	271	1140	421	10
01M	HIV Disease	135	2618	112	15
04B	Diabetes Mellitus	175	6118	267	158
09C	Hypertensive Diseases	246	4332	35	460
15B	Other complications of pregnancy	50	13075	498	311



Pathology group	Participation	CHUK	CHURK	HGPRB
Tuberculosis	POOP	$ 46.93	$ 162.78	$ 351.97
	Insurance	$ 308.54	$ 242.18	$ 15.58
HIV Disease	POOP	$ 45.43	$ 58.83	$ 293.40
	Insurance	$ 77.83	$ 202.75	$ 100.55
Diabetes Mellitus	POOP	$ 25.34	$ 250.68	$ 155.48
	Insurance	$ 116.50	$ 228.67	$ 97.51
Hypertensive Diseases	POOP	$ 14.18	$ 152.01	$ 61.15
	Insurance	$ 69.51	$ 81.37	$ 52.64
Other complications of pregnancy	POOP	$ 67.97	$ 0.19	$ 53.06
	Insurance	$ 227.24	$ 135.09	$ 50.65

Figure 2- Financial burden (in Intl$) per patient and per insurance for five disease groups at CHUK, CHURK and HPGRB (2010-2016)

The financial burden for treating diabetes, tuberculosis and hypertension are the highest in CHURK. Treatment of HIV is most expensive in HPGRB (Intl$394), three times the cost of CHUK (Intl$123). In contrast, the treatment of "other complications of pregnancy" is twice as high in the CHUK (Intl$295) compared to the two other teaching hospitals. There is not much difference in the financial burden of tuberculosis treatment in the three university hospitals. The average treatment costs vary between Intl$355 and Intl$405.

The patient share (POOP) for tuberculosis treatment (Intl$352) and HIV disease (Intl$293) are the highest at HPGRB and represent respectively 95.8% and 74.5% of the total amount of treatment costs. POOPs for the treatment of diabetes (Intl$ 251) and hypertension (Intl$152) are the highest in CHURK and represent respectively 52.3% and 65.1% of the total financial burden for treatment. The POOP for "other pregnancy complications" (Intl$68) is the highest at CHUK, while it is around zero at CHURK thanks to the free health care policy applied in Burundi for children under five years and pregnant women.

Health services coverage per disease group in the three hospitals

Figure 3 illustrates the average level of health service coverage for each disease group in the three university hospitals.

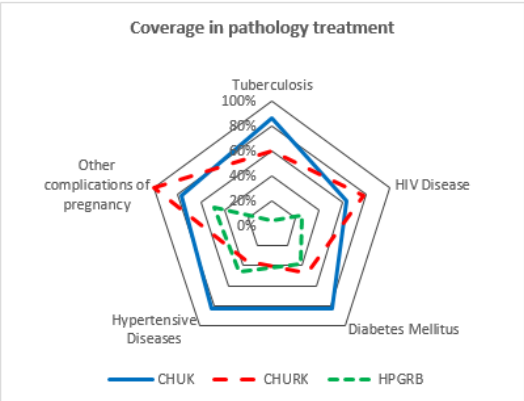
At CHUK, the financial health service coverage of the studied disease groups varies between 63.1% (for HIV) and 86.8% (for tuberculosis). In CHURK, this coverage is close to 100% for "other complications of pregnancy" and remains low (34.9%) for hypertensive diseases. Finally, in HPGRB, the treatment costs of tuberculosis seem to remain predominantly uncovered (only 4.2% coverage) for the patient, and also other groups of diseases have a low coverage rate (between 25.5% and 48.8%) compared to the two other sites.

Discussion

Our study of the financial burden of diseases showed that the situation of case mix in the three university hospitals in the Great Lakes region was not homogeneous both for disease costs and for patient coverage. For tuberculosis for example, the health coverage varied between 87% at CHUK and only 4% at HPGRB.

The patient's financial burden (PHSP) depends on the cost of the health services consumed on the one hand and the level of health coverage on the other hand. It varied between

37% (for HIV) and 13% (for tuberculosis) in CHUK, 65% (for hypertensive diseases) and 0% (for other complications of pregnancy) in CHURK. It was very high in HPGRB (between 96% for tuberculosis and 51% for other complications of pregnancy). Even if the patient share in the health service costs (PHSP) is low in CHUK, it still exceeds the 25%-threshold of health services consumed for the HIV treatment. HIV treatment is not well covered in the CHUK and HPGRB in spite of the free ARVs treatment. The PHSP remains above the 25%-threshold for all disease groups studied in HPGRB



Code	Pathology group	CHUK	CHURK	HPGRB
01B	Tuberculosis	86.8%	59.8%	4.2%
01M	HIV Disease	63.1%	77.5%	25.7%
04B	Diabetes Mellitus	82.1%	47.7%	38.5%
09C	Hypertensive Diseases	83.1%	34.9%	46.3%
15B	Other complications of pregnancy	77.0%	99.9%	48.8%

Figure 3- Health care coverage for five disease groups in CHUK, CHURK and HPGRB

and CHURK, expect for other complications of pregnancy where the PHSP was 0% in CHURK thanks to the policy of free healthcare for pregnant women. Patients may be pushed into poverty if they have to be treated for these diseases not well covered by health insurance schemes.

Treatment of tuberculosis and HIV is supposed to be "free" for the patient, thanks to the specific health programs set up in the countries to fight against these diseases. However, there always remain uncovered costs associated to these treatments. Our findings match the results of several studies that have also found sometimes high associated costs of health care for patients receiving 'free' HIV and tuberculosis treatment [15] [16; 17]. There are indirect costs related to transport, food and waiting time; direct costs due to the treatment of opportunistic infections, self-medication, extra medical exams and traditional healers' services. Chimbindi estimated annually expenditures to USD200 for patients suffering from tuberculosis and USD270 for HIV patients receiving public care in rural KwaZulu-Natal [15] while Laokri estimated tuberculosis associated costs to USD101 in six rural districts of Burkina Faso [16]. On our side, the average POOP associated to these two pathologies varied between USD20 and USD217 for tuberculosis treatment and between USD19 and USD182 for HIV treatment in the three university hospitals of the Great Lakes region. In the HPGRB in DRC, the POOP is more than 50% of the GDP per capita, which could financially ruin the treated patient.

POOPs are related to the treatment of the main disease plus comorbidities that increase the financial burden and add up their disability and mortality weight factors. We have not included the costs of external factors, such as transportation, unproductivity due to time-loss, treatment in the private sector... that may furthermore increase the POOP for the studied diseases.

Although we have noted the important role of health insurance in the coverage of health expenditures in Burundi and Rwanda, appropriate policies should further ensure the management of comorbidities of HIV and tuberculosis to reduce the financial burden for patients and reach total coverage for the two pathologies.

Conclusions

KPGS codes derived from ICD-10 and ICPC-2 classifications integrated into the HMIS software enabled comparison of financial burden and coverage of five disease groups in three university teaching hospitals (CHUK, CHURK and HPGRB) in Great Lakes Africa. Coverage for these diseases is used to monitor UHC progress in countries. The evaluation using case mix methods shows that the situation is quite different in the three hospitals, in terms of total financial burden of disease as well as financial patient coverage.

Our analysis method allowed to detect effects of the free health care policy for pregnant women applied in Burundi in covering pregnancy related problems and the role of CBHI in health services coverage in Rwanda. The method also enabled measuring the financial burden of comorbidity related to treatments for diseases such as HIV and tuberculosis. The same method could also be applied in other low-income countries through an implementation of adequate ICT infrastructure for health information management. Additional broader studies involving more health facilities are needed in order to further evaluate the role of health insurance in covering the financial burden of diseases.

References

[1] WHO and WB. Tracking universal health coverage, First global monitoring report (2015).

[2] Boerma T., Abou Zahr C., Evans T., Monitoring Intervention Coverage in the Context of Universal Health Coverage, *PLoS Med* 11(9) (2014).

[3] Hosseinpoor, A.R., Bergen, N., Koller, T., Prasad, A., Schlotheuber, Valentine, N., Lynch J. and Vega J. (2014). Equity-oriented monitoring in the context of universal health coverage. PLoS Med 11(9): e1001727.

[4] Aguilera, X., Castillo-Laborde, C., Nájera-De Ferrari, M., Delgado, I., Ibañez, C. Monitoring and evaluating progress towards universal health coverage in Chile. *PLoS Med* 11(9*)* (2014).

[5] WHO and WB. Monitoring progress towards universal health coverage at country and global levels: Framework, measures and targets. May 2014.

[6] Roger France, F.H., Mertens, I., Closon, M.C. and Hofdijk, J. Case Mix: Global Views, Local Actions. Evaluation in Twenty Countries. *Studies in Health Technology* Vol 86 (2001).

[7] Mihailovic, N., Kocic, S. and Jakovljevic, M. Review of Diagnosis-Related Group-based financing of hospital care. *Health Services Research and Managerial Epidemiology*, May 12 (2016).

[8] Langenbrunner J.C., Cashin C., O'Dougherty S. Designing and Implementing Provider Payment Systems: How to Manuals. Washington, DC: The World Bank (2009).

[9] Ruff B. Case Mix in South Africa, Pilot projects. In Case Mix: Global Views, Local Actions. Evaluation in Twenty Countries. *Studies in Health Technology* Vol 86 (2009)

[10] Verbeke, F., Murekatete, C., Tran Ngoc, C., Karara, G., Gasakure, E., Nyssen, M. Evaluation of a new method for in-patient co-morbidity analysis based on KHIRI Pathology Group Set codes at the Kigali University Teaching Hospital. *Journal for Health Informatics in Developing Countries*; Vol 3, No 2 (2009).
[11] Verbeke, F., Hategekimana, T., Karara, G., Ly, O., Nimburanira, M., Kayembe, C. and Nyssen, M. Information systems for monitoring the burden of chronic diseases in public reference health facilities in Central Africa. *Journal of Health Informatics in Africa*, Vol 1, No 1 (2013).
[12] Karara, G., Verbeke, F., Ndabaniwe, E., Mugisho, E. and Nyssen, M.). OpenClinic GA Open Source Hospital Information System enabled Universal Health Coverage Monitoring and Evaluation in Burundian Hospitals. *Studies in health technology and informatics*; 245 (2017), 738-742.
[13] Karara, G., Verbeke, F. and Nyssen, M. The Role of Hospital Information Systems in Universal Health Coverage Monitoring in Rwanda. *Studies in health technology and informatics,* 216 (2015), 193-197.
[14] Verbeke, F., De Pauw, F., Tran Ngoc, C., Karara, G., Gasakure, E. and Nyssen, M. Evaluating the relevance of disability weights for adjusting disease-cost and comorbidity calculations at the Kigali University Teaching Hospital. *Studies in Health Technology and Informatics*; 160(Pt 1) (2010):769–73.
[15] Chimbindi, N. et al. Time and money: the true costs of health care utilization for patients receiving 'free' HIV/TB care and treatment in rural KwaZulu-Natal. *J Acquir Immune Defic Syndr. Oct 1*; *70(2): e52–e60* (2015)
[16] Laokri, S., Drabo, M.K., Weil, O., Kafando, B., Mathurin, S., Dujardin and B. Patients Are Paying Too Much for Tuberculosis: A Direct Cost-Burden Evaluation in Burkina Faso. *PLoS One*, Feb 25 (2013)
[17] Hardon, A.P. et al., waiting time and transport costs: time to confront challenges to ART adherence in Africa. *AIDS care*, May;19(5) (2007), 658-665.

Address for correspondence

Gustave Karara, MPH, PhD
Dept of Public Health, Biostatistics and Medical Informatics
Vrije Universiteit Brussel
Laarbeeklaan 103 B-1080 Brussels - Belgium
gustave.karara@vub.ac.be

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190369

Racial Representation Disparity of Population-Level Genomic Sequencing Efforts

Isaac E. Kim, Jr., Indra Neil Sarkar

Center for Biomedical Informatics, Brown University, Providence, RI

Abstract

To develop personalized treatments for diseases, it is essential that they reflect the population of individuals that may be affected by a given disease. Amidst claims that there may be racial disparities in research populations, there have been no direct studies to explore this disparity in disease incidence and research projects that involve genomic sequencing. The precise relationship between underrepresentation of certain races in genomic sequencing studies and health outcomes relative to these races is unknown. Here, we examine the disparities in racial representation of national datasets pertaining to clinical data, mortality rates, and a major initiative involving genomic sequence analysis (The Cancer Genome Atlas [TCGA]). The results suggest that black Americans are underrepresented for most cancers in TCGA compared to clinical and mortality datasets, whereas Asian Americans are overrepresented. These findings accentuate the importance of targeted efforts to recruit representative patient populations into studies involving genomic sequencing.

Keywords:

Precision Medicine; Healthcare Disparities; Cohort Studies

Introduction

Genomic sequencing initiatives such as The Cancer Genome Atlas (TCGA) aim to catalogue cancer-associated genetic mutations towards the overall goal to diagnose, treat, and prevent cancer through better genetic understanding. While different genetic variants may result in similar symptoms, they could lead to diseases that require distinct, "personalized" treatments [1]. Genomic sequencing studies have led to major breakthroughs in the understanding of various types of cancer. In particular, research using TCGA has revealed that the genetic mutations responsible for breast cancer can be categorized into four major subtypes.

As precision medicine initiatives are embarked upon, such as the All of Us Research Program in the United States, it will be essential to be aware of the potential gaps in genetic knowledge. The All of Us Research Program is designed to treat patients based on individual differences in lifestyle, environment, and biology including factors such as race. Studies have shown that genetic makeup across races can impact treatment regimens as well as outcomes. For example, some patients with localized prostate cancer are prescribed active surveillance as opposed to immediate treatment. Other studies have shown, however, that Black American candidates for active surveillance had worse clinicopathological features on final surgical pathology than their white counterparts, suggesting that the criteria for active surveillance should be more rigorous for Black Americans [2].

The advancement of precision medicine requires genetic information on patients of all races [3]. Adequate racial representation in these studies will lead to more effective targeted therapies and at least address the issue of racial disparities in national health measures. Previous studies have shown a disparity between genomic databases and population demographics [4].

This study aimed to quantify the level of representation of racial minorities such as Black Americans and Asian Americans in the genomic sequencing study, TCGA, relative to clinically reported populations in three publicly available epidemiology databases: (1) Health Care Utilization Program [HCUP] from the Agency for Healthcare Research Quality; (2) mortality data from the Centers for Disease Control and Prevention [CDC]; and (3) the National Cancer Institute's Surveillance, Epidemiology, and End Results Program [SEER]). A major goal of this study is to provide data to compel researchers leading cohort-based studies to actively ensure appropriate balance of racial minorities.

This study sought to address the knowledge gap in the representation of racial minorities in genomic studies compared to incidence databases. To quantify and assess the proportional relationships between the racial groups examined in this study, we introduce the Sinuosity Index as a measure to quantify racial discordance.

Methods

Datasets

Four datasets were analyzed and compared: (1) the 2012 National Inpatient Sample (NIS) from the Healthcare Cost and Utilization Project (HCUP), (2) 2012 mortality data from the Centers for Disease Control and Prevention's (CDC) National Vital Statistics Report, (3) genomic sequencing data from TCGA, and (4) 1973-2014 incidence data from the National Cancer Institute's Surveillance, Epidemiology, and End Results Program (SEER). Based on the Clinical Classifications Software (CCS) disease codes that aggregate International Statistical Classification of Diseases and Related Health Problems, patient race information was extracted according to cancer groups from HCUP and SEER.

HCUP included the races White, Black, Hispanic, Asian or Pacific Islander, Native American, Other, and Invalid, while SEER catalogued a patient's race as White, Black, American Indian, as well as a range of ethnicities belonging to the Asian or Pacific Islander demographic, Other, or Unknown. The CDC's mortality data were analyzed from its corresponding website, while TCGA's genomic sequencing data was retrieved from the National Cancer Institute's Genomic Data Commons Data Portal. CDC divided race into White, Black, American Indian or Alaska Native, and Asian or Pacific Islander, whereas TCGA listed the races White, Black or African American, Asian, American Indian or Alaska Native,

Native Hawaiian or Other Pacific Islander, Other, and Not Reported. As a result of the different race variables and data formats across the four databases, the data extraction code required mapping tailored to each database. Tallies were graphed according to overall cancer and cancer, gender, and race for white, black, and Asian Americans using the Plotly.jl Julia package. This study included 16 cancers in total, specifically cancers of the head and neck (CCS code 11), esophagus (CCS code 12), stomach (CCS code 13), colon (CCS code 14), liver and intraheptic bile duct (CCS code 16), pancreas (CCS code 17), bronchus/lung (CCS code 19), skin melanomas (CCS code 22), breast (CCS code 24), uterus (CCS code 25), cervix (CCS code 26), ovary (CCS code 27), prostate (CCS code 29), bladder (CCS code 32), kidney and renal pelvis (CCS code 33), and brain and nervous system (CCS code 35). For certain cancers, data only existed in a single incidence dataset, SEER, leading to the omission of the following codes in the plotted graphs: Other non-epithelial cancer of skin (CCS code 23), Cancer of other female genital organs (CCS code 28), Cancer of other male genital organs (CCS code 31), and Cancer of other urinary organs (CCS Code 34). Analysis of Variance (ANOVA) calculations were done based on the total number of individuals for each race in each dataset using the Tukey honest significant difference test.

Sinuosity Index

Sinuosity index, which is a measure of steepness of a curve commonly used to assess the straightness of geographic features [5] was used as a measure for racial discordance across the examined databases. The relative difference was quantified as the slope between the relative lowest and highest occurrence of a given population group, with higher sinuosity indices suggesting greater variation across the databases.

For each given cancer and race in which at least three datasets were represented, percentage values were sorted in ascending order into an array. The highest value was then assigned an adjusted value of 1, while all other values were divided by the highest value. For the adjusted sinuosity index of each examined cancer and race, in traversing a sorted array of three values, the lowest value was given an x-value of 1 and set to the front index of a new 2D array, the second an x-value of 3 and set to the last, and the third an x-value of 2 and set to the middle index. In traversing a sorted array of four values, the lowest value was given an x-value of 1 and pushed to the front index of a new 2D array, the second an x-value of 4 and set to the last index, the third an x-value of 2 and to the second index, and the fourth an x-value of 3 and set to the third index. The euclidean distance between the first and last points in the sorted 2D array was calculated and designated as the B value. The sum of the euclidean distances between neighboring points (i.e., first and second, second and third, third and fourth, etc.) in the sorted 2D array was calculated and designated as the A value. The sinuosity index was calculated as A/B.

For the adjusted slope angle of each cancer and race, the lowest value was given an x-value of 1, the second lowest an x-value of 2, and so forth. Then, the euclidean distance was calculated between the points with the lowest and highest y-values and designated as the D value. Next, the difference between the points with the lowest and highest x-values was calculated and designated as the C value. The slope angle was

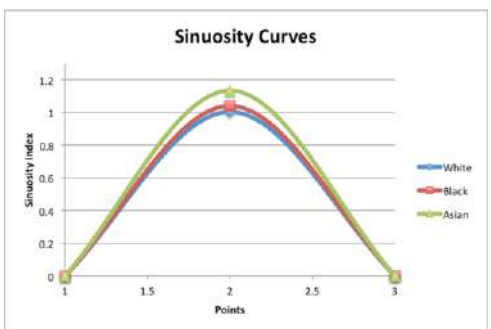


Figure 1 – Sinuosity Curves. Disproportion between sinuosity indices for White, Black, and Asian Americans can be seen as relative differences between the curves.

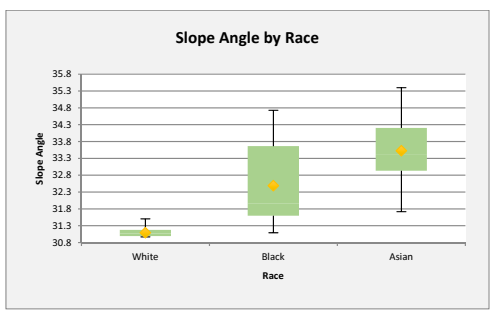


Figure 3 – Distribution of Slope Angle By Race. A distribution of slope angle is shown for each race examined across the datasets: White, Black, and Asian

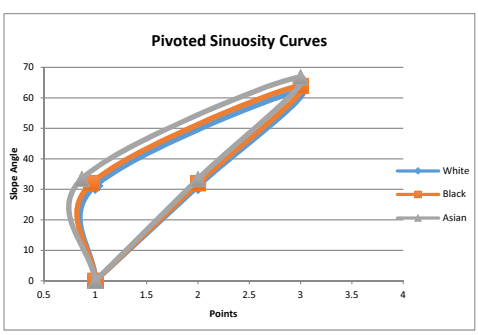


Figure 2 – Pivoted Sinuosity Curves. The relative slope differences between the White, Black, and Asian Americans shows disparity between genomic and epidemiological data sources.

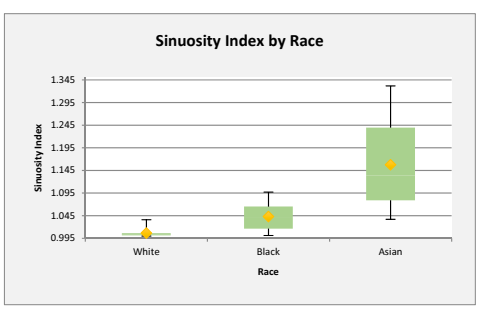


Figure 4 – Distribution of Sinuosity Index By Race. A distribution of sinuosity index is shown for each race examined across the datasets: White, Black, and Asian

calculated using cos(C/D).

The sinuosity curve was plotted for each racial grouping based on the median sinuosity index and then pivoted according to its respective median slope angle. The sinuosity curve was graphed by assigning two points with coordinates (1,0) and (3,0) and a third point with an x-value of 2 and y-value of the median sinuosity index. The sinuosity curves were pivoted so that the line connecting the points (1,0) and (3,0) corresponded to the median slope angle.

Results

For the cancers included in TCGA, Black Americans comprised less of the total sample population in TCGA as compared to HCUP, CDC, and SEER for 15 out of 19 cancers; White Americans and Asian Americans were underrepresented in 11 and 4 out of 19, respectively. Prostate cancer showed the most underrepresentation in TCGA across all examined races, and certain female-dominated cancers such as breast and uterus showed the highest representation of Black Americans in TCGA compared to HCUP, CDC, and SEER. An Analysis of Variance test using Tukey's honest significant difference test for the entire populations in each database revealed statistically significant differences (P=0.0000) between each race group for each dataset, except for Black Americans versus Asian Americans in SEER (P = 0.4001).

For the 16 cancers in which data in TCGA and at least two other datasets were listed, the mean sinuosity indices for White Americans, Black Americans, and Asian Americans respectively were 1.00642±00879 [1.00023-1.03608], 1.04298±02936 [1.00128-1.09715], and 1.15744±09114 [1.03703-1.33156]; the mean slopes were respectively 31.0961±0.1401 [30.9656-31.5034], 32.4931±1.22428

Table 1 – Sinuosity Indices and Slope Angles by Cancer and Race. Percent-adjusted sinuosity indices and percent-adjusted slope angles are shown for each cancer and race.

Cancer (CCS Code)	Race	Percent-Adjusted Sinuosity Index	Percent-Adjusted Slope Angle (Degrees)	Cancer (CCS Code)	Race	Percent-Adjusted Sinuosity Index	Percent-Adjusted Slope Angle (Degrees)
Head and Neck (11)	White	1.00438	31.0737	Breast (24)	White	1.00198	30.9822
Head and Neck (11)	Black	1.01063	31.3418	Breast (24)	Black	1.04249	31.6611
Head and Neck (11)	Asian	1.12017	33.1793	Breast (24)	Asian	1.03703	31.9721
Esophagus (12)	White	1.01167	31.2422	Uterus (25)	White	1.00066	30.9778
Esophagus (12)	Black	1.06297	33.6537	Uterus (25)	Black	1.03717	32.2554
Esophagus (12)	Asian	1.33156	35.3975	Uterus (25)	Asian	1.06960	32.7279
Stomach (13)	White	1.00507	31.1776	Cervix (26)	White	1.00128	30.9941
Stomach (13)	Black	1.07589	34.3908	Cervix (26)	Black	1.06337	31.9875
Stomach (13)	Asian	1.14658	33.5137	Cervix (26)	Asian	1.06368	32.9994
Colon (14)	White	1.00142	30.9756	Ovary (27)	White	1.00344	31.0783
Colon (14)	Black	1.03610	31.5895	Ovary (27)	Black	1.01553	31.7026
Colon (14)	Asian	1.08045	31.7173	Ovary (27)	Asian	1.15474	33.2923
Liver and Intraheptic Bile Duct (16)	White	1.03608	31.5034	Prostate (29)	White	1.01355	31.1755
Liver and Intraheptic Bile Duct (16)	Black	1.04369	33.2769	Prostate (29)	Black	1.08430	33.8490
Liver and Intraheptic Bile Duct (16)	Asian	1.24043	34.5932	Prostate (29)	Asian	1.23926	34.0319
Pancreas (17)	White	1.00770	31.1781	Bladder (32)	White	1.00247	31.0048
Pancreas (17)	Black	1.07147	33.6909	Bladder (32)	Black	1.01656	31.1584
Pancreas (17)	Asian	1.07419	33.1019	Bladder (32)	Asian	1.24861	34.8000
Bronchus; Lung (19)	White	1.00149	31.0110	Kidney and Renal Pelvis (33)	White	1.00571	31.0647
Bronchus; Lung (19)	Black	1.00697	31.0915	Kidney and Renal Pelvis (33)	Black	1.05650	31.9157
Bronchus; Lung (19)	Asian	1.10552	32.3497	Kidney and Renal Pelvis (33)	Asian	1.27815	34.6854
Skin Melanomas (22)	White	1.00023	30.9656	Brain and Nervous System (35)	White	1.00565	31.1334
Skin Melanomas (22)	Black	1.09715	34.7281	Brain and Nervous System (35)	Black	1.02892	31.5980
Skin Melanomas (22)	Asian	1.09512	34.0309	Brain and Nervous System (35)	Asian	1.23394	34.0738

[31.0915-34.7281], 33.5291±1.0569 [31.7173-35.3975]. Figure 1 depicts the sinuosity curves, Figure 2 shows the sinuosity curves pivoted according to their respective slope angles, Figure 3 includes a box-and-whisker plot of the sinuosity indices, and Figure 4 shows a box-and-whisker plot of slope angles. The individual sinuosity indices for each cancer examined are shown in Table 1.

Discussion

The discordance of racial or ethnic groups in genomic sequencing projects potentially exacerbates racial disparities in U.S. health care, especially for complex genetic conditions. As a step towards overcoming these challenges in developing genetically informed healthcare regimens, it is of the utmost importance that the research community actively recruit racially proportional cohorts into national genomic sequencing efforts [3] such as the All of Us Research Program. Efforts such as TCGA require patient consent in the procurement of tissue samples, which implies that non-White Americans tend to withhold their samples from researchers.

The sinuosity index was adapted for this study to understand the relative differences between genomic sequence and epidemiology databases. In geography it is commonly used to study the "straightness" of an earth feature (e.g., a river). In this study, this metric was chosen as a proxy to compare the relative "straightness" of the curves that characterize the difference between the analyzed sources. To our knowledge, this is the first use of the sinuosity index in the context of biomedicine, where higher index values indicate greater dispartity. The differences between sinuosity (which quantifies shape) are more pronounced than slope alone (which measures overall difference) provide a unique persective to compare the general patterns of disparity between the races analyzed in this study.

Previous studies have shown that Black Americans are significantly less likely than their white counterparts to participate in research that used their DNA, share their DNA with a private company, and permit their DNA to be used to generate cell lines for future research [6]. Moreover, as has been noted with clinical trial studies [7], racial minorities are much less likely to want the results of their genetic testing [8] and have diverse views on the utility of genetically targeted treatments [9].

It is important to acknowledge the potential problems associated with using race as a label, especially since racial disparities in genomic sequencing may either be related to or are the direct results of the perception of how healthcare systems and research communities interest. Previous studies connecting race and variable risk for certain diseases have generated considerable controversy, as the correlation between these two factors has long been disputed among members of the scientific community and the general public.

In an effort to direct researchers away from the preconceived notion of race in genetics research, Yudell et al. noted that "the use of biological concepts of race in human genetic research - so disputed and so mired in confusion - is problematic at best and harmful at worst. It is time for biologists to find a better way." [10] In contrast, David Reich suggested that the scientific community legitimize and incorporate these claims into genetic studies while working to provide the same freedoms and opportunities to individuals irrespective of their race [11]. Reich notes that genetic studies have shown distinct differences across populations in traits such as bodily dimensions and susceptibility to diseases. Reich specifically suggests that "it will be impossible - indeed, anti-scientific, foolish, and absurd - to deny the differences [between populations]".

In contemporary medicine, treatment for cancer is usually based on the type, the size, and whether it has metastasized. As our results suggest, however, patients are at higher risk for different cancers based on their race. The results of this study further underscore that the promise of precision medicine must accomodate attributes such as race alongside pre-existing chronic conditions and environmental factors.

This study reinforces the growing wave of support for precision medicine by the government and academic institutions, specifically the United States National Institute for Minority Health and Health Disparities's aim to understand disease mechanisms that lead to differential health outcomes in minorities. This study suggests that the genetic component associated with race that affects the kinds of cancer that patients are at risk for.

For results of initiatives like the All of Us Research Program to have practical meaning to the general population, the scientific community must intently ensure that large-scale genomic sequencing efforts are representative of the range of racial backgrounds. For example, the adequate representation of female Black Americans relative to their male counterparts in TCGA indicates that these recruitment efforts into genomic sequencing studies should especially target male Black Americans. For prostate cancer, the sinuosity indices for White Americans, Black Americans, and Asian Americans were 1.01355, 1.08430, and 1.23926, indicating that Black Americans were more underrepresented in TCGA compared to White Americans (Figure 5). Interestingly, prostate cancer mortality for Black Americans is more than twice the rate observed in White Americans [5]. In contrast, Asian Americans showed overrepresentation in genomic sequencing studies, leading to disproportionately high sinuosity indices across examined cancers. This finding suggests that recruting efforts to target Asian Americans in genomic sequencing studies may not be nearly as vital as those seeking Black Americans.

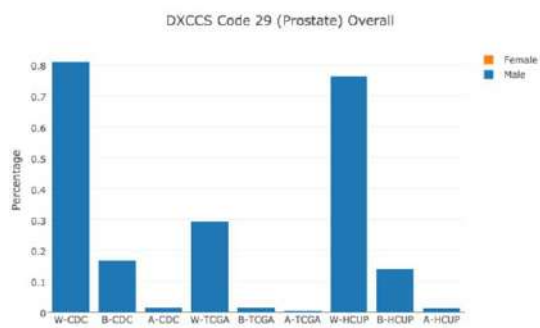


Figure 5 – Percent Representations in CDC Mortality, TCGA, and HCUP By Race for Prostate Cancer. Percent representations of prostate cancer are shown for each race (White [W], Black [B], and Asian [A]) examined across the datasets: CDC, TCGA, and HCUP.

A notable limitation of this study was being unable to examine representation for other races and some cancers due to a lack of sufficient data in at least two of the following datasets: CDC, TCGA, HCUP, and SEER. Such cancers included those of the other GI organs and peritoneum (18) and other urinary

organs (34). Thus, future work should also entail assessing adequate representation of races besides White Americans, Black Americans, and Asian Americans as well as these cancers. Relatedly, there was little to no documentation on how race data was collected in each database. This absence of documentation is relevant in that Mersha et al. found that self-reporting African Americans can have drastically different levels of African or European ancestry. Moreover, genetic analysis of individual ancestry revealed that some self-identified African Americans have significant European ancestry and vice versa [12]. Therefore, that race might be self-reported by patients in these databases would suggest that racial disparities in healthcare are largely due to factors other than genetics. Further work should be done on determining the socioeconomic and genetic basis for why race seems to play an important factor in the onset of different types of cancer. Additionally, data analysis was limited to one genomic database, so future work should incorporate genomic databases other than TCGA.

The challenge in ensuring diversity in large scale initiatives has been acknowledged. For instance, the 1000 Genomes Project analyzed the genomes of 1,092 individuals from 14 populations ranging from people with African ancestry in Southwest United States to Han Chinese in Beijing, China to British from England and Scotland. Given the importance of racial makeup in patient treatment and health outcomes, all genomic sequencing studies should follow suit to initiatives such as the 1000 Genomes Project and contain racial diversity [6]. However, it is important to note that as initiatives like All of Us launch into recruitment efforts, diversity alone is not sufficient. Diversity must be complemented with ensuring that it is with comparable frequencies relative to actual population. Otherwise, the research and clinical community risk the challenge of arriving at putative treatments that are of little utility to significant portions of the population who would benefit the most from personalized medicine approaches.

This study shows that, while race is acknowledged as an important component to consider risk for cancer, cohort studies to date are disproportionate relative to the actual occurrence of cancer by race. Thus, while genomic sequencing studies have led to novel discoveries of disease progression, they may inevitably biased by the races represented in these studies. As our results show, there has been a shortage of minorities in genomic sequencing studies, thereby potentially exacerbating racial disparities in health outcomes. Minimally, from findings of this study, we conclude that researchers must actively recruit Black Americans in genomic sequencing efforts.

The development of personalized treatments for diseases is increasingly plausible due to the increased availability of genomic data. For such efforts to be impactful, it is essential that they reflect the population of individuals that may be affected by a given disease. Despite claims of racial disparities in research populations, there have been no direct studies on such disparities in disease incidence and research projects that involve genomic sequencing. As a result, the relationship between the underrepresentation of certain races in genomic sequencing studies and health outcomes relative to these races is thus unknown.

Conclusions

This study explored the disparities in racial representation of national datasets pertaining to clinical data, mortality rates, and of a major initiative involving genomic sequence analysis (The Cancer Genome Atlas [TCGA]). The results suggest that Black Americans are underrepresented for most cancers in TCGA compared to clinical and mortality datasets, whereas Asian Americans are overrepresented. Additionally, male Black Americans tend to be especially underrepresented in such genomic sequencing studies compared to their female counterparts. These findings highlight the importance of targeted efforts to actively recruit representative patient populations into studies involving genomic sequencing.

Acknowledgements

This study was funded by grant U54GM115467 from the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] S.J. Aronson, H.L. Rehm, Building the foundation for genomics in precision medicine, *Nature* 526 (2015), 336-342.
[2] Y.S. Ha, A. Salmasi, M. Karellas, E.A. Singer, J.H. Kim, M. Han, A.W. Partin, W.J. Kim, D.H. Lee, and I.Y. Kim, Increased incidence of pathologically nonorgan confined prostate cancer in African-American men eligible for active surveillance, *Urology* 81 (2013), 831-835.
[3] L.G. Landry, N. Ali, D.R. Williams, H.L. Rehm, and V.L. Bonham, Lack of diversity in genomic databases is a barrier to translating precision medicine research into practice, *Health Aff* 37 (2018), 780-785.
[4] D.E. Spratt, T. Chan, L. Waldron, C. Speers, F.Y. Feng, O.O. Ogunwobi, and J.R. Osborne, Racial/Ethnic disparities in genomic sequencing, *JAMA Oncol* 2 (2016), 1070-1074.
[5] E.D. Lazarus and J.A. Constantine, Generic theory for channel sinuosity, *Proc. Natl. Acad. Sci. U.S.A.* 110 (2013), 8447-8452.
[6] T. Dye, D. Li, M. Demment, S. Groth, D. Fernandez, A. Dozier, and J. Chang, Sociocultural variation in attitudes toward use of genetic information and participation in genetic research by race in the United States: implications for precision medicine, *J. Am. Med. Inform. Assoc.* 23 (2016), 782-786.
[7] K.L. Pariera, S.T. Murphy, J. Meng, and M.L. McLaughlin, Exploring willingness to participate in clinical trials by ethnicity, *J Racial Ethn Health Disparities* 4 (2017), 763-769.
[8] R.G. Salloum, T.J. George, N. Silver, M.J. Markham, J.M. Hall, Y. Guo, J. Bian, and E.A. Shenkman, Rural-urban and racial-ethnic differences in awareness of direct-to-consumer genetic testing, *BMC Public Health* 18 (2018), 277.
[9] C.H. Halbert, J.A. McDonald, G. Magwood, and M. Jefferson, Beliefs about genetically targeted care in African Americans, *J. Natl. Med. Assoc.* 109 (2017), 98-106.
[10] M. Yudell, D. Roberts, R. DeSalle, and S. Tishkoff, Taking race out of human genetics, *Science* 351 (2016), 564-565.
[11] D. Reich, How genetics is changing our understanding of 'race', *The New York Times* (2018).
[12] T.B. Mersha and T. Abebe, Self-reported race/ethnicity in the age of genomic research: its potential impact on understanding health disparities, *Hum Genomics* 9 (2015): 1-15.

Address for correspondence

Indra Neil Sarkar, PhD, MLIS, FACMI
Center for Biomedical Informatics
Brown University
Box G-R
Providence, RI 02912
neil_sarkar@brown.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190370

Construction of Simulation Platform for Chinese Stroke Economic Burden Based on the National Screening Data

Xuemeng Li[a], Di Bian[b], Mei Li[c], Dongsheng Zhao[a]

[a] *Information Center, Academy of Military Medical Sciences, Beijing, China,*
[b] *School of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an, Shaanxi, China,*
[c] *China Stroke Data Center, Beijing, China*

Abstract

With the characteristics of high incidence, high prevalence and high mortality, stroke has a serious impact on residents' health and imposes a heavy ecomomic burden on China. In order to simulate the economic burden of stroke in the next 20 years, we construct a simulation platform for Chinese stroke economic burden based on the national stroke screening data. We use the Leslie model for population prediction and the equilibrium model to simulate the stroke economic burden in the platform. The platform constructed in this study can dynamically simulate the stroke economic burden during 2020-2040 by computing the incidence, prevelance and mortality rates from the national stroke screening data or as customized by the user. Based on this platform, we can further develop a warning mechanism at the national level, and provide a guide for the planning and allocation of national health resources.

Keywords:

Stroke, Cost of Illness, Computer Simulation

Introduction

In recent years, with the economic development, changes in people's lifestyles and the aging population of China, the incidence of chronic diseases has surpassed that of infectious diseases. With the characteristics of high incidence, high prevalence and high mortality in China, stroke has a serious impact on the health of residents and imposes a heavy burden on society and families. An accurate simulation of stroke economic burden can provide and estimate of stroke burden at the national level, guide the planning and allocation of national health resources—including the required number of beds in hospitals and medical investment, and develop a reasonable intervention strategy for people with high risk of stroke. Researchers have conducted successful studies on disease economic burden. Tomaskova H et al.[1] used system dynamics to predict the number of patients with Alzheimer's disease based on data from the Census Bureau and the U.S. Surgeon General's Smoking Report. Miotto R et al.[2] derived a general-purpose patient representation with unsupervised deep feature learning method to predict disease such as diabetes. Liu J et al.[3] further used the unstructured notes and free-text medical notes to improve the accuracy of the prediction results. Chen M et al.[4] applied machine learning algorithms for prediction of chronic disease based on real-life hospital data collected from central China in 2013-2015. However, these machine learning methods require a long time span cohort data. Keoghbrown M R et al.[5] simulated Alzheimer's disease prevalence, morbidity and mortality for 2011–2050 and analyzed related economic and non-economic impact on Chinese economy. Charlson F J et al.[6] drew upon burden of mental disorders and population data to estimate disability the quality adjusted life years from 2010 to 2050 for the Pacific region. Xu J et al. [7] used regression model and dynamic general disequilibrium model based on data from multiple sources to predict the total annual economic costs of dementia in China in 2020 and 2030. Bommer C et al.[8] predicted the economic burden of diabetes in 180 countries upto 2030, using the epidemiological and demographic data, as well as recent GDP forecasts with three assumed scenarios. Rossi A et al.[9] predicted Parkinson's disease prevalence upto 2040, considering the correlation between Parkinson's disease and smoking. They obtained age and gender.stratified smoking prevalence from the published reports. However, there are very few studies on the simulation platform of the economic burden in China due to stroke. The data used in these simulations are either focused on a certain area [10] or a certain stroke subtype [11], and some studies are based on old data [12]. Most of the research results are presented in a report which is difficult to share and update.

In order to investigate and reduce the incidence, prevalence, recurrence, mortality and disability rate of stroke in China, the Ministry of Health of the People's Republic of China launched a stroke screening and intervention program. The program established stroke centers in the hospitals, which are responsible for screening the risk factors for stroke among residents over 40 years of age in China. Since the launch of the project in 2011, more than 7 million stroke screening cross-sectional data and follow-up cohort data have been accumulated nationwide. Due to the short construction time of the screening cohort, it is not suitable to use regression models, machine learning and other simulation methods that require cohort data over several years. We propose an equilibrium model based on stratified population to simulate the stroke economic burden. We have calculated the age and gender stratified incidence, prevalence, and mortality of stroke based on the national screening big data. Combined with the predicted population structure, we can simulate stroke economic burden and its trends in the next 20 years. At the same time, by adjusting the incidence, prevalence and mortality of stroke, we can simulate the economic burden and provide decision support to develop scientific indicators for stroke prevention and control.

In this paper, we construct the platform for simulating the stroke economic burden, which can dynamically simulate and display the changes in the economic burden of different population groups in China in the next 20 years. The second section describes the main technologies used in the platform,

combined with the usage scenarios, the third section illustrates the results, and the fourth section discusses these results.

Methods

The stroke economic burden is related to the population structure. Based on the sixth national census data in 2010, we predict the population structure of China. And then an appropriate model is selected to simulate the stroke economic burden for the next 20 years.

Population Prediction

The methods for predicting population mainly include the Malthus population model [13], the Logistic model [14], the Leslie model [15], the BP neural network model[16], and the Bayesian model [17]. Economic and health policy in China has impacted the population. The birth and mortality rate have changed greatly in recent decades. The national census is conducted every ten years and as a result, the available national census data for population projections is limited. Among these population prediction models, only the Leslie model does not require long-term cohort data. And we choose the Leslie model to predict the population structure in China. Introduced by Leslie PH. in 1945, the Leslie matrix groups population by age and gender. It overcomes the drawbacks of the Malthus model and the Logistic model, which cannot predict the population structure. Considering the birth rate and mortality of the population, Leslie model can dynamically predict changes in population structure.

Considering the difference in survival rates of different provinces and urban/rural areas, we group the population by the province, age (5-year age as a group), gender, and urban/rural areas for the prediction. The population forecast is based on the 2010 national census data, from which the survival rate of each population group is taken. We assume that the survival rate remains unchanged during the prediction period.

Stroke Economic Burden Simulation

In this paper, equilibrium simulation model is used to predict epidemiological indicators of stroke, including incidence, prevalence, and mortality. The equilibrium model, also known as static model, assumes the epidemiological indicators of stroke in certain population groups do not change over time, and the total changes in the epidemiological indicators are only related to changes of population size and structure. Giving hospitalization rate and average hospitalization expenses as parameters, the changes of stroke economic burden are mainly related to the population structure changes.

Epidemiological Indicators Simulation

Firstly, the population is regrouped by age (10-year age as a group) and sex. We assume that in the next 20 years, China will maintain the current health conditions and the risk factors of stroke will not change, namely, the incidence, prevalence and mortality of stroke will remain at the current level. The default stroke incidence, prevalence, mortality included in the platform are calculated with the national stroke screening cohort data during 2013-2015 and the results from paper [9], which the user can choose as needed. In addition, users can customize the stroke incidence, prevalence, and mortality for each group. Taking the calculation of the incidence cases as an example, the number of incidence cases in each group is the product of the predicted population of that group and the incidence rate. The incidence cases of all groups are increased to obtain the national stroke cases of that year. The calculation method for incidence cases is as follows:

$$\text{cases}_{i,j} = \sum_{j}^{10} pop_{i,j} * inci_{i,j}$$

where i indicates the year, j is the group number, pop is the predicted population, and $inci$ is the incidence of stroke. The incidence of stroke of that year is the incidence cases divided by the total predicted population. The number of stroke patients, the number of deaths, the prevalence, and the mortality rate is calculated in the same way.

Hospitalization Expenses Simulation

The total hospitalization expenses for stroke patients is defined as the product of the number of incidence cases, hospitalization rate, and the average hospitalization expenses (in billion yuan). The hospitalization rate and average hospitalization expenses in a certain year can be customized by users.

Indirect Economic Burden Simulation

The World Bank and the World Health Organization (WHO) proposed the disability adjusted life year (DALY) in 1993 [18] as a comprehensive indicator for assessing indirect economic burden of disease. Widely used in various countries and regions, DALY is used to calculate the premature death caused by various diseases and consequent loss of healthy life. One DALY is defined as a loss of one healthy life year. It consists of two parts: the year of life lost with premature death (YLLs), and the other is the year of lived with disability (YLDs). The calculation formula for DALY is as follows:

$$DALY=YLLs+YLDs$$

The calculation of YLD requires the disability weight (DW) due to stroke. According to GBD, the DW of the first ever stroke is 0.92, and the DW of the surviving case after 28 days is 0.266. The calculation of disability adjusted life years in the platform is based on a computational template provided by the GBD.

GDP Loss Simulation

The GDP loss caused by short-term disability, long-term disability, and premature death is calculated using human resource model. The definition of GDP loss is as follows:

$$GDP\ loss=PPP\ *DALYs*\ pw$$

where PPP indicates GDP per capita (in international dollars, purchasing power parity adjusted), and pw indicates productivity weight, which is related to the productivity of the age of people. The productivity weight of people aged 40-44 years is 0.75, 45-59 years is 0.8, and that of people over 60 years is 0.1. In the platform, users can customize the GDP per capita.

Results

First, we experimented the prediction accuracy of the Leslie model. Using the Chinese fifth census data in 2000, the national population structure in 2010 is predicted. The results are compared with the data of Chinese sixth census in 2010. The error rate of the prediction results is shown in Table 1.

We do not consider the immigration and migration of the population while using the Leislie model, which may bring certain deviation between the prediction results and the actual population data.

Table 1– Prediction error rate (%) of the Leslie model

Age range	Male	Female
40-49	13.84	10.24
50-59	-9.85*	-11.53*
60-69	11.98	10.18
70-79	13.58	13.69
80-	2.70	8.14
Total	6.40	4.77

* Negative sign indicates negative deviation.

At the same time, the increase in the incidence of chronic diseases in the past decade[19-21] may also bring deviations. The population size of people in 40-49 years old is larger than the prediction results of the model. The baby boom in the 1960s may explain this phenomenon. Overall, the population prediction using the Leslie model is basically reliable. Since China has entered an aging society, the actuall population size in 2020-2040 will be theoretially smaller than the prediction results.

The platform is implemented using Visual Studio 2015 based on the MVC framework. The initial interface of the platform is the user input interface, which can be used to set the parameters, as shown in Figure 1.

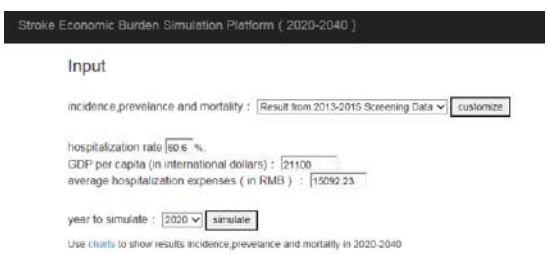


Figure 1– The initial interface of the platform

In the platform, users can choose the incidence, prevalence, and mortality calculated based on the national stroke screening data of a certain year, or use the results of other research papers. The following result acts as an example of simulation using the incidence, prevalence, and mortality calculated using national screening data in 2013-2015. The simulation results of average incidence, prevalence and mortality of stroke in China from 2020 to 2040 is shown in Figure 2-4.

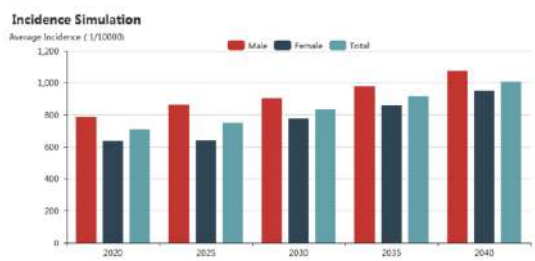


Figure 2– Simulated stroke incidence of China from 2020 to 2040

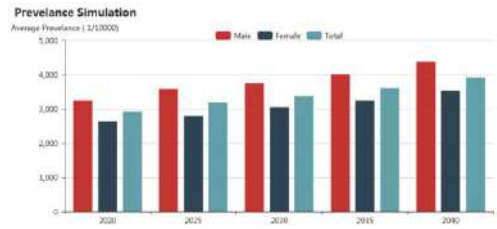


Figure 3– Simulated stroke prevelance of China from 2020 to 2040

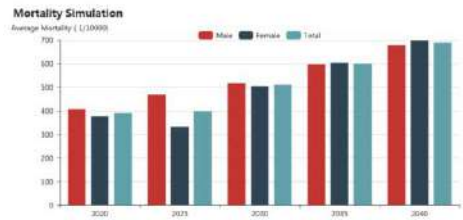


Figure 4– Simulated stroke mortality of China from 2020 to 2040

When simulating the indirect economic burden, we choose the average hospitalization cost and hospitalization rate of stroke patients calculated from the national stroke screening data in 2013-2015. The GDP per capita in 2020 is set to 21,100 international dollars, in reference to the prediction results in paper [22]. The simulation results of the indirect economic burden of stroke in 2020 are shown in Figure 5.

Figure 5– Stroke economic burden simulation results

Stroke incidence, prevalence, and mortality are associated with age [17]. The stroke prevalence is highest in people aged 70-79, and the incidence and mortality are highest in people 80 years and over. The results show that as the aging population in China increases, the stroke incidence, prevalence and mortality will increase anually. In order to further control the economic burden of stroke in China, the risk factors of stroke should be controlled through health education for residents to reduce the incidence and recurrence rate of stroke. And the medical care condition should be improved to reduce the mortality rate of stroke.

Discussion

While using the platform, users can customize the stroke incidence, prevalence and mortality. By changing the parameters, the economic burden of stroke under different conditions can be simulated. The simulation results of

economic burden are more timely and intuitive than the traditional reports. With the advancement of stroke screening and intervention project and increased control of stroke risk factors, the incidence, prevalence and mortality of stroke will change. These values can be entered in this platform to dynamically simulate the future changes in stroke economic burden. Also by setting the warning value of stroke economic burden, the threshold of incidence, prevalence and mortality in each population group can be explored, which can be used as a guide for the national health policy.

In the future, we can predict population structure of various regions based on their population structure and simulate their respective stroke economic burden in order to make different health policies for different regions.

Conclusions

The stroke economic burden simulation platform based on the national screening data constructed in this paper can share the simulation results in time. According to the user's custom input, changes in the stroke economic burden caused by changes in the incidence, prevalence, and mortality will be simulated. Policy makers can further explore the threshold of stroke incidence, prevalence and mortality by calculating the "blowout" or "inflection point" of the economic burden. Rational use of this stroke economic burden simulation platform can help create a guide for planning and national health resource allocation. This would include the required number of beds in hospitals and medical investment needed. Based on the simulation results an early warning mechanism can be developed at the national level and the timing of warning for the onset of stroke can be further explored.

Acknowledgements

Funding: Key Project of Natural Science Foundation of China (71532014).

References

[1] H. Tomaskova, J. Kuhnova, R. Cimler, O. Dolezal, and K. Kuca, Prediction of population with Alzheimer's disease in the European Union using a system dynamics model, *Neuropsychiatric Disease & Treatment* **12** (2016), 1589-1598.

[2] R. Miotto, L. Li, B.A. Kidd, and J.T. Dudley, Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records, *Scientific Reports* **6** (2016), 26094.

[3] J. Liu, Z. Zhang, and N. Razavian, Deep EHR: Chronic Disease Prediction Using Medical Notes, (2018).

[4] M. Chen, Y. Hao, H. Kai, L. Wang, and L. Wang, Disease Prediction by Machine Learning over Big Data from Healthcare Communities, *IEEE Access* **5** (2017), 8869-8879.

[5] M.R. Keogh-Brown, H.T. Jensen, H.M. Arrighi, and R.D. Smith, The Impact of Alzheimer's Disease on the Chinese Economy, *EBioMedicine* **4** (2016), 184-190.

[6] F.J. Charlson, S. Diminic, and H.A. Whiteford, The Rising Tide of Mental Disorders in the Pacific Region: Forecasts of Disease Burden and Service Requirements from 2010 to 2050: The Rising Tide of Mental Disorders in the Pacific Region, *Asia & the Pacific Policy Studies* **2** (2015), 280-292.

[7] J. Xu, J. Wang, A. Wimo, L. Fratiglioni, and C. Qiu, The economic burden of dementia in China, 1990-2030: implications for health policy, *Bulletin of the World Health Organization* **95** (2017), 18.

[8] C. Bommer, V. Sagalova, E. Heesemann, J. Manne-Goehler, R. Atun, T. Bärnighausen, J. Davies, and S. Vollmer, Global Economic Burden of Diabetes in Adults: Projections From 2015 to 2030, *Diabetes Care* **41** (2018), dc171962.

[9] A. Rossi, K. Berger, H. Chen, D. Leslie, R.B. Mailman, and X. Huang, Projection of the prevalence of Parkinson's disease in the coming decades: Revisited, *Mov Disord* **33** (2017).

[10] M.L. Zuo, S. Liu, K.H. Chan, K.K. Lau, B.H. Chong, K.F. Lam, Y.H. Chan, Y.F. Lau, G.Y.H. Lip, and C.P. Lau, The CHADS 2 and CHA 2 DS 2 –VASc scores predict new occurrence of atrial fibrillation and ischemic stroke, *Journal of Interventional Cardiac Electrophysiology An International Journal of Arrhythmias & Pacing* **37** (2013), 47-54.

[11] Moran A , Zhao D , Gu D. The future impact of population growth and aging on coronary heart disease in China: projections from the Coronary Heart Disease Policy Model-China, *BMC Public Health* **8** (2008).

[12] T.R. Malthus, *An essay on the principle of population, or A view of its past and present effects on human happiness, with an inquiry into our prospects respecting the future removal or mitigation of the evils which it occasions (rev. ed.)*, Cambridge University Press, 1992.

[13] P. Meyer, *Growth and Diffusion Phenomena: Mathematical Frameworks and Applications*, World Scientific, 1998.

[14] D.P. Smith and N. Keyfitz, On the Use of Matrices in Certain Population Mathematics, *Biometrika* **33** (1945), 183-212.

[15] C. Yin and L. Chen, Study and Application of Population Forecast Model Based on BP Neural Network, *Population Journal* (2005).

[16] A. Wiśniowski, P.W.F. Smith, J. Bijak, J. Raymer, and J.J. Forster, Bayesian Population Forecasting: Extending the Lee-Carter Method, *Demography* **52** (2015), 1035-1059.

[17] W. Wang, B. Jiang, H. Sun, X. Ru, D. Sun, L. Wang, L. Wang, Y. Jiang, Y. Li, and Y. Wang, Prevalence, Incidence, and Mortality of Stroke in China: Results from a Nationwide Population-Based Survey of 480 687 Adults, *Circulation* **135** (2017), 759.

[18] S. Anand and K. Hanson, Disability-adjusted life years: a critical review, *Journal of Health Economics* **16** (1997), 685-702.

[19] H.U. Sheng Shou, L.Z. Kong, R.L. Gao, M.L. Zhu, W. Wang, Y.J. Wang, W.U. Zhao Su, W.W. Chen, and M.B. Liu, Outline of the Report on Cardiovascular Disease in China, 2010, *Biomedical and Environmental Sciences* **25** (2012), 251-256.

[20] Y. Song, X. Liu, X. Zhu, B. Zhao, B. Hu, S. Xia, L. Chen, M. Yu, T. Yang, and J. Zhao, Increasing trend of diabetes combined with hypertension or hypercholesterolemia: NHANES data analysis 1999–2012, *Scientific Reports* **6** (2016), 36093.

[21] G. Yun, C. Gang, T. Haoming, L. Lixiang, L. Juming, W. Jianping, J. Weiping, J. Linong, X. Jianzhong, and Z. Zhiguang, Prevalence of hypertension in china: a cross-sectional study, *Plos One* **8** (2013), e65938.

[22] G.C. Bernal, The long view: how will the global economic order change by 2050? (2017).

Address for correspondence

Dongsheng Zhao, (email) dszhao@bmi.ac.cn.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190371

Leveraging Patient Safety Research: Efforts Made Fifteen Years Since *To Err Is Human*

Chen Liang[a], Qi Miao[b], Hong Kang[c], Amy Vogelsmeier[d,e], Tina Hilmas[e], Jing Wang[f], Yang Gong[c]

[a] *Arnold School of Public Health, University of South Carolina, Columbia, South Carolina, USA*
[b] *School of Public Health, University of Texas Health Science Center at Houston, Houston, Texas, USA*
[c] *School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, Texas, USA*
[d] *Sinclair School of Nursing, University of Missouri, Columbia, Missouri, USA*
[e] *Center for Patient Safety, Columbia, Missouri, USA*
[f] *School of Nursing, University of Texas Health Science Center at San Antonio, San Antonio, Texas, USA*

Abstract

Despite U.S. federal agencies increasing their investment since 1999's release of To Err Is Human, recent reports suggest there is a lack of measurable outcomes in patient safety research. The present study sought to explore the associations between federal incentives of patient safety research and the outcomes from 1995 to 2014, in which the two historical events – the release of To Err Is Human and the American Recovery and Reinvestment Act – were considered in the analysis. We employed Poisson distribution models to provide a longitudinal picture of (1) how the federal incentives drove sponsored research projects; (2) how hot research topics changed over time. Our findings suggested a positive outcome in patient safety research. We also found trending health information technology (HIT) related topics including "natural language processing", "user-computer interface", and "clinical decision support systems" that are prevalent approaches to patient safety research.

Keywords:
Patient Safety, Medical Errors, Quality of Health Care

Introduction

Patient safety is the first priority for healthcare quality. Its significance had been lost for a long period of time in the public and healthcare professions' attention until the release of the Institute of Medicine's (IOM) report *To Err Is Human* on November 29, 1999 [1]. The report disclosed 44,000 to 98,000 patient deaths of medical errors every year. Nevertheless, errors that resulted in patient harm are preventable, to a great extent [2].

Nationwide collaboration is key to patient safety. Immediately after the IOM's report, several collaborations among hospitals, research institutes, and healthcare organizations were called on. Among these collaborations, patient safety research was an integral component [3]. Meanwhile, financial incentives were offered to motivate the systematic study of better understanding patient safety challenges and solutions. At the direction of Congress and the Agency for Healthcare Research and Quality (AHRQ), strategies were made during the first decade since the IOM's report to (1) develop a solid evidence base, (2) develop useful strategies and tools, and 3) implement the strategies and tools. These efforts made notable progress and were highlighted in a number of AHRQ reports and scholarly publications [4,5].

In 2009, the American Recovery and Reinvestment Act (ARRA) became another notable factor that triggered an influx of research awards to spur patient safety research. Because of this economic stimulus, the National Institute of Health (NIH) received a total of $10.4 billion in funds for patient safety research [6].

Despite the federal government and institutions making more investments than ever, there are recent reports on worsening patient safety [2,7]. Most of the critiques come directly from the statistics of patient harm in the hospitals. As of 2013, the annual patient deaths of preventable medical errors increased to 210,000 ~ 440,000 [2]. Patient safety still poses serious challenges nationwide.

Measuring the efforts and progress of patient safety plays a vital role in the framework of patient safety research [8]. A wide variety of measurements have been used for such a purpose, including clinical study, literature review, and survey on patient safety culture [5]. However, there is a missing perspective on how the activities and outcomes of patient safety research changed under the influence of federal incentives. The awareness of such an association is necessary to sustain the improvement of patient safety. Recommended by IOM, patient safety research was included in the national collaboration on enhancing the knowledge about safety and developing tools for reducing medical errors. Among many other efforts, research is one of the most supported efforts by the federal government and stakeholders.

We investigated the activities of patient safety research in the U.S. from 1995 to 2014, which consisted of fifteen years since *To Err Is Human* and five years before it as a baseline. Data of federally sponsored research projects and publications on health sciences were used as the gauge for the national priority, research activities, and outcomes of patient safety [3]. We included two remarkable events, *To Err Is Human* and ARRA, in the data analysis to assess their influence in a longitudinal view. Specifically, we developed Poisson regression models to provide longitudinal pictures of (1) how federal incentives drove sponsored research projects, reflected in the quantity and cost of grants and the quantity of publications; (2) how hot research topics changed over time.

Methods

Data sources

To track federal funding in patient safety research, we used the Research Portfolio Online Reporting Tools Expenditures and Results (RePORTER) provided by NIH to download research projects funded by U.S. national institutions. These projects were archived in the ExPORTER format. The data underlying

RePORTER contains detailed information for each project. See Table 1 for a dictionary of the data we used. We retrieved the projects from the fiscal year (Oct. 1 – Sep. 30) 1995 to 2014. Note that data for the cost of projects are not available before 1999. To track the publications affiliated with the sponsored projects, we performed a literature search in the Medline database from 1995 to 2014 via PubMed. To extract hot topics of sponsored patient safety publications, we incorporated MeSH major topics and PMID in the search of RePORTER and Medline.

Table 1 – RePORTER data dictionary

RePORTER Database	RePORTER Entity	Definition
Project Data	Application_ID	A unique identifier of the funded project.
	Activity	A 3-character code identifying the funded project.
	FY	The fiscal year appropriation from which project funds were obligated.
	Project_Terms	Prior to the 2008 fiscal year, terms are assigned by NIH CRISP indexers. From the 2008 fiscal year, terms are mined from the project's title, abstract, and specific aims using an automated text mining tool.
	Project_Title	Title of the funded project.
	Total_Cost	Total project funding from NIH for a given fiscal year.
Project Abstracts	Application_ID	A unique identifier of the funded project.
	Abstract_Text	An abstract of the research being performed in the project.
Publication Link Tables	PMID	A PubMed unique identifier.
	Core Project_Num	An identifier of the research project either cited in the publications' acknowledgements section or reported to have provided support in the NIH Public Access manuscript submission system.

Data analysis

Identification of thesaurus terms related to patient safety

Five domain experts developed a list of thesaurus terms related to patient safety. Thereafter, these terms were used to retrieve sponsored projects on patient safety. The decision-making process took three criteria into account: (1) terms that had been documented to determine the patient safety literature were used as a pool of candidates [3], (2) the NIH Computer Retrieval of Information on Scientific Projects (CRISP) thesaurus (1972 – 1995) were used as a source of candidate terms, and (3) MeSH terms were used to substitute candidate terms identified by criteria (1) and (2) when available. We used CRISP because it was a long-lasting database of NIH funded projects before the implementation of RePORTER in 2009.

Finally, we compiled a list of terms: "medical errors" (MeSH ID: D019300), "medical mistakes" (MeSH entry term of "medical errors"), "surgical errors" (MeSH entry term of "medical errors"), "medication errors" (MeSH ID: D008508), "drug use errors" (MeSH entry term of "medication errors"), "patient safety" (MeSH ID: D061214), "patient harm" (MeSH ID: D064406), "iatrogenic disease " (MeSH ID: D007049), and "diagnostic errors" (MeSH ID: D003951).

Trend analysis: Change of grants in patient safety over time

The trend analysis was intended to disclose the change in the number and cost of patient safety grants over time. Within a timeframe from 1995 to 2014, we assumed two historical events may remarkably impact the trend. The first event denotes the release of *To Err Is Human* in 1999. The second event denotes ARRA enacted in 2009.

Sponsored patient safety projects were retrieved by searching the project data from ExPORTER using the thesaurus terms. Table 1 shows the definitions of the fields in ExPORTER we used. We searched in the selected fields of project data consisting of Project_Title, Project_Terms, and Abstract_Text. A project is labeled as a patient safety project if the text in any of these fields contains one or more terms.

We constructed Poisson regression models to evaluate the trends of grant number and cost in patient safety on a longitudinal scale. In the Poisson regression models, two interventions corresponded to the two events [9]. These models accounted for the instant effect and the lasting effect of the interventions. The model for the number and cost of patient safety grants is:

$$\log(Patient\ safety\ grant\ number\ (or\ cost)) = \beta 1 + \beta 2 * time\ + \beta 3 * intervention1 + \beta 4 * time\ after\ intervention1 + \beta 5 * intervention2 + \beta 6 * time\ after\ intervention2 + e$$

where $time$ denotes the year of a grant; $intervention1$ and $intervention2$ are dummy variables denoting the events of *To Err Is Human* and ARRA, respectively; $time\ after\ intervention1$ denotes the year of a grant since *To Err Is Human*; $time\ after\ intervention2$ denotes the year of a grant since ARRA; $\beta 1 \sim 6$ and e are coefficients.

Trend analysis: Change of publications in patient safety over time

Similarly, we accommodated two historical events in the analysis. In the Medline database via PubMed, we retrieved all articles from January 1, 1995 to December 31, 2014, by specifying the following query:

> "Medical errors"[mh:noexp] OR "Medication errors"[mh:noexp] OR "Patient safety"[mh:noexp] OR "Patient harm"[mh:noexp] AND (("1995/01/01"[PDAT] : "2014/12/31"[PDAT]) AND "humans"[MeSH Terms] AND English[lang])

To model the trend of quantity of patient safety publication, we included the number of Medline documented publications per quarter, obtained from Medline Trend [10], and the number of Medline documented patient safety publications per quarter for analysis. The model is:

$$\log(Patient\ safety\ publication\ number) = \beta 1 + \beta 2 * time\ (quarter) + \beta 3 * intervention1 + \beta 4 * time\ after\ intervention1 + \beta 5 * intervention2 + \beta 6 * time\ after\ intervention2 + e$$

Lag effect

To evaluate the possible lag effect of the number of patient safety publications in acknowledgement of the number of

patient safety grants, we assumed the lag time could be 0 ~ 4 year(s). For example, a 2009 grant would start to show its impact on the number of publications in 2012, which indicates 3 years of lag time. We compared the Bayesian Information Criterion (BIC) by applying these candidate lag times [11]. BIC is a measure of the relative quality of models given the existing data. Higher BIC value refers to a better fit of the model.

Hot topic tracking: Change of MeSH major topics in patient safety publications over time

We identified hot topics of the publications sponsored by funded patient safety projects. The MeSH major topics (i.e., MeSH major headings and MeSH major subheadings) were used to track the hot topics since they were designed to identify focusing research subjects, methodologies, fields of interest, etc. To extract MeSH major topics from the patient safety publications, we followed these procedures: (1) we identified PMIDs of publications associated with the patient safety projects in the RePORTER, (2) we retrieved bibliographical information of these publications by searching the PMIDs in Medline, and (3) we extracted MeSH major topics from the bibliographical data.

Results

Identification of patient safety grants and publications

We identified 3,358 (0.28%) patient safety-related projects in a total number of 1,208,188 documented projects from the RePORTER between the fiscal year 1995 and 2014. We identified 21,441 (0.16%) patient safety publications from 13,278,113 publications documented in Medline between January 1, 1995 and December 31, 2014.

Trend of patient safety grants

Table 2 shows the estimated coefficients of the Poisson regression model are statistically significant, which indicates a good fit of the model to the actual number of patient safety grants.

Table 2 – Model parameters for the number of safety grants

Effect	λ	95% CI	z
Time	0.040	[0.026, 0.054]	5.63**
Intervention1	1.280	[1.129, 1.432]	16.58**
Intervention2	0.158	[0.049, 0.262]	2.83*
Time after intervention2	-0.054	[-0.081, -0.027]	-3.93**

* $p < 0.01$; ** $p < 0.001$

Table 3 shows the estimated coefficients of the Poisson regression model are statistically significant, which indicates a good fit of the model to the actual cost of patient safety grants.

Table 3 – Model parameters for the cost of safety grants

Effect	λ	95% CI	z
Time	0.1870	[0.1870, 0.1871]	13000.00**
Intervention2	0.2434	[0.2432, 0.2436]	2847.02**
Time after intervention2	-0.2173	[-0.2174, -0.2173]	-9578.03**

* $p < 0.01$; ** $p < 0.001$

Figure 1 shows the number of grants reveals an overall ascending trend from 1995 to 2014. We observed two rapid momentums in 1999 and 2009, respectively. A similar pattern was observed on the project cost.

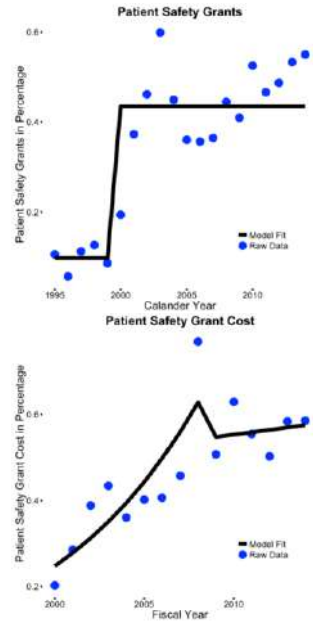


Figure 1 – Trends of patient safety project number (top) and grant cost (bottom)

Furthermore, we categorized the number and cost of awarded projects, respectively, by NIH research award activity codes. Figure 2 shows there is a significant effect on the number of the awarded projects ($F(17,342) = 43.57$, $p < .05$). NIH Research Projects increased by 10.1% ($p < .001$), leading to a significant contribution to the number and cost of the awarded projects.

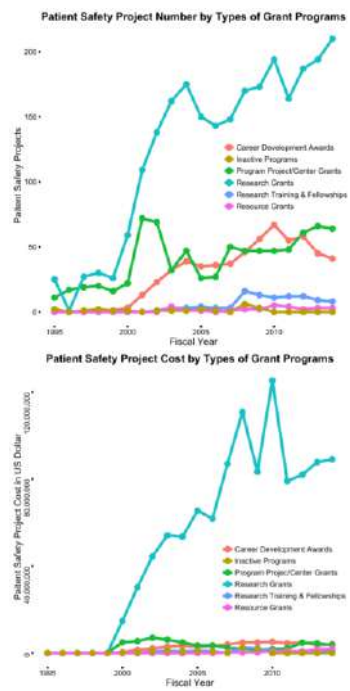


Figure 2 – Trends of patient safety project number (top) and cost (bottom) by NIH activity codes

We presented 24 major MeSH topics that have the highest frequency of occurrences out of 6,472 in total from 1995 to

2014 (see Figure 3). MeSH terms that were used for identifying publications were excluded since they naturally occur in every publication. Due to the limited space, we only demonstrated the hot research topics in 1999 and 2009, respectively, in Table 4.

Figure 3 – Frequencies of hot research topics in the publications of sponsored patient safety projects over time

Table 4 – Examples of hot research topics and frequencies

1999 Hot topics	*f.*	2009 Hot topics	*f.*
Risk Management/ *methods	19	*Safety Management	101
Malpractice/ *legislation & jurisprudence	13	Safety Management/ *organization & administration	93
*Drug-Related Side Effects and Adverse Reactions	10	Safety Management/ *methods	41
Medication Systems, Hospital/*standards	8	*Quality Assurance, Health Care	31
*Quality of Health Care	7	*Attitude of Health Personnel	29
*Clinical Pharmacy Information Systems	6	*Clinical Competence	28
*Truth Disclosure	6	Drug Prescriptions/ *standards	27
*Patient Care Team	6	*Medication Systems, Hospital	24
Nursing Staff, Hospital/*legislation & jurisprudence	6	*Quality of Health Care	24
*Attitude of Health Personnel	6	*Safety	24

The two historical events are associated with the trends of research topics. Notably, we observed an immediate increase of a research topic, "hospital mortality trends", following the release of *To Err Is Human*, suggesting that hospital mortality was one of the earliest research foci since 1999. Immediately following are "adverse drug reaction reporting systems", "needs assessment", "safety management", "quality improvement", "natural language processing", "electronic health records", "medical order entry systems", and "user computer interfaces". Among these topics, health information technology (HIT) related techniques received increasing attention. This finding is in line with the complex origins of HIT-related errors [12,13]. "Natural language processing", "electronic health records", "user computer interfaces", and "medical order entry systems" have maintained a relatively high frequency through the release of ARRA. Research topics such as "quality of healthcare", "computer simulation", and "artificial intelligence" did not show a clear upward trend until the release of ARRA.

Trend of patient safety publications

The estimated coefficients are statistically significant in the model for the number of publications, indicating a fit for the actual data (Table 5). Patient safety publications revealed a similar pattern as compared to the number and cost of patient safety grants, where two rapid momentums were observed in 1999 and 2009 on an ascending trend (Figure 4). The baseline increase was 8.6% by quarter. After the release of the IOM report, patient safety publications instantly increased 11.8% by quarter, and 2.6% thereafter. After the release of ARRA, publications instantly increased 74.5% by quarter, and 0.8% thereafter.

Table 5 – Model parameters for patient safety publications

Effect	λ	95% CI	*z*
Time	1.086	[1.079, 1.093]	24.68**
Intervention1	1.118	[1.035, 1.207]	2.83*
Time after intervention1	0.940	[0.934, 0.947]	-17.99**
Intervention2	1.745	[1.648, 1.848]	18.99**
Time after intervention2	0.982	[0.972, 0.992]	-3.56**

* $p < 0.01$; ** $p < 0.001$

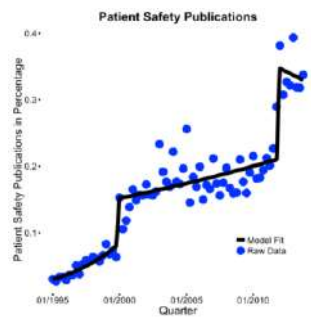


Figure 4 – The trend of patient safety publications

We calculated the BIC for 0 - 4 years of lag time. BIC peaked (BIC = -11.4) when the lag time was 3 years, indicating the trend of patient safety publications has a three-year lag of time as compared to the trend of patient safety grants (Table 6).

Table 6 – Bayesian Information Criterion values

Lag time (year)	0	1	2	3	4	5
BIC	-15.18	-30.37	-13.64	-11.40	-17.26	-13.39

Discussion

Building on a data-driven approach, this study underscores how patient safety research has evolved since 1999. The efforts during the fifteen years since the release of *To Err Is Human* have shown an ascending trend in patient safety research reflected in the volume and cost of federally sponsored projects and scholarly publications. This finding can serve as one of the measurable outcomes that gauges the nationwide endeavor on patient safety research over two decades. It also implies a positive change of priority and cultural attitude toward patient safety. Our findings also confirmed the prominent influence of *To Err Is Human* and ARRA on patient safety research.

The change in topics of patient safety research has provided important insights. One observation is research interests have advanced from the general assessment of preventable medical errors to specific domains (e.g., HIT) that may contribute to errors in the US hospitals. The "hospital mortality trends" was the first topic that caught the researchers' attention since 1999 because hospital mortality rate has been a common measure of healthcare quality. Soon after, researchers started to focus on adverse drug reaction and clinical or patient safety educational needs assessment. Notably, IOM has now identified adverse drug events and medication errors as a national priority in the U.S. [14].

The other observation pertains to the emergence of novel methods. There has been an increasing number of patient safety studies on exploring new knowledge from electronic health records (EHR) [15]. Researchers have identified Human-computer interaction (HCI) as a crucial contributing factor to data quality and data entry during event reporting [16,17]. Natural language processing (NLP) has also emerged in many studies because patient safety event reports consist of considerable free text data [18]. In our view, health informatics is a promising future direction along with new data challenges. Recent studies have shown the application of advanced data analytics, e.g., machine learning, to address the ever-increasing volume of patient safety event reports [19,20].

Our findings need to be discussed in light of limitations. First, our data sources provided only a few facets of patient safety research. *To Err Is Huamn* has led to actions from many governmental agencies and professional groups that were not included in this paper. In 2000, the U.S. congressional hearings directed $50 million funds to establish a patient safety center at AHRQ. Additionally, our study paradigm may be used to gauge efforts on patient safety research in Europe, Asia, etc. when data is available. Secondly, our findings are not intended to explain the latest estimate of increasing patient deaths, between 210,000 and 400,000 [2]. The proliferation of patient safety research and measurable clinical outcomes are separate processes. Nonetheless, we believe the ultimate goal for patient safety research is to reduce preventable deaths. Future studies are needed to investigate the translational outcomes of patient safety research that make measurable changes to patient harm.

Conclusions

We have identified an ascending trend of activities of patient safety research during the 15 years since *To Err Is Human*. We also identified trending research topics in which shifts of research foci, challenges, and future directions were discussed.

Acknowledgements

The study is supported by grants from AHRQ (1R01HS022895), the UTHealth Innovation for Cancer Prevention Research Training Program Post-Doctoral Fellowship (RP160015), and the University of Texas System Grants Program (156374). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality.

References

[1] L.T. Kohn, J.M. Corrigan, and M.S. Donaldson, To err is human: Building a safer health system, National Academies Press, Washington, DC, 2000.

[2] J.T. James, A new, evidence-based estimate of patient harms associated with hospital care, *J Patient Saf* **9** (2013) 122–128.

[3] H.T. Stelfox, S. Palmisani, C. Scurlock, E.J. Orav, and D.W. Bates, The "To err is human" report and the patient safety literature, *Qual Saf Heal Care* **15** (2006) 174–178.

[4] L.L. Leape and D.M. Berwick, Five years after to err is human: What have we learned?, *JAMA* **293** (2005) 2384–2390.

[5] P.J. Pronovost, M.R. Miller, and R.M. Wachter, Tracking progress in patient safety: An elusive target, *JAMA* **296** (2006) 696–699.

[6] H.R. Clinton and B. Obama, Making patient safety the centerpiece of medical liability reform, *N Engl J Med* **354** (2006) 2205–2208.

[7] J.R. Edwards, K.D. Peterson, Y. Mu, et al., National healthcare safety network (NHSN) report: Data summary for 2006 through 2008, *Am J Infect Control* **37** (2009) 783–805.

[8] P.J. Pronovost, C.A. Goeschel, J.A. Marsteller, et al., Framework for patient safety research and improvement, *Circulation* **119** (2009) 330–337.

[9] D. Posada and T.R. Buckley, Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and bayesian approaches over likelihood ratio tests, *Syst Biol* **53** (2004) 793–808.

[10] A.D. Corlan, Medline trend: Automated yearly statistics of PubMed results for any query, (2004). http://Dan. Corlan.net/Cgi-Bin/Medline-Trend (accessed in 2016).

[11] M.A. Makary and M. Daniel, Medical error - The third leading cause of death in the US, *BMJ* **353** (2016) i2139.

[12] D.F. Sittig and H. Singh, Defining health information technology-related errors: New developments since to err is human, *Arch Intern Med* **171** (2011) 1281–1284.

[13] H. Kang, J. Wang, B. Yao, S. Zhou, and Y. Gong, Toward safer health care: A review strategy of FDA medical device adverse event database to identify and categorize health information technology related events, *JAMIA Open* (2018).

[14] P. Aspden, J.A. Wolcott, J.L. Bootman, et al., Preventing medication errors, National Academies Press, Washington, DC, 2007.

[15] D.F. Sittig and H. Singh, Electronic health records and national patient-safety goals, *N Engl J Med* **367** (2012) 1854.

[16] Y. Gong, H.Y. Song, X. Wu, and L. Hua, Identifying barriers and benefits of patient safety event reporting toward user-centered design, *Saf Heal* **1** (2015) 7.

[17] Y. Gong, Data consistency in a voluntary medical incident reporting system, *J Med Syst* **35** (2011) 609–615.

[18] C. Liang, Patient safety narratives as a key to promoting quality measurement, *J Heal Med Informat* **8** (2017) 2.

[19] C. Liang and Y. Gong, Automated classification of multi-labeled patient safety reports: A shift from quantity to quality measure., *Stud Health Technol Inform* **245** (2017) 1070–1074.

[20] C. Liang and Y. Gong, Predicting harm scores from patient safety event reports, *Stud Health Technol Inform* **245** (2017) 1075–1079.

Address for Correspondence

Yang Gong, MD, PhD

7000 Fannin Street, Suite 600, Houston, TX 77030, USA

Yang.Gong@uth.tmc.edu, +1 713 500 3547, +1 713 500 3907 (fax)

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190372

An Integrative Biomedical Informatics Approach to Elucidate the Similarities Between Pre-Eclampsia and Hypertension

Guillermo Lopez-Campos[a], Emma Bonner[a], Lana McClements[a,b]

[a] *Centre for Experimental Medicine, Queen's University of Belfast, Belfast, Northern Ireland, United Kingdom*
[b] *School of Life Sciences, University of Technology Sydney, PO Box 123 Broadway, New South Wales, Australia*

Abstract

Pre-eclampsia is a pregnancy condition affecting 5-10% of pregnancies, and it is the leading cause of death in pregnancy associated with increased risk of cardiovascular disease later in life. Despite research, the pathogenesis of pre-eclampsia is still poorly understood. In this paper, we investigate the overlapping pathogenic mechanisms between pre-eclampsia and adult hypertension using an integrative biomedical informatics strategy that combined text mining techniques to identify genes and proteins, with geneset analyses, generating knowledge on the pathways and mechanisms involved in these conditions. We identified several overlapping pathogenic pathways/systems including metabolic pathways, developmental biology pathways, immune system, haemostasis, tyrosine kinase pathways, extracellular matrix and oxidative stress pathways. This bioinformatics approach could be applied for investigating mechanistic pathways of other disorders.

Keywords:

Pre-Eclampsia, Computational Biology, Data Mining

Introduction

Pre-eclampsia is a condition affecting 5-10% of pregnancies. It is clinically characterised by the new onset of hypertension and proteinuria or other organ damage after 20 weeks' gestation [1]. There are different types of pre-eclampsia diagnosed according to the time of onset in pregnancy, i.e. early (before 34 weeks' gestation) and late (after 34 weeks' gestation) or according to the severity of the symptoms [2]. Pre-eclampsia is a leading cause of mortality and morbidity, claiming approximately 80,000 maternal and 500,000 foetal deaths every year [3]. Long-term complications associated with pre-eclampsia include type 2 diabetes mellitus and cardiovascular disease (CVD) for both mothers and adult offspring, later in life [4,5].

During the early stages of pregnancy, a group of differentiated cells forming blastocysts, become implanted in the lining of the uterus. The outer layer of the blastocyst consists of foetal trophoblast cells which invade the spiral uterine arteries after implantation [6]. The invasion of trophoblast cells into the lining of the uterus results in remodelling of spiral uterine arteries to allow an unlimited supply of oxygen and nutrients to the foetus. These changes involve the replacement of maternal endothelial cells with invasive trophoblasts, which develop to form a large section of the placenta, leading to irreversible dilation of blood vessels and loss of elastic tissue, therefore preventing restriction to blood flow and vasomotor control which regulate blood pressure [7]. Impaired trophoblast invasion has been implicated in the pathogenesis of pre-eclampsia. This results in the absence of uteroplacental remodelling of spiral arteries leading to placental ischaemia and causing insufficient perfusion and delivery of the oxygen and nutrients to the foetus which can affect the foetal growth [8]. In an attempt to overcome the lack of blood and nutrient supply to the foetus, maternal blood pressure becomes gradually elevated. The restricted blood supply to the placenta causes the release of chemical mediators and hormones into the maternal circulation, altering endothelial cell activity and resulting in endothelial cell dysfunction. This suppresses the release of vasodilatory factors such as nitric oxide (NO) and increases the production of vasoconstrictors while inhibiting anticoagulant synthesis and increasing procoagulant production [9]. Dysbalance of vasodilatory and vasoconstrictive factors increases the risk of blood clotting and can lead to an increase in blood pressure. Endothelial dysfunction has been implicated in the pathogenesis of pre-eclampsia [10,11], however endothelial dysfunction following pregnancy complicated by severe pre-eclampsia appears to also persist 10-20 years after pregnancy [12].

As mentioned above, women who have pre-ecplampsia and their children born from such pregnancies have increased risk of CVD later in life, particularly if pre-eclampsia was severe (blood pressure ≥160/110 mmHg, thrombocytopenia, impaired liver function, progressive renal insufficiency, pulmonary oedema, and cerebral or visual disturbances). Based on a meta-analysis of prospective and retrospective cohort studies including 198,252 women with pre-eclampsia, the relative risks of developing hypertension, ischaemic heart disease and stroke were 3.70 after 14.1 years, 2.16 after 11.7 years, 1.81 after 10.4 years, respectively [13]. Furthermore, children born to mothers who suffered pre-eclampsia in pregnancy were also at increased risk of hypertension and stroke in later life [14].

Pre-eclampsia and cardiovascular disease are affected by similar risk factors such as hypertension, obesity, insulin resistance, diabetes, a family history of CVD, and genotypes related to CVD, suggesting that similar pathogenic mechanisms could be present in both diseases [12,15]. Biomarker discovery and characterisation is an important area of research both in pre-eclampsia and CVD that led to the identification of a number of important markers. However, despite research in this field, there is still a lack of knowledge in relation to the mechanisms underpinning the pathogenesis of pre-eclampsia and future increased risk of CVD.

Scientific literature presents the knowledge generated in research and practice. It is a continuously growing wealth of data that is enriched on a daily base with new publications. These data represent an opportunity for the development of research strategies aiming to generate new biomedical knoweldge that is hidden in this vast amount of information, or the generation of new research hypotheses. Therefore, it has

attracted the interest of biomedical informatics, especially the development of methods and tools designed to mine it. An important element of these methodologies has been focused on the identification of diseases, genes or proteins cited in scientific texts and the identification of interaction networks. Biomedical relation extraction is an approach that has shown precedence in similar studies providing cross-sectional and domain-specific views of biomedical research literature [16–18].

In this study, we investigated an *"in silico"* approach, which uses the information embedded in the literature to integrate relevant biomarkers identified in pre-eclampsia and hypertension. Furthermore, we sought to contextualise these in terms of significant biological pathways which are common denominators in both conditions. The approach presented here provides a mechanistic interpretation of the commonalities between pre-eclampsia and hypertension, but it could be applied to study pathogenesis of any other disorder.

Methods

The methodology applied in this work consisted of three sequential stages including: i) document search and retrieval, ii) document annotation and gene/protein extraction and iii) mechanistic analysis. These steps are further detailed below.

The first step in our analyses included the definition of the relevant Pubmed queries, which were used to identify the relevant elements in the literature associated with the two conditions of interest for this study:

- Pre-eclampsia query was built as "("pre-eclampsia"[MeSH Terms] OR "pre-eclampsia"[All Fields] OR "preeclampsia"[All Fields]) AND ("biomarkers"[MeSH Terms] OR "biomarkers"[All Fields] OR "biomarker"[All Fields])". This query aimed to retrieve a broad range of elements that could be associated with this condition.
- Hypertension query was built as "(Hypertension[MeSH] NOT ("pre-eclampsia"[MeSH Terms] OR "pre-eclampsia"[All Fields] OR "preeclampsia"[All Fields])) AND Biomarker". This query focused on the retrieval of documents annotated as hypertension as a MeSH term while excluding pre-eclampsia.

We used the R package *RISMed (2.1.7)* designed to retrieve and download contents from the NCBI databases into R. For this purpose, both of these queries were then passed as arguments to the function "EUtilsSummary" to identify the relevant documents in Pubmed that were subsequently downloaded using the "EUtilsGet" function. This resulted in two R data frames containing information associated with these documents and in particular the relevant Pubmed IDs that were used as input in the second stage for annotation of the literature.

The second stage of the analysis consisted of the annotation of the biological terms of relevance (genes/proteins); for this purpose, rather than developing a new methodology, we relied on the annotations provided by Pubtator [19]. Pubtator annotates Pubmed documents using different "named-entity-recognition" algorithms around four bioconcepts Gene; Chemical; Disease; Species. Although Pubtator was initially developed as a web service at the NCBI ftp site we used the downloadable version of the results from its annotations.

The third and the final stage consisted of the mechanistic analysis of the identified biomarkers. In this analyses we selected two knowledge bases to analyse the enrichment in certain biological processes or functions: DAVID [20], broad and containing multiple annotations such as gene ontology terms, Uniprot keywords or KEGG terms for multiple organisms; and Reactome [21], a highly curated database of biological reactions in humans. Reactome identifies the enrichment in its contents using an overrepresentation analysis based in the hypergeometric test and using a FDR (Benjamini-Hochberg) *p*-value correction to adjust the significance for multiple comparisons. DAVID uses EASE, a modified version of the hypergeometric test and allows the selection of multiple *p*-value correction methods. In both cases the input provided was a list of NCBI Gene IDs that were translated into their own internal IDs to match their contents.

Results

An overall overview of the methodology developed and the results obtained are presented in Figure 1.

The literature search allowed us to identify and retrieve almost 9000 documents in total, 3167 for preeclampsia and 5675 for hypertension. As a quality control we compared the PMID between these two datasets to ensure that there was no overlap

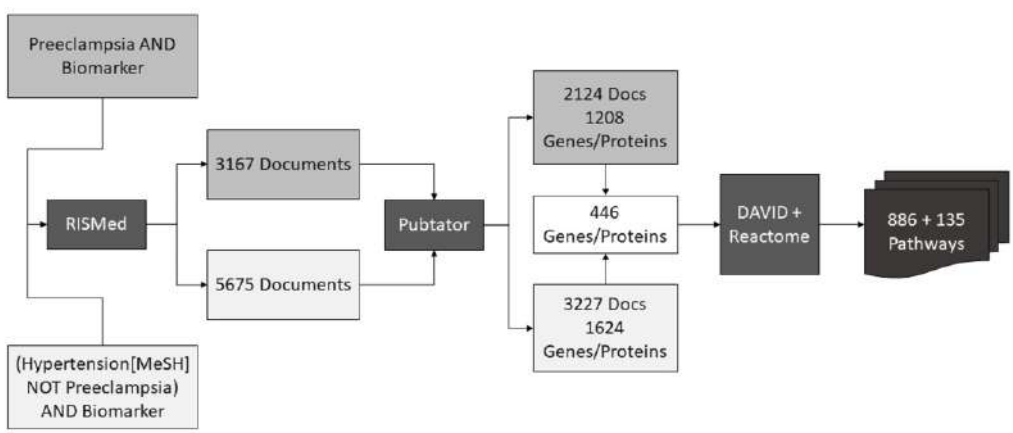


Figure 1 – Overview of the methodology and the results from the the query terms to the final and total number (sum of Reactome and DAVID results) potentially shared pathways and genesets between preeclampsia and hypertension

and that the results would not be misled by publications present in both datasets.

The annotation of these two sets of Pubmed documents allowed us to identify approximately 5,000 different documents that were annotated with at least one gene or protein providing 2,124 in the pre-eclampsia dataset and 3,227 in the hypertension dataset. These documents contained 2,386 genes/proteins of which 446 were present in both datasets (19% of the overall gene list and 37% of the genes identified in the pre-eclampsia dataset; Figure 2).

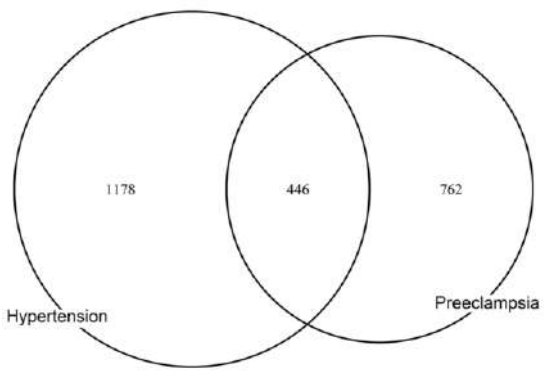


Figure 2 – Number of genes/proteins identified for each dataset and the set of 446 shared genes proteins that were used in the geneset analyses

Once we identified the relevant gene lists the next step was to perform the geneset analyses of the identified genes. As the significance is determined by the number of elements in the genelist we initially run three different analyses using Reactome with the following three subsets: 1- All genes (1624) identified for hypertension (HT.a); 2- All genes (1208) identified for preeclampsia (PE.a); 3- the 446 genes present in both hypertension and preeclampsia (Both). The significance threshold for enriched pathways was set as an adjusted *p*-value < 0.05. The number of significant pathways identified in each of these analyses are presented in the Table 1 and the shared pathways across these comparisons are visualised in Figure 3.

Table 1 – Number of Reactome pathways significantly enriched for each of the three genesets analysed.

Gene Subset (Number of Genes/proteins) (code)	Number of Pathways	Unique Pathways
All genes in Hypertension (1624) (HT.a)	20	7
All genes in Preeclampsia (1208) (PE.a)	107	21
Genes present in both Preeclampsia and Hypertension (446) (Both)	129	42

The set containing the genes/proteins present in both conditions ("Both" set) was the most enriched one with 129 significant pathways (Figure 4) whereas the hypertension set, which was the largest, was the least enriched with 20 different reactome pathways identified. An overlap analysis of these results allowed us to identfy 12 pathways in common for the three different genesets (Table 2).

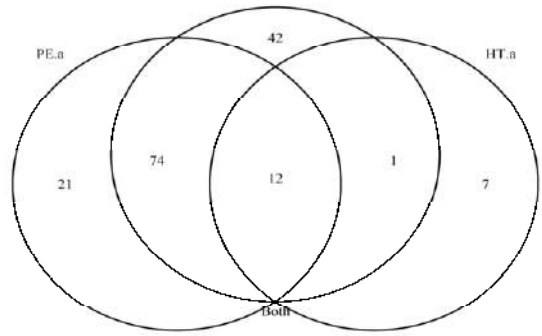


Figure 3 – Pathway enrichment for the three analysed sets, hypertension (HT.a), pre-eclampsia (PE.a) and the intersection between hypertension and pre-eclampsia (both)

Table 2 – List of the 12 statistically significant (adjusted p-value <0.05) pathways identified in Reactome for the three genesets analysed.

Pathway Name
Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)
Platelet degranulation
Response to elevated platelet cytosolic Ca2+
Post-translational protein phosphorylation
Platelet activation, signaling and aggregation
Integrin cell surface interactions
Chemokine receptors bind chemokines
Peptide ligand-binding receptors
Interleukin-10 signaling
Syndecan interactions
Formation of Fibrin Clot (Clotting Cascade)
PI5P, PP2A and IER3 Regulate PI3K/AKT Signaling

As different resources contain different information and annotations for the different genes we submitted the "both" (shared) dataset to DAVID for a complementary analysis. In this case we limit the query to the 446 shared elements between both pathologies. And set a significance threshold of benjamini adjusted *p*-value < 0.05. This resulted in the detection of significant enrichment in 886 different genesets made of pathways, gene ontology terms (Table 3) and other annotations.

Table 3 – List of the top 12 statistically significant (adjusted p-value <0.05) Gene Ontology Biological Process terms identifed in DAVID for the intersected gene list

	Pathway Name		Pathway Name
1	Inflammatory response	7	Response to drug
2	Response to hypoxia	8	Platelet degranulation
3	Leukocyte migration	9	Aging
4	Positive regulation of nitric oxide biosynthetic process	10	Positive regulation of smooth muscle cell proliferation
5	Positive regulation of gene expression	11	Positive regulation of angiogenesis
6	Angiogenesis	12	Regulation of blood pressure

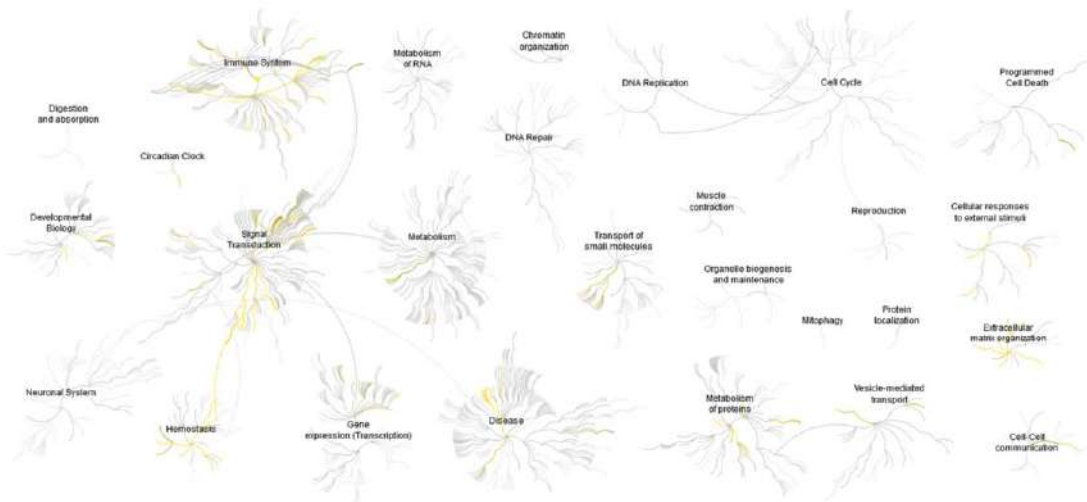


Figure 4 – Results from the Reactome analysis using the shared elements between pre-eclampsia and hypertension. The statistically significant elements are highlighted in yellow in the image.

Discussion

Using an integrated strategy we have been able to retrieve biomarkers (gene/proteins) that are shared between pre-eclampsia and hypertension. Our strategy allowed us to also identify genes from other organisms commonly used to model human diseases such as rats and mice. Although it would have been possible to map those genes to their human orthologues we decided not to do this therefore keeping our focus just on human biomarkers.

The identified genes were transformed into gene lists that were analysed using two different resources. As we mentioned before we were focused on human disease or models and, for this reason, we chose to run initially a more restrictive search using Reactome, as this is a highly curated database used to generate results from high-quality and restricted sources facilitating their biological interpretation. Overrepresentation analyses, similar to what we carried out here, are dependent on the size of the query lists and the size of the background available in the resource they are compared with. Therefore, finding differences in the composition and number of significant pathways identified between the lists containing the whole set of biomarkers for each condition and their intersection, was not completely unexpected. However, it was surprising that the smallest list containing the conserved biomarkers was the one showing the highest enrichment. From a biological perspective, this list, derived from the intersection of biomarkers identified in both pathologies, is expected to contain genes/proteins positively associated with the diseases. These genes would therefore be more biologically meaningful and would better capture the common underlying biological processes in both conditions.

The identified pathways for the shared set of genes are related to metabolism, developmental biology, immune system, haemostasis, tyrosine kinase, extracellular matrix and oxidative stress signalling. All these aspects provide coherent and meaningful biological information in the context of these pathologies. Interestingly, we have identified using DAVID analyses, the GO terms associated with “Regulation of blood pressure” showing up as statistically significant and highly ranked. This is an expected result as both conditions show changes in blood pressure, which validates our methodology as both conditions share a high blood pressure phenotype. In relation to the other enriched pathways identified, for example, it is well recognised that inflammation (associated with the immune system) plays a significant role in hypertension [22,23]. Furthermore, pre-eclampsia is a pregnancy-related condition therefore developmental biology plays a key role [6, 7].

Limitations and future work

The limitations of this study include some aspects of methodology which assume a perfect gene/protein identification in the annotation process not taking into the account the effects of potential false positive and negative biomarker hits. The second limitation is associated with the assumption of treating equally all the terms identified and not considering potential negations contained in the abstracts. An example of this is that we are including a gene in our analysis that could appear in the text as “gene XXX does not have an effect in pre-eclampsia.” However we tried to minimise this by basing our approach on relevant pathways and disease mechanisms.

As part of the future work, we are planning to address specifically the effects and quantity of false positives and negatives potentially present in the current version and we will try to develop new approaches that can help us identify and successfully manage negations found in literature used as input.

In addition to this, as part of the future work, we plan to develop an R package that will seamlessly integrate the methodology and processes that have been described in this work.

Finally, we plan to biologically validate some of the findings derived from this work.

Conclusions

There is a wealth of data and inforamtion available in public repositories that could be mined and analysed to generate new knowledge and research hypotheses. In this work, we combined different biomedical informatics tools, to retrieve

and identify biomarkers in hypertension and pre-eclampsia and then analysed these to identifiy common underlying molecular mechanisms linking these two conditions.

This data-driven approach enabled us to identify the specific aspects and reactions associated with metabolic pathways, developmental biology, immune system, haemostasis, tyrosine kinase pathways, extracellular matrix and oxidative stress pathways as the most prominently involved in the pathogenesis of these two important conditions.

Our bioinformatics approach described in this paper is therefore applicable to any other similar diseases for the purpose of identifying overlapping or individual pathogenic mechanisms.

References

[1] American College of Obstetricians and Gynecologists' Task Force on Hypertension. Hypertension in pregnancy. Report of the American College of Obstetricians and Gynecologists' Task Force on Hypertension in Pregnancy, *Obstet Gynecol* **122** (2013), 1122-1131.

[2] J.M. Roberts, G. Pearson, J. Cutler, M. Lindheimer, and N.W.G.o.R.o.H.D. Pregnancy, Summary of the NHLBI Working Group on Research on Hypertension During Pregnancy, *Hypertension* **41** (2003), 437-445.

[3] K.S. Khan, D. Wojdyla, L. Say, A.M. Gulmezoglu, and P.F. Van Look, WHO analysis of causes of maternal death: a systematic review, *Lancet* **367** (2006), 1066-1074.

[4] G.T. Manten, M.J. Sikkema, H.A. Voorbij, G.H. Visser, H.W. Bruinse, and A. Franx, Risk factors for cardiovascular disease in women with a history of pregnancy complicated by preeclampsia or intrauterine growth restriction, *Hypertens Pregnancy* **26** (2007), 39-50.

[5] T.L. Weissgerber and L.M. Mudd, Preeclampsia and diabetes, *Curr Diab Rep* **15** (2015), 9.

[6] R. McNally, A. Alqudah, D. Obradovic, and L. McClements, Elucidating the Pathogenesis of Pre-eclampsia Using In Vitro Models of Spiral Uterine Artery Remodelling, *Curr Hypertens Rep* **19** (2017), 93.

[7] J. Kieckbusch, L.M. Gaynor, and F. Colucci, Assessment of Maternal Vascular Remodeling During Pregnancy in the Mouse Uterus, *J Vis Exp* (2015), e53534.

[8] P. Kaufmann, S. Black, and B. Huppertz, Endovascular trophoblast invasion: implications for the pathogenesis of intrauterine growth retardation and preeclampsia, *Biol Reprod* **69** (2003), 1-7.

[9] T. Chaiworapongsa, P. Chaemsaithong, L. Yeo, and R. Romero, Pre-eclampsia part 1: current understanding of its pathophysiology, *Nat Rev Nephrol* **10** (2014), 466-480.

[10] T. Clausen, S. Djurovic, J.E. Reseland, K. Berg, C.A. Drevon, and T. Henriksen, Altered plasma concentrations of leptin, transforming growth factor-beta(1) and plasminogen activator inhibitor type 2 at 18 weeks of gestation in women destined to develop pre-eclampsia. Circulating markers of disturbed placentation?, *Placenta* **23** (2002), 380-385.

[11] P.K. Agatisa, R.B. Ness, J.M. Roberts, J.P. Costantino, L.H. Kuller, and M.K. McLaughlin, Impairment of endothelial function in women with a history of preeclampsia: an indicator of cardiovascular risk, *Am J Physiol Heart Circ Physiol* **286** (2004), H1389-1393.

[12] G. Valdes, Preeclampsia and cardiovascular disease: interconnected paths that enable detection of the subclinical stages of obstetric and cardiovascular diseases, *Integr Blood Press Control* **10** (2017), 17-23.

[13] L. Bellamy, J.P. Casas, A.D. Hingorani, and D.J. Williams, Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis, *BMJ* **335** (2007), 974.

[14] B. Arabin and A.A. Baschat, Pregnancy: An Underutilized Window of Opportunity to Improve Long-term Maternal and Infant Health-An Appeal for Continuous Family Care and Interdisciplinary Communication, *Front Pediatr* **5** (2017), 69.

[15] L.J. Alma, A. Bokslag, A. Maas, A. Franx, W.J. Paulus, and C.J.M. de Groot, Shared biomarkers between female diastolic heart failure and pre-eclampsia: a systematic review and meta-analysis, *ESC Heart Fail* **4** (2017), 88-98.

[16] E.K. Mallory, C. Zhang, C. Re, and R.B. Altman, Large-scale extraction of gene interactions from full-text literature using DeepDive, *Bioinformatics* **32** (2016), 106-113.

[17] I. Segura-Bedmar, P. Martinez, and C. de Pablo-Sanchez, A linguistic rule-based approach to extract drug-drug interactions from pharmacological documents, *BMC Bioinformatics* **12** Suppl 2 (2011), S1.

[18] B. Percha and R.B. Altman, A global network of biomedical relationships derived from text, *Bioinformatics* **34** (2018), 2614-2624.

[19] C.H. Wei, H.Y. Kao, and Z. Lu, PubTator: a web-based text mining tool for assisting biocuration, *Nucleic Acids Res* **41** (2013), W518-522.

[20] D.W. Huang, B.T. Sherman, Q. Tan, J. Kir, D. Liu, D. Bryant, Y. Guo, R. Stephens, M.W. Baseler, H.C. Lane, and R.A. Lempicki, DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists, *Nucleic Acids Res* **35** (2007), W169-175.

[21] A. Fabregat, S. Jupe, L. Matthews, K. Sidiropoulos, M. Gillespie, P. Garapati, R. Haw, B. Jassal, F. Korninger, B. May, M. Milacic, C.D. Roca, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusai, G. Viteri, J. Weiser, G. Wu, L. Stein, H. Hermjakob, and P. D'Eustachio, The Reactome Pathway Knowledgebase, *Nucleic Acids Res* **46** (2018), D649-D655.

[22] R. Satou, H. Penrose, and L.G. Navar, Inflammation as a Regulator of the Renin-Angiotensin System and Blood Pressure, *Curr Hypertens Rep* **20** (2018), 100.

[23] R Carnagarin, V. Matthews, M.T.K. Zaldivia, K. Peter, and M.P. Schlaich, The bidirectional interaction between the sympathetic nervous system and immune mechanisms in the pathogenesis of hypertension, *Br J Pharmacol* (2018).

Address for correspondence

Dr Guillermo Lopez Campos
Email: G.LopezCampos@qub.ac.uk

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190373

Patient-Reported Outcomes of Utilising Person-Generated Health Data in Simulated Rehabilitation Technology: Perceptions of Stroke Survivors

Gerardo Luis Dimaguila[a], Kathleen Gray[a], Mark Merolli[b]

[a] Health and Biomedical Informatics Centre, University of Melbourne, Melbourne, Victoria, Australia,
[b] School of Health Sciences, Swinburne University of Technology, Melbourne, Victoria, Australia

Abstract

Patient-reported outcome measures (PROMs) contribute to improving the evidence base in many areas of clinical care. An area where PROMs can help build the evidence base is in person-generated health data (PGHD) that is available to people who engage with many health technologies. An important use case for PGHD outcomes is in simulated poststroke rehabilitation that use body-tracking technologies. This study gathered stroke survivor perceptions—through 2 focus groups and 5 interviews—on PGHD effects from a leading poststroke simulated rehabilitation technology. Deductive thematic analysis was performed. The findings show that PGHD outcomes of utilising PGHD can be measured. Moreover, the findings provided a deeper understanding of PGHD effects and broadened the scope through the perceptions of stroke survivors themselves. This work will further refine and validate the development of a PROM for utilising PGHD for poststroke simulated rehabilitation systems, and has wider relevance and application.

Keywords:

Patient reported outcomes, stroke, telerehabilitation.

Introduction

Patient-reported outcomes are independently self-reported status of a patient's health condition or experience [1-3]. Instruments that measure patient-reported outcomes are patient-reported outcome measures (PROMs). PROMs allow patients to contribute to more precise evaluation of the effects of a variety of health interventions, and their contribution to improving the evidence base in many areas of clinical care is well recognised [4]. The important role PROMs have in assessing and improving health care services are marked by key national projects that sought to develop suites of PROMs for different health domains or conditions. In the US, the National Institutes of Health sought to produce a suite of PROMs for various chronic conditions [5]; in Europe, the Netherlands is leading the push to include outcomes data in clinical registries; and closer to home, the Australian Commission on Safety and Quality in Health Care commissioned scoping studies, towards supporting the use of PROMs to improve quality of health care at a national level [6].

One area where PROMs could contribute to building the evidence base is in understanding person-generated health data (PGHD). PGHD are available to people who engage first-hand with a range of ubiquitous, user-friendly health information technologies such as mobile apps, wearable sensors and interactive computer games [7]. Various kinds of positive effects have been reported on patients who used PGHD in controlled settings [7], e.g., encouraging positive health behaviour change [8] and increasing patients' interest in their care processes [9-11]. However, there is currently no systematic way for patients to measure and report health effects that they experience from utilising their PGHD – whether positive, negative or nil. This hinders building the evidence base about the value of PGHD in conjunction with clinical care. Currently, PGHD outcomes are not integrated into existing clinical workflows [7]. If PGHD outcomes can be systematically measured and clinically integrated, similar to patient-reported outcomes of health conditions [5], PGHD may lead to more efficient and beneficial home-based health care services [12-14].

An important use case for measuring PGHD outcomes is in home-based poststroke rehabilitation, that uses body-tracking technologies [15]. Stroke is a pertinent area for applying rehabilitation systems due to its high global burden, and the complexity of care involved [16; 17]. Rehabilitation can be costly and availability may be scarce for some patients [18], hence the need for convenient and more practical rehabilitation choices for patients. Simulated rehabilitation systems offer a potentially effective option, with some benefits for stroke therapy previously documented [19-22]. These systems offer simulated activities of daily living [20], and – relevant to PGHD – may allow for a semicontrolled, consistent method of recording and assessing patients' therapeutic progress data [20; 21].

Implementing PGHD tools could further optimise these systems [15], by making performance data more accessible to patients at home – and their effects documented and available to therapists. More specifically, a PROM of utilising PGHD (PROM-PGHD) for simulated poststroke rehabiltiation technologies could measure self-reported outcomes of patients' access and use of health data they generate. Similar to how PROMs provide a more precise evaluation of the effects of health interventions [4], a PROM-PGHD could more precisely evaluate the effects of PGHD from poststroke simulated rehabilitation systems, and lead to a better understanding of their impact on the health status of patients.

This study presents the perceptions of stroke survivors, on the effects of PGHD from a poststroke simulated rehabilitation technology in order to refine a PROM-PGHD that has wider relevance and application.

Methods

Focus groups and semi-structured interviews were conducted with stroke survivors who have used a leading poststroke simulated rehabilitation technology [15]. Focus groups and interviews are a key step in a larger study of developing a PROM-PGHD, following the Qualitative Item Review (QIR) process of patient-reported outcome item development [5]. The entire process is described elsewhere.

Ethics Approval

This study was approved by the Deakin University Human Research Ethics Committee (Deakin HREC, ID 2017-087).

Participants

Stroke survivors with experience of utilising PGHD were recruited for convenience from participants in a randomised controlled trial (RCT) called STRIVE, which sought to determine the efficacy of a selected poststroke simulated rehabilitation technology, Jintronix [23]. These participants used Jintronix which utilises Kinect, a leading body-tracking technology [15, 23], and generates PGHD, see example in [15]. Participants were from the Melbourne metropolitan Area, Victoria, and Launceston, Tasmania (Australia). After the RCT lasting 8 weeks, participants in the intervention arm were invited to participate in focus groups of 3-6 people (or semi-structured interviews for participants who could not meet at the same day and time as other participants). There were 3 males and 7 females. All interviewees (5) were females.

Procedure

Participants were asked two categories of semi-structured questions: the first for the purposes of the STRIVE study and the second for the purposes of this study.

The first category for the STRIVE study gauged perceptions for the whole trial in terms of system usability, patient engagement, and intervention impact. Sample questions participants were asked included what parts of Jintronix were easy or difficult to use; whether use of Jintronix resulted in any behavioural changes to daily tasks; and whether Jintronix helped them in understanding and achieving their rehabilitation goals. Participant responses to the questions in this category, if they were also relevant to the second category of questions below, were included in our analysis.

The second category of questions was to conceptualise patient-reported outcomes of PGHD and ask participants to comment on any outcomes their PGHD may have had. Questions included what emotions or reactions they may have from seeing their PGHD; whether they felt more or less engaged after seeing their PGHD; and if any of their rehabilitation goals changed. Participants were also asked to comment on sample patient-reported outcome items. These questions and sample patient-reported outcome items were developed from steps 1-3 of QIR [5], where different effects of PGHD were identified:

1. Influence health-related behavioural changes in patients [8].
2. Change feelings about health status [10].
3. Influence interest in their care processes [9-11].
4. Facilitate personal care goals [10; 24].
5. Influence relationship with care providers [10; 11; 25; 26].

These effects were also used to guide the deductive thematic analysis conducted, as described in the next sub-heading. Steps 1-3 of QIR involved a literature review to scan established PROMs within the areas of interest; and selecting, categorising, and revising relevant PROM items to form initial themes that would guide the discussion with the target cohort.

All focus groups and interviews were held in meeting rooms at Deakin University at days and times that suited the participants. Participants who could meet at similar days and times were organised into focus groups, everyone else was organised into semi-structured interviews. In total there were 2 focus groups with 3 and 2 participants each. In each group, there was one participant who could not make it on the day. Additionally, there were 5 semi-structured interviews. All focus groups and interviews were audio-recorded.

Data Analysis

Nearly 10 hours of focus groups and interviews were transcribed verbatim; the former by GLD and the latter by a research assistant of the STRIVE project. GLD checked all transcriptions for accuracy, and de-identified them. The transcriptions and data management (including coding) were done using NVivo 11 (QSR International Pty Ltd, Melbourne, Australia).

Deductive thematic analysis was initially conducted independently by all authors on a few transcripts. Once common understanding was achieved for how themes are applied, GLD proceeded to analyse the rest of the transcripts; using a coding journal to clarify contentious quotes with the co-authors. Themes were derived deductively from the identified PGHD effects from the literature. The different themes are presented in the results below as subheadings.

Results

The results below show perceptions of stroke survivors on the effects of PGHD from a poststroke simulated rehabilitation technology.

Health-Related Behaviours

While stroke survivors mentioned some improvement with their ability to do some daily tasks as a result of their use of Jintronix, few of them mentioned any behavioural changes as a result of their PGHD. One of them did mention adjusting movement based on the scores they received at the end of an activity: “I automatically adjust what I was doing, if I couldn't burst every balloon I, I had to adjust.. To work out what I was doing wrong” (FG1_STC2-3). Seeing an improvement in performance data encouraged a stroke survivor to do more of the activities: “if I can see the improvement I'm making then it would…encourage me to maybe have more of those sessions” (INT1_STC3). However one stressed that positive health data was not enough, and would not have done the exercises “that long if there hadn't been somebody driving me” (INT1_STC1).

Feelings About Health Status

Most stroke survivors felt either “disappointed” (FG1_STC2-3) or pleased with themselves depending on their PGHD. Seeing their performance data go down, or “wanting to achieve and not achieving, or you not doing it the way you want” (INT2_STC1) resulted in feelings of being “disheartened” (INT2_STC1), with one going as far as describing the experience as “offensive” (INT1_STC1). “Levels of frustration” were also experienced when their data fluctuated: “sometimes you went backwards, sometimes you went forwards” (INT2_STC1). It could also be

discouraging, making one think that they may have reached their limit: “I would think well surely this is the benchmark, and so I probably thought, would have thought nah, I’m probably not gonna to [sic] get any better” (INT1_STC1), with the same survivor cautioning that “it would depress me”. Positive feelings such as “happy” and “satisfying” (INT2_STC3) abound where their PGHD were also positive. One described the experience as feeling “good about yourself” when the data was “improving” (FG1_STC2-1). Positive PGHD was also “reassuring” (INT2_STC3) and “exciting” (INT1_STC1). It could make them feel “hopeful” and “confident” because “it was encouraging,” (INT2_STC3) “saying okay you can perform” (INT3_STC1). A survivor humorously shared that it gave the assurance that “you’re not totally dead” (INT3_STC1). It can therefore be observed that positive or negative PGHD correlated respectively with positive or negative emotions.

There were however a few stroke survivors who were neither disappointed nor pleased based on their PGHD, with one saying “it doesn’t really matter”, and that the data neither “encouraged (n)or discouraged” (FG2_STC1_1). There was a sense of acceptance that their performance can go up or down: “I was very, very upbeat about my progress, I was very pessimistic about my progress. And different times I felt, different” (FG2_STC1_1). Meanwhile one survivor described seeing gains in their PGHD as “kind of surprising how much you can achieve, you know you kind of surprise yourself what you can achieve” (INT3_STC1).

Finally, while some stroke survivors associated their PGHD with their performance: “Yeah you'd know if you didn't get a hundred” (FG1_STC2-3), others relied on their own feelings about their health to assess their performance: “numbers on the screen isn't going to tell me the other, one thing or the other” (INT1_STC1). However one had lesser trust on their PGHD and more on other people describing their improvement for them: “when they say, oh, you're doing this a lot better than you were, that you actually, realise that you have improved” (FG1_STC2-3).

Interest in Care Processes

Stroke survivors reported interest in how the exercises contributed to the progress of their therapy, as a result of seeing their PGHD. They wanted “to see whether there were any gains” (FG1_STC2-3), or some “improvement” (FG2_STC1_2); asserting that “there must be some things that have, that have improved from that, even if” (FG2_STC1_1) they were tired. Others expressed interest in “having an indicator of previous performance” (FG1_STC2-3), a “comparison point” (INT2_STC1), and see how they would “stack up at the end” (FG1_STC2-3). Still others liked “tracking” (FG2_STC1_1) their progress, whether good or bad: “because I understand it, um, some days are bad, some days are good” (FG2_STC1_1). A few survivors also wanted to see when they “weren't so good” (FG1_STC2-1) that week, and when they were: “it told me when you were right it’s, it’s um.. If you are doing okay or or not” (INT2_STC3). They also wanted to know why: “because you realise that you know you're not going as far because you have the disability part of it” (INT3_STC1); and whether “there's a bit more to do” (INT2_STC3) to improve. Some of them especially remarked on things they could have done better, with one remarking that “I should have been standing up, hmm, longer for weeks” (FG1_STC2-1). As such, their PGHD could “be helpful” to know what they needed “to do (to) improve” (INT2_STC3), comparing it with teacher-student feedback: “it’s like uh, you know student getting feedback from…their teacher or something…it's important to get feedback” (INT2_STC3). One described the experience as “a bit like having a physio visit you at home, or having an OT visit you at home” (FG1_STC2-3).

Some stroke survivors also reported interest in understanding where they “were at in the program” (INT2_STC3), to see “how far there was to go if you can, by what percentage was showing” (FG1_STC2-3), and feel better when they were close to finishing an exercise: “I see the percentage, and when I was getting to seventy percent, I would think uh, this is good I'm getting towards the end” (FG1_STC2-1). This could be important for self-management, as described by one survivor: “It was certainly something that I watched, to see where I was at. 'Cause um, y- you need to think about this, we have some sort of a budget of energy that you have to manage yourself, and you can't afford to get to empty” (FG1_STC2-3).

There were a few survivors who didn’t “really care” (FG2_STC1_1), didn’t “take that much notice” (FG1_STC2-1) of their PGHD, and that they were just “going with the flow” (FG2_STC1_1). A survivor described seeing their data but not perceiving any benefit from it: “it was just information…I didn't feel I was involved in any way” (INT2_STC3). This lack of feeling involved was also true even when the PGHD was informative: “You knew what you had to do to improve or not, but I I didn’t feel involved really” (INT2_STC3).

Interestingly, a few stroke survivors reported how seeing their PGHD affected their perceptions towards their therapy. Positive PGHD gave the impression that the therapy was effective, and “the confidence of actually doing the exercises” (FG1_STC2-3); whereas negative PGHD could imply that the therapy was “pretty useless” (INT1_STC1). However there was also one stroke survivor whose perceptions of the exercises was unaffected: “I wouldn’t say I was more engaged or less engaged by it, I took it for what it was” (INT2_STC1). Finally, another survivor found that the data helped with self-assessment: “I'm good in this game but I'm not good in the other game, it gives that sort of feeling you know” (INT3_STC1).

Personal Care Goals

Stroke survivors reported receiving “an extra percentage of motivation” (FG1_STC2-1) and an “incentive to do better” (FG1_STC2-3) when they see their PGHD. Seeing how their data changed over time would make therapy more “motivational”, because it “feels like a bit of a competition? To see if we can do better” (FG1_STC2-3). It helped them develop “short term and long term goals” or a new “benchmark to try and achieve next time”, and kept one “focused to some extent” (INT2_STC1). However, it was cautioned that a downward slope could be detrimental: “you can get hard on yourself or feel ‘Oh I didn’t get that right’ and if you keep doing that and you're not achieving 100% and the score keeps flashing at you, that can be a positive motivator, but can also (be) a negative one” (INT2_STC1). There was one survivor however whose motivation remained unchanged: “I know when I haven’t done well. And numbers on the screen isn't going to tell me the other, one thing or the other” (INT1_STC1). Meanwhile, some survivors reported being motivated to simply finish the therapy, “to get to somewhere in time” and to concentrate on trying to “achieve it” (FG1_STC2-3).

Relationship with Care Provider/s

Some stroke survivors were prompted to ask their therapists when they “haven’t been doing well,” and why so they can improve: “what [sic] reason to it, and then they tell me off for certain things” (INT3_STC1). However, while they “wanted to know why it was so different so…low” (INT2_STC3), they were less prompted to ask about their performance when it was

good: "when it was good I didn't ask questions" (INT2_STC3). Surprisingly good data however prompted one to ask how it could have happened: "a couple of times I was surprised that I had a higher score when I thought I'd done badly" (INT1_STC1). Regardless, one survivor reported being prompted to ask for more assistance: "If they were always bad then I would need uh more assistance and even if they (were) good um, they (are) not perfect, right so I would want to have more, more assistance to um, to to improve" (INT2_STC3).

Discussion

This study is the first to gather perceptions of stroke survivors on the effects of PGHD for PROM development. In fact to our knowledge, this is the first study that gathered patient-reported outcomes of utilising PGHD. While there have been studies presenting patient-reported outcomes of a health condition [3, 27] such as quality of life as a result of using a PGHD-enabled device [28], and patient-reported experiences or satisfaction with health interventions and technologies [29-31], there have been no studies systematically soliciting patient-reported outcomes of utilising PGHD. This study begins to understand how PGHD may affect the health outcomes of people who use them, from their own perspective. More specifically, this contributes to the missed opportunity of the poststroke simulated rehabilitation literature to engage patients in their own health care [15].

The findings of this study demonstrate that we can measure outcomes of utilising PGHD, and have shown some similarities with PGHD effects reported in the literature. Moreover, the findings have also provided a richer understanding of the effects, and at times broadened the scope. This is a strength of conducting qualitative data collection for patient-reported outcomes of a health domain of interest [3].

The findings show that while in general patients are affected positively or negatively based on their PGHD, some are not affected at all. First, while stroke survivors' perceptions confirm previous findings that PGHD may influence positive behavioural changes [8], it has also shown that some patients may only be influenced when encouraged personally by their therapists. Second, the findings confirmed that PGHD could increase patient satisfaction with their health [10], but for the most part only when they are positive. Negative PGHD can, in the stroke survivors' own words, be offensive or depressing. Still, other survivors reported indifference to their PGHD. This emphasises the need to utilise patient-centered design when developing health technologies [32], particularly when PGHD is produced [15].

Third, this study affirms reports that direct access to PGHD can increase patient engagement [9-11]; however, we have learned that for some stroke survivors PGHD is just information that does not make them feel involved. Fourth, development of new goals or benchmarks as a result of their PGHD were also reported by stroke survivors, supporting assertions that PGHD facilitates personalised health strategies [10, 24], but others relied on their own feelings about their health to formulate their goals and were neither motivated nor demotivated by their PGHD. Finally, while the literature showed that PGHD may improve health management coordination between patients and their care providers [10, 11, 25, 26], stroke survivors using a simulated rehabilitation system may only ask questions when their PGHD shows poor results, or when their results surprise them.

These findings have direct implications for PROMs to utilise PGHD. As mentioned before, this paper is part of a larger study to develop a PROM-PGHD with poststroke simulated rehabilitation systems as a use case. For example, we noted that previous literature on effects of PGHD have not addressed "levels of frustration". This will be added as an item to accommodate for stroke survivors' experience of seeing their PGHD go up and down. This study demonstrates how the focus groups and interviews will be used to broaden the coverage of the PROM.

One limitation to our study is that all of the focus groups and interviews were moderated by the same person (GLD), which has the risk of a narrower array of themes and ideas. However, this method's advantage is that it increases the chance of thematic saturation, as GLD would have learned relevant themes as the discussions progressed [3]. Additionally, our sample was never intended to represent all stroke survivors; it was meant to elicit the input of survivors who have used a leading poststroke simulated rehabilitation technology to guide the development of the PROM-PGHD. The PROM will be refined and validated further as per the next step of QIR [5].

Poststroke simulated rehabilitation systems were selected as the use case for measuring PGHD outcomes because of the global burden of stroke and the potential benefits of simulated rehabilitation systems. However research into applicability of a PROM-PGHD for the wider spectrum of health technologies are ongoing. The authors will report elsewhere on how these PROMs are proving their validity in other cases of technology and health domains.

Practitioners and developers of health technologies need to understand how PGHD affect people differently. Therefore, when considering PGHD-enabled devices the patient perspective has to be given priority [15, 32]. Future studies should consider relevant health domain and technology use cases for applying a PROM-PGHD and evaluate their implications on the care process.

Conclusions

This study has described the importance of patients self-reporting the effects of utilising PGHD through a PROM. It has identified a pertinent use case in poststroke simulated rehabilitation systems, and gathered perceptions from stroke survivors. The findings highlight five key themes that demonstrate that PGHD outcomes of utilising PGHD can be measured. Moreover, this study has provided a deeper understanding of those effects and broadened the scope through the words of the stroke survivors themselves. The perceptions gathered here will further refine and validate the development of a PROM-PGHD for poststroke simulated rehabilitation systems and has wider relevance and application.

Acknowledgements

We would like to thank Marie-Louise Bird, Wei-Peng Teo, and Jessica Freeman for assisting with recruitment and administration of the discussions, and Cecily Gilbert for assisting GLD during the discussions. GLD would also like to acknowledge the support of his organisational sponsor, Newman College (University of Melbourne).

References

[1] U.S. Food and Drug Administration, Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims, in: F.a.D.A. U.S. Department of Health and Human Services, ed., 2009.
[2] Canadian Institute for Health Information, PROMs background document, in, Canada, 2015.

[3] T. Walsh, D. Irwin, A. Meier, J. Varni, and D. DeWalt, The use of focus groups in the development of the PROMIS pediatrics item bank, *Quality of Life Research* **17** (2008), 725-735.
[4] E. Basch, J. Spertus, R.A. Dudley, A. Wu, C. Chuahan, P. Cohen, M.L. Smith, N. Black, A. Crawford, K. Christensen, K. Blake, and C. Goertz, Methods for developing Patient-Reported Outcome-Based Performance Measures (PRO-PMs), *Value in Health* **18** (2015), 493-504.
[5] D. DeWalt, N. Rothrock, S. Yount, and A. Stone, Evaluation of item candidates: The PROMIS qualitative item review, *NIH Public Access* **45** (2007).
[6] K. Williams, J. Sansoni, D. Morris, P. Grootemaat, and C. Thompson, Patient-reported outcome measures: Literature review, in, ACSQHC, Sydney, 2016.
[7] S.T. Rosenbloom, Person-generated health and wellness data for health care, *Journal of the American Medical Informatics Association* **23** (2016), 438-439.
[8] R. Shaw, D. Steinberg, J. Bonnet, F. Modarai, A. George, T. Cunningham, M. Mason, M. Shahsahebi, S. Grambow, G. Bennett, and H. Bosworth, Mobile health devices: will patients actually use them?, *Journal of American Medical Informatics Association* **23** (2016), 462-466.
[9] P. Sanger, A. Hartzler, R. Lordon, C. Armstrong, W. Lober, H. Evans, and W. Pratt, A patient-centered system in a provider- centered world: challenges of incorporating post-discharge wound data into practice, *Journal of American Medical Informatics Association* **23** (2016), 514-525.
[10] S. Woods, N. Evans, and K. Frisbee, Integrating patient voices into health information for self-care and patient-clinician partnerships: Veterans Affairs design recommendations for patient-generated data applications, *Journal of American Medical Informatics Association* **23** (2016), 491-495.
[11] R. Kumar, N. Goren, D. Stark, D. Wall, and C. Longhurst, Automated integration of continuous glucose monitor data in the electronic health record using consumer technology, *Journal of American Medical Informatics Association* **23** (2016), 532-537.
[12] A. Da Gama, P. Fallavollita, V. Teichrieb, and N. Navab, Motor rehabilitation using Kinect: A systematic review, *Games for Health* **4** (2015), 123-135.
[13] H.M. Hondori and M. Khademi, A review on technical and clinical impact of Microsoft Kinect on physical therapy and rehabilitation, *Journal of Medical Engineering* (2014).
[14] D. Webster and O. Celik, Systematic review of Kinect applications in elderly care and stroke rehabilitation, *Journal of Neuroengineering and Rehabilitation* **11** (2014).
[15] G.L. Dimaguila, K. Gray, and M. Merolli, The role of person-generated health data in simulated rehabilitation for stroke: A literature review, *JMIR Rehabilitation and Assistive Technologies* **5** (2018).
[16] S. Mendis, Stroke disability and rehabilitation of stroke: World Health Organization perspective, *International Journal of Stroke* **8** (2012), 3-4.
[17] World Health Organization, Cardiovascular diseases (CVDs), in, 2016.
[18] N. Mayo, M. MacKay-Lyons, S. Scott, C. Moriello, and J. Brophy, A randomized trial of two home-based exercise programmes to improve functional walking post-stroke, *Clinical Rehabilitation* **27** (2013), 659-671.
[19] K. Lohse, C. Hilderman, K. Cheung, S. Tatla, and H.F.M. Van der Loos, Virtual reality therapy for adults post-stroke: A systematic review and meta-analysis exploring virtual environments and commercial games in therapy, *PLoS ONE* **9** (2014).
[20] A. Darekar, B. McFadyen, A. Lamontagne, and F. Joyce, Efficacy of virtual reality-based intervention on balance and mobility disorders post-stroke: a scoping review, *Journal of Neuroengineering and Rehabilitation* **12** (2015).
[21] M.C. Moreira, A.M. de Amorim Lima, K.M. Ferraz, and M.A.B. Rodrigues, Use of virtual reality in gait recovery among post stroke patients – a systematic literature review, *Disability and Rehabilitation: Assistive Technology* **8** (2013), 357-362.
[22] T. Ogourtsova, W.S. Silva, P. Archambault, and A. Lamontagne, Virtual reality treatment and assessments for post-stroke unilateral spatial neglect: A systematic literature review, *Neuropsychological Rehabilitation* (2015), 1-46.
[23] L. Johnson, M.-L. Bird, M. Muthalib, and W.-P. Teo, Innovative STRoke Interactive Virtual thErapy (STRIVE) online platform for community- dwelling stroke survivors: a randomised controlled trial protocol, *BMJ Open* **8** (2018).
[24] L. Mamykina, M. Levine, P. Davidson, A. Smaldone, N. Elhadad, and D. Albers, Data-driven health management: reasoning about personally generated data in diabetes with information technologies, *Journal of American Medical Informatics Association* **23** (2016), 526-531.
[25] C. Petersen, Patient-generated health data: a pathway to enhanced long-term cancer survivorship, *Journal of American Medical Informatics Association* **23** (2016), 456-461.
[26] E. Murnane, D. Cosley, P. Chang, S. Guha, E. Frank, G. Gay, and M. Matthews, Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: implications for the design of technologies to manage mental health, *Journal of American Medical Informatics Association* **23** (2016), 477-484.
[27] D. Cella, S. Yount, N. Rothrock, R. Gershon, K. Cook, B. Reeve, D. Ader, J. Fries, B. Bruce, and M. Rose, The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH roadmap cooperative group during its first two years, *NIH Public Access* **45** (2007).
[28] K. Johnson, B. Patterson, Y.-X. Ho, Q. Chen, H. Nian, C. Davison, J. Slagle, and S. Mulvaney, The feasibility of text reminders to improve medication adherence in adolescents with asthma, *Journal of American Medical Informatics Association* **23** (2016), 449-455.
[29] C. Thompson, J. Sansoni, D. Morris, J. Capell, and K. Williams, Patient-reported outcome measures: An environmental scan of the Australian healthcare sector, in, ACSQHC, Sydney, 2016.
[30] K. Williams, J. Sansoni, D. Morris, and C. Thompson, A Delphi study to develop indicators of cancer patient experience for quality improvement, *Support Care Cancer* **26** (2018), 129-138.
[31] A. Coulter, R. Fitzpatrick, and J. Cornwell, The point of care - Measures of patients' experience in hospital: Purpose, methods and uses, in, The King's Fund, United Kingdom, 2009.
[32] M. Constand, J. MacDermid, V. Dal Bello-Haas, and M. Law, Scoping review of patient-centred care approaches in healthcare, *Biomed Central* **12** (2014).

Address for correspondence

Gerardo Luis Dimaguila (Email: dgl@student.unimelb.edu.au)

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190374

IoT, Cloud Computing and Big Data: Integrated Framework for Healthcare in Disasters

Samaneh Madanian[a], Dave Parry[a]

[a] Department of Computer Science, AUT University, Auckland, New Zealand

Abstract

Currently, healthcare in disaster management context faces a number of challenges mostly due to the lack of availability of reliable data from diverse sources required to be accessible by appropriate authorities. Therefore, the main objective of this study is the introduction of a framework based on the integration of three technologies, Internet of Things (IoT), cloud computing and big data to solve this issue in all disaster phases and provide precise and effective healthcare. This framework supports healthcare managers by enabling data sharing among them and assists them in performing analytical calculations to discover meaningful, logical and accurate trend(s) required for strategic planning and better preparedness in the face of disasters. Also, the outcome of the framework may help decision makers to identify and predict the health consequences of the disasters for any specific geographical location in any country based on its geographical properties and disaster background.

Keywords:
Medical Informatics, Disasters, Disaster Planning

Introduction

Disasters (natural, manmade, or epidemic) have detrimental effects and annually force huge expenses on governments nationwide in terms of economical, social, environmental, and human conditions and life. According to the statistics from Centre for Research on the Epidemiology of Disasters, the number of disasters and their complexity and severity (damage to life and property) shows an upward trend over recent decades. Only in 2017, 536 disasters occurred causing 16,458 deaths, affecting 95,979,187 people and costing USD 337,542,768,000 globally [1].

Regardless of disaster types and sources, their consequences mostly threaten people's health and wellbeing to varying degrees. This includes injuries or deaths of the affected people or even disrupting the healthcare systems that affects both disaster injuries and people with chronic healthcare conditions who still require care. Therefore, the greatest challenge to healthcare systems is probably when it is to deal with disasters; they should change swiftly from their everyday activities to adjust to the conditions of great uncertainty that, at the same time, may exceed their available resources and their ability to attend to a large number of casualties in a short time [2].

Although most disasters are not preventable, having a proper healthcare preparedness plan in advance can significantly lessen the harmful effects of disasters and diminish their adverse consequences. For such a plan and addressing citizens' healthcare requirements more effectively, reliable data from diverse sources are needed to be shared among different authorities and parties in disaster management. In this regard, effective data gathering and sharing is the building block leading to more accurate data analysis and information extraction required for a disaster healthcare plan. This issue is a challenging task and has not been effectively addressed yet.

The attempt in this paper is to introduce an integrated framework of three technologies. This framework aims to address challenges in data collection and sharing that are faced by different authorities, including healthcare, within Disaster Management Cycle (DMC). The framework enables automatic data collection and sharing that can be categorized for each region and the whole country before, during and after disasters. Also, it can further be used for real-time decision making, discovering trends, and possibly forecasting healthcare consequences of disasters on communities, or predicting epidemics outbreak. Moreover, it is expected that authorities will be able to prepare customized healthcare disaster management plans based on the geographical location properties, disaster history and the healthcare background of their inhabitants.

The rest of this article is organized as follows. First, the research methods and motivations are explained. Then, a brief background of the utilized technologies is provided. After that, the proposed model is presented followed by detailed explanations of its layers and its implications.

Research Methods and Motivation

As a part of Disaster e-Health (DEH) project (see [3] and [4]), a comprehensive scoping study in three fields of disaster management, disaster medicine, and healthcare was conducted. The scoping study was carried out in two different phases of controlled and uncontrolled search in 13 different databases including Scopus, CINAHL, ProQuest, IEEEXplore, Science Direct, and PubMed. In these databases, keywords or combination of keywords such as disaster, healthcare, technology, e-health and challenges, were used to extract relevant studies. Among 3956 captured articles in uncontrolled search and 925 articles in controlled, after considering the exclusion factors, 264 and 328 articles in each method, respectively, were selected for review.

Citizens' health requirements and demands following disasters are varied and highly depend on the disaster types, location, and intensity. Therefore, efficiency and effectiveness of healthcare disaster response activities are of prime importance due to finite resources compared to the number of affected population. In most disasters, preventing further life loss has a direct relation to the level of preparedness and readiness of different organizations in a community. Thus, pre-disaster phases, that is, preparedness and mitigation, are key elements of effective

response; however, their importance may be hidden or underestimated unless a major disaster occurs [5]. By managing pre-disaster phases wisely and properly in their planning, there will be no, or less need to run response and recovery phases as the disaster risks are prevented or reduced and appropriate systems, procedures, and resources are in place to assist affected population [6]. Consequently, rather than a reactive system, governments can shift to a proactive system.

However, most healthcare centers are still struggling to prepare an appropriate disaster management plan. According to the results of the scoping review, some of the identified root cause challenges are:

- Data gathering (considering data quality and reliability)
- Data sharing (which data, how to share and with whom)
- Data analysis (healthcare risk analysis to prepare for disaster threats, respond to and recover from them)

Data gathering and availability can be the greatest challenge for preparing an effective disaster management plan for healthcare centers most of which suffer from a lack of reliable and relevant data. This deficiency forces most countries, especially developing ones, to have a general, or national, disaster management plan mostly prepared and designed based on the assumptions for the whole country with no or a very few customizations for different regions [7]. However, based on regions' backgrounds, having some customization is a must. It is expected that disaster healthcare plans would be more effective if created for each geographical region by considering the number of its inhabitants and its disaster background information, including type, frequency, severity, and consequences. This information helps healthcare authorities to prepare region based and customized disaster healthcare plans, which are possibly more effective and efficient, and they may result in a better and effective healthcare response.

For a precise and effective disaster healthcare plan, it is necessary to have information regarding past disasters, and the effectiveness, success or failure rates of healthcare response activities. However, many countries have little information regarding the real effectiveness and efficiency of their healthcare response activities or their challenges and deficiencies [8, 9]. Therefore, healthcare response shortcomings have not been identified properly, and consequently, learned lessons from past disasters are poorly considered for future healthcare planning. This issue results in having mitigation and preparedness plans mostly based on the assumptions and not on the real data or information. This not only wastes a lot of valuable resources, but it also results in not being able to meet some of the demands of disaster casualties, specifically the healthcare related demands [10].

Technology Background

Information Technology (IT) utilization could be a viable solution for the identified challenges. It has the potential to improve overall disaster management, better prepare against disasters, facilitate response when disasters occur, enhance support after disasters, and keep records for better future preparedness.

In the proposed framework, three different technologies are utilized. These technologies have already revolutionized other fields/industries such as supply chain, healthcare and military. Therefore, it is believed that they may offer a huge potential for healthcare within disaster DMC. Although these technologies by themselves are powerful, it is expected that they may exploit further potential if they integrate with each other.

Internet of Things (IoT)

IoT is a technological phenomenon that stems from new advances and ideas in ICT, and it is related to ubiquitous communication/connectivity, ambient intelligence, and pervasive computing [12]. It comprises sensing devices and technologies such as sensors, Radio Frequency Identification (RFID), and Global Positioning System (GPS) [12] and network and communication components such as Wireless Sensor Network (WSN), 4/5G communication, and Wide Area Network (WAN). IoT integrates humans, physical objects and digital devices by making use of distributed sensors in environments [13]. IoT enables people to automatically gather all sorts of data and information such as location, movement, sound, and heat. Since IoT encompasses different types of technologies and components, its scope covers a broad range of applications in different fields such as healthcare, military, supply chain, and agriculture. IoT Devices are already starting to appear in the consumer space in the form of wearable devices and smart appliances so that it may be possible to both use data from these devices and the infrastructure being built to support them.

Cloud Computing

Cloud computing is a framework of "abstracted, virtualized, dynamically scalable, managing, computing, power storage platforms and services for on-demand delivery over the Internet." [14]. It consists of both hardware and software that offer various services in different levels of infrastructure, platform and software [15]; it has revolutionized IT resources utilization in different industries. Cloud services are available anywhere at anytime and bring flexibility and scalability of IT services and resources while enhancing their employment in different levels and organizations. Cloud computing provides a series of advantages for the organizations as mentioned by [16] including scalability, reliability, efficiency, and data storage.

Big Data

The "Big data" concept has emerged as a result of a massive increase in data with regard to its quantity, type and speed of generation. Big data can be defined as "high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization" [17]. Big data is used in different organizations for analytical purposes. Its data are from diverse sources, and sometimes in real-time manner, fueled by the latest development in technologies such as mobile devices adoption, social media, IoT, and RFID [18].

Related Studies And Research Objectives

Although the applications of the discussed technologies in other fields are increasing, only in recent times have the researchers started exploring the role of these technologies for disaster management purposes. However, there are some gaps in these studies. A significant number of solutions concentrate on utilizing only one technology: ([19, 20] on IoT, [21, 22] on big data, and [23] on cloud computing). Some others have utilized these technologies for a specific disaster phase: ([24] on disaster recovery, [22, 25-27] on disaster response, [28] on disaster response and preventive monitoring). Still, some other studies look into a specific activity in a specific disaster phase ([21] on big data for disaster relief activities and supply chain

management). Only a few research studies proposed their solutions based on technology integration for addressing the disaster management challenges ([29] integrated big data and cloud computing, [30] proposed big data and IoT integration, and [25] integrated cloud and IoT). Among the identified studies, only, [31] integrated all the mentioned technologies for the purpose of event detection and sending alerts, and coordinating rescue and medical operations. Therefore, it can be said that among all the recent research, little attention has been paid to technology integration and healthcare operations, specifically for pre-disaster phases.

In this context, in this research, the proposed framework has the potential to be used in supporting and assisting healthcare and governmental authorities in their decision making and disaster healthcare planning for any disaster type, in different phases and for different regions. The objectives of the framework are:

- automatic gathering precise and real-time or near real-time data at all disaster phases
- aggregating the captured data to facilitate data sharing among different parties
- identifying and revealing any meaningful trends in the captured data that can be utilized for forecasting future disasters and the required healthcare actions resulting in accurate disaster healthcare planning and precise response
- enabling authorities to customize and localize disaster healthcare planning based on regions' precise data.

Technology Integration and Proposed Model

One of the greatest challenges for preparing an effective disaster management plan for healthcare is the lack of reliable and relevant data. It is believed that a great number of issues can be prevented in healthcare by having better and quality data. These data can be a great source for healthcare disaster planning and management as we cannot manage something when we cannot measure it.

Therefore, the proposed framework in this research integrates and utilizes three advanced technologies in three layers (Figure 1) to facilitate data gathering, sharing and analysis which are necessary for an effective disaster healthcare plan and management, as these areas have been identified by several researchers as barriers [10, 32-34].

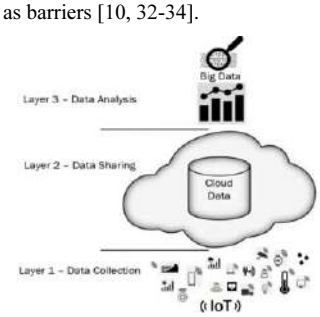


Figure 1 – Proposed Technology Integration Model

Layer One: IoT

The aim of this layer is to solve the problem of data availability and reliability by utilizing different IoT devices. The IoT devices are able to precisely capture different types of data for a particular case (object or human) with no or minimum human intervention in a periodic or continuous form in a specific location. These data can be people's vital signs or location, environment temperature, humidity, location of the objects, etc. The variety in the data types makes IoT suitable for monitoring purposes and transmitting or receiving real-time data regarding different conditions in the implemented environment. Furthermore, as most of the IoT devices have embedded memory and are battery powered, if network and power are disrupted by disasters, they can save data in their memory and upload them whenever their connection reestablishes. Therefore, the effects of disruptions in network communication after disaster could be mitigated.

Based on IoT functionalities and by giving digital identity to the objects, IoT sensors are able to be context-aware, sense the monitoring environmental conditions, communicate with neighboring nodes and even sometimes perform some basic computations on the collected data [35, 36]. These unique features of IoT can integrate the objects in the real world with the digital environment which, eventually, brings an integration, and relevant and quality data/information for healthcare planning at all phases of disasters. IoT also provides a cooperative monitoring environment either for environmental conditions or physical objects. Moreover, as IoT sensors require no or little human intervention, they can be used in harsh or infectious environments where conditions are not suitable for human work [35]. In such situations, the sensors can be set to execute processes to activate specific actions or create services [36].

Layer Two: Cloud Computing

Although in disaster healthcare planning, data availability and quality are crucial, sharing these data with right authorities is vital. Otherwise, even having quality datasets, may lead us to thousands of silos of fragmented data. While data sharing is the other common problem in healthcare planning. Therefore, the second layer of the framework attempts to address this problem by utilizing cloud computing technology. Furthermore, to apply the full potential of IoT, it seems that utilizing cloud computing in parallel with IoT is vital [37].

Cloud computing can provide the required virtual infrastructure for integrating the monitoring devices [37]. As it is required to store a wide range of data types, cloud facilitates deployment of a system for data sharing across different platforms and through a unified framework for a fast and easy access with unlimited data storage [16, 37]. This technology is an efficient and reliable solution for data sharing among different parties and authorities and supports collaborative work and access anywhere. Moreover, it is potentially more resilient and scalable in terms of physical locations.

In this layer, the captured IoT data available in a single cloud repository ultimately facilitate sharing of data and enhance collaboration and cooperation of people, programs, processes and services [36]. This issue is particularly important in healthcare as all the parties will be able to work on a common and aggregated data and establish a network to coordinate and exchange data more effectively and efficiently that, in turn, also supports strategic planning and research. Moreover, different authorities and governmental agencies, including healthcare, monitor different types of objects, locations and environmental conditions in a way that if any significant issues happen, they can respond swiftly.

Layer Three: Big Data

The outcome of the lower two layers is an integrated repository of data of differing quality, provenance and or reliability which disaster management and planning can use for evidence-based

plans and responses. For this purpose, the analysis and validation of all the captured data from monitoring devices are required. In this context, big data algorithms and analytics can be considered since they have the potential to make use of those data and extract the available trends and evidences. As in the previous layer cloud computing has been used, big data sharing and exchange becomes possible without further investing in big computing assets [38]. Big data algorithms and analytics can be used to visualize the connected IoT objects, make sense of them and help IoT for acceptance, usage and influence [18]. For this purpose, big data analytics tools and applications can be used to exploit the full potential of the proposed framework by converting the data into information and knowledge. In other words, big data "takes the amount of intelligence within the network to another level where devices reason and take action in ways that aren't necessarily preprogrammed or even initially understood by humans" [18].

This layer's outcome could facilitate supplying information and knowledge required for disaster healthcare planning as well as enhancing the quality of the plans and decisions in both normal and disaster situations. The generated information can be real-time for immediate decision making in any alarming situations [2], such as epidemics outbreak, or about the past disasters to support healthcare resilience and prevention research.

Discussion

Appropriate and precious healthcare planning, for normal or disaster situations, relies on past together with real-time data from various sources. If these data are available and reliable, their aggregation in a single repository with further processing and analysis may result in extracting important information and trends that have the potential to enhance the quality of disaster healthcare planning and preparedness. Hereof, it is believed that integrating the discussed technologies can support governments and authorities to provide a more accurate, robust and customized disaster preparedness and response plans based on the geographical locations by considering their geographical conditions, healthcare and disaster backgrounds. The proposed framework can be seen as a possible solution to the identified challenges in the disaster healthcare. The following paragraphs discuss the contribution of each technology.

IoT connects the physical objects with the digital world and makes objects identifiable and traceable. It also can collect and disseminate data from the monitoring areas to the control centers. Besides this offered convenience, these connected and intelligent devices could have the life-saving applications during disasters and be considered critical applications [39] as they can capture and then disseminate accurate data with minimal human interventions.

However, for a complete realization of the IoT potential, an efficient, secure and scalable storage is required. Consequently, in the proposed framework, cloud computing is suggested that can also provide a reliable and ubiquitous access to the dynamic IoT objects [37]. Therefore, the large amount of the collected data from IoT devices are aggregated and shared among different centers through cloud computing. In this way, collaboration among parties is improved, and critical data are swiftly distributed to the top decision makers [36].

This integrated repository can be considered as an easily accessible platform for using data mining models to discover new facts and trends. For this purpose, big data approaches such as large-scale visualizations, stream processing, machine learning, etc. can be used to generate knowledge and discover trends. This information can be employed by decision makers and other authorities for a quality strategic planning and predicting future requirements and demands in the face of disasters.

As a result of adopting this framework of interconnected sensing devices, sharing data across different platforms through a unified framework becomes possible. Therefore, healthcare decision-makers will be able to use the real-time and precise data for real-time decision making [36, 37] and for better understanding the impact of disasters and reacting more appropriately [40] with respect to the healthcare consequences. Additionally, every healthcare center is able to be integrated into the disaster plans, along with its critical components of the response efforts [41], and its situations can be assessed by using big data and cloud computing. This framework can also make data and information available for the wider and authorized access that may enhance data processing and sharing.

This approach may also enhance and facilitate decision making in the healthcare response missions as different objects in the monitoring environment are inter-connected via their properties [13]. Therefore, IoT for the disaster management application not only can be seen as a data acquisition technology, but it can also facilitate establishing a more effective, intelligent and goal-oriented communication [11]. Moreover, information regarding medical resources availability and location can be tracked and shared among related organizations automatically and without any human intervention that, in turn, results in enhancing tasks efficiency and accuracy. Consequently, matching the available medical resources and staff with the region population can be performed in disaster preparedness, and the chance of overwhelming different systems and organizations can be decreased in the disaster response.

Conclusions

IoT, cloud computing and big data are powerful on their own but together can potentially improve disaster healthcare. The combination of these technologies will require work to assure privacy and safety. However, research in these areas may produce synergistic benefits. Frameworks for integration of these approaches are vital for effective implementation.

References

[1] Centre for Research on the Epidemiology of Disasters, The International Disaster Database in, CRED, 2018. https://www.emdat.be/emdat_db/ (accessed October 12, 2018)

[2] Y. Bar-El, S. Tzafrir, I. Tzipori, L. Utitz, M. Halberthal, R. Beyar, and S. Reisner, Decision-support information system to manage mass casualty incidents at a level 1 trauma center, *Disaster Medicine and Public Health Preparedness* **7** (2013), 549-554.

[3] A.C. Norris, S. Martinez, L. Labaka, S. Madanian, J.J. Gonzalez, and D. Parry, Disaster e-Health: A new paradigm for collaborative healthcare in disasters, in: *The 12th ISCRAM International Conference*, Agder University, Kristiansand, Norway, 2015.

[4] D. Parry, S. Madanian, and A.C. Norris, Disaster e-health: Sustainability in the extreme, in: *2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology*, Auckland, New Zealand, 2016, pp. 943-946.

[5] PAHO, Natural disasters: Protecting the public's health, in, Pan American Health Organization, Washington, D.C, 2000. https://www.mona.uwi.edu/cardin/virtual_library/docs/1226/1226.pdf (accessed October 9, 2018)

[6] A. Ahmed and L.F. Sugianto, A 3-tier architecture for the adoption of RFID in emergency management, in: *Proceedings*

of the International Conference on Business and Information, Tokyo, Japan, 2007, pp. 1-14.
[7] A. Unlu, N. Kapucu, and B. Sahin, Disaster and crisis management in Turkey: a need for a unified crisis management system, *Disaster Prevention and Management: An International Journal* **19** (2010), 155-174.
[8] M.L. Birnbaum, E.K. Daily, A.P. O'Rourke, and J. Kushner, Research and evaluations of the health aspects of disasters, part VI: Interventional research and the disaster logic model, *Prehosp Disaster Med* **31** (2016), 181-194.
[9] T.D. Kirsch and E.B. Hsu, Disaster medicine: What's the reality?, *Disaster Medicine & Public Health Preparedness* **2** (2008), 11-12.
[10] A. Boyd, N. Chambers, S. French, D. Shaw, R. King, and A. Whitehead, Emergency planning and management in health care: priority research topics, *Health Systems* **3** (2014), 83-92.
[11] A. Dohr, R. Modre-Opsrian, M. Drobics, D. Hayn, and G. Schreier, The internet of things for ambient assisted living in: *Information Technology: 7th International Conference on New Generations,* IEEE, Las Vegas, NV, 2010, pp. 804-809.
[12] K. Zhao and L. Ge, A Survey on the Internet of Things security, in: 9*th International Conference on Computational Intelligence and Security*, 2013, pp. 663-667.
[13] A. Zelenkauskaite, N. Bessis, S. Sotiriadis, and E. Asimakopoulou, Interconnectedness of complex systems of Internet of Things through social network analysis for disaster management, in: *4th International Conference on Intelligent Networking and Collaborative Systems*, IEEE, Bucharest, Romania, 2012, pp. 503-508.
[14] E. AbuKhousa, N. Mohamed, and J. Al-Jaroodi, e-Health cloud: opportunities and challenges, *Future Internet* **4** (2012), 621-645.
[15] B. Ludwig and S. Coetzee, A comparison of platform as a service (PaaS) clouds with a detailed reference to security and geoprocessing services, in: *Proceedings of the First International Workshop on Pervasive Web Mapping, Geoprocessing and Services,* Como, Italy, Citeseer, 2010.
[16] L. Jun and Z. Song, *E-health web application framework and platform based on the cloud technology*, Master, Kristianstad University, 2013.
[17] Gartner, The importance of 'Big Data': A definition, in, 2012. https://www.gartner.com/doc/2057415/importance-big-data-definition (accessed October 9, 2018)
[18] F.J. Riggins and S.F. Wamba, research directions on the adoption, usage, and impact of the Internet of Things through the use of big data analytics, in: *48th Hawaii International Conference on System Sciences*, 2015, pp. 1531-1540.
[19] C. Yang, Q. Huang, Z. Li, K. Liu, and F. Hu, Big Data and cloud computing: Innovation opportunities and challenges, *International Journal of Digital Earth* **10** (2017), 13-53.
[20] M. Kamruzzaman, N.I. Sarkar, J. Gutierrez, and S.K. Ray, A study of IoT-based post-disaster management, in: *2017 International Conference on Information Networking*, 2017, pp. 406-410.
[21] T. Papadopoulos, A. Gunasekaran, R. Dubey, N. Altay, S.J. Childe, and S.F. Wamba, The role of Big Data in explaining disaster resilience in supply chains for sustainability, *Journal of Cleaner Production* **142** (2017), 1108-1118.
[22] S. Choi and B. Bae, The real-time monitoring system of social big data for disaster management, in: *Computer Science and its Applications,* J.J. Park, I. Stojmenovic, H.Y. Jeong, and G. Yi, eds., Springer Berlin Heidelberg, Berlin, Heidelberg, 2015, pp. 809-815.
[23] N. Jangid and B. Sharma, Cloud computing and robotics for disaster management, in: *7th International Conference on Intelligent Systems, Modelling and Simulation*, 2016, pp. 20-24.
[24] V. Chang, Towards a Big Data system disaster recovery in a Private Cloud, *Ad Hoc Networks* **35** (2015), 65-82.
[25] K. Chung and R.C. Park, P2P cloud network services for IoT based disaster situations information, *Peer-to-Peer Networking and Applications* **9** (2016), 566-577.
[26] L. Yang, S.H. Yang, and L. Plotnick, How the internet of things technology enhances emergency response operations, *Technological Forecasting and Social Change* **80** (2013), 1854-1867.
[27] A. Rauniyar, P. Engelstad, B. Feng, and D.V. Thanh, Crowdsourcing-based disaster management using fog computing in Internet of Things paradigm, in: *2016 IEEE 2nd International Conference on Collaboration and Internet Computing*, 2016, pp. 490-494.
[28] J.P. de Albuquerque, B. Herfort, A. Brenning, and A. Zipf, A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management, *International Journal of Geographical Information Science* **29** (2015), 667-689.
[29] R. Ranjan, Streaming big data processing in datacenter clouds, *IEEE Cloud Computing* **1** (2014), 78-83.
[30] Y. Sun, H. Song, A.J. Jara, and R. Bie, Internet of Things and Big Data analytics for smart and connected communities, *IEEE Access* **4** (2016), 766-773.
[31] D. Puthal, S. Nepal, R. Ranjan, and J. Chen, A secure big data stream analytics framework for disaster management on the cloud, in: *2016 IEEE 18th International Conference on High Performance Computing and Communications*, 2016, pp. 1218-1225.
[32] S. Underwood, Improving disaster management, *Communications of the ACM* **53** (2010), 18-20.
[33] A. Ahmed and L.F. Sugianto, RFID in emergency management, in: *Auto-identification and ubiquitous computing applications,* J. Symonds, J. Ayoade, and D. Parry, eds., IGI Global, Hershey, PA, 2009, pp. 137-155.
[34] D.W. Callaway, C.R. Peabody, A. Hoffman, E. Cote, S. Moulton, A.A. Baez, and L. Nathanson, Disaster mobile health technology: Lessons from Haiti, *Prehospital and Disaster Medicine* **27** (2012), 148-152.
[35] M.S. Aliyu, H. Chizari, and A.H. Abdullah, Integrated sensor and RFID network node placement in disaster monitoring applications, in: *RFID-Technologies and Applications (RFID-TA), 2013 IEEE International Conference on*, IEEE, 2013, pp. 1-5.
[36] S. Agrawal and M.L. Das, Internet of Things-A paradigm shift of future Internet applications, in: *Nirma University International Conference on Engineering*, IEEE, Ahmedabad, Gujarat, India, 2011, pp. 1-7.
[37] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, Internet of Things (IoT): A vision, architectural elements, and future directions, *Future Generation Computer Systems* **29** (2013), 1645-1660.
[38] J. Moock, Cloud Computing creates climate change in teleradiology, in, *Modern Medicine Network*, 2011.
[39] P. Johnson How the Internet of Things could become a critical part of disaster response, 2014.
[40] H. Petersen, E. Baccelli, M. Wählisch, T.C. Schmidt, and J. Schiller, The role of the Internet of Things in network resilience, *arXiv preprint arXiv:1406.6614* (2014).
[41] A.H. Kaji and J.F. Waeckerle, Disaster medicine and the emergency medicine resident, *Annals of Emergency Medicine* **41** (2003), 865-870.

Address for correspondence

Samaneh Madanian, Department of Computer Science, Auckland University of Technology, Auckland, New Zealand, email: Sam.madanian@aut.ac.nz.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190375

Algorithm Formalization for Decision Making in Influenza Vaccination

Enrique Monsalvo San Macario[a,b], Marta Fernández Batalla[a,c], Alexandra González Aguña[a], José María Santamaría García[a,c], Adriana Cercas Duque[a,d], Virginia Díaz Teruel[a,e], Blanca Gonzalo de Diego[a,c], Sara Herrero Jaén[a,f], Diego Cobo González[a,g]

[a] Research Group MISCK (Alcala University), Alcalá de Henares, Spain
[b] Health Center La Garena, Alcalá de Henares, Spain
[c] Health Center Meco, Meco, Spain
[d] La Zarzuela Hospital, Madrid, Spain
[e] 12 de Octubre Hospital, Madrid, Spain
[f] Severo Ochoa Hospital, Leganés, Spain
[g] La Paz Hospital, Madrid, Spain

Abstract

Influenza is an important public health problem with consequences on the health of people, but also at the state level with social and health costs due to the morbidity and mortality produced. Vaccination is an act of care of high clinical complexity, which can be learned and trained, so that decision-making in vaccination requires elements of judgment that act logically in a cascade. The creation of algorithms and their implementation in computer systems will identify susceptible people more quickly and improve the competence in the administration of vaccines. For the creation of the algorithm, the variables and their relationships were identified through mathematical formulation. As a result, the nine variables for vaccination that result in 47 different clinical situations were identified (29 "non-vaccination" and 18 "vaccination" situations). The formalization of algorithms in vaccine administration allows to represent the process by which the professional carries out the decision making process.

Keywords:

Algorithms; Decision Support Systems, Clinical; Influenza Vaccines

Introduction

The present study starts from the following reseach aim: to design an algorithm that describes the decision-making process in influenza vaccination. Influenza is an acute respiratory infection caused by RNA-virus that encompasses different groups (mainly groups A, B and C) with type A being the most prevalent [1]. This infection appears in the form of epidemic outbreaks and because of the high rate of virus mutation (especially H3N2) [2], annual vaccination is necessary (with an effectiveness of 70-90% if there is concordance between the strains) [3,4]. The virus is transmitted by air, and transmission period may begin from one day before the symptomatic phase to seven days after symptoms start [5]. Influenza is characterized by high fever, cough, muscle and joint pains, headacheand and intense discomfort. In people with poor health, the flu and its complications can be fatal, due to pneumonias caused by secondary bacterial invaders [1].

Worldwide, influenza accounts for about 300,000 to 600,000 seasonal influenza-associated respiratory deaths annually [2]. In Spain, the flu surveillance report indicates that 5,977 severe hospitalized confirmed cases of influenza were reported, of which 21.8% were admitted to the ICU and 17.3% died in the 2017/2018 season [6]. In the Community of Madrid, an accumulated incidence of 1,540.33 cases per 100,000 inhabitants was estimated in the same season [7]. In conclusion, influenza is a major public health problem worldwide related to a high rate of morbidity and mortality, increase in social and health costs, potentially preventable and whose management and control of risk should continue to evolve and improve to increase the level of health of the community [2,7].

Vaccination is one of the main acts of care in primary prevention, definid as health behaviors aimed at avoiding the health problem, or reducing the probability of it appearing [8]. In particular, it is essential for the prevention of health problems of infectious origin worldwide [9]. Fernandez Batalla et al. indicate that vaccination is an act of care of high clinical complexity, which can be learned and trained, so that decision-making in vaccination requires elements of judgment that act logically in a cascade. The first assessment should be based on the characteristics of the individual (vaccine selection) and then the administration decision according to the context [10]. In addition, the computational implementation of decision-making in care allows a better identification of the population to which vaccination is recommended. This allowed us to design tools for specific distance recruitment and increase vaccination rates, which has been progressively decreasing in recent years [11].

Methods

For the identification and description of variables, the methodology used was deductive, based on:

- Extraction of knowledge through text analysis
  - Scientific and technical regulations of the Public Health services, in relation to the indication of influenza vaccination for the 2017-2018 influenza campaign of the Community of Madrid.
  - Instructions for use and technical data sheets of the influenza vaccines available in said campaign.
- Education throught expert's knowledge.

The expert group consists of a nurse doctor in computer science, mathematical doctor in computer science, a nurse specialist in community health, and two master's degree nurses in multidisciplinary informatics.

Establishment of existing relationships between variables: algorithm creation.

- Mathematical formulation through bivalued logic: Identification of relevant clinical variables and definition of possible values for each variable[12-14].

Results

Extraction and Definition of the Variables.

A total of 9 variables relevant to influenza vaccination were selected.

Age

Time a person has lived at the time of the valuation. It was classified into 6 stages: less than 6 months (e0), 6 - 35 months (e1), 3 - 8 years (e2), 9 – 59 years (e3), 60 - 64 years (e4) and 65 years or more (e5).

Life Process

Indicates the presence of factors that indicate risk group for vaccination. Three main groups were identified:

- Processes that increase the risk of complications from the flu
- People who can pass the flu to those who are at high risk of complications
- Essential public service workers

With the presence of one factor it will be considered posive (r) while the absence of all of them will be considered negative (¬r)

First Vaccination

Indicates that the person has not been vaccinated against the flu in previous years (p).

Interval Between Doses

When vaccination with several doses is necessary, it indicates that 4 weeks have passed after the first dose (i).

Initial Dose

When vaccination with several doses is necessary, it indicates that the vaccine proposed constitutes the first dose to be administered in this anti-flu campaign.

Influenza Vaccination Completed

Indicates the correct vaccination in the current anti-flu campaign (c)

Health Situation

Indicates the presence of factors that contraindicate the administration of the vaccine at that time. Three main situations were identified: fever, acute infection, allergy to some components of the vaccine.

With the presence of one factor it will be considered posive (s) while the absence of all will be considered negative (¬s).

Anticoagulation

Indicates that the patient is at increased risk of hemorrhage by intramuscular puncture (a).

Vaccination Action

It is the result variable. Indicates whether the vaccine is indicated (v) or not (¬v) and complementary information (type of vaccine, dose, route of administration, cause of non-vaccination).

Order of Variables

For decision making in vaccination, a hierarchy of the variables is necessary to indicate when to ask for each variable.

The final formalization of the variables is presented in Table 1.

Table 1 – Description and Formalization of Variables

Variable	Order	Code	Range
Age	1	e	e0 –e5
Life process	2	r	r/¬r
Influenza vaccination completed	3	c	c/¬c
First vaccination	4	p	p/¬p
Initial dose	5	d	d/¬d
Interval between doses	6	i	i/¬i
Anticoagulation	7	a	a/¬a
Health situation	8	s	s/¬s
Vaccination action	9	v	v/ ¬v

Establishment of the Relationships Between Variables

With the identified variables, a total of 47 different clinical cases were obtained leading to a specific vaccination action (Table 2).

- 29 "non-vaccination" situations. The algorithm has five different causes that indicate "no vaccination" (distribution in Figure 1).
- 6 situations in which to vaccinate with inactivated vaccine (0.25ml)
- 11 situations in which to vaccinate with inactivated vaccine (0.5ml)
- 1 situation in which to vaccinate with inactivated vaccine with adjuvant (0.5ml)

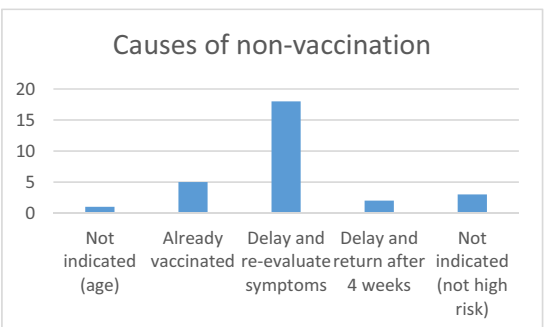


Figure 1 – Distribution of Reasons for Non-Vaccination

Table 2 – Coding of the Algorithm

Caso clínico		Acción vacunal
e0	→	¬v1
e1 ∧ r ∧ c	→	¬v2
e1 ∧ r ∧ ¬c ∧ p ∧ d ∧ a ∧ s	→	¬v3 ∧ ¿s?
e1 ∧ r ∧ ¬c ∧ p ∧ d ∧ ¬a ∧ s	→	¬v4 ∧ ¿s?
e1 ∧ r ∧ ¬c ∧ p ∧ ¬d ∧ i ∧ a ∧ s	→	¬v5 ∧ ¿s?
e1 ∧ r ∧ ¬c ∧ p ∧ ¬d ∧ i ∧ ¬a ∧ s	→	¬v6 ∧ ¿s?
e1 ∧ r ∧ ¬c ∧ p ∧ ¬d ∧ ¬i	→	¬v7 ∧ ¿i?
e1 ∧ r ∧ ¬c ∧ ¬p ∧ a ∧ s	→	¬v8 ∧ ¿s?
e1 ∧ r ∧ ¬c ∧ ¬p ∧ ¬a ∧ s	→	¬v9 ∧ ¿s?
e1 ∧ ¬r	→	¬v10
e2 ∧ r ∧ c	→	¬v11
e2 ∧ r ∧ ¬c ∧ p ∧ d ∧ a ∧ s	→	¬v12 ∧ ¿s?
e2 ∧ r ∧ ¬c ∧ p ∧ d ∧ ¬a ∧ s	→	¬v13 ∧ ¿s?
e2 ∧ r ∧ ¬c ∧ p ∧ ¬d ∧ i ∧ a ∧ s	→	¬v14 ∧ ¿s?
e2 ∧ r ∧ ¬c ∧ p ∧ ¬d ∧ i ∧ ¬a ∧ s	→	¬v15 ∧ ¿s?
e2 ∧ r ∧ ¬c ∧ p ∧ ¬d ∧ ¬i	→	¬v16 ∧ ¿i?
e2 ∧ r ∧ ¬c ∧ ¬p ∧ a ∧ s	→	¬v17 ∧ ¿s?
e2 ∧ r ∧ ¬c ∧ ¬p ∧ ¬a ∧ s	→	¬v18 ∧ ¿s?
e2 ∧ ¬r	→	¬v19
e3 ∧ r ∧ c	→	¬v20
e3 ∧ r ∧ ¬c ∧ a ∧ s	→	¬v21 ∧ ¿s?
e3 ∧ r ∧ ¬c ∧ ¬a ∧ s	→	¬v22 ∧ ¿s?
e3 ∧ ¬r	→	¬v23
e4 ∧ c	→	¬v24
e4 ∧ r ∧ ¬c ∧ a ∧ s	→	¬v25 ∧ ¿s?
e4 ∧ r ∧ ¬c ∧ ¬a ∧ s	→	¬v26 ∧ ¿s?
e5 ∧ c	→	¬v27
e5 ∧ r ∧ ¬c ∧ a ∧ s	→	¬v28 ∧ ¿s?
e5 ∧ r ∧ ¬c ∧ ¬a ∧ s	→	¬v29 ∧ ¿s?
e1 ∧ r ∧ ¬c ∧ p ∧ d ∧ a ∧ ¬s	→	v1
e1 ∧ r ∧ ¬c ∧ p ∧ d ∧ ¬a ∧ ¬s	→	v2
e1 ∧ r ∧ ¬c ∧ p ∧ ¬d ∧ i ∧ a ∧ ¬s	→	v3
e1 ∧ r ∧ ¬c ∧ p ∧ ¬d ∧ i ∧ ¬a ∧ ¬s	→	v4
e1 ∧ r ∧ ¬c ∧ ¬p ∧ a ∧ ¬s	→	v5
e1 ∧ r ∧ ¬c ∧ ¬p ∧ ¬a ∧ ¬s	→	v6
e2 ∧ r ∧ ¬c ∧ p ∧ d ∧ a ∧ ¬s	→	v7
e2 ∧ r ∧ ¬c ∧ p ∧ d ∧ ¬a ∧ ¬s	→	v8
e2 ∧ r ∧ ¬c ∧ p ∧ ¬d ∧ i ∧ a ∧ ¬s	→	v9
e2 ∧ r ∧ ¬c ∧ p ∧ ¬d ∧ i ∧ ¬a ∧ ¬s	→	v10
e2 ∧ r ∧ ¬c ∧ ¬p ∧ a ∧ ¬s	→	v11
e2 ∧ r ∧ ¬c ∧ ¬p ∧ ¬a ∧ ¬s	→	v12
e3 ∧ r ∧ ¬c ∧ a ∧ ¬s	→	v13
e3 ∧ r ∧ ¬c ∧ ¬a ∧ ¬s	→	v14
e4 ∧ ¬c ∧ a ∧ ¬s	→	v15
e4 ∧ ¬c ∧ ¬a ∧ ¬s	→	v16
e5 ∧ ¬c ∧ a ∧ ¬s	→	v17
e5 ∧ ¬c ∧ ¬a ∧ ¬s	→	v18

Discussion

In relation to the use of Information and Communication Technologies, the CDC indicates logic specification for ACIP Recommendations in Clinical Decision Support for Immunization (CDSi). The variables used in this research work resemble the variables defined by the CDC (Target dose, patient series)[15].

In the context of primary health care, Martín-Ivorra indicates that the development of this type of tools would improve to identify the people of being vaccinated who attend the scheduled consultations in Primary Care[16]. On the contrary, it has found different publications that indicate that the use of ICTs are useful to determine and increase vaccination coverage[17,18].

Gerard et al. developed a system of influenza vaccination reminders in a hospital's information systems to increase the rate of influenza vaccination in patients hospitalized in internal medicine. Although the system was effective, it did not discriminate the indication of the vaccine based on their clinical features[19]. In relation to decision making in vaccines, Jacobson describe the development, of a software tool, introduced to assist health care professionals and public health administrators in managing pediatric vaccine purchase decisions and making economically sound formulary choices[20]. In relation to the methodology, Shiffman and Greenes used this methodology with with logic and decision-table techniques to improve cinical guidelines applicates to prevention of perinatal transmission of hepatitis B by immunization[12].

Conclusions

The formalization of algorithms in the vaccination allows to represent the process by which the professional carries out the decision making process, in this case, in the influenza vaccination. In addition, the algorithms serve as a guide for the professional. The hierarchy of variables is not the most computationally efficient but it represents in a better way the process of decision making. To measure progress in vaccination decision-making training, it is proposed to incorporate a case generator into the application in the future. Thanks to the development of the algorithm, all the variables have been coded to generate these cases.

Acknowledgements

This study was not financed by public or private company. Management about Information and Standard Knowledge of Care Research (University of Alcala) supported this work.

References

[1] C. J. Woolverton, J. M. Willey, and L. M. Sherwood, *Microbiología de Prescott, Harley y Klein, 7th Edition*, McGraw-Hill-Interamericana, España, 2013.

[2] Lancet, Editorial. Preparing for seasonal influenza, The Lancet 391(10117), 2018.

[3] A. Domarus, C. Rozman, and P. F. Valentí. *Medicina interna. 17th Edition*, Elsevier Health Sciences Spain, Barcelona, 2012.

[4] World Health Organization (WHO). *Vacunas antigripales: Documento de posición de la OMS,* (2012) http://www.who.int/immunization/position_papers/WHO_PP_Influenza_Nov_2012_Spanish.pdf?ua=1

[5] I. Hernández-Aguado, A. Gil de Miguel, M. Delgado Rodríguez, F. Bolúmar Montrull, F. G. Benavides, and M. Porta Serra, *Manual de Epidemiología y Salud Pública para grados en ciencias de la salud. 2nd Edition,* Médica Panamericana, Madrid, 2012.

[6] Instituto de Salud Carlos III. Informe de Vigilancia de la Gripe en España. Temporada 2017-2018, (2016). http://vgripe.isciii.es/documentos/20172018/InformesAnuales/Informe_Vigilancia_GRIPE_2017-2018_27julio2018.pdf

[7] Servicio de Prevención de la Enfermedad. Dirección General de Salud Pública. Vacunación frente a la Gripe estacional 2018-2019, Technical document, (2018)

[8] L. Mazarrasa Alvear, *Salud pública y enfermería comunitaria*, 2nd Edition, McGraw-Hill Interamericana de España, 2003.

[9] A. Martín Zurro, *Atención Primaria: conceptos, organización y práctica clínica*, 6th Edition, Elsevier, 2008.
[10] M. Fernández Batalla, M. L. Jiménez Rodríguez, J. M. Santamaría García, J. L. Gómez González, A. González Aguña, and E. Monsalvo San Macario. Conceptualización de la toma de decisiones en el cuidado: acercamiento desde la vacunación. *ENE, Revista de Enfermería*, 9(3) (2015).
[11] Ministerio de Sanidad Servicios Sociales e Igualdad. Evolución coberturas de vacunación antigripal en población ≥ 65 años. España, temporadas 2006-2007 a 2015-2016. http://www.msssi.gob.es/profesionales/saludPublica/prevPromocion/vacunaciones/docs/CoberturasVacunacion/Tabla10.pdf
[12] R. N. Shiffman, and R. A. Greenes, Improving clinical guidelines with logic and decision-table techniques: application to hepatitis immunization recommendations, Med Decis Making, 14(3)(1994),245-254.
[13] M. L. Jiménez Rodríguez. Sistema Basado en el Conocimiento para ayuda en el diagnóstico del cansancio del desempeño del rol de cuidador. PhD. University of Alcalá de Henares, Madrid, Spain. 2006.
[14] B. Gonzalo de Diego. S. Herrero Jaén, E. Monsalvo San Macario, L. Madariaga Casquero, M. L. Jiménez Rodríguez, J. M. Santamaría García, et. al. *Robotic Implementation of the Necessary Mechanism for the Human Characteristics Simulation: Approach from Self-Care Conceptual Modeling.* Stud Health Technol Inform. 2018;250:115-120.
[15] National Center for Immunization and Respiratory Disease (NCIRD), Immunization Information Systems Support Branch (IISSB). Clinical Decision Support for Immunization (CDSi): Logic specification for ACIP recommendations, (2019). https://www.cdc.gov/vaccines/programs/iis/interop-proj/downloads/logic-spec-acip-rec-4.0.pdf
[16] Martín-Ivorra R, Alguacil-Ramos AM, Lluch-Rodrigo JA., Pastor-Villalba E, Portero-Alonso A. *Actividades para captar y vacunar a la población susceptible en la Comunidad Valenciana*. Rev. Esp. Salud Pública 2015 89(4): 419-426.
[17] I. N Souza Interaminense, S. Costa de Oliveira,L. Pedrosa Leal, F. M. Pereira Linhares, M. Pontes Cleide. *Educational technologies to promote vaccination against human papillomavirus: integrative literature review*, (2016). 25(2). http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-07072016000200502&lng=en.
[18] Pastor Villalba E, Cremades Bernabeu A, Alguacil Ramos AM, Martin Ivorra R, Portero Alonso A, Lluch Rodrigo JA et al. *Recaptación activa de la 1a y 2a dosis de triple vírica en la comunidad valenciana. Utilización y rendimiento de la herramienta del Sistema de Información Vacunal (SIV).* 6o Congreso de la Asociación Española de Vacunología 23 al 26 de Nov 2011. Santiago de Compostela: Asociación Española de Vacunología; 2011.
[19] M. N. Gerard, W. E. Trick, K. Das, M. Charles-Damte, G. A. Murphy, I. M. Benson. *Use of clinical decision support to increase influenza vaccination: multi-year evolution of the system*. J Am Med Inform Assoc. (2008), 15(6):776-9.
[20] S. H. Jacobson, and E. C. Sewell, *A Web-based Tool for Designing Vaccine Formularies for Childhood Immunization in the United States*, J Am Med Infor Asoc, 15(5), (2008), 611-619.

Address for correspondence

Enrique Monsalvo San Macario
Research Group MISKC
University of Alcalá.
Alcalá de Henares, Madrid (Spain).
Email: enrique.monsalvo@hotmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190376

A Knowledge-Based Platform for Assessing Potential Adverse Drug Reactions at the Point of Care: User Requirements and Design

Pantelis Natsiavas[a,b], Marie-Christine Jaulent[b], Vassilis Koutkias[a]

[a] Institute of Applied Biosciences, Centre for Research & Technology Hellas, Thermi, Thessaloniki, Greece
[b] Sorbonne Université, INSERM, Univ Paris 13, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances pour la e-Santé, LIMICS, F-75006 Paris, France

Abstract

Even though Adverse Drug Reactions (ADRs) constitute a significant public health issue, there is a lack of Information & Communication Technologies (ICT) tools supporting Pharmacovigilance activities at the point of care. In this paper, we present the rationale of a Web-based platform to address this need. The driving user scenario of the proposed platform refers to a clinician who investigates information for a possible ADR as part of a specific patient treatment. The goal is to facilitate this assessment through appropriate tools for searching various relevant data sources, analysing the acquired data, aggregating the obtained evidence, and offering follow-up ADR monitoring over time in a systematic and user-friendly way. In this regard, we describe the adopted user requirements engineering methodology and illustrate the use of Knowledge Engineering (KE) as the platform's main technical paradigm to enable heterogeneous data integration and handle the complexity of the underlying information processing workflow.

Keywords:

Pharmacovigilance; Drug-Related Side Effects and Adverse Reactions; Knowledge Management

Introduction

Pharmacovigilance (PV) is defined as "the science and activities related to the detection, assessment, understanding and prevention of adverse effects or any other possible drug-related problems" [1]. As Adverse Drug Reactions (ADRs) cause a significant social and financial burden [2], PV is widely recognized as an important public health priority. An estimation by the US Office of Disease Prevention and Health Promotion has recently calculated that Adverse Drug Events (ADEs)[1] are responsible for 1 in 3 of all hospital adverse events, related to about 2 million hospital stays each year, and increased hospitalization by 1.7 to 4.6 days[2]. Thus, the detection and prevention of ADRs at the point of care rise as a major clinical issue as the probability of benefit should balance the possibility and cost of potential harm [4].

The assessment of potential new or incompletely documented ADRs (called "signals") is typically performed by national and international drug monitoring/regulatory organizations (e.g., the Food and Drug Administration [FDA] in the US, the Uppsala Monitoring Centre [UMC], World Health Organization collaborating centre for international dug monitoring). These organizations perform statistical analysis of individual case safety reports (ICSRs) gathered in Spontaneous Reporting Systems (SRSs), in order to identify indications of a causal relationship between the drug administration and the adverse effect based on the measures of disproportionality [5], taking also into account other sources of evidence (e.g., scientific literature, clinical trial databases, etc.).

While SRSs are the dominant data source for PV, recent advances in Information and Communication Technologies (ICT) enable the exploitation of new, emerging data sources that can expand the real-world evidence base for PV (e.g., observational healthcare databases, social media, internet search logs). Thus, the need for comprehensive and knowledge-intensive ICT tools supporting the systematic and efficient exploitation of diverse data sources for PV is evident, in order to accommodate the entailed big data challenges [6].

To this end, we develop a Web-based platform aiming to facilitate the early identification and assessment of potential ADRs at the point of care. The main objective is to contribute at "active", post-marketing drug safety surveillance [7], focusing on the timely assessment of potential drug safety risks, supporting clinicians (as well as PV experts and researchers) to explore diverse data sources of interest and obtain actionable insights via knowledge-intensive analytics [8]. The proposed platform is currently in its "user requirements analysis" and "design" phase, which is driven by the real-life scenario according to which a clinician investigates information for a possible ADR as part of a specific patient treatment. The ultimate goal is to facilitate the integration of ADR assessment in routine clinical practice (Figure 1), by introducing tools that facilitate the search of diverse data sources, the analysis of the acquired data, the aggregation of the evidence to conclude with ADR assessment, and follow-up ADR monitoring over time in a systematic and user-friendly way.

In this paper, we present the methodology applied and the main challenges identified during the "user requirements analysis" phase of our development, which were in turn mapped to relevant user goals. We also illustrate the main elements of the platform design through the respective information processing workflow. In this regard, we elaborate on how a Knowledge Engineering (KE) based approach can accommodate the re-

[1] ADEs include side-effects that may or may not have a causal relationship with the respective drug, also referring to cases of adherence failure. ADRs refer only to side-effects caused after legitimate drug use, therefore implying a possible causal relationship between the drug and the adverse effect [3].

[2] https://health.gov/hcq/ade.asp

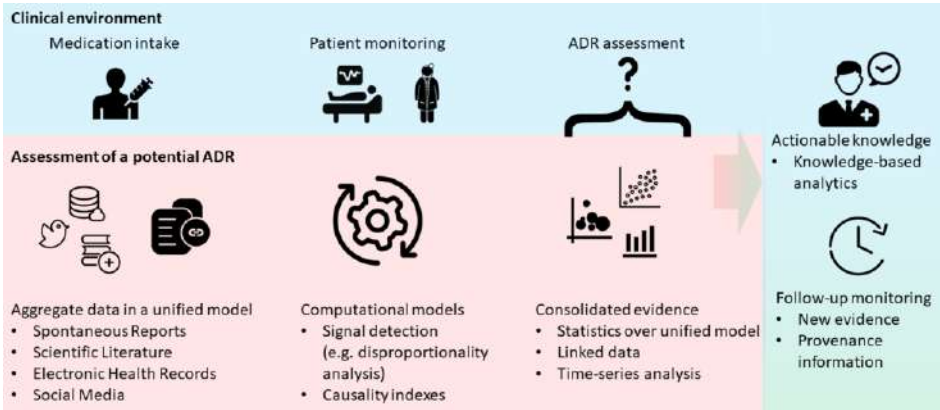


Figure 1 - Integrating the ADR assessment process in the clinical context.

spective design and development challenges. We further discuss practical implications of our work and outline directions for future work, aiming to support a comprehensive learning health system for active, post-marketing drug safety surveillance at the point of care [9].

Compared to relevant works, like the SALUS platform [10], which focused on exploiting Electronic Health Records (EHRs) for ADR detection, or the platform developed in WebRADR [11], which exploited social media for new insights on drug safety, our work relies on multiple, diverse data sources, increasing the search space for real-world evidence with an explicit focus on the clinical environment.

Methods

The employed "user requirements analysis" process is an adjusted version of the methodology described in [12]. It can be summarized as follows:

1. Analysis of the currently applied *Business Processes* (BPs) based on the respective user scenarios.
2. Definition of *User Goals* upon the elaborated BPs based on end-user input.

A *Business Process* (BP) is defined as a collection of relevant and ordered structured activities/tasks aiming to produce a specific outcome [13]. ADR assessment can also be considered as a BP conducted in the context of a hospital, as part of other parallel BPs (e.g., patient treatment, administrative processes). We envisage that the use of the proposed platform could reform the current process of ADR assessment, typically conducted manually and without systematic ICT support, to a well-defined sequence of information processing steps, supporting the overall clinical treatment processes. The ultimate goal is to optimize this BP model by satisfying the so-called *User Goals*. User goals are defined as "abstract user requirements, not directly referring to specific technical solutions or components" [12], associated with specific user actors or roles and facilitating timely identification and resolution of potential conflicts between actors. For the optimization of the ADR assessment process, user goals were elaborated based on feedback provided by clinicians and PV experts in the "user requirements analysis" and the "design" phases.

Regarding the presented platform design, the main BP of interest refers to the assessment of a potential ADR by a clinician. However, other BPs could also interact with it, e.g., concerning the patient's treatment. Given that patient treatment is the topmost priority in the clinical environment and that it is a personalized process, highly dependent on the local context (e.g. the way the specific clinic/hospital is organized), the modelling of these "interacting" BPs is very important and could be rather complex. The proper modelling and early identification of such BP interactions could be critical as they may substantially affect the identified goals and, thus, the platform design.

We model the identified BPs using flowcharts based on a notation similar to Business Process Management Notation (BPMN), in order to identify decision points, possible information processing bottlenecks, interactions with other BPs, etc. These flowcharts are further refined collectively by ICT experts, healthcare professionals (i.e., clinical doctors) and PV professionals (i.e., scientists who investigate potential ADR signals). Furthermore, interviews and workshops among researchers and end users were conducted as part of the overall "user requirements analysis" phase to analyse the established BPs, identify the *User Goals* and refine the platform's information processing workflow accordingly. Meetings were also held in the clinical environment (i.e., in the two hospitals which will host the platform in its pilot phase) to validate these goals and also address deployment issues in practice.

All types of current and emerging data sources considered in PV [6], were found interesting to explore by the end users. These include the local EHR systems, national and international SRSs, reference bibliographic databases as well as social media. From a technical viewpoint, programmatic data access is considered of high priority, as it enables systematic data gathering (e.g., spontaneous reports from the FDA Adverse Event Reporting System via openFDA [14], articles via the PubMed Central Application Programming Interface).

In order to successfully accommodate the imposed challenges regarding the synthesis and analysis of the vast data available, *Knowledge Engineering* (KE) is adopted as the main technical paradigm for our platform development. KE refers to methods, tools and theories for developing knowledge-intensive applications [15], and includes *knowledge extraction*, *knowledge integration*, *knowledge representation*, *knowledge dissemination*, and *knowledge elicitation* as its subdomains.

In the scope of our work, which concerns the systematic exploitation of all the available evidence from multiple PV data sources for the assessment of possible ADRs, we employ two technology artefacts tightly related with KE, namely, "*Linked Data*" [16] and "*Semantic Web*" [17]. Linked Data refer to a group of standards which facilitate the interconnection of data

over the existing Internet infrastructure, while Semantic Web refers to the vision of semantically annotated publicly available data interlinked via Linked Data standards.

The use of Semantic Web and Linked Data standards provides two main technical benefits: (a) *Interoperability:* The use of the Linked Data paradigm provides syntactic and semantic interoperability tools to interlink heterogeneous data sources and unify them in one processing realm. (b) *Reasoning capabilities:* The well-defined semantics upon a robust mathematical infrastructure, i.e., Description Logics [18], enable automatic reasoning through specific software, a.k.a. "reasoners".

Results

Based on this methodology and the choice of KE as the main technical paradigm, several challenges were identified from the end user perspective, leading to concrete user goals and the design of the platform's information processing workflow. In particular, given the characteristics of the clinical environment, the following challenges (enumerated with Cx) regarding the adoption of an ICT-based ADR assessment process were identified:

C1 | *Lack of time*: While the assessment of potential ADRs is identified as an important task, it is often neglected by clinicians due to lack of time.

C2 | *Lack of expertise*: PV entails specialized knowledge, which may not be available in clinical settings. While this argument supports the need for ICT-based support tools, it could also be conceived as a barrier for their adoption as their value might not be evident for the end-users.

C3 | *Adaptation to the clinical workflow*: Workflow diversity among various clinical environments (different hospitals, or even different clinics in the same hospital apply different BPs) could hinder the definition and the adoption of a "one-size-fits-all" workflow of PV information processing.

C4 | *Inadequate evidence*: While spontaneous reports are the dominant source of evidence for PV, other data sources such as EHRs, bibliographic databases, and even social media platforms are interesting for clinicians during ADR assessment. However, systematic access to multiple data sources shall be facilitated through appropriate tools.

C5 | *Coping with "big data"*: Expanding the search space for PV does not only provide a broader evidence space, but it also imposes "big data" challenges.

Overall, challenges C1-C5 have been discussed in the conducted workshops and the following user goals (enumerated with Gx) were identified and mapped to the respective challenges:

G1 | *Flexibility and Unobtrusiveness (mapped to C3):* The ADR assessment process should be flexible and tolerant to interruptions by tasks directly related with patient treatment. Thus, an important feature would be the ability to easily recover from such interruptions. Practically, this can be interpreted as the need to "save" the ADR assessment workflow and continue later. In addition, clinicians stressed that the patient's treatment should not be disrupted. Thus, the designed process should be as unobtrusive as possible, minimizing "alerts"/ "warnings".

G2 | *Balance between assessment depth and speed (mapped to C1 and C5):* As the clinician's time is valuable, the platform should enable both "in depth" assessment capabilities, while also supporting a "quick look" which could provide rigorous information. Although the information provided this way would obviously be more superficial than an "in depth" assessment, it could still provide value for clinicians.

G3 | *Semantic enhancement (mapped to C1, C2 and C5)*: Since the expression of the drug and condition of interest can be ambiguous (e.g., active substances, trade names, synonyms etc., could be used as drug terms), the overall process should be supported by curated standard terminologies and lexicons (e.g., with automatic synonym matching) to accelerate and facilitate information search and synthesis.

G4 | *Heterogeneous data synthesis (mapped to C4 and C5):* Clinicians identified the need to synthesize various and heterogeneous data sources (e.g., scientific literature, drug-information databases, clinical trial information, SRS data, observational healthcare databases). Overloading the end user with incomprehensible data was identified as a major risk and, thus, the need for knowledge-based analytics emerged.

G5 | *Data sharing (mapped to C2):* The need to share data to further elaborate on the collected ADR information and assessment results was also identified. Moreover, the value of data provenance was highlighted, especially for the process of reporting assessment outcomes to regulatory organizations.

G6 | *Follow-up monitoring over time (mapped to C4):* Typically, an ADR assessment produces a report with the conclusion and the supporting evidence. However, the time dimension is critical in PV, especially regarding newly marketed drugs. Thus, a follow-up mechanism for monitoring potential ADRs over time is important and is currently missing.

The main information processing workflow supported by the proposed platform (Figure 2) that was defined based on the abovementioned challenges and user goals are organized in 5 steps. These steps are summarized next, describing the use of the KE methods that are applicable in each case:

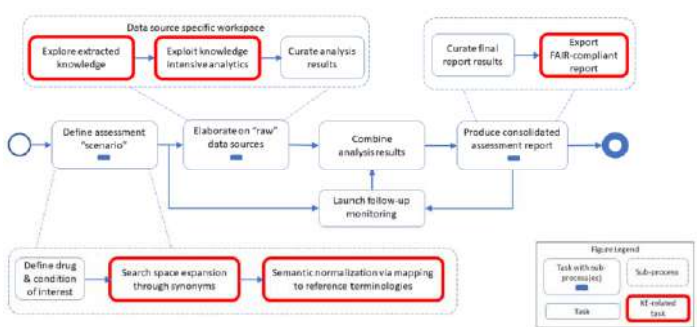


Figure 2 – Outline of the main information processing workflow supported by the proposed platform (tasks related with Knowledge Engineering are highglihted in red).

Step 1 | *Definition of "ADR assessment scenario"*: When the user launches an ADR assessment, the platform shall support the definition of the drug and the condition of interest by automatic suggestions of synonyms and relevant terms obtained from reference terminologies, e.g., the Anatomical Therapeutic Chemical (ATC) classification for drugs, and the Medical Dictionary for Regulatory Activities (MedDRA®) and the International Classification of Diseases (ICD) for the conditions of interest. The use of such well-defined knowledge structures expands the search space, enables the semantic normalization of the overall process, prevents ambiguities, and facilitates automatic information interlinking in the next analysis steps.

Step 2 | *Browse/analyse raw data from each data source*: Each raw data source has its own characteristics. For example, SRS data could be used for disproportionality analysis, while EHR data could be explored through observational healthcare data analytics [14]. KE approaches will be exploited for the analysis of each data source, e.g., text mining techniques can be used to extract and semantically annotate information from unstructured data sources such as the literature or social media. Thus, the end user can browse or analyse the respective data source in a dedicated workspace, providing suitable features and analysis capabilities.

Step 3 | *Combine analysis results from raw data sources*: The results/analysis outcome obtained from each data source workspace shall be integrated in one common processing realm, where all the analysis results could be integrated, compared and evaluated by the end user. Knowledge integration is based on semantic annotations produced in the previous steps and the use of Linked Data standards. Moreover, semantic reasoning can be applied to further elucidate knowledge from the already extracted analysis results.

Step 4 | *Produce a consolidating assessment report*: The overall analysis outcome shall be generated as a consolidated report, facilitating further analysis in collaboration with other clinicians, or even reporting to PV regulatory agencies. The produced report shall be available in both human-readable (e.g., in text form as a PDF document) and machine-readable (e.g., an RDF document) formats. Knowledge dissemination approaches can be used to facilitate the respective information exchange in a way that could promote the automatic reuse of this information. For example, the recently developed OpenPVSignal model [19] could be used in this regard, to enable compliance with the FAIR data principles [20].

Step 5 | *Launch follow-up monitoring*: The end user can launch a monitoring follow-up process, in order to receive potentially new information regarding the assessed ADR from the available data sources. This process would notify the end-user based on his/her notification preferences to avoid over-alerting. Ontology models such as the *Time Ontology* [21] and the *PROV-O Ontology* [22] can be used to enrich the obtained information with semantically enhanced time and provenance annotations, and thus, facilitate further processing/reasoning regarding the time aspects and the origin of the information collected regarding the ADR under assessment.

It should be noted that the presented workflow defines an independent BP, which can be adapted to each organizational context. Furthermore, in each assessment process, the end-user may decide the time spent on each BP step, either selecting to use the automatically retrieved information, drill-down to investigate further or manually curate the produced outcomes, and save or share his/her work with others at any time.

Discussion

Drug safety is an important issue in the clinical environment. Among the common tasks that are routinely performed in PV centres/departments in hospitals, the collection and review of all the available data for a potential ADR of interest is vital. However, there is a lack of comprehensive tools to support PV activities, specifically tailored for use at the point of care. For example, a clinician may ask for a timely evaluation of the respective patient case after a new drug administration, and the PV centre/department shall provide a documented answer with a medical advice about the case, after assessing the eventuality of an ADR [8]. To respond to this challenge, current available sources of information about the drug–event pair have to be searched by PV experts separately and in many cases without using appropriate support tools.

To address this need, we are currently developing a Web-based, knowledge-intensive platform aiming to support the assessment of potential ADRs, experienced during routine patient treatment. In the current paper, we presented the entailed challenges and the goals for such a development from the user perspective. We also presented the platform's main information processing workflow (Figure 2). Its design relies on exploiting various KE-based methods, employed in each step of the workflow.

In particular, the use of Linked Data and Semantic Web technologies provides the following key benefits:

- *Information linking* can be improved and automated by reducing the need for manual data exploration and discovery as data could be automatically retrieved.
- Rich *semantics* enhance information processing capabilities, which are typically limited in the PV domain to statistical measures of disproportionality. The already established statistical methods could be combined with semantically-enhanced knowledge sources to improve outcomes via automatic reasoning capabilities (e.g. regarding causality assessment).
- *Evidence can be strengthened* through the knowledge-intensive, concurrent exploitation of multiple data sources, eliminating false positive findings [6].
- KE-based automatic information linking enables the use of *provenance information* to annotate the generated analysis outcomes. This is important as full supporting evidence shall be explicitly available, e.g. when reporting results to regulatory organizations.

Despite the abovementioned benefits, the implementation of KE-oriented techniques entails complex challenges, both in methodological and technical terms, e.g.:

- Automatic reasoning capabilities based on the Description Logic defined semantics are one of the most prominent features of Semantic Web technologies. However, "reasoners" require significant computational resources and their efficient use in large datasets remains a challenge.
- Various reference knowledge sources (e.g. terminologies/thesauri/vocabularies) are available and can be applicable in the scope of this work. However, since these sources are constantly evolving and refined, their alignment is a complex task as it can lead to semantic inconsistencies.
- Integrating all the collected evidence under one unified knowledge model can be very challenging. This

process engages many heterogeneous data sources, which could be available via standard data exchange interface or not. For example, using proprietary EHRs to retrieve observational healthcare data would typically require specific interface implementations.

To this end, besides the ultimate goal of delivering a robust and evaluated ADR assessment platform, our mid-term goals are: (a) the design of a unifying semantic model enabling the integration of heterogeneous data sources in one information processing realm, and (b) the modelling of ADRs in one ontological model, facilitating advanced reasoning operations upon the collected information.

Conclusions

There is a clear need for comprehensive tools to support PV activities at the point of care. The proposed platform aims to support the assessment of potential ADRs in routine clinical practice, relying on the concurrent exploitation of multiple data sources for appropriate evidence. This entails the analysis of the acquired data, the aggregation of the obtained evidence, and the support of follow-up ADR monitoring over time in a systematic and user-friendly way. In this paper, we presented the main challenges and the goals from the end-user perspective for such a development, identified during the "user requirements analysis" phase of our development. We also presented the main elements of the platform design, i.e. its main information processing workflow, and illustrated the use of KE as the platform's main technical paradigm. Our work contributed to the development of a learning health system for active, post-marketing drug safety surveillance at the point of care.

Acknowledgements

This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneur-ship and Innovation, under the call RESEARCH – CREATE - INNOVATE (project code: T1EDK-03789).

References

[1] W.C.C. for I.D.M. World Health Organization, The importance of pharmacovigilance., World Health Organization, 2002. http://apps.who.int/medicinedocs/en/d/Js4893e/ (accessed Nov. 24, 2018).

[2] D. Formica, et al., The economic burden of preventable adverse drug reactions: a systematic review of observational studies, *Expert Opin Drug Saf.* **17** (2018) 681–95.

[3] M. Lindquist, The need for definitions in pharmacovigilance, *Drug Saf.* **30** (2007) 825–30.

[4] R.E. Ferner, and P. McGettigan, Adverse drug reactions, *BMJ*. **363** (2018) k4051.

[5] J.-L. Montastruc, et al., Benefits and strengths of the disproportionality analysis for identification of adverse drug reactions in a pharmacovigilance database., *Br J Clin. Pharmacol*. **72** (2011) 905–8.

[6] V.G. Koutkias, and M.-C. Jaulent, Computational Approaches for Pharmacovigilance Signal Detection: Toward Integrated and Semantically-Enriched Frameworks, *Drug Saf.* **38** (2015) 219–32.

[7] J.J. Gagne, J.A. et al., Active safety monitoring of new medical products using electronic healthcare data: selecting alerting rules, *Epidemiology* **23** (2012) 238–46.

[8] V.G. Koutkias, et al., Exploiting heterogeneous publicly available data sources for drug safety surveillance: computational framework and case studies, *Expert Opin Drug Saf.* **16** (2017) 113–24.

[9] L.B. Ramsey, et al., Learning Health Systems as Facilitators of Precision Medicine., *Clin Pharmacol Ther.* **101** (2017) 359–67.

[10] M. Yuksel, et al., An Interoperability Platform Enabling Reuse of Electronic Health Records for Signal Verification Studies, *Biomed Res Int.* **2016** (2016) 6741418.

[11] R. Gosh and D. Lewis, Aims and approaches of Web-RADR: a consortium ensuring reliable ADR reporting via mobile devices and new insights from social media, *Expert Opin Drug Saf.* **14** (2015) 1845–53.

[12] P. Natsiavas, et al., Comprehensive user requirements engineering methodology for secure and interoperable health data exchange, *BMC Med Inform Decis Mak.* **18** (2018) 85-101.

[13] M. Weske, *Business process management: concepts, languages, architectures*, *2nd Edition*, Springer, 2012.

[14] T.A. Kass-Hout, et al., OpenFDA: an innovative platform providing access to a wealth of FDA's publicly available data, *J Am Med Inform.* **23** (2016) 596–600.

[15] G. Schreiber, Knowledge Engineering, in: F. Van Harmelen, V. Lifschitz, and B. Porter (Eds.), *Handb Knowl Represent.*, Elsevier, 2008: pp. 929–46.

[16] T. Heath, and C. Bizer, Linked Data: Evolving the Web into a Global Data Space, *Synth Lect Semant Web Theory Technol.* **1** (2011) 1–136.

[17] N. Shadbolt, et al., The Semantic Web Revisited, IEEE Intell Syst. **21** (2006) 96–101.

[18] F. Baader, et al., Description Logics, in: *Handb Ontol.*, Springer, Berlin, Heidelberg, 2004: pp. 3–28.

[19] P. Natsiavas, et al., OpenPVSignal: Advancing Information Search, Sharing and Reuse on Pharmacovigilance Signals via FAIR Principles and Semantic Web Technologies, *Front Pharmacol.* **9** (2018) 609-24.

[20] M.D. Wilkinson, et al., The FAIR Guiding Principles for scientific data management and stewardship, *Sci Data.* **3** (2016) 160018.

[21] S. Cox, et al. Time Ontology in OWL, 2017. https://www.w3.org/TR/owl-time/ (accessed Nov 24, 2018).

[22] Y. Gil, et al., W3C PROV Model Primer, 2013. https://www.w3.org/TR/prov-primer/ (accessed Nov 24, 2018).

Address for correspondence

Pantelis Natsiavas, Institute of Applied Biosciences, Centre for Research & Technology Hellas, 6th Km. Charilaou-Thermi Road, P.O. BOX 60361 GR – 57001, Thermi, Thessaloniki, Greece, email: pnatsiavas@certh.gr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190377

Same Goals, Yet Different Outcomes: Analysing the Current State of eHealth Adoption and Policies in Austria, Germany, and Switzerland Using a Mixed Methods Approach

Laura Naumann[a*], Moritz Esdar[a*], Elske Ammenwerth[b], Dieter Baumberger[c], Ursula Hübner[a]

[a] Health Informatics Research Group, University AS Osnabrück, Osnabrück, Lower Saxony, Germany
[b] Institute of Medical Informatics, UMIT – University for Health Sciences, Medical Informatics and Technology, Hall in Tyrol, Tyrol, Austria
[c] Research and Development, LEP AG, St. Gallen, St. Gallen, Switzerland
**Contributed equally to this work.*

Abstract

Despite similar policy goals, the adoption of eHealth practices took different paths in Austria (AT), Switzerland (CH), and Germany (GER). We seek to provide a rigorous analysis of the current state of hospitals by focusing on three key eHealth areas: electronic patient records (EPR), health information exchange (HIE), electronic patient communication. For validation and in order to gain better contextual insight we applied a mixed method approach by combining survey results from clinical directors with qualitative interview data from eHealth experts of all three countries. Across countries, EPR adoption rates were reported highest (AT: 52%, CH: 78%, GER: 50%), HIE-rates were partly lower (AT: 52%, CH: 14%, GER: 17%), and electronic patient communication was reported lowest overall (AT: 17%, CH: 8%, GER: 19%). Amongst others, results indicate patient awareness about eHealth to be equally weak across countries, which thus may be an important focal point of future policy initiatives.

Keywords:

Electronic Health Records, Health Information Exchange, Health Policy

Introduction

Widespread diffusion and usage of electronic health records across care settings are a major issue on health policy agendas worldwide [1–3]. Also, Austria (AT), Switzerland (CH), and Germany (GER) aspire to improve continuity of care by fostering eHealth. Although, there are similarities between the major German-speaking regions in Europe, there are also crucial differences (Tab. 1).

Table 1 – Country characteristics

	AT	CH	GER
Population (2013)*	8.5 Mill.	8.1 Mill.	80.6 Mill.
Federal States	9	26	16
Welfare type	SHI	Public/Private	SHI
Hospitals (2016)*	273	283	1,951**
Expenditure on hospitals (2016)*	4.0% of GDP	4.3% of GDP	3.2% of GDP

SHI: Social Health Insurance, GDP: Gross Domestic Product
*see www.stats.oecd.org, **see www.destatis.de (accessed 10/25/2018)

Correspondingly, eHealth legislation took different paths in the three countries (Fig. 1). For instance, Germany has seen a rather long process of eHealth legislation with changing goals and approaches, dating back to 2003 but medically useful applications are not available up to this point. In Austria, the introduction of the Electronic Health Record ("ELGA") has already started, allowing health care providers and patients to access selected structured patient documents. In Switzerland, the federal government regulations of the Electronic Patient Dossier (EPD) stipulate that health professionals in hospitals are technically able to store essential patient information required for further treatment until 2020 (see Tab. 2).

Table 2 – Recent eHealth legislation

AT	**ELGA** *(Electronic Health Record Act)*
	• Focus on the "Elektronische Gesundheitsakte" (ELGA) to exchange discharge letters, laboratory data, medical imaging, medication data
	• Mandatory participation for health care providers
	• Citizens participate unless they object (Opt-Out)
	• Defined structure, format and standards for ELGA data
CH	**EPDG** *(Federal law on the electronic patient dossier)*
	• Focus on the "Electronic Patient Dossier" (EPD)
	• Mandatory participation for in-patient care providers
	• Voluntary participation for out-patient health care providers and citizens (Opt-In); patients themselves determine access rights
	• National subsidies to fund and build the necessary preconditions for the EPD
	• Defined monetary penalties in case of misuse
	• Defined standards, which are to be used to get certified (legal obligation)
GER	**E-Health-Gesetz** *(Act for Secure Digital Communication and Applications in the Healthcare Sector)*
	• Planned: Medication summary, telemedical applications, emergency data management, electronic patient records
	• Subsidies for sending and receiving medical eSummaries
	• Penalty for out-patient health care providers in case the insurance data is not up to date
	• Implementation of an interoperability register

Recent studies show that Germany is lagging behind Austria and Switzerland when it comes to diffusion and use of health IT applications in hospitals [4–7]. Spreading medical innovation in health care, hospitals are crucial hubs also for national eHealth infrastructures [8]. While there is information about the current state in terms of numbers, little is known about how the stakeholders perceive and evaluate this situation against the background of the national eHealth legislation and

the respective healthcare ecosystem. In order to find out how high-level survey data go along with the perceived reality, a combination of quantitative and qualitative methods is helpful to yield the full picture and cross-validate findings.

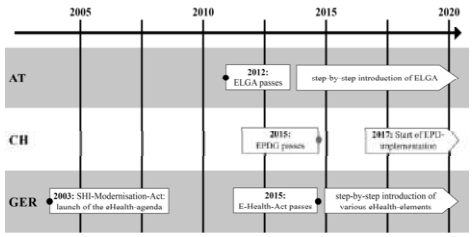


Figure 1 – Timeline of eHealth-laws

This study thus aims at investigating the current state of development and the combined views from a broad range of stakeholders in Austria, Germany, and Switzerland in the key areas: a) hospital internal electronic patient records (EPR)[1], b) health information exchange (HIE) across settings, and c) electronic communication with patients.

RQ1: Do quantitative data and qualitative findings provide a similar picture in the three key eHealth areas?

RQ2: Can qualitative findings explain potential quantitative differences between the three countries?

Methods

We used a parallel mixed method study design (Fig. 2) in which the data of both sets were collected simultaneously, hence providing a point in time observation [9]. Quantitative data were obtained from clinical directors (either nursing or medical directors) as hospital representatives using the standardised online survey *IT Report Healthcare 2017* [5] that measured various aspects of IT adoption [4]. Qualitative data were obtained from 59 health care experts representing a broad field of expertise (Fig. 2): health care delivery (hospitals, out-patient-care, nursing, telemedicine), industry (IT-provider, pharma industry), health care policy, and others (academia, data protection, patient organisations). Phone interviews were conducted using a guideline covering the national eHealth initiatives. Using the software *MAXQDA*®, interview data was screened systematically for statements regarding the three key eHealth areas in each country. Relevant quantitative data from the survey respondents (Fig. 2) was tested for country differences using logistic regression models in each area. In order to synthesise both data sets, the quantitative data was then complemented by selected quotes from the expert interviews. The screening of the qualitative interview material for hospital related statements lead to 547 initial hits of which we retained 97 statements for further analysis that were relevant to the research questions (AT=28, CH=29, GER=40,). The quantitative survey yielded a response rate (RR) of 17.8%. Out of 2,421 hospitals contacted (contact data were missing for some hospitals), we received 430 responses – 32 from Austria (RR = 12.3%), 43 from Switzerland (RR = 20.4%), and 355 from Germany (RR = 18.2%). Bigger and non-private hospitals were slightly overrepresented in our samples.

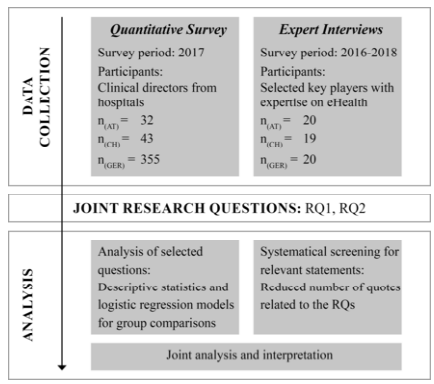


Figure 2 – Research process

Results

Research question 1:

Do quantitative data and qualitative findings provide a similar picture in the three key eHealth areas?

Key area a) Hospital internal electronic patient records

Adoption rates of EPR systems within hospitals showed to be similar in the German and Austrian sample while Swiss hospitals indicated a significantly higher EPR adoption rate (Tab. 3).

Table 3 – Adoption rates of electronic patient records with 95% confidence intervals (CI) and test for group differences

Question	Country	% of Hospitals with an EPR
Q1: Does your hospital have an Electronic Patient Record (EPR)?	AT (n=29)	52% (±18%)
	CH (n=41)	78%*(±13%)
	GER (n=338)	50% (±5%)

*significantly higher adoption rate (p<0.01), GER as reference category

As summarized in Table 4, experts in all three countries provided a mixed picture, with some recognizable trends: Austrian experts pointed to the progress in building the clinical IT-infrastructure but also showed to be aware of the still existing deficiencies. In Germany, experts mostly confirmed the deficiencies and only alluded to progress made in selected institutions. The Swiss experts clearly perceived some progress and an advanced state of developments without neglecting some deficiencies. The similarity in adoption rates between Austria and Germany were referred to by one Austrian expert as follows:

"In the hospital sector, I think we have approximately the same IT status as, for example, in Germany or other comparable EU-countries."

The German experts pointed at only modest maturity levels, stressing for instance that

"talking about EPRs, hospitals are already somewhat advanced, although there are still many blank spots".

Swiss hospitals appeared to be better off with EPR adoption rates close to 80% based on the quantitative survey. These results were supported to some extent by expert opinions in the interviews:

[1] We provided the following definition of an EPR for all respondents: "The EPR is an electronically generated and based institution-specific collection of patient information on current and previous stays in the institution. It is supported by clinical decision-making systems and replaces paper based medical documentation as the primary source of information."

"So many hospitals are now becoming much more active in this domain."

However, seen collectively, the interviewees also added more sceptical assessments and pointed to difficulties (Tab. 4):

"One would simply have to show much more the benefits, wouldn't one? If you are looking at the hospital processes and you are a clinical director and you know this system offers me benefits (presumably in monetary terms as well) and the processes are so much better and so on… then I consider using the system – but usually it´s not like that right now."

Table 4 – Expert statements on the electronic patient record and clinical IT-infrastructure within hospitals

Expert assessments	AT	CH	GER
Existing deficiencies	6	3	6
Selected progress among individual hospitals	2	2	4
Progress in the clinical IT-infrastructure is recognisable	1	5	2
Advanced status	6	2	1
Included interview statements in total*	15	12	13

*representing 40 out of 97 or 41.2% of all included expert statements

Key area b) Health information exchange

When extending the focus to information exchange with other health care institutions, the quantitative survey data indicated lower implementation rates in German and Swiss hospitals and somewhat higher implementation rates in Austrian hospitals (Q2). Austria is also using more sophisticated technology (Q3), i.e. portals, compared to Swiss hospitals, which were using primarily email to communicate (Tab. 5).

Table 5 – Health Information Exchange with 95% CI's and tests for group differences.

Question	Country	Respondents indicating "yes"		
Q2: Are external health data usually transferred using a portal?†	AT (n=12)	52%** (±28%)		
	CH (n=20)	14% (±15%)		
	GER (n=123)	17% (±5%)		
Q3: Is the medical discharge letter (doctor's letter) provided electronically for outside practitioners?		**No**	**Via email**	**Via Portal**
	AT (n=26)	31%** (±18%)	11% (±11%)	58%** (±19%)
	CH (n=39)	33%** (±15%)	54%** (±16%)	13% (±11%)
	GER (n=315)	87% (±5%)	5% (±2%)	8% (±3%)

*significantly higher adoption rates (p<0.05), GER as reference category
**significantly higher/lower adoption rates (p<0.01), GER as reference category
†This question was addressed only to those who indicated that they electronically integrate data from previous care stages into their systems. Almost all other respondents that didn't use a portal, answered that they scan in paper documents.

The experts pointed at comparably few deficits and more progress in Austria than in the other two countries, though structural barriers were reported in all three of them (Tab. 6).

"I think we're on the right track with this IT-infrastructure, which we're currently setting up in the course of introducing ELGA." one Austrian expert reported.

The modest adoption rates of HIE in Switzerland (see Tab. 5) were reflected by existing deficits and structural barriers as expressed by one Swiss expert:

"The possibilities for health information exchange across settings are still very limited and the patient record, i.e. the EPD, does not yet exist."

German hospitals were significantly poorer developed with regard to transferring discharge letters (Tab. 5) – a well-known drawback in Germany´s eHealth landscape which was reiterated by many of our interviewees (Tab. 6). One expert stated:

"Even across sector boundaries, from hospitals to out-patient care. We are still, I don't know, 20 years behind."

Table 6 – Expert statements on HIE

Expert assessments	AT	CH	GER
Existing deficits in HIE	1	4	9
Structural barriers for HIE	4	4	7
Progress in HIE recognisable	4	1	1
Structural facilitators for HIE	1	3	2
Included interview statements in total*	10	12	19

*representing 41 out of 97 or 42.3% of all included expert statements

Key area c) Electronic communication with patients

Looking at IT-functions that allow for direct communication between patients and providers, all countries still operated on a rather low level (Tab. 7). Swiss hospitals reported the lowest adoption rates while rates in Austria and Germany were slightly, but not significantly, higher. This pattern was mostly in line with the experts´ comments (Tab. 8). However, deficits were more often voiced by German experts. In total, this topic was not addressed all too often by the experts.

Table 7 – "Communication with patients" with 95% CI and test for group differences.

Item	Country	% of Hospitals indicating availability
Q4: Availability of IT function for communication with patients (e.g. via patient portals)*	AT (n=30)	17% (±13%)
	CH (n=40)	8% (±8%)
	GER (n=319)	19% (±4%)

*no significant group differences, GER as reference category

Table 8 – Expert statements on electronic communication with patients

Expert assessments	AT	CH	GER
Deficits present	1	3	5
Increasing expectations	1	1	2
Progress discernible	1	1	1
Included interview statements in total*	3	5	8

*representing 16 out of 97 or 16.5% of all included expert statements

Research question 2:

Can qualitative findings explain potential quantitative differences between the three countries?

Key area a) Hospital internal electronic patient records

The quantitative data pointed to a more advanced situation with regard to EPRs in Switzerland in comparison to the other countries. However, the Swiss experts did not offer a comprehensive explanation but rather pointed to the strengths of the Swiss hospitals, as one expert stated:

"And that is why in-patient structures, such as hospitals, are of course good carriers for ICT-innovations and for the promotion of the EPD, because with their central structures they have the necessary power (human and financial resources) to carry out such projects much better."

The qualitative interviews provided some background information on the mixed results among the Austrian hospitals. One participant indicated that ELGA initiated positive stimuli:

"I also think that such topics are very good drivers for innovation in a hospital. I see it this way: When you introduce a new system, when you deal with processes, then you always have the opportunity to clean up old things and think about how processes can be streamlined. From my point of view, the ELGA system has also brought us something positive."

However, criticism was also expressed particularly with regard to advanced functionalities:

"What is missing is a real innovation, like automatically creating summaries, displaying trends, abstracted from the concrete data. [...] What used to be known as a medical expert system or as clinical decision support is now completely lacking. We are currently at the level of medical raw data."

Reasons for the rather modest EPR adoption rates in Germany included the following explanation given by one expert:

"Of course, there are reasons for that, as I already mentioned, the financial situation: Half of the hospitals generate a deficit and they have to try to buy IT with the resources they have. There is only little support, financial funding in other words."

Key area b) Health information exchange

The Austrian hospitals showed the greatest progress in HIE in comparison to Switzerland and Germany based on the quantitative survey results. This progress was reflected in the interviews particularly in reference to ELGA:

"And ELGA has actually started to standardise all the documents in the hospital. This means that the doctor´s summary looks the same throughout Austria: it has the same structure, the same layout, it is generated in the same way. From my point of view, this is something that will help the health care system to move forward."

However, as one expert expressed, implementation was not yet completed among all stakeholders and some unsolved problems remain:

"But what's still a problem for us, is the representation and integration into local information systems."

The survey results indicated limited HIE capabilities of Swiss hospitals. This could be related to the ongoing introduction of the EPD as the following statement illustrates:

"Due to the obligation of hospitals and other in-patient providers, it will inevitably happen within a time horizon of three to five years, all in-patient providers will become part of this system."

Despite similar HIE adoption rates in Switzerland and Germany, the German experts described the situation differently and expanded on the missing incentives for collaboration across care settings:

"Our health care system has a silo mentality. […] One worries if something works in one´s own system. As soon as it comes to cross-sectorial issues, it doesn´t work because there is no incentive, no financial incentive, to do so."

They also criticized the *E-Health-Gesetz* for its one-sided focus on the out-patient sector and for its missing strategic approach.

Overall, experts from all three countries addressed similar barriers and facilitators for hospitals (Tab. 6):

- missing, insufficient, or inadequate funding,
- lack of interoperability,
- lack of willingness to cooperate across sectors,
- resistance of physicians,
- less technically advanced out-patient sector.

In total, more barriers than facilitators were mentioned in this context. Only Austrian and Swiss experts stated that the national eHealth laws serve as facilitators.

Key area c) Electronic communication with patients

The quantitative survey results suggested that the capabilities to communicate with patients were equally poor in all three countries. The interviews provided some background information on this issue. For instance, interviewees stressed that citizens did not yet make use of the power they possess – there was a clearly lacking demand on the patients´ end. One Austrian expert commented on the role of citizens within the ELGA initiatives as follows:

"I think it's true that many people may not even know that they have access to ELGA, they don't care about it."

Similarly, a Swiss hospital representative stated:

"On the other hand, we have not had a single request from a patient in recent years: Can I access my data? In this respect, interest in effective access to the data: zero. Really, zero. And that also tells me, how active are we there as a hospital at the moment? And the answer is: not at all. Because there would be no balance between effort and yield."

However, according to one German expert patients and health policy makers actually are desirable facilitators of electronic data exchange with patients:

"If you follow the treatment process and realise that hospitals send their documents via mail to the GP or that you yourself are walking around with a letter, then that is very absurd. In my opinion, the driver can only be the citizen or policy by changing certain laws."

The lacking demand and the resulting unwillingness to offer an electronic communication service was summarised by another German expert as follows:

"Well, I think there´s going to be a lot of adjustments. […] Because there are a lot of patients who are ignorant of this; or patients who don´t want it and don´t request it at all. Then, the hospitals notice: Oh, we don´t necessarily have to provide this service, people don´t want to have it anyway, and there is no one who keeps track if we are providing these things."

This statement highlighted the need of health policy to get involved and to set up mechanisms of informing the citizens.

Discussion

This study offered unique insights into key eHealth areas by combining cross-country surveys with qualitative expert interviews from the countries involved. To our knowledge, it is the first study of this kind. Many of the interview statements corroborated the survey findings and thus contributed to their validation. There is some support from other studies as well [4,7,10] that overlap in parts.

Overall, adoption rates were the highest for EPRs, followed by HIE, and electronic patient communication ranking lowest. This finding largely matched the patterns of statements on deficiencies, progress and advanced status provided by the experts. Background information provided by the experts shed light into the "whys". Among the most salient reasons given were increased expectations towards the rather new legislative frameworks in Austria as well as Switzerland that might have sparked some advances of hospital EPRs. Moreover, particularly Switzerland spends more on secondary care, thus allowing more operational flexibility. In contrast, missing incentives in Germany could have dulled the motivation of hospitals to invest in EPR systems. A clear story of the benefits can work as a strong motivator. ELGA is seen as an important lever for HIE in Austria. Due to the fact that the Swiss eHealth act, the EPDG, is a rather young law, the technical developments for HIE are still in their infancy and effects do not materialise yet. In Germany, where HIE exists only in few places, lack of real incentives and a preponderance of the out-

patient sector in the eHealth law seem to function as strong barriers. In general, wherever funding, willingness to cooperate and interoperability are missing, the odds are rather low to have HIE in place. Electronic communication with patients is not well developed in all three countries. Experts spoke about the chicken or egg problem in this context: Because patients do not voice a strong demand hospitals do not offer it. Because there are no offers from providers, patients do not know about them and do not ask for it.

Policy recommendations

Based on results of this study, some massages might be of interest for health policy considerations: All three countries are facing similar structural barriers for HIE when it comes to funding, interoperability, and willingness for collaboration across settings. Despite these communalities eHealth took a different path in Austria and Switzerland than in Germany that could be due to the design of eHealth legislation that better integrates hospitals in Austria and Switzerland. Large healthcare organisations are well known for being able to drive and spread IT innovation [4,8]. Although patient centred approaches are claimed to be pursued in all three, the citizen is not really part of the digital agenda yet, raising efforts by health policy. Effects of pertinent new laws that are under way, e.g. in Germany, need to be observed.

Findings from other countries

Especially many Nordic countries such as Finland, Norway and Denmark, that compare to Austria and Switzerland at least in terms of population, have made noticeably greater progress towards national eHealth infrastructures. Following early political commitment, all three countries approached market saturation of hospital EPR's about 5-10 years ago [11–13] and policy makers have since moved on to establish HIE capabilities across sectors and connect all citizens. Other, larger OECD countries, like the United Kingdom made some more troublesome experiences. Despite early advances through the "National Programme for Information Technology" (NPfIT) in 2004, aiming (amongst others) at digitizing secondary care, the program was terminated unsuccessfully in 2011 – essentially due to inadequate management [14]. However, in contrast to Germany, the UK has and still does acknowledge secondary care to be a crucial determinant for fostering eHealth and is thus launching new funding initiatives [15]. In essence, the different approaches seen internationally show that it is not only about doing it but also about doing it right.

Limitations

Due to the study design, some limitations need to be considered: Selection bias might occur in the survey data in light of the modest response rates (volunteer bias) and in the qualitative data (purposive sampling). Furthermore, this study provided a point in time analysis only. Follow-ups are planned and an in-depth analysis of the qualitative data is currently in progress. However, we tried to mitigate these limitations by pooling the two data sets, thereby mutually validating the findings and by blending a broad scope with contextual information.

Conclusions

The mixed methods study offered a new approach, contributed to a validation of the findings per country and provided a better insight into the overall context than with a single method alone. Hospitals and large care providing organisations must be well integrated into a national eHealth strategy before all sectors can benefit. The patients' awareness of the potential of eHealth still needs to be developed by health policy in conjunction with providers offering tangible solutions. Cross-national studies yield a good and rich basis to leverage the science-politics dialogue.

Acknowledgements

This study is funded by Lower Saxony, Germany (ZN 3062).

References

[1] A.K. Jha, D. Doolan, D. Grandt, T. Scott, D.W. Bates The use of health information technology in seven nations. *Int J Med Inform* **77**, 2008, 848–854.
[2] J. Adler-Milstein, C.E. Green, D.W. Bates A survey analysis suggests that electronic health records will yield revenue gains for some practices and losses for many. *Health Aff* (Millwood) **32**, 2013, 562–570.
[3] J.D. Halamka, M. Tripathi The HITECH Era in Retrospect. *N Engl J Med* **377**, 2017, 907–909.
[4] J. Hüsers, U. Hübner, M. Esdar, E. Ammenwerth, W.O. Hackl, L. Naumann, J.-D. Liebe Innovative Power of Health Care Organisations Affects IT Adoption: A bi-National Health IT Benchmark Comparing Austria and Germany. *J Med Syst* **41**, 2017, 33.
[5] U. Hübner, M. Esdar, J. Hüsers, J.D. Liebe, J. Rauch, J. Thye, J.-P. Weiß *IT-Report Gesundheitswesen: Wie reif ist die IT in deutschen Krankenhäusern?,* Schriftenreihe Nds. Ministerium für Wirtschaft, Arbeit, Verkehr und Digitalisierung., Hannover, 2017.
[6] W. Gall, A.-F. Aly, R. Sojer, S. Spahni, E. Ammenwerth The national e-medication approaches in Germany, Switzerland and Austria: A structured comparison. *Int J Med Inform* **93**, 2016, 14–25.
[7] R. Haux, E. Ammenwerth, S. Koch, C.U. Lehmann, H.-A. Park, K. Saranto, C.P. Wong A Brief Survey on Six Basic and Reduced eHealth Indicators in Seven Countries in 2017. *Appl Clin Inform* **9**, 2018, 704–713.
[8] D.E. Leidner, D. Preston, D. Chen An examination of the antecedents and consequences of organizational IT innovation in hospitals. *The Journal of Strategic Information Systems* **19**, 2010, 154–170.
[9] J.W. Creswell, V.L. Plano Clark *Designing and conducting mixed methods research,* Sage, Los Angeles, 2011.
[10] J. Zelmer, E. Ronchi, H. Hyppönen, F. Lupiáñez-Villanueva, C. Codagnone, C. Nøhr, U. Huebner, A. Fazzalari, J. Adler-Milstein International health IT benchmarking: learning from cross-country comparisons. *J Am Med Inform Assoc* **24**, 2017, 371–379.
[11] I. Winblad, P. Hämäläinen, J. Reponen What is found positive in healthcare information and communication technology implementation?-the results of a nationwide survey in Finland. *Telemed J E Health* **17**, 2011, 118–123.
[12] P. Kierkegaard eHealth in Denmark: a case study. *J Med Syst* **37**, 2013, 9991.
[13] V. Heimly, A. Grimsmo, A. Faxvaag Diffusion of Electronic Health Records and electronic communication in Norway. *Appl Clin Inform* **2**, 2011, 355–364.
[14] T. Justinia The UK's National Programme for IT: Why was it dismantled? *Health Serv Manage Res* **30**, 2017, 2–9.
[15] H.S. Sood, K. McNeil How is health information technology changing the way we deliver NHS hospital care*? Future Healthcare Journal* **4**, 2017, 117–120.

Address for correspondence

Prof. Dr. Ursula Hübner, Health Informatics Research Group, University AS Osnabrück, u.huebner@hs-osnabrueck.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190378

Network Analysis of Citation in Hypertension Clinical Guidelines

Youjin Park[a], Hyung Woo Kim[b], Seng Chan You[c], George Hripcsak[d,e], Han Eol Cho[f,g], Ji Hyuk Han[h], Seo Jeong Shin[a], Rae Woong Park[a, c]

[a] Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Korea,
[b] Department of Internal Medicine, College of Medicine, Institute of Kidney Disease Research, Yonsei University, Seoul, Korea,
[c] Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea,
[d] Department of Biomedical Informatics, Columbia University Medical Center, New York, NY USA,
[e] Medical Informatics Services, NewYork-Presbyterian Hospital, New York, NY USA,
[f] Department of Rehabilitation Medicine, Gangnam Severance Hospital, Rehabilitation Institute of Neuromuscular Disease, Yonsei University College of Medicine, Seoul, Korea,
[g] Department of Rehabilitation Medicine and Research Institute of Rehabilitation Medicine, Yonsei University College of Medicine & Severance Hospital, Seoul, Korea,
[h] Armed Forces Capital Hospital, Seongnam-si, Gyeonggi-do, Republic of Korea

Abstract

Recently the two most influential clinical guideline were published for diagnosing and treating hypertension in US and Europe: 2017 American College of Cardiology/American Heart Association (ACC/AHA) and 2018 European Society of Cardiology/European Society of Hypertension (ESC/ESH) Guideline. Though both of them have most in common, the differences in details between guidelines have confused many clinicians in the world. Because guidelines were evidence-based literature, through the analysis of articles cited in guidelines, these similarities and differences could be explained. Bibliometric analysis is a method of quantifying the contents of literature to analyze literature. So using the bibliometric analysis including co-citation network analysis, articles cited in guideline were analyzed. As a result, we figured out that bibliometrics can analyze the influence of the countries, authors and studies on the guidelines, which might affect on the similarities and the differences between both guidelines.

Keywords:

Hypertension; Bibliometrics; Guideline

Introduction

In clinical areas, guidelines serve as standards to diagnose diseases and to give proper treatment to patients. Guidelines rely on evidence-based literature [1], citing articles about relevant diseases and treatments, and they are continually updated based on newly identified clinical research [2]. Therefore, guidelines can be used to identify key literature that contains the most important and essential information in the clinical domain from past to present.

Recently, two most influential guidelines for hypertension, the 2018 European Society of Cardiology/European Society of Hypertension (ESC/ESH) Guideline and the 2017 American College of Cardiology/American Heart Association (ACC/AHA) guideline have resulted in vigorus debate because of common ground in a broad view and differences in details [3]. Since both guidelines have been established on thorough review of hugh amount of high-quality evidences, the complex network of cited articles might reflect and help to understand the inherent intention of the guidelines.

Bibliometric analysis is a method of quantifying the contents of literature through statistical methods to analyze literature [4]. Co-citation analysis is one of the bibliometric analysis, which focuses on the documents cited, by calculating the frequency of two documents that are co-cited by specific documents [5]. We conducted bibliometric analysis using articles which were cited in guidelines to identify the contributions according to the countries, authors, cited articles [6]. We can present the network between articles' references, co-citation network [7,8].

To our knowledge, there has been no previous evidence of a comparative bibliometric analysis of articles cited between two clinical guideline for the same theme. We explored the impact of the articles that make up the guideline and how these influences explain the similarities and differences between two guidelines.

Methods

Extracting data

Data was gathered by using the Web of Science (WoS) database. The references cited in the 2018 ESC/ESH hypertension guideline and the 2017 ACC/AHA hypertension guideline were searched. Published years, author(s), countries, and cited references were obtained. 629 articles for 2018 European guideline and 991 for 2017 US guideline were included.

Bibliometric analysis

The countributions of countries and author participation in guideline production were obtained with the open-source Bibliometrix R-package (http://www.bibliometrix.org/) [9]. We conducted the co-citation analysis to identify highly cited references and productive authors. The results of co-citation analysis was presented as network visualization maps using the VOSviewer version 1.6.9 (Leiden University, Leiden, Netherlands) techniques [10].

Results

Through the WoS, total 609 references for the European guideline and 860 for the US guideline were extracted out of the total cited papers 629, and 991, respectively. Articles cited in the 2018 European guideline were published from 1950 to 2018, and articles cited in the 2017 US guidelin were from 1967 to 2018. Guidelines tended to cite more recent referneces (Figure 1a,b).

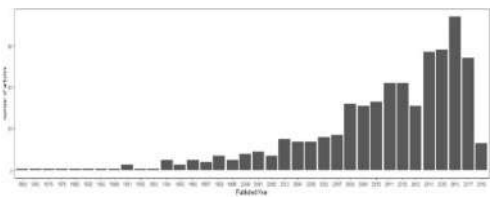


Figure 1a – Annual articles cited in the 2018 European guideline

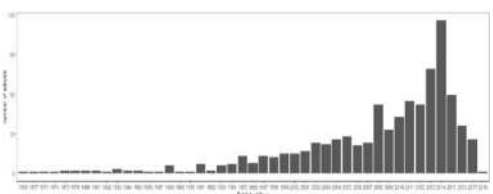


Figure 1b – Annual articles cited in the 2017 US guideline

Table 1 lists the top ten countries that contributed to each guideline using citation analysis. The US was the country with the highest production in both guidelines. In the European guideline, the articles published by the US was total 133, 22.2 %, followed by the UK 102 articles, 16.9 % and the Italy 99, 16.4 %. In the US guideline, the articles published by the US was 395, 47.0 %, followed by the UK (68, 8.1 %), Australia (48, 5.7 %), Italy (37, 4.4 %).

Table 1 – The contribution of countries

Europe guideline		US guideline	
Country	Publication, n (%)	Country	Publication, n
USA	133 (22.1)	USA	395 (47.0)
UK	102 (16.9)	UK	68 (8.1)
Italy	99 (16.4)	Australia	48 (5.7)
Canada	29 (4.8)	Italy	37 (4.4)
France	24 (4.0)	Canada	28 (3.3)
Germany	22 (3.6)	China	24 (2.9)
Australia	20 (3.3)	Netherlands	24 (2.9)
Spain	20 (3.3)	Belgium	18 (2.1)
Belgium	19 (3.2)	Japan	17 (2.0)
Netherlands	19 (3.2)	Spain	17 (2.0)

The contributions of authors published articles cited in each guideline were analyzed, and the top 10 most cited author were listed in the table 2. In the European guideline, G. Mancia published 83 articles and it took 13.6 % in total articles cited in the European guideline. A. Zanchetti and B. Williams followed next published 44 (7.2 %), 36 (5.9 %). In the US guideline, P.K. Whelton contributed most, 40 articles 4.7 %. B. Neal, M. Woodward were 2nd and 3rd most contributed author, 23 (2.7 %), 23 (2.7 %). In case of G. Mancia, ranked 10th in the US guideline, published 19 articles (2.2 %).

Table 2 – The contribution of authors

Europe guideline		US guideline	
Author	Publication, n (%)	Author	Publication, n (%)
G. Mancia	83 (13.8)	P.K. Whelton	40 (4.8)
A. Zanchetti	44 (7.3)	B. Neal	23 (2.7)
B. Williams	36 (6.0)	M. Woodward	23 (2.7)
G. Parati	35 (5.8)	J. He	22 (2.6)
J. Redon	26 (4.3)	S. Oparils	22 (2.6)
S.E. Kjeldsen	23 (3.8)	J. Chalmers	21 (2.5)
S. Yusuf	22 (3.6)	L.J. Appel	20 (2.4)
J.A. Staessen	21 (3.5)	J.A. Cutler	20 (2.4)
G. Grassi	20 (3.3)	J.T. Wright	20 (2.4)
B. Dahlof	19 (3.2)	G. Mancia	19 (2.3)

In co-citation network analysis, the most frequent appeared author in articles cited in the European guideline's references was G. Mancia. His citation count was nearly 3 times more than the second highest citation count, P. Verdecchia. The thrid was S. Yusuf, similar with P. Verdecchia (Figure 2a). In the US guideline, the first most frequently cited author, G. Mancia, and the second most, A.V. Chobanian. G. Mancia and A.V. Chobanian had almost similar citation counts. The third most frequently cited author was S. Yusuf (Figure 2b).

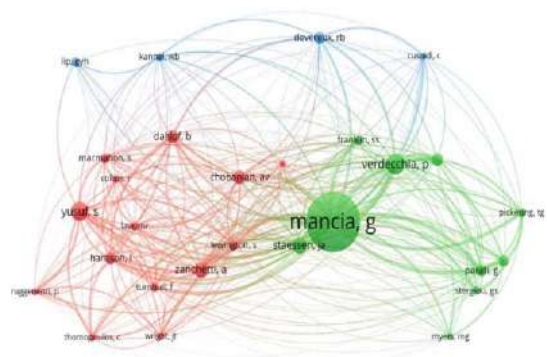


Figure 2a – co-citation network map presenting cited authors in references of the 2018 European guideline's references

The size of each circle reflects the number of citations an article has received. So citation count could be estimated and compared by circles' size. And circles that are located close to each other in the visualization tend to be more strongly related

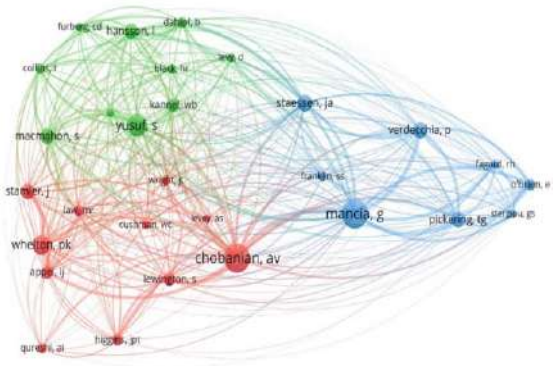


Figure 2b – co-citation network map presenting cited authors in references of the 2017 US guideline's references

Except for the previous versions of hypertension guidelines, the most frequently cited study in the European guideline's references was *S. Lewington, 2002, LANCET* which was meta-analysis. The next most frequently cited study was Syst-Eur trial which targeted people aged more than 60, followed by *F. Turnbull, 2003, LANCET* (meta-analysis), HOT trial, HOPE trial, ACCORD-BP trial, *M.R. Law, 2009, BMJ* (meta-analysis), SHEP trial, HYVET trial, SPRINT trial. In the US guideline, except for previous version of guidelines, *S. Lewington, 2002, LANCET* was the most frequently cited study same as the European guideline. The next was the HOPE trial, followed by *S. MacMahon, 1990, LANCET* (meta-analysis), ALLHAT trial, *M.R. Law, 2009, BMJ* (meta-analysis), *F. Turnbull, 2003, LANCET* (meta-analysis), SEHP trial, Syst-Eur trial, HOT trial, PROGRESS trial, ACCORD-BP trial. The list of top 15 studies most higly cited were almost same between two guidelines except for 4 studies.

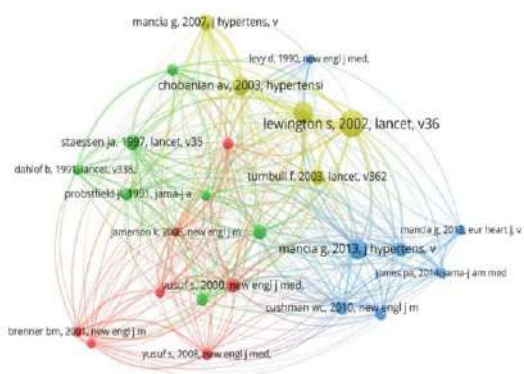


Figure 3a – co-citation network map present cited articles in references of the 2018 European guideline's references

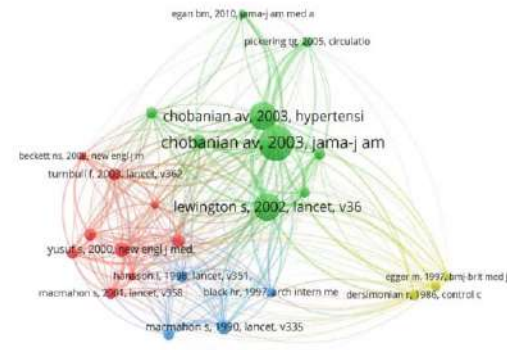


Figure 3b – co-citation network map present cited articles in references of the 2017 US guideline's references

Discussions

Our study figured out that the results of the bibliometric analysis using articles cited in the guideline had similarities and differences between two clinical guidelines. The result of measuring published countries of articles citied in the two guidelines was that the US guideline chose the articles published in the US, and the european guideline chose the articles published in the US, UK, Italy evenly. Because Europe consists of various countries, the European guideline had tendency to choose multi-country published articles.

The list of top 10 most contributed author, there were no authors that ranked at the same time except for G. Mancia. In the terms of the contributed authors, the composition of the two guidelines looked different. However, there were more similarities between two guidelines in the influence of the author obtained by author co-citation analysis. It had something in common with the results of reference co-citation analysis. Studies most cited in the articles used in the guideline were almost same in the Europe and the US guideline. Compared with the 15 highest cited articles in both the European and the US guideline, the list of studies were same except 4 articles. The bases of the two guidelines's evidences were similar. It could explain the European guideline and the US guideline were similar in general. And these similarities showed that the contents of the actual guidelines are similar to each other. For example, *S. Lewington, 2002, LANCET* was the most cited study in both guidelines except for previous guidelines. It is meta-analysis which said even in the normotensive people lowering systolic and diastolic blood pressure lowers the risk of vascular death. It could reflect the tendency of the two guidelines treats hypertension more tightly. The SHEP trial and the Syst-Eur trial which had high citation count in both guidelines could explain to set more strict target blood pressure for elder people who aged and over 65 years in the 2018 European guideline and the 2017 US guideline. Also, the HOPE trial ranked high in both guidelines, and it could explain the reason why angiotensin converting enzyme inhibitors (ACE inhs) remain first line drug in both guidelines. There were also differences between the results of reference co-citation analysis. The ALLHAT trial was ranked high in the US guideline, but not in the European guideline. It could make differences between two guidelines' drug choice. While the European guideline recommended single pill combination consisting of ACE inhs or angiotensin receptor blockers (ARB) plus Calcium channel blockers (CCB) or thiazide-like diuretics to start anti-hypertensive medication, the US guideline recommended to start with thiazide-like diuretics ahead of other

classes. In this way, analyses of articles cited in guideline could explain the similarity and differences between the European guideline and the US guideline.

But the influence of previous guidelines could not be excluded. JNC 7, 2013 European guideline, 2007 European guideline ranked high in the results of reference co-citation analyses. Actually studies of hypertension are inevitably affected by the previous guidelines, because their settion followed previous guidelines. It could cause overestimation of the power of some articles or authors. The result of author co-citation analysis was that G. Mancia had the most frequently cited count in both guidelines. Espetially in the Europe guideline, his cited count was almost three times more than P. Verdecchia. This overwhelming number of citation count was seen because G. Mancia was overestimated by previous guidelines.

Conclusions

In the bibliometric analysis, the contributions of countries in the European guideline and the US guideline were found. There were differences the list of most highly contributed author between two guidelines, but almost same in the list of the most highly cited in the guideline's references. Also, the studies highly cited in the guideline's references were almost similar in both but slightly different. In these results, we could find two guidelines looked different but its bases were most in common. Through the co-citation analysis, we figured out how much the whose or which articles contributed to the references used in guidelines. These contributions affected guideline's direction. As a result, analysis of references used in guideline could explain the similarity and differences between two guidelines.

Acknowledgements

This work was supported by the Bio Industrial Strategic Technology Development Program (20001234) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea) and a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea [grant number: HI16C0992].

References

[1] E.-B.M.W.G.J. Jama, Evidence-based medicine. A new approach to teaching the practice of medicine, *The Journal of the American Medical Association* 268 (1992), 2420.

[2] P. Shekelle, M.P. Eccles, J.M. Grimshaw, and S.H.J.B. Woolf, When should clinical guidelines be updated?, *British Medical Journal* 323 (2001), 155-157.

[3] P.K. Whelton and B. Williams, The 2018 european society of cardiology/european society of hypertension and 2017 american college of cardiology/american heart association blood pressure guidelines: More similar than different, *JAMA* 320 (2018), 1749-1750.

[4] N. De Bellis, *Bibliometrics and citation analysis: from the science citation index to cybermetrics*, scarecrow press, 2009.

[5] K.W. Boyack, R.J.J.o.t.A.S.f.I.S. Klavans, and Technology, Co citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?, *Journal of the American Society for Information Science and Technology* 61 (2010), 2389-2404.

[6] A.F.J.T.A.-T. Van Raan and Practice, The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments, *Theorie und Praxis* 1 (2003), 20-29.

[7] H.J.J.o.t.A.s.f.I.S. Small, Visualizing science by citation mapping, *Journal of the American Society for Information Science* 50 (1999), 799-813.

[8] N.J. Van Eck and L. Waltman, VOS: A new method for visualizing similarities between objects, in: *Advances in data analysis*, Springer, 2007, pp. 299-306.

[9] M. Aria and C.J.J.o.I. Cuccurullo, bibliometrix: An R-tool for comprehensive science mapping analysis, *Journal of Informetrics* 11 (2017), 959-975.

[10] N.J. van Eck and L.J.S. Waltman, Software survey: VOSviewer, a computer program for bibliometric mapping, *Scienctometrics* 84 (2010), 523-538.

Address for correspondence

Rae Woong Park, MD, PhD

Department of Biomedical Informatics, Ajou University School of Medicine, 164, World cup-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 16499, Republic of Korea

Tel: +82-31-219-4471, Fax: +82-219-4472

E-mail: veritas@ajou.ac.kr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190379

Using Simulation Modeling to Inform Policy Makers for Planning Physician Workforce in Healthcare System in Croatia

Danko Relić[a], Kristina Fišter[a], Jadranka Božikov[a]

[a] *Andrija Štampar School of Public Health, University of Zagreb School of Medicine, Zagreb, Croatia*

Abstract

The objective of this paper is to show how a simple but powerful simulation model can be build up using standard spreadsheet program and used to simulate future, needs and supply of physicians in order to inform policy makers at national level when deciding on enrollment to medical schools and immigration quotas for physicians. The Republic of Croatia is facing a serious shortage of physicians in the healthcare system and simulation results have shown that the gap between needs and supply will even increase if current enrollment qoutas to medical schools would persist. Increasing enrollment quotas, adjusting immigration policy, re-directing physicians from other professions to the healthcare system, task shift and skill mix options are just some of the measures needed to be taken promptly in order to prevent a huge deficit of physicians in the future. Simulation modeling is certainly a method for predicting changes within healthcare systems with a possibility to examine multiple different scenarios and suggest interventions.

Keywords:

Physicians; Decision Support Techniques; System Analysis

Introduction

Imbalance in the health workforce is a major concern in both developed and developing countries [1]. Shortage of proper policy planning and human resources management in many countries has resulted in an imbalance of needs and supply with multiple effects on employees in health care as well as on population health. Although the determination of the ratio between number of health workers and population size including age structure has become an aim of national policy on the human resources development in health care, in most cases remains a gap between plans and their realization [2]. The number of physicians is continuously changing because new physicians enter the system almost daily, and many in different ways come out of it. During 10-year period starting by 1 January 2007 the number of physicians in Croatia increased from 12,065 to 14,394 (i.e. by 19% or from 2.8 to 3.5 per 1,000 inhabitants) while in the same period number of inhabitants decreased by 4%. In the same time population ageing continued and a shift of age distribution towards older ages resulted in the increased needs for health care services and surge of their use [3].

The increase of the number of elderly people in Croatia in the upcoming period could become one of the major generators of increasing need for physicians, despite the expected stagnation or even decline in the total population. Folowing the country's accession to the European Union (EU) in 2013, the outward migration of health workers to other Member States has increased and contributed to the workforce shortage in the Croatian health care system. The majority of physicians are concentrated in major cities and their density varies regionally within country which is administratively organized in 21 counties. Some counties do not have enough physicians in primary health care, particularly in remote poorly populated areas and islands. In some counties, more than 35% of physicians are aged 55 and over, what raises concerns about their number in the near future when they will be retired, especially having in mind that those counties are less developed [3].

With around 3.3 full-time physicians per 1,000 inhabitants including those without specialization and interns, Croatia was at the end of 2016 below the EU average being 3.6 physicians per 1,000 inhabitants, despite the fact that the number of physicians in the last three decades has increased by about 57%. A large increase in number of physicians was recorded by mostly developed EU member states, while among Central and Eastern European countries an increase in number of physicians was less pronounced, especially after their accession to the EU (e.g. Estonia and Romania) [3].

Croatia has kept publicly funded health care services accessible to its population. Besides concerns about financial sustainability of health care system due to large share of expenditures spend on pharmaceuticals and low efficiency of hospital sector accumulating debts, shortage of health care personel is seen as a main threat to the sustainability of health care system, particularly after EU accession. Number of physicians, and particularly nurses, is below the EU average and thus the strategic planning of human resources for health is urgently needed [4].

Simulation modeling has been used for a long time primarily for understanding of complex phenomena but it is also a suitable tool to present a problem and enhance a systematic debate between the stakeholders in many areas of human activity including healthcare. It allows to imitate a complex system behaviour in situations wich are unfit for experiment and prediction of future events under assumed circumstances. Generally, simulation models have been successfully used in research, instruction and training as well as an aid to decision making why they have been long ago named also policy simulations. The main advantage of policy simulations is their capacity to deal with considerable complexity of the system and the ability to explore "What if …?" type of questions [5].

There are comercialy available application-oriented simulation packages but a special program can also be developed in any programing language including standard spreadsheet as a platform for representing model structure with a set of formuals, performing simulation experiments and charting results [6].

The objective of this paper is to show how a simple but powerful simulation model can be build up using standard spreadsheet program and used to simulate future needs and supply of physicians in order to inform policy makers at national level when deciding on enrollment to medical schools

and immigration quotas for physicians entering the public healthcare system.

Methods

Two main types of simulation models are discrete events simulation and continuous simulation also known as system dynamics which proved to be powerful method not only when used in research but also as a tool in decision making providing necessary information to policy makers [6].

Already in late 1980s, an original method based on the use of standard spreadsheet program was developed in authors' Department at the Andrija Štampar School of Public Health and used to simulate population dynamics of infectious diseases. The approach was general and method enables simple and intuitive use of simulation models including model implementation, changes of parameters and initial values in an easy manner, performing simulation experiments and display of results in tabular and graphical form [7]. At that time the most used spreadsheet program was Lotus 1-2-3 like nowadays is Microsoft EXCEL. It proved to be powerful tool for development of system dynamics models and running of simulations. Same approach was employed in simulation modeling of health personnel dynamics aimed for decision making on enrollment policy to medical schools and higher education programs in health sciences in Croatia [6].

The working group established in 1989 by the Association of higher education schools in medicine and health sciences as a consultative body at national level (at that time of the Federal Republic of Croatia), whose founding was initiated by discussions on enrollment quotas at schools of medicine and proposals for their reduction, accepted the proposal to use the simulation modeling as an aid for decision making. A simulation model was developed with the aim to simulate needs and supply of physicians as well as other health professionals (nurses, laboratory technicians, radiology technicians and other). A series of simulation experiments has been performed with different assumptions about the number of enrolled students and other relevant aspects (e.g. introduction of private practice, employment outside health care and differences in the presumed number of health care users), and the results served as a basis for decisions on enrollment policy [8,9]. Based on that results authorities adopted a decision to decrease the enrollment quotas starting from the year 1993 because at the time there were a surplus of physicians resulting in rise of their unemployment rates.

Adaptation of the previously developed model structure and parameters has been made in order to use the simulation again as a tool to inform decision makers in current situation when the Republic of Croatia is faced with an opposite situation, a huge lack of physicians.

Simulation model is structured as multicompartmental (or multistate) model in which variable values are numbers of physicians in one-year age groups and simulation is run by calculating changes in one-year steps (calculated values for successive years are stored in rows). We can consider data as a table (or matrix) with rows being consecutive years (here 2016-2040) and columns are containing number of persons of the same age (25-65). We can denote these values with $X_{i,j}$, i=2016, …,2040; j=25, …, 65 having in mind that the youngest age at which one can graduate in medicine is 25 years and 65 is age for retirement. Thus model simulation correspond to calculation of state variables (or compartments) and intuitively results in a move of one compartment in diagonal direction (from left to right means ageing for one year, from one row to the next one is move to the next year value). Of course, it is not only about moving because of ageing, but other different calculations are taking place (mortality, disability, emigration and immigration as well as flows reproducing graduation and retirement). Model parameters (mortality and disability rates, immigration and emigration and other parameters and input values) are entered in different areas of spreadsheet and used in calculations.

Parameters' estimation

Population projections by 2040 were taken from the publication of the Croatian Bureau of Statistics [10]. Currently there are 3.3 full-time physicians employed in health care system per 1,000 population while in the more developed EU Members States there are 3.6 physicians per 1,000 inhabitants [4]. It can be noticed that Croatia lags behind the European average that we strive for. The target value is defined as the EU average (3.6 practicing physicians per 1,000 population).

In developing this model and assessing the needs for physicians in Croatia, it is important to consider the following factors: the number of physicians working in the healthcare system and their age distribution, the number of inhabitants, the number of enrolled students at the medical schools and their years of studying span, as well as the mortality, disability and emigration rates of physicians. The rate of emigration is particularly significant after Croatia joined the EU, but one has to take in account also the employment opportunities for physicians within the country but outside the healthcare system (for example in pharmaceutical industry, insurance companies or even outside the biomedical profession). We estimated that about 11% of physicians are currently employed outside the healthcare system and it is expected that this percentage will rise up to 15% by 2040. According to available data there are 14,394 practicing physicians and it is estimated that another 1,600 are out of the health system [3].

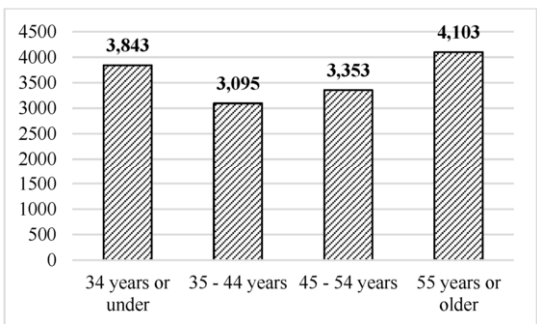


Figure 1 Age distribution of Croatian physicians

Age distribution of physicians employed in Croatian healthcare system is shown in Figure 1. Data are for the year 2016 and it is obvious that the age group 55 years and older is the largest one and having in mind that according to present regulations retirement age is 65 years, it means that in the next 10 years 28.5% of working force is going to be out of employment.

There are four medical schools in Croatia and medical study lasts 6 years. We estimated that about 5% of enrolled students never complete their studies while 80% of students graduate six years since enrollment, 10% after seven years, 3% after eight years, while approximately 2% of students complete their studies at the age of 28 years i.e. after nine years of study.

Estimated rates of mortality and disability are spanning from 0.12% for physicians at 25 years of age up to 2.39% for those aged 61 to 65. Apart from mortality and disability, which affect drain outside profession, migration outside Croatia is also supposed to be significant. It is estimated that the number of

those who will leave the country will increase. Simulations were made with the assumption that 3% of those aged 25 to 27 years will leave the country every year, 2% of those aged 28 to 29 years, and 1% of physicians aged 30 to 45 years. No outflow is expected for those over 45 years of age who are already acknowledged in their profession and have established their social and family surrounding.

Simulation was made for three potentially achievable scenarios: (1) increase of the total enrollment quota at the existing four medical schools by 50 students per year starting from 2020; (2) introduction of immigration quota for of 50 physicians per year and (3) combination of previous two scenarios - increase of the enrollment quota by 50 students per year and immigration of 50 physicians per year. Year 2020 has been taken as the earliest year when potential changes could be implemented with respect to all political and administrative procedures.

Results

In the case of an increased enrollment quota by 50 medical students starting from the year 2020 (a total of 650 instead of current 600 freshmen per year at the national level) there will be a huge lack of 3,511 physicians in the health care system in 2025 if we want to reach most developed EU countries which have 3.6 physicians per 1,000 population; in 2040 there will be shortage of 1,769 physicians. Second scenario with introduction of immigration quota of 50 physicians per year would give a little bit faster raise of supply with a shortage of 3,467 physicians in 2020 and 1,542 in 2040. Third (combined) scenario would result in the growing number of physicians from the year 2020 and the lack of 1,029 physicians in 2040. Not only with current enrollment quota, but even with higher quota of 650 freshmen per year it is obviously not possible to meet the needs for physicians in the Republic of Croatia by the year 2040.The results of all three scenarios are presented in one chart (Figure 2).

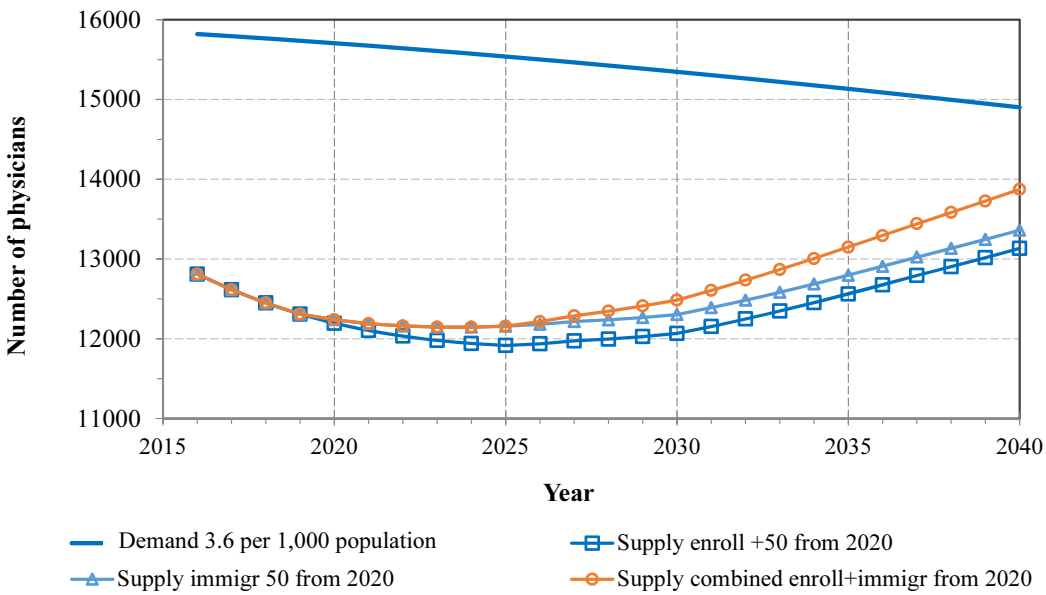


Figure 2 Results of simulation experiment

Table 1 Predicted number of physicians in the Republic of Croatia – results of simulation with 650 freshmen medical students and immigration quota of 50 physicians per year

Year	Enrolled	Graduated	Health care	Total	Drain				
					Total	%	Retirement	Emigration	Death+Disability + Emigration
2020	650	570	12,239	13,930	633	4.53	313	109	320
2025	650	570	12,157	14,066	547	3.90	236	124	311
2030	650	618	12,485	14,688	524	3.60	212	138	312
2035	650	618	13,148	15,468	494	3.23	176	149	318
2040	650	618	13,873	16,321	492	3.04	164	154	328

Discussion

Use of spreadsheet program for simulation modeling

Although the application-oriented programing languages and packages for simulation modelling appeared very early and became commercially available already during 1980s and 1990s, different software tools can be used to represent structure of simulation model and perform simulation experiments. Many applications were developed and used within certain academic or other institutions, among them widely used standard spreadsheet programs like Microsoft EXCEL proved to be a suitable platform for development and use of simulation models of both type's system dynamics and discrete events simulation [6, 11].

"Spreadsheet simulation" refers to the use of a spreadsheet as a platform for representing simulation models and performing simulation experiments. Despite nowadays advanced commercial program packages for modeling and simulation would be a method to choose, standard spreadsheet programs as a tool for model implementation and simulation run is worthy to use due to many reasons. Spreadsheets are widely used by many people for different tasks on everyday basis and the spreadsheet's table structure allows the developer to organize the computations and outputs in a natural and intuitive manner. A large number of functions to do mathematical, statistical, database, date/time, financial and other calculations are available and, very important, it is easy to make charts and present simulation results either in table or graphical form. It is very important that end users are either able to develop simulation models by themselves or understand and use models developed by others, control parameters and perform simulation experiments [11].

Informativeness of simulation results for policy makers and applicability to other circumstances and countries

Although the enrollment quotas have been increased at Croatian medical schools since the year 2009, this increase is not sufficient to meet the current needs for physicians. Unfortunately, this increase in quotas was not followed by the adequate staff and financial needs of medical schools. As we see in the projections made in this research - if the enrollment quotas were increased by 50 students per year, what is challenging in an organizational and financial way; it would be insufficient if Croatia were to reach the current EU average of 3.6 physicians per 1,000 inhabitants.

In recent years the interest of students in studying medicine is decreasing, there are less and less applicants competing for one enrollment place [3]. Reasons may be different, from the quite a long lasting and demanding study program, difficulties in getting the desired specialization at the desired place upon graduation to the better opportunities in other professions. The diploma of physician still does not provide income and the position adequate to the invested efforts and time. The fact is that most physicians are still at the beginning of their careers in their middle age, while other professions are at the peak of their careers.

Since increasing quotas is not an only option that can contribute good results, a potential option is a better organized and coordinated immigration policy - recruitment of physicians trained in other countries to whom work environment and life in Croatia are acceptable. Since Croatia joined the EU (1st July 2013) till the end of 2016 only 30 foreign-trained physicians have been employed in Croatia. The largest number came from Macedonia (32.3%), Bosnia and Herzegovina (19.4%), Serbia (16.1%), Syria (9.7%) and the rest of physicians migrated from Slovakia, Montenegro, Lithuania, Russia, Slovenia and Kosovo [3]. Current quotas for physician's immigration in Croatia are negligible. A significant number of domestically trained physicians work outside of their profession (around 1,600) so another possibility might be to return some of them to the healthcare system. A harmonized combination of increasing enrollment quotas for medical students and increasing immigration of physicians will lead to an EU average of over the next 20 years.

There is a problem of organization and insufficient planning of healthcare staff in Croatia visible in the fact that there are 116 physicians per 100,000 inhabitants at the tertiary level of health care, 107 at the secondary level and 92 physicians in primary health care. This is also manifested by a huge number of overtime working hours and imbalanced geographical distribution of physicians [3].

It is necessary to take into account all that provide skill mix – the mix that produces the maximum number of health care services at a given quality and cost – more health care services are going to be accessible and affordable to populations seeking care and task shifting - a policy option that should be considered to help achieve productive efficiency and provide access to services that otherwise might not be available. A more productively efficient skill mix will partially dampen the effect of health workforce needs-based shortages [12].

According to estimates of population trends in Croatia, reduction in the number of inhabitants is forecasted along with population aging, what means that the number of patients and those who need health care will increase.

Other groups of researchers also used system dynamics modeling for the forecasting of needs/demand and supply of physicians at national or regional level. Similarly to Croatian situation in the late 1980s, Japan had also reduced enrollment quotas to medical schools due to surplus in number of physicians and has faced absolute and relative shortage of physicians twenty years later. Of course, Japanese absolute numbers are more than tenfold higher than Croatian ones, but the problem and approach employed in order to inform policy makers are the same [13]. Majority of other problems inherent to healthcare system are similar: maldistribution of physicians at both the regional and departmental level and others. Regarding calculation of needs/demand similar approach was taken: the number has been evaluated by a comparison with OECD average. Spain has also gone from a surplus to a shortage of physicians in a very few years what initialed a development of a system dynamics model for simulation of supply and demand/need of specialist of 43 medical specialties [14]. Problems and debates within the medical profession and in public at general are very similar in many countries. Besides planning of enrollment quotas to medical schools and immigration quotas for foreign trained physicians, an important question that needs to be tackled is how to plan specialist education [2, 14]. Another example is simulation model of the allocation of residency grants for medical specializations given from regional and national level [15]. System dynamics model was developed with the aim to project and optimize supply of medical specialist under three different supply scenarios in Emilia-Romagna Region in Italy. Demography, service utilization rates and hospital beds were taken into account as demand drivers while a potentially effective assignment of medical specialization grants for each year in 20-year period was targeted.

System dynamics modeling is an excellent method for simulation of health workforce supply and demand. A proactive approach to the planning and management of health workforce, involving continued monitoring, might facilitate more frequent and less dramatic adjustments to needed supply. Many issues

and challenges of public health can also be presented and analyzed using simulation modeling [16].

Limitations

This study did not take into account the number of medical students in English study programs (Medical Studies in English) that are running by the Croatian medical schools. Students enrolling in English programs are largely foreigners who plan their employment after graduation in their homelands.

Furthermore, the study did not take the geographical distribution of physicians and it has not targeted adequate distribution according to the levels of health care – these are very important determinants of quality planning of a public healthcare system.

Although official data and projections of the Croatian Bureau of Statistics on the population trends have been used, we are already witnessing of a decline in the total population that happened by the year 2018 which is far greater than it was forecasted in 2015. Population aging has not been taken into account, but only projected total number of inhabitants.

Conclusions

The Republic of Croatia is facing a serious shortage of more than 3,000 physicians in the healthcare system.

Increasing enrollment quotas, defined and agreed immigration policy, re-directing physicians from other occupations to the healthcare system, task shift and skill mix options are just some of the measures that need to be taken now in order to improve the situation in the near future.

Simulation modeling is certainly a method for predicting flow of physicians and other health professionals within healthcare system capable to simulate different scenarios that can potentially take place in future. Standard spreadsheet program proved to be powerful platform for model implementation, simulation experiments and presentation of results.

References

[1] Zurn P, Dal Poz MR, Stilwell B, Adams O, Imbalance in the health workforce, Hum Resour Health2(1) (2004),13. doi: 10.1186/1478-4491-2-13

[2] Relić D, Development of a Model for Planning Specialist Education of Medical Doctors in Croatia, Stud Health Technol Inform 228 (2016),798-800. doi: 10.3233/978-1-61499-678-1-798

[3] Čipin I, Smolić Š, Vlah Jerić S, Demographic atlas of Croatian practicing physicians, Croatian Medical Chamber, Zagreb, 2017.

[4] Organisation for Economic Co-operation and Development (OECD), Health at a glance. OECD indicators, OECD Publishing, Paris, 2017. doi: 10.1787/health_glance-2017-en

[5] Schultz RL, The use of simulation for decision making, Behavioral Science 19(1974),344-350.

[6] Bozikov J, Relic D, Dezelic G, Use of System Dynamics Modeling in Medical Education and Research Projects, Stud Health Technol Inform 247(2018),830-834. doi: 10.3233/978-1-61499-852-5-830

[7] Bozikov J, Dezelic G, A microcomputer method for continuous system simulation in health care, Computer Methods Programs Biomed 34(1991),17-25. doi: 10.1016/0169-2607(91)90078-8

[8] Jakšić Ž, Božikov J, Kljajić K, Pokrajac N, Medical school admission: Zagreb experience, Croat Med J, 34(1993),66-77.

[9] Budak A, Božikov J, Are there too many doctors in Croatia?, Lijec Vjesn 115(1993),261-267.

[10] Grizelj M, Akrap A, Population projections of the Republic of Croatia 2004-2051, Croatian Bureau of Statistics, Zagreb, 2006.

[11] Seila AF. Spreadsheet simulation. In: Kuhl ME, Steiger NM, Armstrong FB, Joines JA eds, Proceedings of the 2005 Winter Simulation Conference, Orlando: IEEE Computer Society, 2005;33-40. doi: 10.1109/WSC.2005.1574237

[12] Fulton BD, Scheffler RM, Sparkes SP, Auh EY, Vujicic M, Soucat A, Health workforce skill mix and task shifting in low income countries: a review of recent evidence, Hum Resour Health 9 (2011),1. doi: 10.1186/1478-4491-9-1

[13] Ishikawa T, Ohba H, Yokooka Y, Nakamura K, Ogasawara K. Forecasting the absolute and relative shortage of physicians in Japan using a system dynamics model approach, Hum Resour Health, (2013);11:41. doi: 10.1186/1478-4491-11-41

[14] Barber P, López-Valcárcel BG, Forecasting the need for medical specialists in Spain: application of a system dynamics model, Hum Resour Health, (2010),8-24. doi: 10.1186/1478-4491-8-24

[15] Senese F, Tubertini P, Mazzocchetti A, Lodi A, Ruozi C, Grilli R. Forecasting future needs and optimal allocation of medical residency positions: the Emilia-Romagna Region case study, Hum Resour Health, (2015);13:7. 10.1186/1478-4491-13-7

[16] Homer JB, Hirsch GB. System dynamics modeling for public health: background and opportunities, Am J Public Health, (2006); 96(3): 452-458. doi: 10.2105/AJPH.2005.062059

Address for correspondence

Danko Relić, Andrija Štampar School of Public Health, Rockefeller St. 4, 10000 Zagreb, Croatia, e-mail: danko.relic@mef.hr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190380

Towards a Systematic Construction of a Minimum Data Set for Delirium to Support Secondary Use of Clinical Routine Data

Michael Schaller[a], Werner O. Hackl[a], Bogdan Ianosi[a,b], Elske Ammenwerth[a]

[a] *Institute of Medical Informatics, UMIT - University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria*
[b] *Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria*

Abstract

Patient safety is an important topic but non-trivial to measure, as it comprises very different phenomena. One important example of patient safety issues is delirium. Many approaches for the detection and prediction of delirium are described in the literature. However, additional effort is often needed for a comprehensive data collection and for the avoidance of potential biases. To systematize a process for the detection and prediction of delirium reusing available clinical routine data, we aim to develop a minimum data set (MDS) for delirium. By combining a top-down and bottom-up approach, we compiled a comprehensive delirium map containing potential delirium elements based on evidence. The alignment with clinical routine data led to a specific version of an MDS for delirium and revealed that most of the delirium elements could be identified within available nursing routine data.

Keywords:
Delirium, Data Collection, Patient Safety

Introduction

Patient safety is an important topic receiving increasing attention [1,2]. The World Health Organization defines patient safety as the reduction of risk of unnecessary, preventable harm to an acceptable minimum [3]. Issues of patient safety and patient safety research have a big impact on the health care sector. However, patient safety features a high complexity characterized by a whole series of interacting factors and dimensions [1,4].

Hence, measuring patient safety is not a trivial task. Several methods for the detection of patient safety incidents are already described in the literature, e.g., patient safety indicators for the detection of adverse events, chart reviews, incident reporting systems, or trigger tools [2,4,5].

However, various weaknesses of those methods and instruments are evident [2,4–7]: Incident reporting systems tend to have a lack of sensitivity and an increased risk of (reporting) bias; manual review of patient records (as necessary for global trigger tools) requires a vast amount of resources; patient safety indicators based on administrative data are often biased, as documentation follows accounting and billing instead of clinical purposes.

In order to overcome some of those disadvantages, especially to reduce the resources needed to measure patient safety accurately, the secondary use of available clinical routine data within an institution seems to be a helpful approach [2,8].

Secondary use of clinical routine data is defined as usage of available clinical routine data for other purposes than patient care, e.g., for health care research or quality management [9,10].

Examples of secondary use of clinical routine data can be found in the literature [9,11]. However, secondary use projects exhibit a high level of complexity and issues to be resolved, such as inadequate standardization and poor data quality. This demands the structured and goal-oriented planning of such projects. The SPIRIT framework (Systematic Planning of Intelligent Reuse of Integrated Clinical Routine Data), which is designed for systematic planning of secondary use of clinical routine data, can help [9].

Every incident and measuring method requires a clearly defined set of well-described specific data, and available clinical routine data often diverge from one institution to another. Thus, it seems necessary to individually align patient safety issues of interest with the available clinical routine data. For such an alignment, the issues of interest must first be selected. Then a priori knowledge and expert knowledge on all dimensions of these issues must be assessed to define all relevant data elements for the respective issues. These elements would constitute the theoretic “maximum data set”.

In a following step, the clinical routine data set has to be screened for these elements. If available, the routine data elements are included in a special data set which can be understood as the “greatest common divisor” between all available routine data and the maximum data set. It thus represents the “minimum data set” for the patient safety incident of interest. Implicit is that minimum data sets can vary between institutions.

In this study, we want to demonstrate how this approach is applied to a very important patient safety issue: delirium. Delirium is an acute disturbance of mental functioning which occurs quite often within the population of older hospitalized persons. It is associated with an enormous burden to the affected people and to the health care systems. It is regarded as preventable in up to 40% of the delirium events and is considered to be an important patient safety indicator [12,13].

Delirium features a high complexity and is often overlooked by clinical staff. It is estimated that up to two out of three of all delirium cases are overlooked in hospitals despite available assessment and screening instruments [12–14].

Several predictive models for the detection of delirium are proposed in the previous studies [15], some of which are based on routine data [16,17]. However, the available clinical routine data often diverge from one institution to another,

which makes the generalizability and usability of these predictive models challenging. Thus, it is necessary to specify the data types necessary for detection of delirium as well as data availability.

Our objective is to systematically compile a Minimum Data Set for the Detection of Delirium (MDS-DD). We will achieve this aim through the systematic development of a generally applicable and comprehensive view of relevant delirium elements (maximum data set) as a basis for the alignment with already available clinical routine data in hospitals with divergent hospital information systems.

Methods

Based on the SPIRIT framework as a systematic approach for the secondary use of clinical routine data [9], we combined a top-down with a bottom-up approach to construct the MDS-DD. Figure 1 depicts in a summarizing manner the undertaken actions (grey boxes) and the expected results (white boxes).

Top-Down Approach: Literature Work and Expert Review

A literature and guideline review was conducted to obtain a thorough overview of relevant elements of delirium as an important patient safety incident in order to generate the maximum data set for delirium. The maximum data set thus comprises all those elements associated with delirium. Thus, it is necessary to incorporate and collect all elements related to risk factors , diagnosis, and treatment for delirium measurement tools. In order to systematize our findings and to gain a classification of delirium, the identified elements were assigned to different classes based on common features. A classification is the compilation of classes (e.g., risk factors) with their semantic relationships [3]. The classes for outlining delirium were developed inductively and deductively out of the body of literature.

Furthermore, previously developed relevant classification systems were screened systematically to collect information on delirium, including the International Statistical Classification of Diseases and Related Health Problems (ICD), the Diagnostic and Statistical Manual of Mental Disorders (DSM), the nursing diagnosis classification of the North American Nursing Diagnosis Association International (NANDA-I), and the Nursing Outcome Classification (NOC).

To validate the usability of the comprehensive maximum data set, the first draft was reviewed by clinical and IT experts and within our working groups in addition to literature.

During the iterative developing process, we found that the alignment of the classes of the maximum data set with the International Classification for Patient Safety (ICPS) key concepts in PS-CAST (Patient Safety – Categorical Structure) [3,18] expedient as a first step for knowledge management and representation.

In order to obtain a visual representation of the maximum data set, PS-CAST was used to structure a "delirium map", including the identified classes and particular elements of delirium.

Bottom-Up Approach: Case Study

The maximum data set, a comprehensive collection of relevant elements for delirium, was then used as a "dragnet" to screen in a clinical information system of Tirol Kliniken, a health care institute in western Austria, for available delirium-specific routine data elements.

In cooperation with an IT expert from Tirol Kliniken, first all (nursing) data elements were screened, as they were easily accessible in a comprehensive nursing data mart [19]. In the next step, the clinical and medical data elements were screened together with a chief physician from a psychiatric clinic within Tirol Kliniken.

The Minimum Data Set for the Detection of Delirium (MDS-DD) was then described as a simple spreadsheet which contains the linkage between the elements of the maximum data set and the available clinical routine in the clinical information system of Tirol Kliniken. This MDS-DD can now be used as a basis for all following data extraction processes.

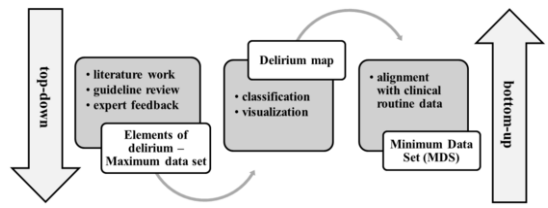


Figure 1 – Systematic approach for the development of a Minimum Data Set

Results

With this approach, we were able to collect, classify, and visualize relevant elements associated with delirium. The alignment of the maximum data set with the available clinical routine data revealed that several elements could be identified within various data sources. That means, indications for the risk factor such as *infection* could be identified within the laboratory information system, administrative data, nursing assessment, etc., although other elements, e.g., the training of the staff, were not identifiable within the routine data. Most of the elements could be identified in available nursing data. The medical data often remained on a less structured data quality level.

Elements of Delirium - Maximum Data Set

We identified eight classes: *circumstance, action, outcome, detection, care setting, agent, data,* and *classification.*

The class *circumstance* has two subclasses: *risk factors* and *organizational factors*. *Risk factors* are divided into *contributing factors (pathophysiological factors), predisposing factors* and *precipitating factors*. The *predisposing factors* can be further divided into *demographic characteristics* (with 14 concrete elements, e.g., age, sex, language), *cognitive status* (four elements, e.g., dementia, cognitive impairment), *functional status* (eight elements, e.g., functional dependence, pain), *sensory impairment* (two elements, e.g., visual impairment) and *coexisting medical conditions* (ten elements, e.g., comorbidity). The *precipitating factors* are constituted of *primary neurologic diseases* (six elements, e.g., stroke, head trauma), *intercurrent illnesses* (14 elements, e.g., infection, hypoxia), *surgery* (nine elements, e.g., orthopedic surgery), *environmental factors* (17 elements, e.g., setting, social isolation) and *physiological factors* (five elements, e.g., electrolyte disturbance). Additionally, *drugs* are related to the class *risk factors* (28 elements, e.g., drug abuse, analgesic agents). We also found 34 *pathophysiological factors* (e.g., proinflammatory markers) and genetic factors (e.g., Apolipoprotein E). Altogether, 151 risk factors were identified. The distinction of the elements was not always clear; for the purpose of integrity,

therefore, if it was possible to assign the same risk factors to two or more classes within the literature. Those factors were also represented as many times in the maximum data set. The *organizational factors* comprise five elements, including occupancy rate and technical resources; they were mainly mentioned in the expert review.

We screened four classification systems thoroughly: ICD, DSM, NANDA-I, and NOC. We identified 237 delirium-related nursing outcomes according to the Nursing Outcome Classification (NOC), e.g., 0905-Concentration (the ability to focus on a specific stimulus). These can be linked to the NANDA-I diagnoses "acute confusion" (00128) and "risk for acute confusion" (00173).

Furthermore, features and symptoms constitute additional delirium elements. They were assigned to the class detection. According to the literature, relevant criteria of delirium in ICD-10 and DSM-5 are: consciousness disturbance, inattention/ disturbance in attention, perceptual disturbance, disturbance in day-night rhythm, increased or decreased activity/agitation; disturbance in psychomotor activity, disorientation, memory disorder, cognitive disturbance, thought disorder, communication disorder, disturbance in visual-spatial capabilities, and/or disturbance in emotion/affect, characterized by an acute outbreak, a remarkable trigger, and changes during the course of the day and in the degree of severity.

Additionally, 51 assessment and screening instruments as relevant elements for delirium were identified for differential diagnostics and the *detection* of delirium, respectively. Examples are the Confusion Assessment Method (CAM) and the Delirium Observation Screening Scale (DOSS). Several tools for the measurement of delirium (patient safety indicators, trigger tools) were added to the maximum data set.

Concrete elements concerning the *outcome* class were identified which could be divided into *patient outcome* and *organizational outcome*. Relevant elements are the consequences for hospitalized patients with delirium, e.g., higher mortality during hospitalization, an increased length of stay, and higher expenditure.

Delirium management (6 elements, e.g., drug adjustments), therapy/treatment (pharmacological vs. non-pharmacological therapy), prevention (12 elements, e.g., prevention of cognitive impairment/disorientation), and diagnostics constitute the class *action*, i.e. interventions and actions to handle the topic were subsumed under this class.

Moreover, we identified the class *data*. Here, we screened the literature in order to find the first data group where the relevant data elements of delirium could be stored in general. The data are categorized as: demographic data, administrative data, nursing assessment, laboratory data, diagnosis, and procedures (health interventions).

The classes *care setting* and *agent* are not addressed at this point since they have been already discussed above.

Delirium Map

Figure 2 shows the structure of these classes and their subclasses without the concrete elements. Furthermore, the first-level relationships between delirium and the eight main classes are shown. This visualization can be regarded as the first step for knowledge management and representation for delirium.

Minimum Data Set for the Detection of Delirium (MDS-DD)

The bottom-up approach led to a specialized version of a minimum data set for delirium and revealed that most of the delirium elements could be identified within the available nursing data.

Based on the comprehensive maximum data set, we were enabled to screen the available clinical routine data in a goal-oriented manner. Every single element was considered. Within this screening process, we were able to compile a special data set, the "greatest common divisor" between all available routine data and the maximum data set: the Minimum Data Set for the Detection of Delirium (MDS-DD).

As it turned out within the case study, the minimum requirement for future data extraction and analyses is the access to the following data groups: nursing assessment data, care planning data (including nursing diagnoses, nursing goals, nursing interventions), administrative data, laboratory data, surgery data, medication data, medical data, several

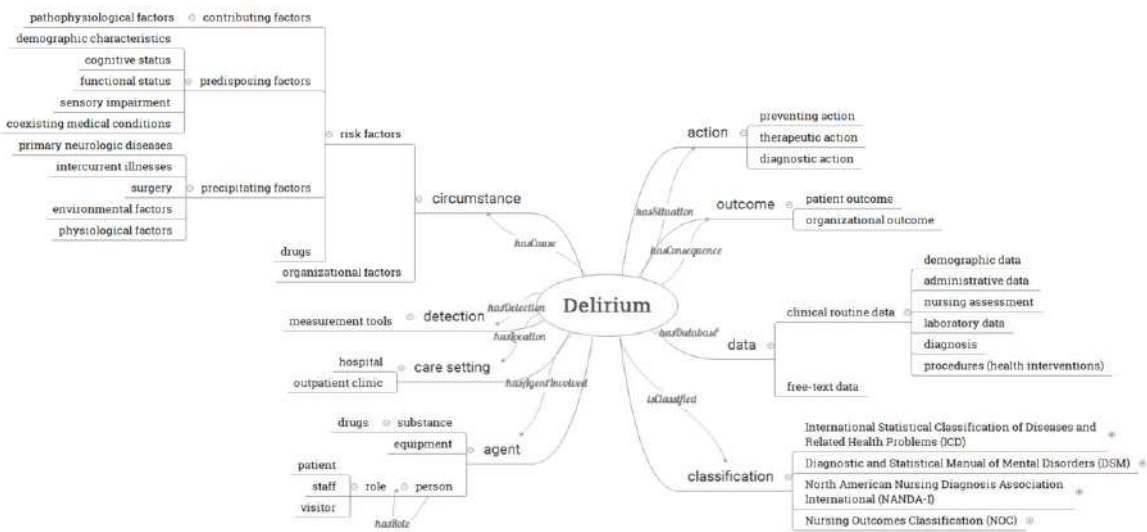


Figure 1– Delirium map following the PS-CAST with classes, subclasses and relationships (arrows)

additional forms (vital signs, pain, stool excretion, physical restraints, falls, decubital ulcers).

Within the class of risk factors, we ascertained that approximately 50% (75 elements) of the risk factors could be identified within the nursing routine data.

In table 1, we present a short extract of the MDS-DD.

Table 1 – Extract from MDS-DD with elements, classes, and the data source where it is stored; "n.a." denotes "not available in clinical routine data".

Element	Class	Data source
Age	Risk factors	Nursing assessment/ administrative data
Cognitive impairment	Risk factors	Nursing assessment/care planning data/medical data/administrative data
Electrolyte disturbance	Risk factors	Laboratory data
Pain	Risk factors	Nursing assessment/care planning data/forms
Technical resources	Organizational factors	n.a.
Apolipoprotein E	Pathophysiological factors	n.a.
Inattention/ disturbance in attention	Detection	Nursing assessment/care planning data
Normalize sleep-wake cycle	Action	Care planning data (nursing interventions)
Acute confusion (00128)	Classification	Care planning data (nursing diagnoses)

Discussion

The first version of a Minimum Data Set for the Detection of Delirium (MDS-DD) was developed at Tirol Kliniken. The MDS-DD consists of eight main classes which were further divided into subclasses. The content of the classes constitutes the relevant elements associated with delirium.

The compiled data set has not yet been validated. This circumstance must be regarded as a limitation for now.

One strength of our approach is the reproducibility with regard to further patient safety phenomena and especially to various hospital information systems that are designed in different ways. Evidence within our development procedure are several similarities to the existing indicator development approaches [20, 21], such as the combination of literature work, guideline reviews, and expert feedbacks with regard to available clinical routine data.

Relevant measuring methods for delirium, such as delirium indicators and assessment methods, were also incorporated into the delirium map. This comprehensive compilation of the different delirium classes offers the possibility to link several approaches in order to detect delirium events with one automated tool. This the secondary use of clinical routine data to avoid the disadvantages of the particular existing methods and instruments within future stages of the project.

The delirium map can be used as the foundation for a concept map to represent concepts and relationships and describe the safety-related phenomenon of delirium. Furthermore, the map's elements and classes could be aligned in further steps in directed acyclic graphs to attempt to find causal effects and identify confounders. The cross mapping between delirium map and the International Classification for Patient Safety (ICPS) key concepts in PS-CAST provides the advantage of incorporating the ontology with the indicator measurement [22]. One of the next steps in the patient safety project could therefore be the formalization of the delirium map into a context-embedded and secondary data-based ontology to automatically detect and compute delirium indicators both retrospectively and prospectively.

There might be potential elements missed in our method. It is conspicuous, however, that social factors such as *family absent* or *social isolation* are named as risk factors in the literature but concrete mention of risk factors such as *language, religion* or *socioeconomic status* could be hardly identified, likewise *organizational aspects*, e.g., light conditions, the organizational occupancy rate or available technical resources. Such potentially relevant delirium elements were added after expert feedback with regard to the meaningfulness of the comprehensive delirium element set.

Another possibility for the execution of the bottom-up approach would be a comprehensive analysis of the existing information systems and available data within an organization, i.e. a thorough description of the whole system, for example, with a so-called three-layer graph-based meta model [9]. This will make it possible to identify unknown risk factors and enable in-depth exploratory analyses.

According to our results, the maximum data set seems to be a generally applicable and comprehensive overview of relevant delirium elements as a basis for the alignment with already available clinical routine data in hospitals with divergent hospital information systems. It must be clarified, however, which data elements can be used for which analyses and whether this makes sense. This means, for example, that a symptom as an element of delirium cannot be used as a predictor variable.

Up to this point, it was clarified which data groups are relevant for subsequent analyses. The next steps on the way to these analyses are the development of data extraction and transformation routines.

Conclusions

In this study, a comprehensive maximum data set for relevant delirium elements was developed systematically. This maximum data set shows its applicability for the alignment with publically available data and the construction of MDS for delirium in various organizations. We tried to confirm the issue in a first case study. As a next step, we will use the MDS-DD as the basis to develop a model for automated delirium detection and prediction and validate our data model.

Overall, the approach we applied for the systematic construction of concrete MDSs seems to be usable for the development of other patient safety phenomenon element and data sets. This circumstance should be considered for further studies.

Acknowledgements

We would like to thank Tirol Kliniken GmbH for the possibility to cooperate and especially Stefanos Rauchegger and Prof. Josef Marksteiner for their support. The authors did not report any conflicts of interest. This work forms part of the PATIS Project (A PATient Safety Intelligence System and framework

for the secondary use of multimodal clinical data to assess and improve patient safety) funded by the Austrian Science Fund (FWF – P 29076).

References

[1] S.P. Rodrigues, N.J. Van Eck, L. Waltman, and F.W. Jansen, Mapping patient safety: A large-scale literature review using bibliometric visualisation techniques, *BMJ Open*. **4** (2014). doi:10.1136/bmjopen-2013-004468.

[2] S.N. Musy, D. Ausserhofer, R. Schwendimann, H.U. Rothen, M.-M. Jeitziner, A.W. Rutjes, and M. Simon, Trigger Tool–Based Automated Adverse Event Detection in Electronic Health Records: Systematic Review, *J. Med. Internet Res.* **20** (2018) e198. doi:10.1016/0002-9343(80)90114-X.

[3] World Health Organization (WHO), Conceptual Framework for the International Classification for Patient Safety, 2009. http://www.who.int/patientsafety/taxonomy/icps_full_report.pdf [accessed November 23, 2018].

[4] A. Karpov, C. Parcero, C.P.Y. Mok, C. Panditha, E. Yu, L. Dempster, and C.M. Hohl, Performance of trigger tools in identifying adverse drug events in emergency department patients: a validation study, *Br. J. Clin. Pharmacol.* (2016) 1048–1057. doi:10.1111/bcp.13032.

[5] M. Hanskamp-Sebregts, M. Zegers, C. Vincent, P.J. Van Gurp, H.C.W.D. Vet, and H. Wollersheim, Measurement of patient safety: A systematic review of the reliability and validity of adverse event detection with record review, *BMJ Open*. **6** (2016) 1–9. doi:10.1136/bmjopen-2016-011078.

[6] Health Quality & Safety Commission, The Global Trigger Tool : A Review of the Evidence, 2013. https://www.hqsc.govt.nz/assets/GTT/GTT-evidence-review-Oct-2013.pdf [accessed November 24, 2018].

[7] M. Govindan, A.D. Van Citters, E.C. Nelson, J. Kelly-Cummings, and G. Suresh, Automated detection of harm in healthcare with information technology: A systematic review, *Qual. Saf. Heal. Care*. **19** (2010) 1–12. doi:10.1136/qshc.2009.033027.

[8] H.J. Murff, V.L. Patel, G. Hripcsak, and D.W. Bates, Detecting adverse events for patient safety research: A review of current methodologies, *J. Biomed. Inform.* **36** (2003) 131–143. doi:10.1016/j.jbi.2003.08.003.

[9] W.O. Hackl, and E. Ammenwerth, SPIRIT: Systematic planning of intelligent reuse of integrated clinical routine data: A conceptual best-practice framework and procedure model, *Methods Inf. Med.* **55** (2016) 114–124. doi:10.3414/ME15-01-0045.

[10] Canadian Institutes of Health Research, Secondary use of personal information in health research: Case studies, 2002. http://www.cihr-irsc.gc.ca/e/documents/case_studies_nov2002_e.pdf [accessed October 23, 2018].

[11] D.R. Schlegel, and G. Ficheur, Secondary Use of Patient Data: Review of the Literature Published in 2016, *Yearb. Med. Inform.* **26** (2017) 68–71. doi:10.15265/IY-2017-032.

[12] S.K. Inouye, R.G.J. Westendorp, and J.S. Saczynski, Delirium in elderly people, *Lancet*. **383** (2014) 911–922. doi:10.1016/S0140-6736(13)60688-1.

[13] National Clinical Guideline Centre, DELIRIUM: diagnosis , prevention and management. Clinical Guideline 103, 2010. https://www.ncbi.nlm.nih.gov/books/NBK65558/pdf/Bookshelf_NBK65558.pdf [accessed October 12, 2018].

[14] S.P.K. Veeranki, D. Hayn, D. Kramer, S. Jauk, and G. Schreier, Effect of nursing assessment on predictive delirium models in hospitalised patients, *Stud. Health Technol. Inform.* **248** (2018) 124–131. doi:10.3233/978-1-61499-858-7-124.

[15] H. Lindroth, L. Bratzke, S. Purvis, R. Brown, M. Coburn, M. Mrkobrada, M.T. V Chan, D.H.J. Davis, P. Pandharipande, C.M. Carlsson, and R.D. Sanders, Systematic review of prediction models for delirium in the older adult inpatient, *BMJ Open*. **8** (2018) e019223. doi:10.1136/bmjopen-2017-019223.

[16] D. Kramer, S. Veeranki, D. Hayn, F. Quehenberger, W. Leodolter, C. Jagsch, and G. Schreier, Development and validation of a multivariable prediction model for the occurrence of delirium in hospitalized gerontopsychiatry and internal medicine patients, *Stud. Health Technol. Inform.* **236** (2017) 32–39. doi:10.3233/978-1-61499-759-7-32.

[17] C.M. Gonzalvo, H.A.J.M. De Wit, B.P.C. Van Oijen, D.S. Deben, K.P.G.M. Hurkens, W.J. Mulder, R. Janknegt, J.M.G.A. Schols, F.R. Verhey, B. Winkens, and P.H.M. Van Der Kuy, Validation of an automated delirium prediction model (DElirium MOdel (DEMO)): An observational study, *BMJ Open*. **7** (2017) 1–5. doi:10.1136/bmjopen-2017-016654.

[18] J.M. Rodrigues, N. Dhingra-Kumar, S. Schulz, and J. Souvignet, A patient safety information model for interoperability, *Stud. Health Technol. Inform.* **223** (2016) 77–84. doi:10.3233/978-1-61499-645-3-77.

[19] W.O. Hackl, F. Rauchegger, and E. Ammenwerth, A Nursing Intelligence System to Support Secondary Use of Nursing Routine Data., *Appl. Clin. Inform.* **6** (2015) 418–28. doi:10.4338/ACI-2015-04-RA-0037.

[20] H. Hyppönen, E. Ammenwerth, and N. De Keizer, Exploring a methodology for eHealth indicator development, *Stud. Health Technol. Inform.* **180** (2012) 338–342. doi:10.3233/978-1-61499-101-4-338.

[21] M. Van Engen-Verheul, H. Kemps, R. Kraaijenhagen, N. De Keizer, and N. Peek, Modified rand method to derive quality indicators: A case study in cardiac rehabilitation, *Stud. Health Technol. Inform.* **169** (2011) 88–92. doi:10.3233/978-1-60750-806-9-88.

[22] M.D.C. Legaz-García, K. Dentler, J.T. Fernández-Breis, and R. Cornet, Combining Archetypes, Ontologies and Formalization Enables Automated Computation of Quality Indicators, *Stud. Health Technol. Inform.* **235** (2017) 416–420. doi:10.3233/978-1-61499-753-5-416.

Address for correspondence

Michael Schaller, UMIT, Eduard-Wallnöfer-Zentrum 1, 6060 Hall in Tirol, Austria, E-mail: michael.schaller@umit.at

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190381

Comparing Information Needs of Diabetes Patients in Chinese and American Health Communities of Questions and Answers

Jing Shi[a], Yuxing Qian[a], Chenlu Li[a], Liqin Zhou[a], Bin Zhang[b]

[a] Center for Studies of Information Resources, Wuhan University, Wuhan, Hubei, China,
[b] Center of Traditional Chinese Cultural Studies, Wuhan University, Wuhan, Hubei, China

Abstract

This study selects the users generated content in diabetes groups of ManYouBang and DailyStrength which are two representative HCQA (Health Communities of Questions and Answers). Theme coding was applied to identify information needs while social network analysis was used to compare Chinese and American HCQA. In theory, we combine Social Network Analysis and content analysis to capture and contrast the pattern of Q&A in HCQA. On the one hand, the core needs of HCQA are "how to treat the disease" and the "diet" theme is closely related to other themes. On the other hand, the Chinese diabetes communities have "question and answer attributes" while the American diabetes communities have both "question and answer attributes" and "social support attributes." In practice, the founding provides enlightenment for the development of HCQA including improving content quality and strengthen emotional connection and disease treatment infrastructure.

Keywords:

Diabetes Mellitus; Diet; Social Support

Introduction

Health information refers to information related to people's physical and mental diseases, nutrition, health care, etc. [1] Health Information Seeking Behavior (HISB) is described as "verbal or nonverbal behavior used to obtain, clarify or confirm knowledge or information about a specific event or situation" [2]. Recently, most HISB researches have adopted questionnaire and structured interview methods to conduct quantitative analysis[3]. Themes cover searching behaviors, motivations, purposes, influence factors, information quality assessment, impact on health practice, and so on [5]. Many people, either adolescents or older adults, reported that the Internet was an important channel to reach health information source available [4]. HISB can be affected by external and internal factors, such as uses and gratifications (U&G), Media System Dependency (MSD) [6], user's goals, motivations, emotions and so on [7]. Betsch [8] studied online vaccine information's impact on patient's vaccine receiving behaviors based on TPB theory. It appears information plays an important role in patient's vaccine receiving behaviors, especially for lack-of-knowledge individuals.

HCQA is different from users' online health information browsing. It is characterized by two main functions: informational support and social support. With the number of people with chronic diseases skyrocketing, more and more users joined HCQA to seek and share treatment experience and health knowledge. They offer views, conduct emotional communication, and seek emotional support from others in the community. It is considered to be critical in patient's recovery [9]. Researches analyze more based on encoding topics including content analysis and quality evaluation [10,11]; user behaviors and psychological analysis [12, 13]; topic identification [14-16].

Kim et al. [10] and Oh et al. [11] studied the content of forums in Yahoo！Answers. The former analyzes linguistic properties of different types of questions about eating disorders while the latter investigates evaluation criteria people use about the online health information. Factor and regression analyses conducted among 231 breast cancer patients were used to examine the role of social support perception and emotional wellbeing on online information seeking within the context of CHESS [12]. Ahadzadeh [13] used PLS-SEM to test the moderating effect of health consciousness on attitude towards internet usage in HISB. Coulson [14] and Arden [15] captured three main themes of alcohol discussion forums and guidance on public forums respectively by qualitatively thematic analysis. The current body of research has shed light on the HCQA related landscape. Notwithstanding, researches using hybrid methods on user online health information need are limited. Only a few studies focused on understanding information needs. Min [16] used both content and semantic network to analyze forums which revealed two dominant topics: breast cancer tests and treatment.

Few researchers have applied Social Network Analysis combined with content analysis to capture and contrast the pattern of Q&A in HCQA. To fill the gap, this study focuses on understanding the patterns and users' needs of HISB in Chinese and American communities, taking diabetes for example.

Methods

Data were obtained from diabetes-patient-group in DailyStrength (https://www.dailystrength.org/) and Man Youbang (http://www.manyoubang.com/), one of the largest online communities in the USA and China. DailyStrength was established in 2006 comprised of 421 patients-groups and 1096 members in diabetes-group; while Man Youbang was established in 2013 comprised of 84 patients-groups and 7,995 members in diabetes-group. Although Man Youbang also raises money for patients, we only focus on its Q&A which is a relatively separate and important section in fact.

Users post messages and replies on a variety of topics referring to diabetes. A total of 3,553 replies associated with 777 questions from Man Youbang and 11,754 replies associated with 2,115 questions from DailyStrength related to diabetes posted in Jan. 1st, 2013 to Dec. 3st, 2017 were collected. Through manual review, meaningless entries are deleted. Then, we get 2,556 replies associated with 777 English questions fr

Table 1 – Topic Coding Scheme of Dataset

Level 3	Level 2	Level 1	Connotation
Disease specific information	Prevention	How to prevent diabetes	How to prevent diabetes from genetic, environmental, diet, exercise, and other factors.
	Diagnose	Symptom	Anything that accompanies diabetes and is regarded as an indication of diabetes.
		Cause	Events that lead to diabetes.
		Typing	Diabetes is divided into type 1 and type 2
	State of an illness	Describe the state of an illness.	Description of the patient's condition.
	Treatment	insulin therapy	Treatment with insulin injection.
		Other treatments	Other treatments, such as oral medication.
	Complication	Diabetic foot	A diabetic foot is a foot that exhibits any pathology that results directly from diabetes.
		Ketoacidosis	Acidosis with an accumulation of ketone bodies; occurs primarily in diabetes mellitus.
		Retinopathy	The most important manifestation of diabetic microangiopathy is specific changes in the fundus.
	Blood glucose monitoring	Self-monitoring of blood glucose	Patients observe and record their blood glucose by instruments, test paper, etc.
	Special cases	Gestational diabetes	Pregnant women develop diabetes before or during pregnancy.
		Juvenile diabetes	Patients are diagnosed with diabetes between the ages of 13-25.
		Senile diabetes	Patients are diagnosed with diabetes after age 60.
		Others	Patients have other diseases besides diabetes.
	Drug's side effects	Side effects caused by drugs	Symptoms, such as nausea, vomiting, hair loss and fatigue, caused by drugs.
Non-disease specific information	Diet	Dietary precautions	Food that helps or deteriorates with treatment.
	Exercise	Exercise to lose weight	Exercise during illness to lose weight.
	Insurance	Medical insurance	Obtain compensation or security for treatment.
		Other insurance	Other insurance.
	Finance	Financial concerns	Financial pressure during treatment.
	Emotional support	Spiritual support	Seek comfort and encouragement during treatment.
	Social support	Charity relief	Property and other charitable assistance during treatment.

Man Youbang and 9,179 replies associated with1,640 questions from DailyStrength. In this study, we explore hot topics and the evolution in different periods by coding content to reflect users' information need.

Manual annotation and word frequency statistics are often used to summarize the characteristics of information needs. Topics are identified manually by two coders considering the coding scheme adapted by Oh, Zhang, and Park [17]. Table 1 shows the coding results. 50 English and 50 Chinese questions and associated replies were selected randomly to test inter-rater reliability. For the level 2 categories, Cohen's kappa scores were 0.892 and 0.913 for the English questions and Chinese questions respectively. For the level 1 categories, Cohen's kappa score were 0.812 and 0.791 for them. It indicated that the two coders achieved good agreement on categories.

Results

Hot Topics Comparison of Chinese and American HCQA

Topics in domestic and foreign have similar performance. The most popular topics are treatment, diet, and state of illness. It is surprising that "community health service" is mentioned merely as an important auxiliary treatment for chronic diseases. It reflects that patients still pay more attention to disease-specific information. Most of them expect to obtain practical and reliable treatment and health management information, which can guide medical treatment and drug selection.

We also found some differences. To make numbers comparable, we calculate the proportion of each topic instead of absolute value. In Man Youbang, participants sought information about treatment most frequently, followed by diet, state of illness, complication, and diagnose, which reflects obvious practicability. Especially, "treatment" accounted for more than a quarter of all topics, which was the top priority for patients in Chinese communities. The least were prevention, insurance and, social support. Regarding DailyStrength, information need focuses on treatment, state of illness, and diet. Beyond the disease-specific information, an important non-disease specific term, emotional support, accounts for 15% in DailyStrength. It appears that American users pay great attention to the impact of emotion and gain comfort and encouragement from others in the community. Figure 1 present the top 5 hot topics.

Comparing the two communities, it becomes apparent that the proportion of "emotional support" in DailyStrength is about twice that of Man Youbang. Considering the content of relevant posts, American users intend to use emotional words more frequently out of religious belief, humanistic concern, and so on. Most questioners not only describe conditions but show their moods. At the same time, answers deliver their concern, sympathy, and support. It reflects the fact that DailyStrength not only possesses the attribute of "Q&A" but the attribute of "social contact." Patients express emotions and seek support frequently because they trust and depend on other members. It becomes apparent that western countries pay attention to emotional support in addition to drugs and physical ways.

Figure 1– Comparison of the top 5 hot topics in HCQA

Evolution of Hot Topics

Combined with the time label, Figure 2 and Figure 3 show the topics' composition of diabetes forums' in 2013-2017.

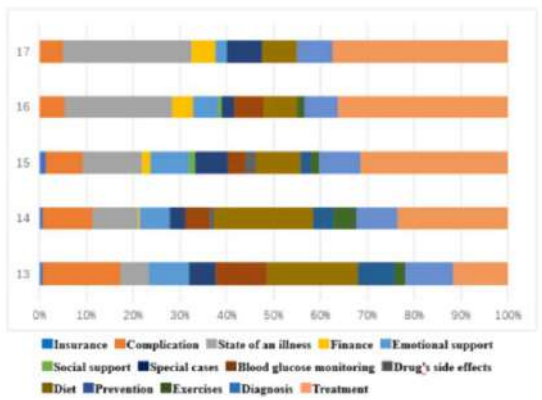


Figure 2– Theme Tags of Man Youbang

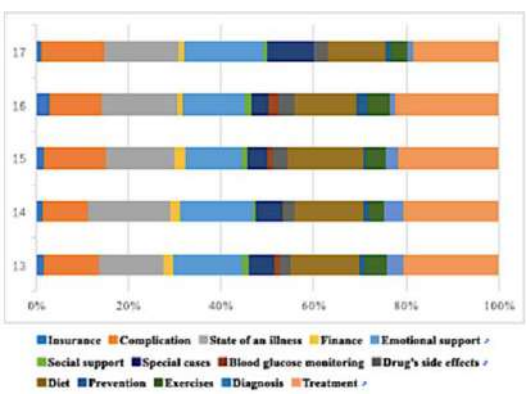


Figure 3– Theme Tags of DailyStrength

First, the overall trend of topics evolution is different. Annual topics' composition changes obviously in Man Youbang while not in DailyStrength. It is related to the development of HCQA in Chinese and American. Founded in 2006, DailyStrength is at a stable stage now. Users of various conditions and courses are balanced. By contrast, the heats have a larger fluctuation with members in different stages altering in Man Youbang.

Second, specific topics have different trends. In Man Youbang, the proportion of "treatment," "state of illness," and "finance" increase while "complication" and "diet" decrease. "Emotional support" and "diagnosis" are stable. By contrast, topics proportion is more balanced in DailyStrength. "Complication" usually posted by severe patients while "treatment" and "state of illness" may be issued by newcomers. In China, changing topics' proportion reflects Diabetes incidence and treatment burden are growing.

Comparison of Co-occurrence Network of Topics

Figure 4 shows the co-occurrence networks of Man Youbang and DailyStrength, which demonstrates the overall patterns and holistic properties. The larger the node, the higher the betweenness centrality. It shows that these topics are more likely to co-occurrence with others. The thicker the line, the higher the co-occurrence frequency between two topics.

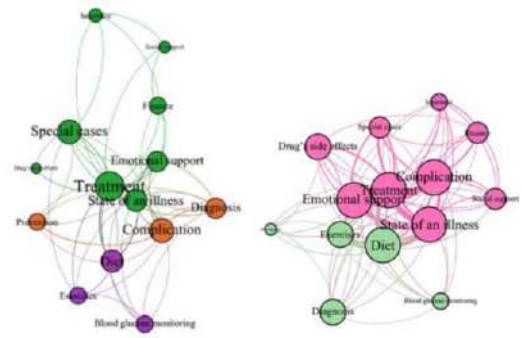


Figure 4– Co-occurrence Network of Topics in Man Youbang(left) and DailyStrength(right)

Figure 4 indicates the heats distribution is similar. We also focus on detailed information on different topics.

First, the top degree centrality topics are "treatment," "special cases," "state of illness," and "diet" in both Man Youbang and DailyStrength. All of them appear alongside more than 10 themes. "Treatment" is the most critical theme. Even if the main purpose of the question is not "treatment," the related narrative emerge. Second, we found that although "diet" does not co-occur with many topics, its average co-occurrence rate is high. It indicates "diet" has a closer correlation with its co-occurrence.

Diabetes is closely related to dietary habits for improper diet leading to the fluctuation of blood sugar. Therefore, patients discuss whether certain food can be eaten and how to control blood sugar. Food therapy is also used in the treatment of diabetes, which explains why "diet" is frequently associated with "treatment" and "state of illness." Also, obesity is one of the causes, and losing weight is a hot topic for diabetic patients. As two major ways to lose weight, "exercise" and "diet" are connected closely.

On the other hand, there are also significant differences. First, the overall network attributes are different. Topics in DailyStrength are connected more closely than Man Youbang. The co-occurrence network of Man Youbang includes 14 nodes and 50 lines. Average degree is 7.143, that is each topic co-occurs with other 7 topics on average. The density is 0.549, and average weighting degree is 48.571. Overall, the topic themes are not closely connected. The co-occurrence network of DailyStrength includes 14 nodes and 77 lines. Average degree is 11. The density is 0.846, and average weighting degree is 208.571. Overall, these themes are closely connected.

Second, topics play different roles. The high-degree nodes in DailyStrength are more than that in Man Youbang. In DailyStrength, nodes with a degree centrality above 10 accounts for 79% where "complication," "state of illness,"

"emotional support," "diet," and "treatment" co-occur with all other topics. We found that DailyStrength has a distinct role in "sharing." Patients ask and answer questions, also share their conditions, moods, and even matters in life. This increases the number of words and the probability of topics occurrence. While obvious attribute of "practice" in Man Youbang, patients ask and answer with short words involving fewer topics.

In addition, members have different emotional needs. As shown in Figure 4, "emotional support" not only co-occurs with all topics but appear a higher average co-occurrence rate In DailyStrength. Diabetes has a long treatment period accompanied by serious complications, which may cause fatigue and negative emotions. Diabetes is closely related to living habits, especially diet and exercise. Most patients have to reduce living quality and take on a greater economic burden.

At last, Topic clustering of Man Youbang is relatively scattered but more concentrated in DailyStrength. Table 2 presents the clusters. Cluster 1 refers to the main direct-related topics. Other topics are divided into 2 categories in Man Youbang. Cluster 2 refers to few disease-specific topics about diabetes or complications, while cluster 3 refers to non-disease specific topics about living habits. In DailyStrength, the remaining topics do not be divided. Therefore, we can infer American HCQA users pay more attention to the impact of diet and exercise on prevention. Improper living habits is an important cause of diabetes, such as high-calorie eating habits.

Table 2–Topics clusters results

Man Youbang		**DailyStrength**	
Clusters	**Topics**	**Clusters**	**Topics**
Treatment and protection of disease	Treatment, Special cases, State of an illness, Emotional support, Finance, Drug's side effects, Insurance	Treatment and protection of disease	Emotional support, Treatment, State of an illness, Complication, Drug's side effects, Special cases, Social support, Insurance, Finance
Prevention and complications	Prevention, Complication, Diagnosis	Diagnosis and management of disease	Diet, Exercises, Diagnosis, Blood glucose monitoring, Prevention
Daily management of disease	Diet, Exercises, Blood glucose monitoring		

Discussion

Using topic analysis and social network analysis, we examine hot topics evolution, related terms, and co-occurrence of HCQA in China and the USA. Then we will summarize the main similarities and differences of user information needs.

Common characteristics

First, the information needs of users in HCQA are similar. The main purpose is to provide patients with sharing disease and treatment experience. The core need is "how to treat diseases," which involves such topics as "state of illness," "treatment," "complication," and so on. It mainly depends on the function of HCQA. HCQA are working to provide information exchange channel for participants. Even if users do not intend to ask questions about treatment, most of the narratives relate to the core. Second, "diet" and "finance" are the most popular non-disease related topics. This is related to characteristics of chronic diseases, such as long treatment period, the healing process, multiple complications, high cost, etc. which brings greater financial burden to patients. Most patients expect to get practical and reliable information about treatment and health management in the community.

Different Characteristics

First, communities stand at different stages. American communities started early and develop stable now. The composition of members and demands are relatively stable, obvious, and connected more closely. Chinese communities started late and possess development period. Users' composition and demands fluctuate widely and are not closely related to each other. Second, there are different attributes. The attribute of "Q&A" is remarkable in Chinese communities. Patients prefer direct and short narrative for paying more attention to practicality. But they seldom seek mood expression and emotional support. However, American communities also show the attribute of "social connect" in addition to "Q&A." The former is even more significant. According to the content of related posts, it is related to differences in the social environment and living habits between Chinese and American communities. Due to religious beliefs and humanistic care, American HCQA users are better at expressing emotions and communicating with people. They intend to use more emotional words in expressions and eager to be comforted by others. Chinese users implicitly express their emotions and do not express their emotions to strangers or seek encouragement. It also reflects low trust and dependence in community members.

Although interesting findings were made, this study has its limitations. First, correlation degree and the quality of Q&A have not been considered. It remains to be studied to take correlation degree into the multi-dimensional analysis framework. Second, researches on topic and content evolution are not enough. We should consider time slice analysis by season, quarter, etc.

Conclusions

In this study, we analyze hot topics evolution and co-occurrence on the topic of diabetes. On the one hand, the information needs of HCQA in Chinese and American are similar focusing on "how to treat diseases." Other living-habits-specific needs, such as "diet," are clearly expressed in chronic disease groups. On the other hand, Chinese communities appear the attribute of "Q&A," and members pursue practicality. By contrast, American communities have both the attribute of "Q&A" and "social connect." Furthermore, topics fluctuate frequently in Chinese communities while stable in American communities.

In theory, we combine Social Network Analysis and content analysis to capture and contrast the pattern of Q&A in HCQA. Some clear conclusions are obtained referring to the similarities and differences of users' information needs of HCQA in China and the USA, which has certain theoretical significance for following researches. In practice, the founding provides enlightenment for the development of HCQA. Chinese communities should focus on improving the content quality and strengthen emotional connection among users to retain them. In addition, the rising incidence of diabetes must be taken seriously by government departments, which can take certain measures to further improve disease treatment infrastructure. For example, they can promote the role of community service and strengthen medical security to reduce the financial burden of patients.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Nos. 71661167007).

References

[1] E. Manafo and S. Wong, Exploring older adults' health information seeking behaviors, *Journal of Nutrition Education & Behavior* **44** (2012), 85-89.
[2] V.M. Conley, Beyond knowledge deficit to a proposal for information-seeking behaviors, *International Journal of Nursing Terminologies & Classifications* **9** (2010), 129-135.
[3] J. Wu, X. Huang, K. Dong, A review of online medical research based on knowledge graph, *Journal of Information Resources Management* **2** (2016), 4-21. [4] N.J. Gray, J.D. Klein, P.R. Noyce, T.S. Sesselberg, and J.A. Cantrill, Health information-seeking behaviour in adolescence: the place of the internet, *Social Science & Medicine* **60** (2005), 1467-1478.
[5] Y. Li and W. Cai. A review of the studies on health information seeking behavior overseas, *Library & Information Service* **19** (2012), 128-132.
[6] Tustin N. The role of patient satisfaction in online health information seeking. *Journal of Health Communication* **1** (2010), 3-17.
[7] Y. Zhang, Contextualizing consumer health information searching: an analysis of questions in a social Q&A community, in: *ACM International Health Informatics Symposium*, 2010, pp. 210-219.
[8] C. Betsch, N.T. Brewer, P. Brocard, P. Davies, W. Gaissmaier, N. Haase, J. Leask, F. Renkewitz, B. Renner, and V.F. Reyna, Opportunities and challenges of Web 2.0 for vaccination decisions , *Vaccine* **30** (2012), 3727-3733..
[9] N. Priya, Information seeking and social support in online health communities: impact on patients' perceived empathy, *Journal of the American Medical Informatics Association* **18** (2011), 298-304.
[10] S. Kim, S. Oh, and J.S. Oh, Evaluating health answers in a social Q&A site, *Proceedings of the Association for Information Science & Technology* **45** (2010), 1-6.
[11] J.S. Oh, D. He, J. Wei, E. Mattern, and L. Bowler, Linguistic characteristics of eating disorder questions on Yahoo! Answers – content, style, and emotion, in: *Asis&t Meeting: Beyond the Cloud: Rethinking Information Boundaries*, 2013, p. 87.
[12] S.C. Kim, D.V. Shah, K. Namkoong, F.M. Mctavish, and D.H. Gustafson, Predictors of online health information seeking among women with breast cancer: the role of social support perception and emotional well-being, *Journal of Computer-Mediated Communication* **18** (2013), 98–118.
[13] A.S. Ahadzadeh, S.P. Sharif, and F.S. Ong, Online health information seeking among women: the moderating role of health consciousness, *Online Information Review* **42** (2017), 58-72.
[14] N.S. Coulson. Sharing, supporting and sobriety: a qualitative analysis of messages posted to alcohol-related online discussion forums in the United Kingdom, *Journal of Substance Use* **19** (2014), 176-180.
[15] M.A. Arden, A.M.S. Duxbury, H. Soltani. Responses to gestational weight management guidance: a thematic analysis of comments made by women in online parenting forums. *BMC Pregnancy and Childbirth* **14** (2014), 216.
[16] M.S. Park, H. Park. Topical network of breast cancer in formation in a Korean American online community: a semantic network analysis. *Information Research: An International Electronic Journal* **21** (2016), n4.
[17] J.S. Oh, D. He, W. Jeng, et al. Linguistic characteristics of eating disorder questions on Yahoo! Answers – content, style, and emotion. *Proceedings of the American Society for Information Science & Technology* **50** (2014), 1-10.

Address for correspondence

Corresponding author: Liqin Zhou, PhD student. Center for Studies of Information Resources, Wuhan University, Wuhan, Hubei, China

E-mail: zhoulq92@163.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190382

Using Health Information Exchange: Usability and Usefulness Evaluation

Mauricio Soto[a,b], Claude Sicotte[b,c], Aude Motulsky[b,d]

[a] Department of Family Medicine, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile,
[b] Department of Management, Evaluation & Health Policy, School of Public Health, Université de Montréal, Montreal, Canada.
[c] Healthcare Organization Management Host Team (EA7348 MOS – Management des organisations de santé - Healthcare Organization Management), EHESP – École des hautes études en santé publique, France
[d] Research Center, Centre hospitalier de l'Université de Montréal, Montreal, Canada

Abstract

Health information exchange (HIE) is implemented in Quebec, Canada to improve the exchange and use of clinical information for decision making in the province's health care system. The objective of this mixed-method approach study was to evaluate the usage, usability, and usefulness of an electronic health record (EHR) to access and import clinical information from a centralized HIE. First, a heuristic analysis was performed of three different commercial EHRs, using end users to analyze the integration feature through think aloud protocols. Second, interviews were conducted with advanced users to describe the usefulness. Third, usage data were analyzed to describe the level of use of each data domain, per EHR and user. The results show that usefulness is high for medication data; leading to greater use of this domain, its relation to integration, and meets the needs of general practitioners. Difficulties in the implementation processes reduced the potential of this system.

Keywords:

Health Information Exchange, Electronic Health Records, Systems Integration

Introduction

Health Information Exchange (HIE) has been defined as the "electronic sharing of clinical information among users to facilitate care coordination and transitions between different settings of care" [1]. The use of HIE has the potential to generate significant improvements in patient safety, efficiency, and productivity of healthcare organizations [2]. In emergency departments (ED), the HIE can improve the quality of health care [3]. For example, sharing clinical information with a HIE has reduced the number of patient readmissions in seven hospitals in Israel [4]. In primary care, the use of HIE has led to improvements in efficiency: better access to test results, as well as reduced staff time for managing referrals and processing complaints [5]. Despite the achievements, several barriers to the exchange of health information have been recognized [1-2]: not enough information in the HIE to justify its use, differences in the way HIE is integrated into the clinical workflow and a poorly designed interface.

HIE usage has become an important issue in achieving the desired effects. Several studies have indicated that usefulness and usability are the most important variables for ensuring use of information technology[6-8]. The perceived usefulness of a health information technology depends on two types of impact: perceived improvement in work efficiency and perceived improvement in quality of care [6], two variables that have been evaluated by either quantitative or qualitative approaches [7].

Usability is the degree to which a system can be successfully integrated to perform tasks in a planned work environment [9]. As a result, usability is evaluated by user, system, and task interaction in a specified context. A review of usability assessments of electronic health records (EHR) revealed that the most common methods were surveys, think-aloud protocols and heuristic methods [10]. Heuristic methods involve the examination of an interface by a human evaluator looking for usability problems [11]. Such problems represent violations of the usability principles developed by Molich and Nielsen [12]. Scapin and Bastien developed ergonomic criteria as dimensions in order to synthesize best practices in the field of the person-machine interface[11].

In Canada, HIE initiatives have been developed in different provinces, resulting in a significant increase in the number of active monthly users of HIE from 2006 to 2017[13]. However, the same report indicates that adoption and utilization rates fluctuated between 22% and 42% of the total number of potential users, which is still far from optimal. The main barriers identified were at the level of system usability, project management, and how to integrate the HIE into clinical practice.

In Quebec, a HIE, the "Dossier Santé Québec" (DSQ), has been proposed as "a highly secure technological environment that allows the collection, storage, and consultation of certain health information" [14]. The law governing the exchange of health information stipulates that the HIE collects, stores and allows the consultation of health information in six domains: medication, laboratory tests, medical imaging, immunization, the hospitalization summary, and allergy and intolerance. In 2018, the information domains available are prescribed medications dispensed by community pharmacies, results of laboratory tests carried out in public and private labs, and reports and images from medical imaging examinations (radiography, computed tomography [scans], MRIs, etc.) carried out in public and private organizations.

The aim of our study was to assess the usage, perceived usefulness, and usability of the Quebec HIE in order to identify factors that may enhance or reduce its potential in clinical practice in primary care.

Methods

This mixed method study was conducted to evaluate three EHR systems integrated through HIE. The three chosen EHRs represented more than two thirds of all EHR users in the province of Quebec. The HIE integration features evaluated are

vendor-mediated HIE, with a pull-type mechanism to access information (clinicians needed to actively request or "pull" from their EHR the patient information from the HIE). The Quebec HIE applies an “opt-out” patient consent model, in which patients need to opt out if they want to make their information inaccessible to clinicians.

This study was conducted in three parallel parts. First, a descriptive analysis of HIE usage was conducted using audit trails of HIE accesses provided by the Ministry of Health from January 1 to December 31, 2017. The data provided included the user’s role, the data accessed, and the commercial EHR used to access the HIE. The median number of accesses per general physician (GP) was calculated per week, for each EHR-HIE interface (EHR 1, EHR 2, and EHR 3) and for each clinical data domain (Drugs, Laboratory, and Images).

Second, a modified heuristic analysis was performed using observation sessions with end-users, because in Quebec HIE access is not allowed if you are not a clinician. Experts had no access to the EMHR/HIE integrated interface without a clinician actually accessing it for a given patient. Usability problems were identified using the ergonomic criteria developed by Scapin and Bastien for assessing the usability of interfaces [11] (Table 1).

Table 1 Ergonomic criteria developed by Scaping and Bastien [7]

Criteria	Description
Guidance	Available to advise, orient, inform, instruct, and guide the user throughout their interactions, including from a lexical point of view.
Workload	Concerns all interface elements that play a role in reducing the user’s perceptual or cognitive load.
Explicit control	Concerns both the system processing of explicit user actions and the control that users have over the processing of their actions.
Adaptability	Refers to the system’s capacity to behave contextually and according to the users’ needs and preferences.
Error management	Refers to the means available to prevent or reduce errors and to recover from them when they occur. Errors are defined in this context as invalid data entry, an invalid format for data entry, incorrect command syntax, etc.
Consistency	The way interface design choices (codes, naming, formats, procedures, etc.) are maintained in similar contexts and are different when applied to different settings.
The significance of codes	Between a term and/or a sign and its reference. Codes and names are significant to the users when there is a strong semantic relationship.
Compatibility	Refers to the match between the real task and system inputs or outputs.

Think-aloud observation sessions were organized for each EHR. Clinicians had to perform specific tasks. The screen and the voice of the user were captured during the sessions. For each EHR, three physician sessions were recorded around the following subtasks:

1. Accessing the HIE for a given patient
2. Visualizing and importing drug-related information
3. Visualizing and importing laboratory-related information
4. Visualizing and importing clinical imaging-related information

Finally, a qualitative analysis was conducted of user perceptions of usefulness, benefits, and barriers. Twelve semi-structured individual interviews were conducted with the aim of identifying perceptions of usefulness as well as barriers and facilitators to the use of the HIE system. The participants were general practitioners (GP) and family medicine residents (FMR).

The distribution of the participants in the different HIE systems was as follows:

- EHR 1: 6 participants (4 GPs and 2 FMRs)
- EHR 2: 3 participants (3 GPs)
- EHR 3: 3 participants (2 GPs and 1 FMR)

An analysis was performed of the content of the interviews using a qualitative analysis software program (ATLAS TI) based on the following dimensions:

1. Completeness of clinical data: the underlying reasons for incomplete information can include that HIE systems do not integrate all essential sources of data, perhaps over a concern for losing referrals to competitors[1],
2. Organization and workflow: the perception of changes in work efficiency and quality of care,
3. Technology and user needs: forms of HIE technology (direct exchange or push vs. query-based exchange) and information needs (more information vs. meaningful information),
4. Perceived usefulness.

Results

HIE use

Medication-related information was the most used clinical data domain. EHR1 presented the highest level of HIE usage, followed by EHR 2, while EHR 3 showed the lowest level of use (Figure 1). In the laboratory test domain, EHR 1 and EHR 2 showed the highest level of HIE usage, while EHR 3 showed the lowest level of use. In the imaging test, the level of usage was similar among the three EHRs.

Figure 1 presents the differences in HIE usage among the EHRs by clinical data domain.

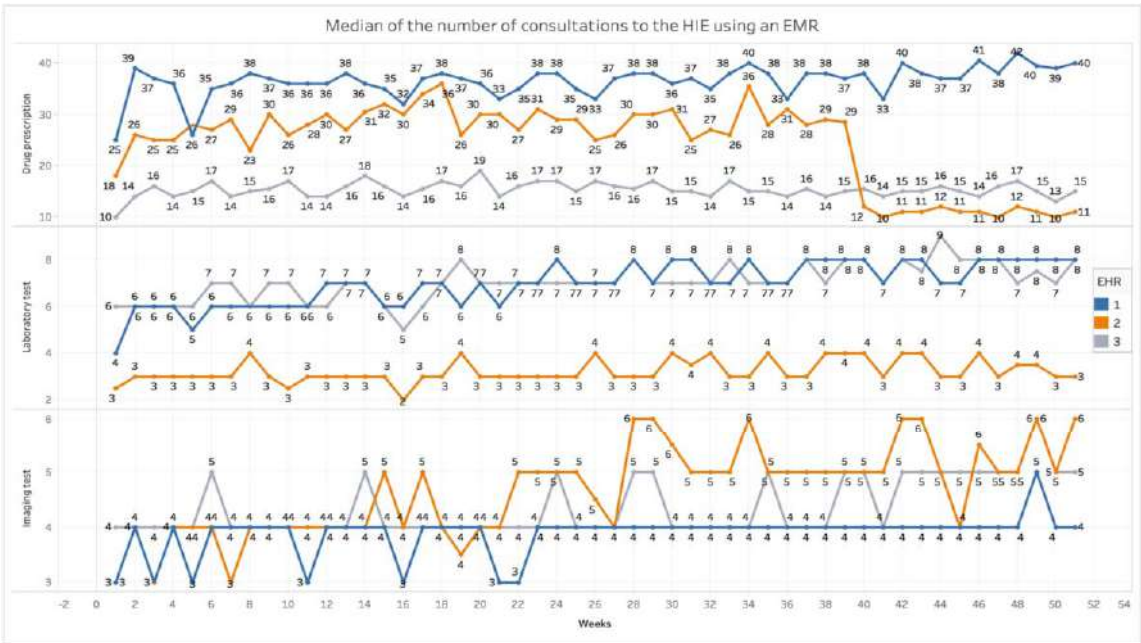


Figure 1 HIE-EHR integration usage: medians of the number of consultations of HIE using an EHR (different accesses) per week, by EHR

Usability problems

The fewest usability problems were observed in the EHR1 (Figure 2)

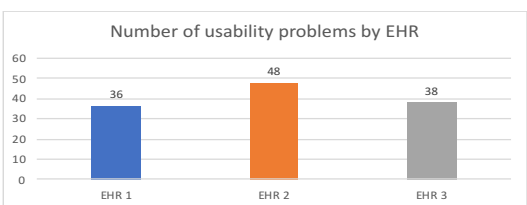


Figure 2 Usability issues by EHR

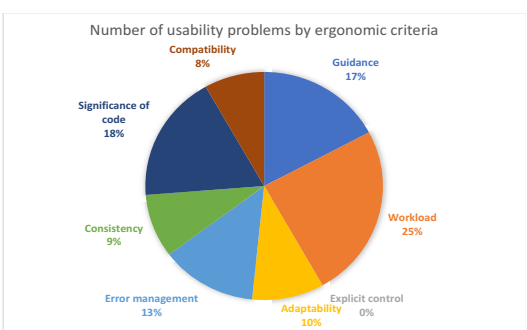


Figure 3 Usability issues by the problem domain (ergonomic criteria)

The main usability issues were code significance, workload, and guidance problems (Figure 3).

Regarding the severity of the identified problems, minor severity problems were the most frequent, followed by major severity problems (Figure 4). High severity, potentially catastrophic problems were rare.

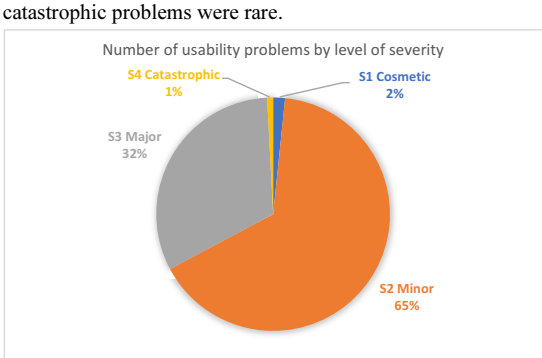


Figure 4 Analysis of the severity of usability problems

Drug prescription was the domain with the highest number of usability problems (43%), and visualization of the medication list was the most problematic feature, with the issues the most important to usability, especially guidance, workload, and code significance problems (Figure 5).

EHR 1 is the drug prescription integration that have fewer usability problems compared to the other two EHRs. We must also remember that drug prescription is the domain with the largest use.

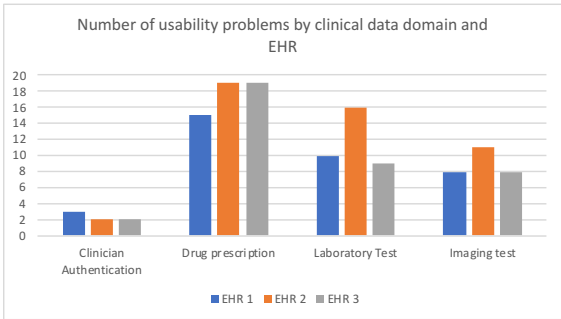


Figure 5 Usability problems according to EHR-HIE integration functionalities

Clinician authentication presented a common challenge. Different logins were needed for the EHR and HIE systems. This generated an extra burden for clinicians.

Perception of usefulness: advantages and barriers

All the physicians perceived that the EHR-HIE integration was meeting, for each clinical data domain, the objectives for which it was designed:

- Drugs: Access to information about prescriptions and dispensations.
- Laboratory: Access to laboratory results.
- Medical imaging: Access to information on imaging results, the status of the examination and the report.

However, there were differences in perceptions of the utility and the number of barriers to its use, and they had an impact on the level of utilization of EHR-HIE integration.

Also, some aspects of the implementation process are important to consider. Lack of familiarity and a lack of knowledge of some functionalities of integration were observed. On the other hand, the progressive availability of the different domains of clinical data in the HIE may have had an impact on perceptions of utility and use.

Medication

The EHR-HIE integration used the most by the physicians was the medication domain.

General practitioners identified groups of patients for which integration had been useful:

- Patients with multiple morbidities and multiple medications or for whom there were doubts about compliance.
- Frequent users of narcotics.
- New patients, patients without an appointment or former patients who had not consulted for a long time.
- Patients who did not know their medications.
- Patients who had seen other specialists or who had medicines prescribed by another doctor.

The most critical barriers were lack of certainty about the information available (the medication list) and the duplication of drugs. One of the main concerns was about whether the information was complete (i.e., the need to have all the information about the medications that the patient is receiving or has ever received). In all three clinics, certainty about having access to an updated medication list was a common theme; especially in relation to all the extra tasks and workarounds and the resulting overload of clinical work needed to develop such a medication list.

Interestingly, there seems to be a threshold beyond which perceived benefits can lower the number of reported barriers, a state which was observed for EHR 1. In contrast, when there was a higher level of barriers, as in the case of EHR 3, there were fewer reported benefits. EHR 2 showed an intermediate position between the other two EHRs. In this domain, greater options for integration functions generated fewer barriers to usage and better perceptions of usefulness.

Laboratory

Although more requests for laboratory tests were generated per patient than for imaging tests, this was the domain that was the least used by the physicians, who preferred the legacy local laboratory systems and data accumulated in their own EHR.

The lack of adaptation to clinical needs was an important factor affecting EHR-HIE use. The functionalities valued in clinical practice included having the ability to observe trends in graphs and the need to record whether the doctor had visualized the exams, and these features were only available in the legacy local laboratory systems and local EHR. These functionalities were not available with the HIE integration.

Some benefits were observed only in the case of EHR 1 (efficiency gains in clinical practice and avoiding repetitions of laboratory tests). The lack of knowledge of some functionalities and a deficit in the availability of some laboratory tests were mentioned as factors diminishing the use of EHR-HIE integration.

Imaging

Consulting reports or results of imaging exams are essential for informed clinical decision-making. Visualizing medical images is not as important a need, especially for complex images, for which a radiologist's opinion is needed anyway. For less complex images, it can be useful to see the images before receiving the radiologist report. In any case, the inaccessibility of images in the HIE remained a barrier, more so since other systems offered such access.

Perceived benefits included avoiding the duplication of imaging exams and developing medical notes on the basis of a copy-and-paste function.

Discussion

The most important finding of this study is the importance of the value added for clinicians by EHR-HIE integration. The medication domain was where we observed the largest number of usability problems, but it was also the domain in which we observed the greatest use of EHR-HIE functionalities.

The aspects of using HIE integration that were the most valued for their usefulness were:

- The opportunity to improve adherence to treatment in the case of chronic conditions.
- The usefulness of prescription and dispensing information on medications for better management of patients who were frequent users of narcotics.
- Avoiding duplicated image and laboratory tests.

The importance of getting access to an updated medication list that was observed in our study is consistent with a study that evaluated the accuracy of drug data in Quebec [15], in which a total of 38% of the information represented discrepancies when the list of references was compared with the list obtained from the HIE. From this perspective, we underline that HIE medication data was the most important source of added value for clinical practice as perceived in the clinics we studied.

In relation to laboratory tests and images, the positive appreciation of EHR-HIE integration as a tool to reduce the duplication of tests is another finding consistent with the literature. A systematic review on the benefits of HIE published in 2018 listed: the reduced duplication of imaging exams, therapeutic medical procedures, total number of orders, and reduced total cost of care, lab test costs and imaging test costs [16].

Our study was focused on the experience of primary care users with EHR-HIE integration. We did not directly study the implementation processes. However, according to the interviews and observations of EHR-HIE integration in use, we observed some barriers resulting from the implementation processes, such as a lack of knowledge of the EHR-HIE integration functionalities and insufficient follow-up activities after the initial training provided when the system was introduced. The literature has shown those factors to be essential to the use of HIE.

This study has some limitations. Given that we did not directly observe the use of the EHR-HIE integration functions, the method chosen to conduct the heuristic analysis can limit the quality of usability problem evaluations. As well, the difficulties encountered balancing the recruitment of participants for each EHR can induce some variations in the analysis of the perceptions of utility, barriers and facilitators associated with the EHR systems studied. Finally, we underline that the dissemination of HIE in Quebec, Canada is not complete. For instance, the availability of data in different domains was still a work in progress. This phenomenon could have affected perceptions of usefulness and usage in the clinics studied.

Conclusions

The deployment of HIE is still in a work in progress in Quebec. Even if usage levels need to improve, clinicians have already identified areas in which system usage proved useful to their clinical practice. There are grounds for hope. To support this hope, it is important to address the problems of usability and barriers to utilisation that we have identified. It is only by improving the design of EHR-HIE integration or its implementation processes that this initiative will be an unqualified success for the province's health system.

References

[1] K. Eden, A. Totten, S. Kassakian, P. Gorman, M. McDonagh, B. Devine, M. Pappas, M. Daeges, S. Woods, and W. Hersh, Barriers and facilitators to exchanging health information: a systematic review, *International Journal of Medical Informatics* **88** (2016), 44-51.

[2] C. Sicotte and G. Pare, Success in health information exchange projects: solving the implementation puzzle, *Soc Sci Med* **70** (2010), 1159-1165.

[3] O. Ben-Assuli, Electronic health records, adoption, quality of care, legal and privacy issues and their implementation in emergency departments, *Health Policy* **119** (2015), 287-297.

[4] O. Ben-Assuli, I. Shabtai, M.J.B.M.I. Leshno, and D. Making, The impact of EHR and HIE on reducing avoidable admissions: controlling main differential diagnoses, **13** (2013), 49.

[5] P. Fontaine, S.E. Ross, T. Zink, and L.M. Schilling, Systematic Review of Health Information Exchange in Primary Care Practices, *The Journal of the American Board of Family Medicine* **23** (2010), 655.

[6] J.M. Palm, I. Colombet, C. Sicotte, and P. Degoulet, Determinants of user satisfaction with a Clinical Information System, *AMIA Annu Symp Proc* (2006), 614-618.

[7] M. Soto, *Facteurs liés à l'adoption du dossier médical électronique (DME): une étude de cas sur le processus d'implantation d'un DME dans un groupe de médecine de famille.*, M. Sc., Université de Montréal, 2012.

[8] G. Pare, C. Sicotte, and H. Jacques, The effects of creating psychological ownership on physicians' acceptance of clinical information systems, *J Am Med Inform Assoc* **13** (2006), 197-205.

[9] P.-Y. Yen and S. Bakken, Review of health information technology usability study methodologies, *Journal of the American Medical Informatics Association* **19** (2011), 413-422.

[10] M.A. Ellsworth, M. Dziadzko, J.C. O'Horo, A.M. Farrell, J. Zhang, and V. Herasevich, An appraisal of published usability evaluations of electronic health records via systematic review, *Journal of the American Medical Informatics Association* **24** (2016), 218-226.

[11] D.L. Scapin and J.C. Bastien, Ergonomic criteria for evaluating the ergonomic quality of interactive systems, *Behaviour & information technology* **16** (1997), 220-231.

[12] J. Nielsen and R. Molich, Heuristic evaluation of user interfaces, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Seattle, Washington, USA, 1990, pp. 249-256.

[13] Inforoute Santé du Canada, Étude d'évaluation des avantages de l'information de santé connectée au Canada, 2018.

[14] MSSS, Dossier Santé Québec. Ministère de la Santé et des Services sociaux, 2018.

[15] A. Motulsky, D.L. Weir, I. Couture, C. Sicotte, M.-P. Gagnon, D.L. Buckeridge, and R. Tamblyn, Usage and accuracy of medication data from nationwide health information exchange in Quebec, Canada, *Journal of the American Medical Informatics Association* **25** (2018), 722-729.

[16] N. Menachemi, S. Rahurkar, C.A. Harle, and J.R. Vest, The benefits of health information exchange: an updated systematic review, *Journal of the American Medical Informatics Association* **25** (2018), 1259-1265.

Address for correspondence

Corresponding author: Mauricio Soto, email: msotod@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190383

Developing Customizable Cancer Information Extraction Modules for Pathology Reports Using CLAMP

Ergin Soysal[a], Jeremy L. Warner[b,c,d], Jingqi Wang[a], Min Jiang[a], Krysten Harvey[c], Sandeep Kumar Jain[e], Xiao Dong[a], Hsing-Yi Song[a], Harish Siddhanamatha[a], Liwei Wang[f], Qi Dai[b], Qingxia Chen[g], Xianglin Du PhD[h], Cui Tao[a], Ping Yang[f], Joshua Charles Denny[b,c], Hongfang Liu[f], Hua Xu PhD[a]

[a] *School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, Texas*
[b] *Department of Medicine, Vanderbilt University, Nashville, Tennessee*
[c] *Department of Biomedical Informatics, Vanderbilt University, Nashville, Tennessee*
[d] *Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, Tennessee*
[e] *Vanderbilt School of Medicine, Vanderbilt University, Nashville, Tennessee*
f *Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, Minnesota*
[g] *Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee*
[h] *School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas*

Abstract

Natural language processing (NLP) technologies have been successfully applied to cancer research by enabling automated phenotypic information extraction from narratives in electronic health records (EHRs) such as pathology reports; however, developing customized NLP solutions requires substantial effort. To facilitate the adoption of NLP in cancer research, we have developed a set of customizable modules for extracting comprehensive types of cancer-related information in pathology reports (e.g., tumor size, tumor stage, and biomarkers), by leveraging the existing CLAMP system, which provides user-friendly interfaces for building customized NLP solutions for individual needs. Evaluation using annotated data at Vanderbilt University Medical Center showed that CLAMP-Cancer could extract diverse types of cancer information with good F-measures (0.80-0.98). We then applied CLAMP-Cancer to an information extraction task at Mayo Clinic and showed that we can quickly build a customized NLP system with comparable performance with an existing system at Mayo Clinic. CLAMP-Cancer is freely available for academic use.

Keywords:

Natural Language Processing, Information Storage and Retrieval, Electronic Health Records

Introduction

Cancer is a highly heterogeneous group of related diseases with a growing body of evidence that cancer initiation, progression, metastasis, and response to treatment are directly associated with variability at the molecular level [1]. Technological advances have enabled us to identify an increasing number of molecular biomarkers, leading towards personalized oncology [2]. Personalized cancer research relies heavily on large clinical series of well annotated and high-quality data. One emerging research direction is to leverage large, real-world practice data stored in electronic health records (EHRs) to facilitate clinical and translational research [3]. However, much of this information (e.g., tumor characteristics) can only be found in narrative or semi-structured text in EHRs, such as pathology reports. Despite the move towards synoptic reporting by the pathology community [4], uptake of this formatting is not complete and synoptic reports are usually only issued for definitive resections, as mandated by the American College of Surgeon's Commission on Cancer Standard 2.1. Natural language processing (NLP) technologies that can automatically extract and structure information from narrative documents have been extensively investigated in the medical domain and cancer sub-domain [5, 6]. In particular, many studies have focused on named entity recognition (NER), a fundamental NLP task that locates and classifies named entities to pre-defined categories (e.g., diseases, drugs, and lab tests).

Several NLP systems have been developed to process pathology reports, which contain rich information about tumor specimen characteristics [7]. An earlier study by Xu et al. extended the existing MedLEE system to extract information from pathology reports to support cancer studies [8]. Crowley et al. reported the Cancer Tissue Information Extraction System (caTIES), which focuses on de-identification of cancer specimens for research purposes, along with information retrieval and concept coding functionalities [9]. MedKAT (MedTAS/P) is a rule-based system, which aims to extract cancer characteristics from pathology reports including anatomic site, histology, and grade [10]. Recently, Savova et al reported another rule-based system named DeepPhe, which can extract summary of cancer phenotypes including morphology, topology, procedure, and staging information from clinical documents [11].

Despite the success of reported use cases on cancer information extraction from clinical text, NLP has not been widely used for routine EHR-based cancer research, likely due to implementation barriers. Because of the diversity and complexity of clinical documentation, significant effort is required to develop custom clinical NLP systems or to extend existing systems for individual applications. In a study that aimed to locally adopt a smoking module in the cTAKES system [12], users had to write extensive code to re-train modules and re-define rules [13], which could be challenging for programmers without much NLP expertise. Additionally, completeness and the accuracy of the information remain as major obstacles in computational processing of clinical texts.

CLAMP (Clinical Language Annotation, Modeling, and Processing) [14] is a general clinical NLP system that provides not only high-performance NLP components but also user-friendly interfaces for building customized NLP pipelines for individual needs. In this study, we describe our work on developing a library of information extraction modules for

important tumor informiton in pathology reports within CLAMP. Using these built-in modules and user-friendly interfaces of CLAMP, users can quickly build customized NLP solutions to extract cancer information from their local pathology reports, with minimal programming effort. We believe such a tool would greatly benefit cancer researchers to leverage clinical text for their studies.

Methods

CLAMP follows the pipeline-based architecture as defined by the Apache UIMA™ (Unstructured Information Management Applications) framework, where an NLP system consists of multiple components in a specific order. The development of CLAMP Cancer Modules involved building cancer-specific NLP components and extending CLAMP interface that allows end users to customize and assemble different components into an NLP pipeline for their specific needs.

Developing Cancer-Specific NLP components

As the primary goal is to extract important information related to cancer, we limited the scope to the "diagnosis" section of pathology reports and followed these steps to build CLAMP Cancer Modules:

Define an Information Model for Pathology Reports

We reviewed available models including the Cancer Disease Knowledge Representation Model (CDKR) used in the MedKAT [10] and the information model in caTIES [9]. As our goal is to support clinical research, we specified data elements for pathology reports as recommended by the College of American Pathologists (CAP) (Figure 1). We reviewed all available CAP cancer templates (available at http://www.cap.org/cancerprotocols) as of January 2017, including the biomarker templates, and defined a list of minimum elements and relations to be included in CLAMP-Cancer Modules.

We define *Primary-site* (site of the source of the material extracted from the body) as the root concept, with relations to *Specimen* (often numbered in the reports), *Sub-site* (detailed sub-anatomical sites related to the primary sites), *Procedure* (methods for specimen extraction from the body, e.g., biopsy), and *Histology* diagnosis (the morphology of cancer cells, e.g., "squamous cell carcinoma"). In addition, we also extract several important attributes defining histologic behavior and malignant potential, including *Tumor Grade*, *Tumor Size*, *Invasiveness* (invasion status and sites infiltrated by the tumor), *Tumor Margin* (tumor involvement in excisional margins), as well as *Tumor Biomarkers and Values* (e.g., estrogen and progesterone receptors).

Annotate a Corpus of Pathology Reports following the Information Model

Based on the information model, we developed an annotation guideline, which specifies the types of entities and relations for annotation, as well as different examples. We randomly selected 400 pathology reports from patients with a cancer ICD9 code at Vanderbilt University Medical Center (VUMC) in 2010. Four domain experts manually reviewed each pathology report and annotated all the entities and relations following the guideline, using CLAMP. Following a typical evaluation design for NLP studies, the annotated corpus was then divided into a training set (200 reports) and a test set (200 reports).

Develop Various Entity and Relation Extraction Components of CLAMP Cancer Modules

As CLAMP Cancer Modules extract a broad range of entities and their relations from pathology reports, different approaches were implemented to achieve optimal performance. For named entity recognition, we implemented regular expression-based, dictionary lookup-based, as well as machine learning-based approaches. For relation extraction, both rule-based and machine learning approaches were developed. Hybrid approaches that combine rules and machine learning were also used to optimize the performance. All the resources collected for developing these components (e.g., dictionaries, rules, and models) are available in CLAMP Cancer Modules and users can use them as-is or customize them to improve performance on their local data.

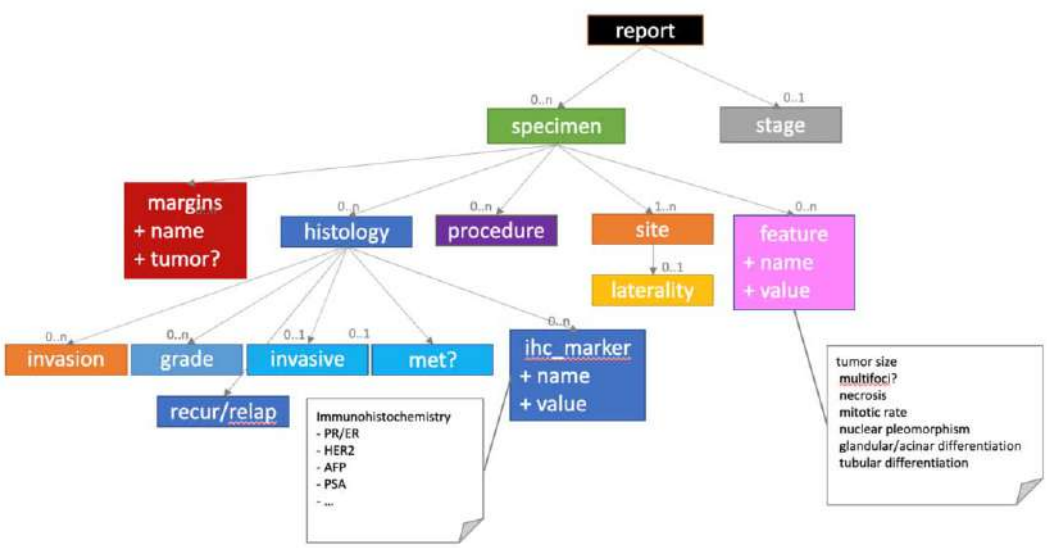


Figure 1. CLAMP information model is based on data element suggestions of College of American Pathologists

Evaluating CLAMP Cancer Modules

The evaluation of CLAMP Cancer Modules has two aspects. First, we developed the default CLAMP Cancer Modules using the training set from VUMC and evaluated them using the test dataset from VUMC. For each type of entity, we report precision, recall, and F-measure on recognizing entities only, using both exact and partial (relaxed) matching criteria [15]. We also report the system performance by considering both entity and relation information (e.g., a detected "sub_site" entity is correct only if both the entity and its relation to the linked "primary_site" are correct). Then, we apply CLAMP Cancer Modules to the same information extraction task as the MedKAT system [10]. Using CLAMP Cancer Modules, we assembled a customized NLP pipeline and compared it with MedKAT, using the same annotated corpus from Mayo Clinic.

Results

The developed CLAMP Cancer Modules consist of several components with machine learning models, dictionaries and rule sets to extract various types of cancer specific information from pathology reports following the CAP recomendations. Figure 2-a shows the cancer-specific modules in CLAMP and how they can be used to build customized NLP pipelines for local pathology reports. In Figure 2-b shows document annotation for cancer specific entities, and relationships among them.

Table 1 – Evaluation results of CLAMP Cancer Modules

Type of information	# of ent.*	Entity only (Exact/Relaxed Matching)			Entity and Relation		
		Precision	Recall	F-measure	Precision	Recall	F-measure
Specimen	310	0.99/0.99	0.99/0.99	0.99/0.99	0.97	0.98	0.98
Primary-site	351	0.98/0.99	0.98/0.99	0.98/0.99	0.98	0.98	0.98
Sub-Site	187	0.96/0.98	0.82/0.83	0.89/0.90	0.88	0.78	0.83
Procedure	339	0.98/0.99	0.98/0.99	0.98/0.99	0.97	0.97	0.97
Histology	553	0.91/1.00	0.85/0.93	0.88/0.97	0.90	0.85	0.86
Tumor Grade	92	0.96/1.00	0.88/0.91	0.92/0.96	0.91	0.83	0.86
Tumor Size	60	0.96/0.96	0.90/0.90	0.93/0.93	0.88	0.83	0.85
Tumor Margin	93	0.92/0.99	0.91/0.98	0.92/0.98	0.80	0.79	0.80
Invasion	71	0.92/1.00	0.83/0.90	0.87/0.95	0.86	0.78	0.82
Biomarker	107	0.95/0.99	0.90/0.94	0.92/0.96	0.88	0.84	0.86

(a) Performance of CLAMP-Cancer components on the VUMC test corpus
* The number of each type of entities in the test corpus of 200 notes.

	CLAMP-Cancer			MedKAT		
	Precision	Recall	F-1	Precision	Recall	F-measure
Tumor size	1.00	0.99	0.99	1.00	1.00	1.00
Dimension Extend	1.00	0.99	0.99	0.99	1.00	0.99
Dimension Unit	1.00	1.00	1.00	1.00	1.00	1.00
Tumor Site	0.94	0.89	0.92	0.96	0.95	0.96
Histology	0.91	0.92	0.92	0.96	0.98	0.97
Grade	1.00	0.88	0.94	0.93	0.97	0.99
Date	1.00	1.00	1.00	1.00	1.00	1.00

(b) Performance comparison of CLAMP Cancer Modules and MedKAT on the same information extraction task from 302 pathology reports at Mayo Clinic

Table 1 (a) shows the evaluation results of CLAMP Cancer Modules for different types of entities and relations. The default components of CLAMP Cancer Modules achieved good F-measures for entity recognition, with a range of 0.87 to 0.99, for exact matching criterion. A quick manual review shows that a fair number of relevant entities are missed by the system (e.g., a low recall in "invasion"), probably due to the complexity of expression patterns of these entities and the relatively limited size of the training corpus. When the relaxed matching criterion was used, CLAMP Cancer Modules achieved higher performance for some entities (e.g., Histology improved from 0.88 to 0.97), indicating some errors were about entity boundary only. When the relationships of entities were taken into consideration, performance dropped as expected (F-measures within 0.82-0.98), because of some known challenges for relation extraction (e.g., long-distance dependency). Table 1 (b) shows the results of the customized NLP pipeline built for Mayo Clinic's pathology reports using CLAMP Cancer Modules, as well as the results from MedKAT. It took 12 hours for a developer to build the customized NLP pipeline for the specified task using CLAMP Cancer Modules, and it achieved comparable performance as MedKAT, indicating high effectiveness of CLAMP-Cancer for building customized NLP pipelines for pathology reports. Although MedKAT has slightly higher performance in certain entities, the production time and the adaptability remain as the major advantage of using CLAMP.

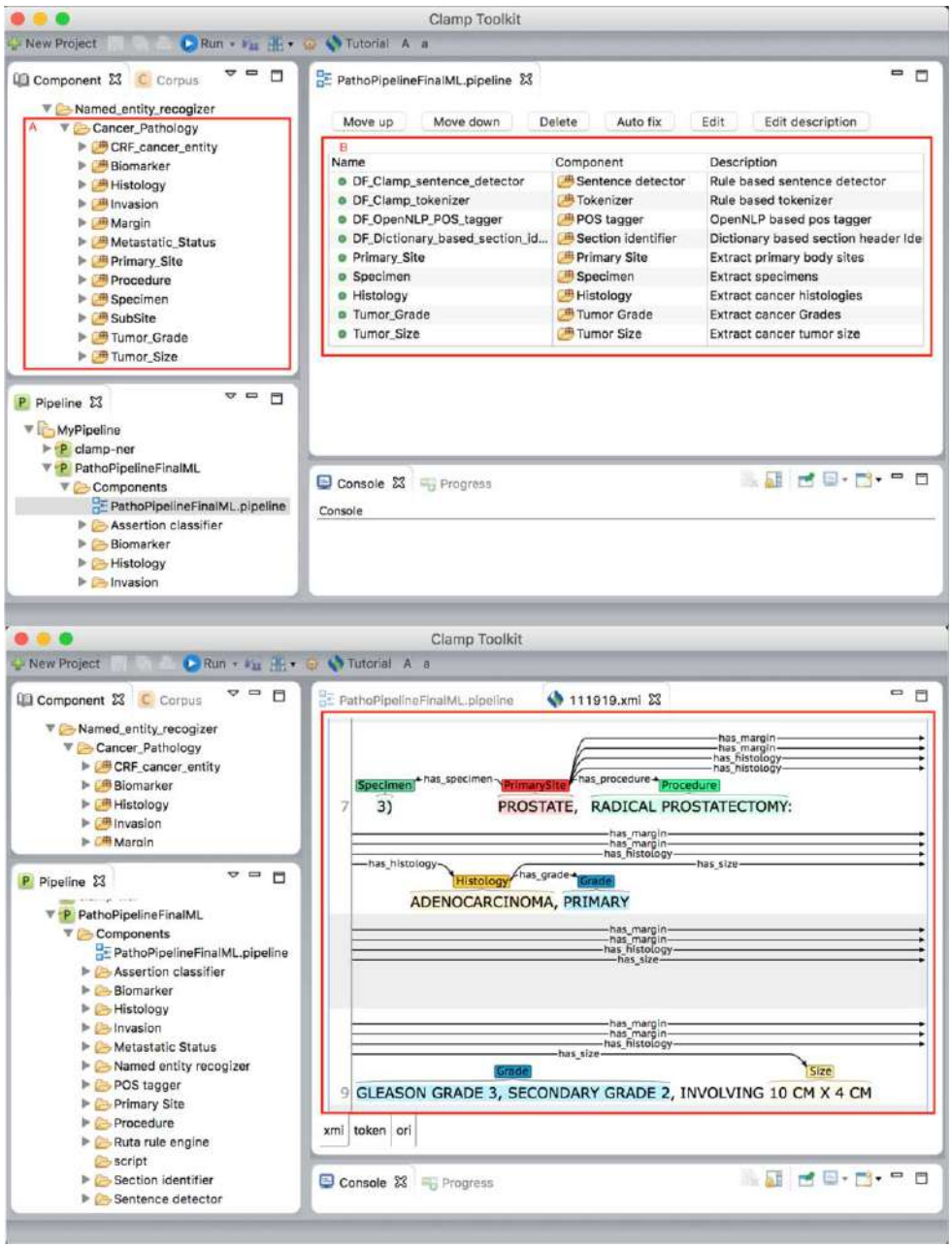


Figure 2. a. CLAMP Cancer module implements several cancer specific components (A), can be used to build customized pipelines (B). b. Pipeline annotates pathology reports to extract entities as well as their relationships.

Discussion and Conclusions

The contribution of CLAMP Cancer Modules is two-fold. First, it is a new NLP system that extracts comprehensive cancer information from pathology reports, including some that are not included by previous systems, e.g., biomarkers. Moreover, CLAMP Cancer Modules allow users to quickly build customized NLP pipelines by leveraging existing components and user-friendly interfaces. To the best of our knowledge, this is the first attempt to build cancer-focused NLP system that integrates into a drag and drop graphical user interface (GUI) for building customized NLP solutions for individualized needs. As pathology reports contain important cancer information, we believe that CLAMP Cancer will facilitate wide adoption of NLP in cancer research.

We acknowledge that cancer knowledge is constantly evolving and new standards, such as the AJCC 8th Edition, may require

re-annotation for additional concepts. As a full-featured NLP system, CLAMP Cancer Modules ease such tasks. To further accelerate EHR-based cancer research, we plan to extend CLAMP Cancer Modules in several directions. First, we will develop an encoding function to further map extracted cancer information to existing terminologies, e.g., ICD-O-3, or to standard value sets. We will then expand CLAMP Cancer Modules to include other types of clinical reports containing rich cancer information, e.g., radiology reports. We also plan to conduct formal usability studies of CLAMP Cancer Modules and iteratively improve the user interface based on users' feedback.

Acknowledgements

This study is partially supported by grants from NCI U24 CA194215, NIGMS R01 GM103859, NIGMS R01 GM102282, R01 LM011829, CPRIT R1307, and UTHealth Innovation for Cancer Prevention Research Training Program Pre-doctoral Fellowship (CPRIT # RP160015).

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the Cancer Prevention and Research Institute of Texas.

Disclosure

Dr. Soysal, Mr. Wang, Dr. Xu, and UTHealth have research-related financial interests in Melax Technologies, Inc.

References

[1] N. Goossens, S. Nakagawa, X. Sun, and Y. Hoshida, "Cancer biomarker discovery and validation," (in eng), *Transl Cancer Res,* vol. 4, no. 3, pp. 256-269, Jun 2015.

[2] M. Kalia, "Biomarkers for personalized oncology: recent advances and future challenges," (in eng), *Metabolism,* vol. 64, no. 3 Suppl 1, pp. S16-21, Mar 2015.

[3] P. Yu, D. Artz, and J. Warner, "Electronic health records (EHRs): supporting ASCO's vision of cancer care," (in eng), *Am Soc Clin Oncol Educ Book,* pp. 225-31, 2014.

[4] J. R. Srigley *et al.*, "Standardized synoptic cancer pathology reporting: a population-based approach," (in eng), *J Surg Oncol,* vol. 99, no. 8, pp. 517-24, Jun 15 2009.

[5] W. W. Yim, M. Yetisgen, W. P. Harris, and S. W. Kwan, "Natural Language Processing in Oncology: A Review," (in eng), *JAMA Oncol,* vol. 2, no. 6, pp. 797-804, Jun 01 2016.

[6] I. Spasić, J. Livsey, J. A. Keane, and G. Nenadić, "Text mining of cancer-related information: Review of current status and future directions," *International Journal of Medical Informatics,* vol. 83, no. 9, pp. 605–623, September 2014 2014.

[7] G. Burger, A. Abu-Hanna, N. d. Keizer, and R. Cornet, "Natural language processing in pathology: a scoping review," (in en), *Journal of Clinical Pathology,* vol. 69, no. 11, pp. 949-955, 2016-07-22 2016.

[8] H. Xu, K. Anderson, V. R. Grann, and C. Friedman, "Facilitating cancer research using natural language processing of pathology reports," *Studies in health technology and informatics,* vol. 107, no. Pt 1, pp. 565-572, 2003.

[9] R. S. Crowley, M. Castine, K. Mitchell, G. Chavan, T. McSherry, and M. Feldman, "caTIES: a grid based system for coding and retrieval of surgical pathology reports and tissue specimens in support of translational research," *J Am Med Inform Assoc,* vol. 17, no. 3, pp. 253-64, May-Jun 2010.

[10] A. Coden *et al.*, "Automatically extracting cancer disease characteristics from pathology reports into a Disease Knowledge Representation Model," *Journal of Biomedical Informatics,* vol. 42, no. 5, pp. 937–949, October 2009 2009.

[11] G. K. Savova *et al.*, "DeepPhe: A Natural Language Processing System for Extracting Cancer Phenotypes from Clinical Records," *Cancer Research,* vol. 77, no. 21, pp. e115-e118, 2017.

[12] G. K. Savova, P. V. Ogren, P. H. Duffy, J. D. Buntrock, and C. G. Chute, "Mayo clinic NLP system for patient smoking status identification," (in eng), *J Am Med Inform Assoc,* vol. 15, no. 1, pp. 25-8, 2008 Jan-Feb 2008.

[13] M. Liu *et al.*, "A Study of Transportability of an Existing Smoking Status Detection Module across Institutions," presented at the AMIA Annu Symp Proc, 2012. Available: http://dx.doi.org/

[14] E. Soysal *et al.*, "CLAMP - a toolkit for efficiently building customized clinical natural language processing pipelines," (in eng), *J Am Med Inform Assoc,* Nov 2017.

[15] Ö. Uzuner, B. R. South, S. Shen, and S. L. DuVall, "2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text," (in eng), *J Am Med Inform Assoc,* vol. 18, no. 5, pp. 552-6, 2011 Sep-Oct 2011.

Address for Correspondence

Hua Xu, PhD

The University of Texas Health Science Center at Houston

7000 Fannin St, Houston, TX 77030, USA

hua.xu@uth.tmc.edu

Hongfang Liu, PhD

Mayo Clinic College of Medicine

200 First St SW, Harvick 3rd, Rochester, MN, 55905, USA.

Liu.Hongfang@mayo.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190384

Bridging Documentation and Metadata Standards: Experiences from a Funding Initiative for Registries

Jürgen Stausberg[a], Sonja Harkener[a]

[a] University of Duisburg-Essen, Faculty of Medicine,
Institute for Medical Informatics, Biometry and Epidemiology (IMIBE), Essen, Germany

Abstract

Within a funding initiative for patient registries in Germany, data specifications from 15 projects were collected in a structured format. Subsequently, the data specifications were transferred into a structure following the ISO/IEC 11179-3 standard for metadata registries. The data specifications included a median of 16 documentation objects and 196 data elements per project. Only Common designations were used. The bridging to ISO/IEC 11179-3 revealed several overlaps between the specifications. For example, bridging can be used to align the multifold designations of sex and sex categories as well as to harmonize the respective value lists. Further work intends to create a metadata repository based on a community-driven approach as part of the IT-infrastructure of this funding initiative. Without this infrastructure, comparability of the data specifications is unlikely.

Keywords:

Documentation, metadata, registries

Introduction

From the perspective of medical informatics, registries are a type of medical documentation system aiming at the analysis of research questions regarding a group of individuals, typically patients. From this perspective, registries could be differentiated from health records supporting health care for individuals. In a broader view, registries are projects that maintain staff, rooms, and IT infrastructure and sometimes have a specific legal status [1]. However, even in the broader view, data recorded on the observational units are the assets of registries. The definition of those data in terms of variables and value lists is therefore a main task in the development of registries [1, 2, 3]. Ideally, this definition is supported by documentation standards. Documentation standards guarantee that the data on the one hand cover all information needed to answer the research questions. On the other hand the data must support the analysis methods applied in statistical evaluations in reports and in scientific and public publications. Proposals for those documentation standards are available [4].

On a level going beyond a single registry, the interaction between projects and between different types of documentation systems gains importance. This interaction requires a description of the data independent from individual research questions and individual statistical analyses. Use cases on that level can include the identification of projects sharing the same approach to quality of life or addressing the same population, for example patients suffering from dementia [5]. ISO/IEC 11179 "Information technology - Metadata registries (MDR)" [6] is a standard that relates to such kinds of use cases. ISO/IEC offers an approach to maintain data about data, i.e. metadata, in a systematic and unambiguous approach. Although ISO/IEC 11179 is noticed in health care [7, 8], the relationship between documentation standards and the metadata approach is complex [9].

With the goal of supporting both the development of individual registries and the interaction between different registries, a German funding initiative defined a bridging between documentation standards and metadata standards that will be introduced below. The bridging will be used to establish a metadata repository for the funding initiative in the future.

Methods

Registries

The German Ministry for Research and Education launched a funding initiative for registries in health services research. Sixteen projects were accepted for a concept development phase lasting nine months. At the end of this phase, the projects were invited to submit a proposal for an implementation phase lasting up to five years. In parallel to the funding of individual registries, an accompanying project (abbreviated REGISVF-AP) was accepted that supports the registries in development and implementation. REGISVF-AP started one month later than the registries which began in September 2017 and will last two years. Thus, it covered eight of the nine months of the concept development phase. The decision about funding of the registry implementation will be announced at the end of 2018.

The 16 projects of the development phase covered different areas of health care: rare diseases, oncology, acute conditions, chronic diseases, interventions, and other conditions. Further information about the projects is provided in table 1.

Table 1 – Projects of the development phase

Acronym	Institution	Area
Rare diseases		
sLEGER	University clinics Düsseldorf	Systemic lupus erythematosus in Germany
PAREMIS	University clinics Leipzig	Prader-Willi-Syndrome
Oncology		
HerediCaRe	University clinics Cologne	Hereditary breast and ovarian cancer
BRE-4-MED	University clinics Würzburg	Breast cancer care for patients with metastases
Acute conditions		
RADaR	University clinics Regensburg	Acute respiratory distress syndrome
FieberApp-Register	Private university Witten/Herdecke	Fever in childhood
HIRB	Brandenburg Medical School	Heart attacks in Brandenburg
EMBO-Lung	Ludwig-Maximilians-Universität Munich	Pulmonary embolism
RECUR	University clinics Freiburg	Recurrent urolithiasis of the upper urinary tract
Chronic diseases		
TOFU	University clinics Bonn	Treatment exit options for uveitis
GeCeR	Competence network bowel diseases	German celiac registry
ParaReg	University clinics Heidelberg	Lifelong monitoring of paraplegic patients
Interventions		
INDICATE-TKR	Technical University Dresden	Appropriateness of total knee replacement for osteoarthritis
IISAAR	Saarland University	Vaccination information system Saarland
Other conditions		
SoLKID-GNR	University clinics Münster	Safety of living kidney donors in Germany
CDDD_ENMR	Center for Population and Health Wiesbaden	National mortality registry

Documentation Standard

REGISVF-AP suggested the use of the catalogue of attributes from Leiner and Haux [10] as a format for data definition in the registries' concepts. The registries agreed to use this approach in an adapted structure. The catalogue of data elements offers a view on data definitions that is communicable to domain experts on the one hand and usable for computer scientists to implement the data definitions in the data management infrastructure on the other hand. The catalogue of data elements differentiates between documentation objects like "patient" or "quality of life" and data elements like "sex" or "VAS" (= visual analogue scale). Data elements are assigned to one and only one documentation object. Documentation objects and data elements share several attributes like designation and description. Both could be related to predefined research questions, to project modules, to statistical methods defined in a statistical analysis plan, to health care related categories such as therapies, and to project-related events such as follow-ups. It was furthermore possible to enrich each data element with a) a value set that is either a list or a data type, b) a coding of values within a value list, c) a unit in case of a numerical data type, d) a differentiation between single and multiple occurrences, and e) plausibility checks.

The registries were advised to start the development with the agreement on research questions, to extract target variables, influencing variables and confounders as potential data elements, and to check any proposed data elements according to the following aspects:

- Is the data element needed to answer a research question?
- Does the data element adequately cover the information?
- Is the recording of the data element feasible for the study sites?
- Could the data element be used in the process of analysis?

Templates for the registries' catalogues of data elements were provided as Microsoft Excel and Microsoft Access files. The templates were available in German and English. A short explanation supported the use of the templates. Additionally, a few integrity constraints were implemented in the Access file.

Metadata Standard

The Data Description metamodel and the Concepts metamodel parts of the ISO/IEC 11179 third edition (ISO/IEC 11179-3) were used [6]. In short, a DATA_ELEMENT is defined by a DATA_ELEMENT_CONCEPT (e.g. sex of a patient) which uses a CONCEPTUAL_DOMAIN (sex) combined with a specific VALUE_DOMAIN (e.g. consisting of the PERMISSIBLE_VALUES male, female, diverse, undetermined, and unknown, representing the legislative definition of sex for citizens in Germany). A VALUE_DOMAIN has two possible subclasses, namely the ENUMERATED_VALUE_DOMAIN (represented by a list of distinct values) and the DESCRIBED_VALUE_DOMAIN (represented by a textual definition).

Bridging Approach

The documentation object of the data element catalogue was mapped to the OBJECT_CLASS of ISO/IEC 11179-3. Each data element of this documentation object establishes a PROPERTY leading to a DATA_ELEMENT_CONCEPT combining both. The value set of the data element catalogue was defined as ENUMERATED_VALUE_DOMAIN in case of a list of values and as DESCRIBED_VALUE_DOMAIN otherwise. Values of the value list became PERMISSIBLE_VALUES of ISO/IEC 11179-3. These transformations were automatically performed based on the received specification of the registries. To close the loop between the DATA_ELEMENT_CONCEPT and the VALUE_DOMAIN at the conceptual level of the Data Description metamodel, a CONCEPTUAL_DOMAIN was created in a stepwise-approach. Firstly, all designations of data elements that occur at least twice established automatically a CONCEPTUAL_DOMAIN. This CONCEPTUAL_DOMAIN was assigned to all data elements that share the designation. In a second step, all remaining data elements were assigned to a preexisting CONCEPTUAL_DOMAIN if possible. Otherwise a new CONCEPTUAL_DOMAIN was created. Finally, the CONCEPTUAL_DOMAINS were manually corrected for duplicates and overlaps. To support understandability and retrieval of the results, each CONCEPTUAL_DOMAIN was

related to a single broader CONCEPT via a LINK of the ISO/IEC Concept metamodel region. In the preliminary stage of the presented work, the CONCEPT was arbitrarily defined by the authors.

Table 2 shows an overview of the bridging concept. The bridging was performed semi-automatically using Microsoft Access.

Table 2 – Bridging concept

Catalogue of data elements	ISO/IEC 11179-3	Comment
Documentation object	OBJECT_CLASS	
Data element	PROPERTY	
Value set	ENUMERATED_ VALUE_DOMAIN	in case of a list of values
Value set	DESCRIBED_ VALUE_DOMAIN	in case of data types
	PERMISSBLE_ VALUE	items in the list of values
	CONCEPTUAL_ DOMAIN	semi-automatically determined
	CONCEPT	manually created

Results

Fifteen out of the 16 projects submitted a specification of their data. Two projects submitted a Word file with free text; 13 projects made use of the template of the catalogue of data elements: seven submitted an Excel file, five submitted an Access file, one submitted an Excel and an Access file in parallel. Six projects defined their data in German, six in English, and three in both languages. Nine projects primarily used phrases for the designation of data elements, one used questions, one used phrases and questions, and four used labels.

Catalogues of Data Elements

The catalogues of data elements defined a total of 352 documentation objects. Excluding a project that used 126 documentation objects, the number of documentation objects ranged between 8 and 27 (median: 16, variation coefficient: 0.31). Fifty-one documentation objects were additionally described in a second language (German or English) leading to 403 designations. The most frequent designation was “basis data” (four in English, six in German as “Stammdaten”). A unique phrase or question was present in 385 out of the 403 designations.

The documentation objects included 3,935 data elements (range per project: 48 to 756, median: 196, variation coefficient: 0.73), 842 additionally available in a second language. Excluding duplicates for the combination of designation and language, 4,068 different designations remained from 4,777. From 4,068 designations, 3,929 occurred only once, and 53 occurred several times. The most frequent designations were “Geschlecht” (=sex, six times), “Geburtsdatum” (=date of birth, six), “Pseudonym” (German, five), “date of birth” (five), “Vorname” (=first name, four), “gender” (English, four), and “sex” (four). The range of data elements per documentation object was 1 to 191 (median: 16, variation coefficient: 1.62).

Example - The Data Element Sex

All projects defined at least one data element related to the sex of a patient. In total, 25 data elements occurred for sex with eight different designations. The data elements for sex were assigned to 16 different designations of documentation objects, most frequently to “Stammdaten” (=basis data, five times) or “basis data” (four times). The 25 data elements shared 14 different value lists. The languages and acronyms were mapped to the following seven values: männlich/male/m, weiblich/female/w, intersexuell/intersexual, unbekannt/unknown, unbestimmt/not determined/undecided, anderes/other/a, and unclassified. Only one data element used the value list defined on a national level for routine data in secondary care, and not one used the definition defined in Germany for primary care.

Mapping to ISO/IEC 11179-3

1. Identified by using a well formatted structure as the value list, 1,197 ENUMERATED_VALUE_DOMAINS out of 3,935 value sets were identified (30%) in the preliminary stage of the bridging process.
2. Those ENUMERATED_VALUE_DOMAINS included 4,162 different PERMISSIBLE_VALUES.
3. The remaining 70% of the value sets were regarded as DESCRIBED_VALUE_DOMAIN.
4. With a minimum of two occurrences of data element designations independent from languages or project assignments, 670 CONCEPTUAL_DOMAINS were extracted. Those covered 1,709 data elements out of 4,777.
5. By now, another 761 data elements have been manually assigned to an existing or newly created CONCEPTUAL_DOMAIN.
6. After clearing of duplicates, 464 CONCEPTUAL_DOMAINS remained covering 2,470 data elements.
7. Preliminarily, 325 of the 464 CONCEPTUAL_DOMAINS covering 2,143 data elements were aggregated to 21 CONCEPTS with drug (533 data elements) and disease/symptom (302) as the most frequent ones.

Discussion

Fifteen projects submitted their specifications of variables and values as part of the registry development. Thirteen used the recommended format of a catalogue of data elements for the submission. According to the coefficient of variation, the projects agreed concerning a reasonable number of documentation objects. To the contrary, the total number of data elements varied strongly between the projects. Additionally, the number of data elements that were condensed to a documentation object was quite different. In agreement with a survey including 30 already implemented

registries [11], most of the registries intended to record between 100 and 499 data elements (cf. figure 1).

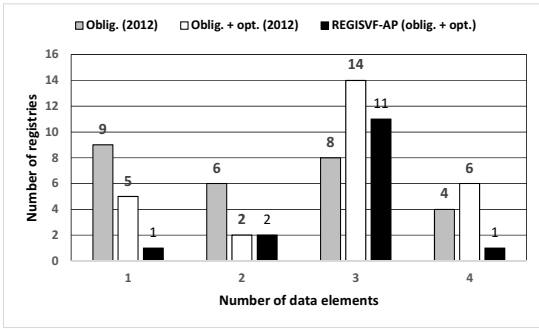


Figure 1 – Number of of data elements per registry (references for 2012 from [11]). Data elements: oblig.=obligatory, opt.=optional.

However, the projects strongly disagreed concerning the designations of documentation objects, data elements and values. They also disagreed concerning the value lists that arise through the aggregation of single values. To some extent this could be explained through the diverse use cases like uveitis, fever, or metastasizing breast cancer. But the example of the values found for the data element sex showed that the projects did not even consider national conventions. An evidence-based recommendation about an appropriate documentation standard was not sufficient to establish homogenous metadata.

Therefore, the results clearly emphasize the need for metadata repositories. A metadata repository can be used for several use cases [11]:

- it makes the creation of case report forms and data models easier,
- it improves the quality of case report forms and data models,
- it harmonizes variables and value lists,
- and it supports the mapping of metadata and data.

Nevertheless, the relationship between metadata approaches and documentation standards is ambivalent [4]. For both, the modelling of health-related information could be different. One might consider “disease” as the main entity and “heart attack” as attribute or “heart attack” as main entity and “existence” as attribute. The implementation of models for complex information objects like “blood pressure” or “quality of life” is unclear for both metadata standards and registry definitions.

The interim results of the bridging between the projects’ specifications of data elements and the structure of ISO 11179-3 underline the value of a metadata repository. For many data elements, counterparts in other projects can be identified. The authors expect a significant stimulus for harmonization of variables and value lists through feedback of this information to the projects. However, the time-consuming effort for the mapping demonstrated the necessity of direct involvement of registry staff in maintaining their metadata themselves. Additionally, the high abstraction level of ISO 1179-3 must be concretized to reach acceptance by those users.

Conclusions

A metadata repository is a core component of an IT-infrastructure for registries as well as for health services and health science research in general. The experiences gained from merging data specifications from 15 registries demonstrate impressively that harmonization and comparability will not arise without explicit metadata services. The high resource consumption needed to align data specifications subsequently underpin that a community-driven approach is essential involving the individual projects in the maintenance of the content. REGISVF-AP will further pursue the objective to implement this approach in the funding initiative of registries for health services research in Germany.

Acknowledgements

The work was funded by the German Federal Ministry of Education and Research under contract 01GY1720B. REGISVF-AP is implemented together with the Technology, Methods, and Infrastructure for Networked Medical Research (Berlin, Germany) and the German Network for Health Services Research. We would like to thank all projects funded within REGISVF who submitted their specifications.

References

[1] R.E. Gliklich, N.A. Dreyer, M.B. Leavy, eds., *Registries for Evaluating Patient Outcomes: A User’s Guide. Third edition. Two volumes,* Agency for Healthcare Research and Quality, Rockville, MD, 2014.

[2] D. Muller, M. Augustin, N. Banik, W. Baumann, K. Bestehorn, J. Kieschke, R. Lefering, B. Maier, S. Mathis, S.J. Rustenbach, S. Sauerland, S.C. Semler, J. Stausberg, H. Sturm, C. Unger, and E.A. Neugebauer, [Memorandum registry for health services research], *Gesundheitswesen* **72** (2010), 824-839.

[3] M. Zaletel, M. Kralj, eds., *Methodological guidelines and recommendations for efficient and rational governance of patient registries,* National Institute of Public Health, Ljubljana Slovenia, 2015.

[4] S.M. Ngouongo and J. Stausberg, Terminology-based documentation systems: a systematic comparison of four different approaches, *Stud Health Technol Inform* **190** (2013), 97-99.

[5] J. Stausberg, M. Lobe, P. Verplancke, J. Drepper, H. Herre, and M. Loffler, Foundations of a metadata repository for databases of registers and trials, *Stud Health Technol Inform* **150** (2009), 409-413.

[6] ISO/IEC 11179-3:2013(E), *Information technology – Metadata registries (MDR). Part 3: registry metamodel and basic attributes,* Third edition, 2013-02-15.

[7] D.B. Warzel, C. Andonaydis, B. McCurry, R. Chilukuri, S. Ishmukhamedov, and P. Covitz, Common data element (CDE) management and deployment in clinical trials, *AMIA Annu Symp Proc* (2003), 1048.

[8] Australian Institute of Health and Welfare, *National Health Information Standards and Statistics Committee. Creating nationally-consistent health information: Engaging with the national health information committees,* AIHW, Canberra, 2014.

[9] S.M. Ngouongo, M. Lobe, and J. Stausberg, The ISO/IEC 11179 norm for metadata registries: does it cover healthcare standards in empirical research? *J Biomed Inform* **46** (2013), 318-327.

[10] F. Leiner and R. Haux, Systematic planning of clinical documentation, *Methods Inf Med* **35** (1996), 25-34.

[11] J. Stausberg, Register, Kohorten und Data Repositories, in: *IT-Infrastrukturen in der patientenorientierten Forschung. Aktueller Stand und Handlungsbedarf - 2012/2013,* J. Drepper, S.C. Semler, eds., AKA GmbH, Berlin, 2013, 31-42.

Address for Correspondence

Prof. Dr. med. Jürgen Stausberg, Institute for Medical Informatics, Biometry and Epidemiology, University Hospital of Essen, Hufelandstraße 55, D-45147 Essen, Germany
Email: stausberg@ekmed.de.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190385

Evaluation of Treatment Success Rate Among Antihyperuricemic Using Real-World Data

Kei Teramoto[a], Toshihiro Takeda[b], Yasushi Matsumura[b], Akira Ohtahara[c] , Ichiro Hisatome[d] , Hiroshi Kondoh[a]

[a] The Division of Medical Informatics, Tottori University Hospital, Tottori, Japan
[b] Medical Informatics, Osaka University Graduate School of Medicine, Osaka, Japan
[c] Division of cardiology San-in Rosai Hospital, Tottori, Japan
[d] Division of Regenerative Medicine and Therapeutics, Department of Genetic Medicine and Regenerative Therapeutics, Institute of Regenerative Medicine and Biofunction, Tottori University Graduate School of Medical Science, Yonago, Japan

Abstract

In this study, treatment and the serum uric acid (UA) level were compared using medication history generated by prescription order records of antihyperuricemic to examine the treatment success rate. We examined the treatment success rate among these patients based on the serum UA level during 120–180 days after the initiation of treatment, which was set as the endpoint. The number of patients whose UA level before the start of treatment was > 8.0 mg/dL but decreased to < 6.0 mg/dL after the treatment, which is the target treatment success, was 92 (success rate of 14.2%), 50 (53.2%), 76 (41.5%), 35 (31.9%), and 45 (37.8%) in the allopurinol 100 mg/day (A1) and 200 mg/day (A2), febuxostat 10 mg/day (F1) and 20 mg/day (F2), and benzbromarone 50 mg/day (B), respectively. Compared with that of the other drugs, the treatment success rate was high with A2 and low with A1. From the generated medication history, the treatment success rate with antihyperuricemic can be extracted mechanically.

Keywords:
Electronic health records, Pharmacoepidemiology, Antihyperuricemics.

Introduction

Clinical studies include crossover studies in which different drugs are administered to the same patient in an order and the respective results are compared. This study represents one of the methods that can be used in the treatment of chronic diseases [1]. As the effect of a target drug can be compared among different individuals, individual differences in the effects can be eliminated, which is an advantage. However, with long-term administration, a washout period is necessary and that can prolong the study period. Consequently, the economic costs and burden on patients are high; even for established standard drugs, evidence of treatment efficacy that is required in a clinical setting is not adequate.

Pharmacoepidemiological studies that use real-world data stored in electronic medical charts as target are drawing attention. If patients who are eligible for a crossover study can be extracted automatically from the drug ordering information stored in the electronic medical charts using a computer, then, data similar to those of interventional studies can likely be obtained at a low cost. In drug intervention studies, the relationship between a patient's treatment period and the endpoint is important. However, drug information stored in the electronic medical chart system is a record of prescription orders generated during a physician's prescription. Therefore, it is difficult to select subjects while estimating the treatment period for each patient. In this study, we developed a system for estimating the treatment period of patients by drug and dose using a computer to correct overlapping periods and gaps present in prescription order records [2]. We evaluated the treatment efficacy of antihyperuricemic from the treatment periods that were generated using this system, and herein report our findings [3].

Methods

Study environment

The EMR of Tottori University Hospital (TUH) was introduced on September 1, 2002, and by September 2015, the treatment data of 307,944 patients had been recorded. In this study, the treatment efficacy of antihyperuricemic used at the TUH was the observation subject. The target drugs of evaluation were allopurinol (A), febuxostat (F), and benzbromarone (B). All of them tend to be prescribed as antihyperuricemic for long periods. Both A and F inhibit the production of uric acid (UA), whereas B promotes the excretion of UA. Although A and B are present in the prescription ordering database since the start of EMR operations, the use of F in the hospital started in June 2011 and therefore data only from that time are available.

The outline of processing flow of the system that automatically predicts the outcomes of antihyperuricemic from the EMR data is elucidated in Fig. 1. The prescription records are extracted from the EMR prescription database, and then a record of medication history is generated. The generated record of medication history contains the start and end dates of medication. Subsequently, the generated record of medication history and the record of labottory test data that includes the serum UA level are combined. The UA laboratory test data corresponding to before and during (endpoints) treatment with antihyperuricemic become the subject. By this process, a record by which changes in the UA level before and during treatment can be observed is generated. The details of respective processes are described below.

Method of generating medication history

Prescription records are used in generating medication history. Generally, a prescription record contains the patient ID, drug name, prescription start and end dates, unit dose, and method of administration. The outcome of treatment with a drug has a major effect on the dose of the drug, and even with the same

drug, when the daily doses are different, the endpoints have to be distinguished.To distinguish the same drug name present in a prescription record by daily dose, the information on daily dose is combined with the drug name in the prescription record. For example, if A 200 mg/day is prescribed, the drug name will be "allopurinol-200mg." By this process, the same drug that is prescribed at differing daily doses can be distinguished mechanically.The medication history generation program developed in this study was used to generate and process the medication history. With this program, prescription records of the same patient with the same drug name are extracted, and the overlapping periods from records with old prescription dates and remnant drugs that the patient forgot to take are calculated. If there are gaps in the prescription periods, a decision is made regarding whether or not a correction can be made with the remaining drugs. The operations for combining the records are then performed (Fig. 1-1). The record on medication history comprises the patient ID, name of the drug, and dates of start and end of the treatment. The generated records of medication history are stored in the medication history database. The treatment start and end dates of a patient can be obtained from the record of medication history.

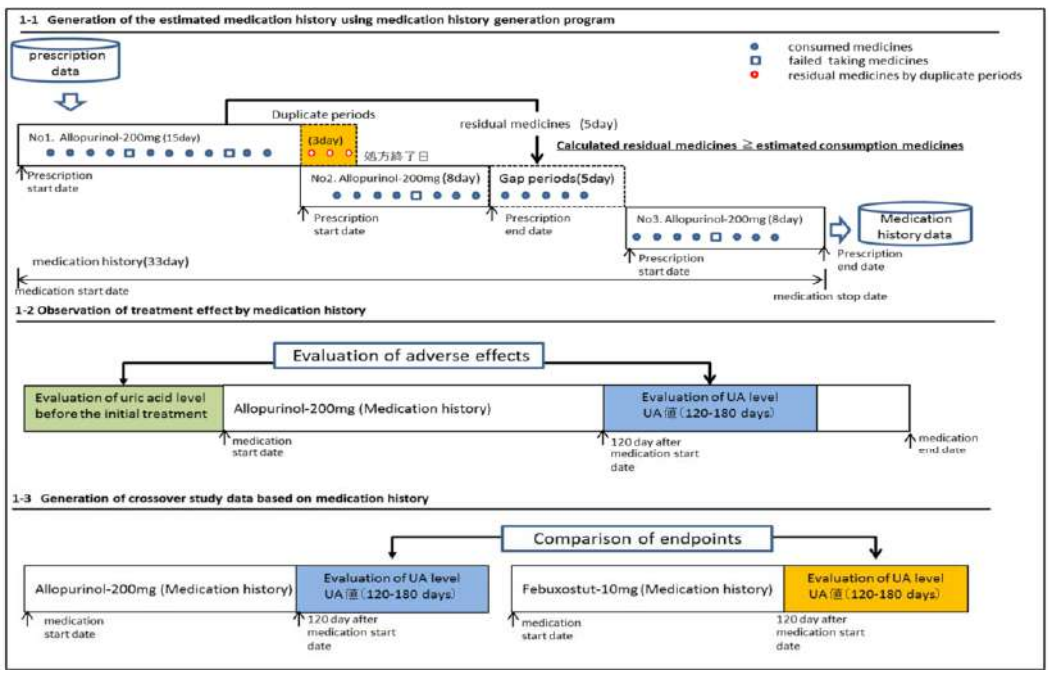


Figure 1 - RCT study based on medication history

Evaluation of treatment outcome based on medication history

The treatment objective of hyperuricemia drugs is to lower the serum UA level to < 6.0 mg/dL between 90 and 180 days by repeated administration. With the treatment start and end dates in the record of medication history as the standard, the treatment efficacy can be observed by comparing data of the serum UA level before and during the treatment, which can be obtained from the EMR sample test data (Fig. I1-2). The criteria presented in Table 1 were used to screen the subject patients.

Table 1 - Screening criteria of subject patients for the endpoint evaluation

	Criteria for extracting subject patients
I	Serum uric acid (UA) level is ≥ 8 mg/dL from 30 days prior to the start of treatment to the day when the treatment is started.
II	From the day when the treatment was started to at least 180 days of repeated treatment, with the same drug and the same dose of hyperuricemia medication.
III	Test records on the UA level during 120‐180 days after the start of treatment are available.

With medication history as the standard, patients on repeated treatment with different antihyperuricemic at different periods were extracted, and a crossover intervention study was performed using the EMR data (Fig. I 1-3). Table 2 shows the screening of subject patients.

Table 2 - Extraction of patients to be subjects of the crossover intervention study with medication history as the standard

	Criteria for extracting the subject patients
I	Serum uric acid (UA) is ≥ 8 mg/dL from 30 days prior to the start of treatment to the day when the treatment is started.
II	From the day when the treatment was started to at least 180 days of repeated treatments, with the same drug and the same dose of hyperuricemia medication.
III	Test records on the UA level during 120‐180 days after the start of treatment are available.
IV	At least two patients meet criteria II and III and have a medication history.
V	During the treatment period between the first evaluation drug and the next evaluation drug, there is no period of concomitant use of another hyperuricemia medication.

To exclude effects due to other antihyperuricemic, with the start and end dates of medication history as the standard, patients on treatment with different antihyperuricemic between the first and next endpoints were excluded from the observation subjects. Moreover, to reduce the effect of treatment with the first drug, as much as possible, the period to the next endpoint evaluation was set at 120–180 days after the initiation of treatment.

Results

Extraction of subject patients for the verification by generating medication history

The number of prescription records of antihyperuricemic extracted from the electronic medical charts was 52,797 (number of patients = 2687), 7400 (675), 5530 (655), 2304 (316), and 10,997 (519) prescriptions of A 100 mg/day (A1) and 200 mg/day (A2), F 10 mg/day (F1) and 20 mg/day (F2), and B 50 mg/day (B), respectively (Table 1). The mean treatment period was 898, 398, 438, 419, and 897 days with A1, A2, F1, F2, and B, respectively. Compared with that of A2, F1, and F2, A1 and B tended to have a long period of administration ($P < 0.01$). A and B were commercialized more proactively than F, and they have a prolonged treatment period. As A2 is commonly a dose increase from A1, its administration period is shorter than that of A1. The number of patients who met the screening criteria in Table 1 and those who could be in the endpoint evaluation were 649 (24%, observation subject patients/prescription subject patients), 94 (14%), 183 (28%), 69 (22%), and 119 (23), for A1, A2, F1, F2, and B, respectively. We examined the treatment success rate in these patients based on the serum UA level during 120–180 days after the initiation of treatment, which was set as the endpoint. Furthermore, 37 patients were confirmed to have received repeated treatments with another drug at a different period for at least 180 days. Among these patients, the serum UA level during 120–180 days after the initiation of treatment was compared.

Comparison of the uric acid lowering effect of hyperuricemia medication based on the estimated medication history

The endpoint of each drug was examined with medication history as the standard. The mean UA level before the initiation of treatment decreased after the initiation of treatment with all the drugs ($P < 0.01$). The number of patients whose UAlevel reached < 7.0 mg/dL by treatment with A1, A2, F1, F2 and B was 346 (53.3%), 78 (83%), 119 (65%), 35 (50.7%), and 63 (75%), respectively. The number of patients whose UA level reached < 6.0 mg/dL by treatment with A1, A2, F1, F2, and B was 92 (14.2%), 50 (53.2%), 76 (41.5%), 35 (31.9%), and 45 (37.8%), respectively. Compared with that of the other drugs, the treatment success rate was high with A2 and low with A1.

Results of the crossover study based on the EMR data

Among patients on antihyperuricemic, those whose treatment was switched to another drug or the daily dose was changed were extracted from the medication history database. The endpoints in the initial treatment and those after the switch in the same patient were compared (Fig. 2).

Table 3 - Medication history generation results

Allopurinol 100 mg/day (A1)	Allopurinol 200 mg/day (A2)	Febuxostat 10 mg/day (F1)	Febuxostat 20 mg/day (F2)	Benzbromarone 50 mg/day (B)
Number of prescription subject patients				
2687	675	655	316	519
Number of prescription records				
52,797	7400	5530	2304	10,997
Number of treatment histories				
4825	1055	860	368	729
Number of observation subject patients				
649	94	183	69	119
Mean number of treatment days				
898	398	438	419	897
Comparison of mean number of treatment days; P value (Fisher's exact test) P < 0.01				
A2, F1, F2	A1, B	A1, B	A1, B	A2, F1, F2

Table 4 - Evaluation of endpoints of the antihyperuricemic

Allopurinol 100 mg/day (A1)	Allopurinol 200 mg/day (A2)	Febuxostat 10 mg/day (F1)	Febuxostat 20 mg/day (F2)	Benzbromarone 50 mg/day (B)
Number of examination target patients (male/female ratio)				
649 (493/156)	94 (87/7)	183 (124/59)	69 (45/24)	119 (82/37)
Mean ± SD of age				
64.8 ± 16.5	65.3 ± 16.3	65.4 ± 17.8	64.2 ± 15.6	58.1 ± 20.2
Mean ± SD serum UA level of > 8 mg/dL before treatment				
9.6 ± 1.9	9.0 ± 0.9	9.7 ± 1.2	9.8 ± 1.7	9.8 ± 1.7
Mean ± SD serum UA level 120–180 days after the initiation of treatment				
7.1 ± 1.3	6.0 ± 1.5	6.4 ± 1.6	7.1 ± 1.8	6.7 ± 1.9
Comparisons before and after the initiation of treatment P value (Paired t-test) P < 0.01				
UA level < 7.0 mg/dL; No./total no. (%) *P < 0.01and **P < 0.05				
346/649 (53.3)	78/94 (83.0)	119/183 (65.0)	35/69 (50.7)	63/512 (75)
Comparison of treatment success rate of UA level < 7.0 mg/dL P value (Fisher's exact test) *P < 0.01 and **P < 0.05				
A2*, F1*	A1*, F1*, F2*, B**	A1*, A2*, B**	A2*	A2*, F1**
UA value < 6.0 mg/dL; No./total no. (%)				
92/649 (14.2)	50/94 (53.2)	76/183 (41.5)	35/69 (31.9)	45/512 (37.8)
Comparison of treatment success rate of UA level < 6.0 mg/dL P value (Fisher's exact test)				
A2*, F1*, F2*, B*	A1*, F2**, B**	A1*, A2**	A1*, A2**	A1*, A2**

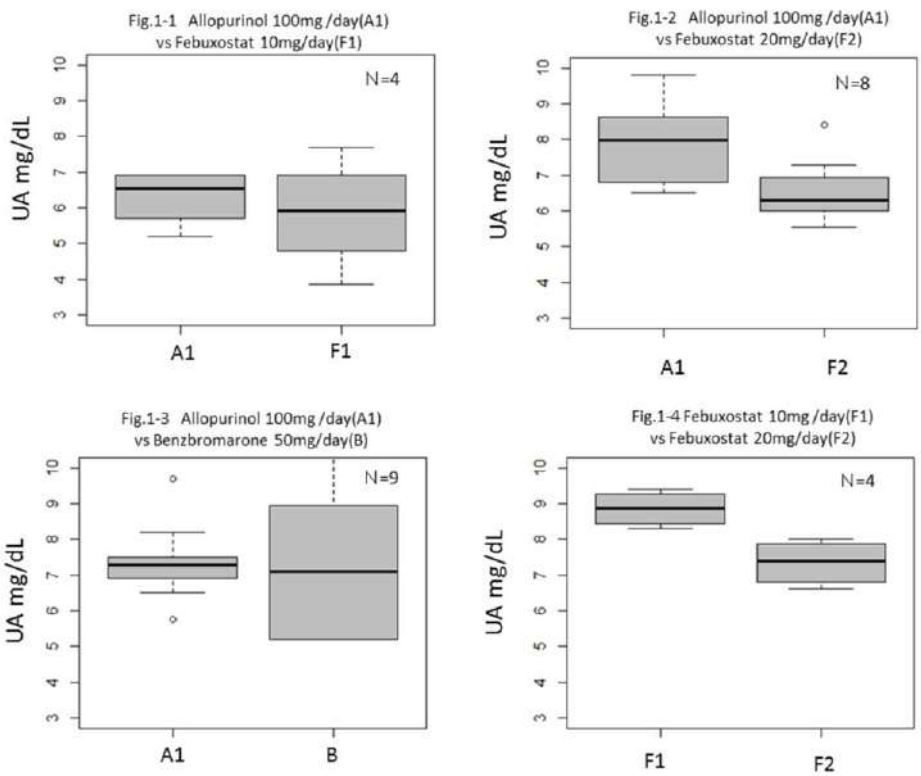


Figure 2 - Results of the crossover intervention study based on the EMR data

With the drug change from A1 to F1, the mean UA level decreased from 6.30 ± 0.8 mg/dL to 5.85 ± 1.6 mg/dL. With the drug change from A 1 to F2, the mean UA level decreased from 7.8 ± 1.1 mg/dL to 6.5 ± 0.9 mg/dL(paired t-test, $P < 0.05$). With the drug change from A1 to B, the mean UA level increased from 7.3 ± 1.1 mg/dL to 7.5 ± 2.5 mg/dL. Among patients in whom the dose of the drug was doubled, with the drug change from F1 to F2, the mean UA level decreased from 8.6 ± 0.5 mg/dL to 6.5 ± 0.9 mg/dL (paired-*T* test, $P < 0.05$).

Discussion

When clinicians observe the effect of drugs from retrospective studies, it is necessary to estimate the medication history of patients. In such a case, clinicians need to estimate the treatment period for each drug from the prescription order history in the electronic medical charts. The number of first-time screening subjects used in this study was 4852 and the number of target prescription records was 79,028. With only a few subjects, it is possible to estimate the treatment period by manual operations. However, with a large number of patients as in this study, it is difficult to extract patients who are required for the endpoint from such a large number. Besides the cost associated with processing human data, there is also the issue with gathering and managing the data. In this study, medication history that was automatically generated using a program that generates such information using prescription orders was used. This enabled us to perform the relevant process of gathering observation data automatically.

With the OMOP method, the effect of drug exposure period can be examined from the prescription records. If the gap between the end of a first prescription record to the start of the next prescription record is within 30 days when the prescription records are consecutive, that gap is also considered a period of drug exposure. Therefore, even when there is a washout, there is a possibility that as an endpoint, it is considered an observation period. Meanwhile, in medication history, by correcting for the remnant drugs in the overlap period, the gap in the prescription order is corrected for, and treatment start and end dates that are close to the actual treatment period of the patient are generated. In this study, this is accompanied by the evaluation of endpoint; the method of generation of medication history is effective.

As shown in Table 2, the number of patients with improved UA level of < 7 mg/dL was 641 of 1108 (57.9%). The percent of patients in whom gouty tophi disappeared as the UA level improved to < 6 mg/dL was 25.7% (285/1108). The rapid decline in the UA level due to antihyperuricemic is the cause for the onset of gouty arthritis, and the sudden increase in UA excretion is the cause of UA calculi and renal failure. Therefore, clinicians start prescription from a low dose to minimize the effect of adverse effects as much as possible. Thus, once there is improvement (UA < 7 mg/dL) from the state of hyperuricemia, fixing the drug and dose is commonly considered. In a report on the outcomes of treatment with antihyperuricemic in a phase III clinical study conducted in Japan, the patients on A2 and F2 whose UA level improved to < 6 mg/dL were 35.2% and 29.6%, respectively. In this study, the results were 33.5% and 27.4%, respectively, which were

similar to those of the previous clinical study. With B 50 mg/day, a postmarketing surveillance was performed and the rate of patients with decrease in UA level was 37% (31.6% in this study). This also supports the data of this study. In other clinical studies, A at a dose of ≥ 200 mg/day and F at a dose of ≥ 40 mg/day are generally used. Therefore, study data showing treatment results of long-term use of low-dose A, F, and B are beneficial while formulating treatment plans.

In a crossover intervention study, a subject is selected, administered various drugs at different treatment periods, and the treatment effects are compared. In this study, highly reliable data were obtained even with a fewer number of patients by removing the diversity among patients. However, with this method, once the final treatment is completed, a return to baseline is necessary. Moreover, a washout period is necessary based on the time of switching the drugs. It is a challenge to completely meet these two criteria when the data of patients attending hospital for the purpose of receiving antihyperuricemic are to be used as subject. Thus, in this study, the endpoint used for making comparisons before and after switching the drugs was set at 120–180 days. As such long-term data were used and as patients on concomitant treatment with other antihyperuricemic were removed, the effect of previous drugs was decreased as much as possible. Therefore, among the 4852 patients, only 37 (0.76%) were used as observation subjects. Of the 37 patients, 28 were patients who had switched from A to other medications. Five were switched from B to F and four were on increased F dose. Febuxostat was launched approximately 30 a fter A. Even in patients with moderate renal impairment, F can be administered with dose reduction. The administration is once daily, and it can be indicated even in the absence of complication of hypertension. This makes it a readily switchable drug. Allopurinol and B are medications that have been in use for a long time, and several studies on their treatment efficacy have also been performed. On the contrary, the fact that a study equivalent to a crossover study using real-world data with F as the subject could be performed is of significance.

Clinical data stored in clinical facilities are an important intellectual resource. These data include diseases with various characteristics. By consolidating these data, the causal relationship between interventions and outcomes can be examined.

In this study, the effect of antihyperuricemic on renal function and diuretics, which are related to adverse events, was not observed. These data that are linked to the unique ID of patients are stored in electronic medical charts, making it easy to extract those patients who can be the subjects of an evaluation. Patients can be classed according to the status of renal impairment and the status of diuretic treatment. By doing so, the relative evaluation of treatment effect and the onset of adverse events can also be performed. One can also expect that the factors affecting treatment efficacy can be investigated using patients who are extracted by this method. With the system developed in this study, data of multiple medical facilities could be handled. Moreover, we were successful in generating medication history from prescription order data using the electronic medical chart of various makers. If the criteria of patients, who will be the subject of observation by physicians based on medication history, can be set using a computer, clinical epidemiology studies will improve dramatically.

Conclusions

In this study, medication hostory for hyperuricemia and the serum uric acid (UA) level were compared using mdeication history database generated by prescription order records. From the generated medication history, the treatment success rate with antihyperuricemic can be extracted mechanically. This method will be useful for comparing drug effects retrospectively.

References

[1] S.J. Nolan, I. Hambleton, and K. Dwan, The Use and Reporting of the Cross-Over Study Design in Clinical Trials and Systematic Reviews: A Systematic Assessment, *PLoS One* 2016 Jul **13**;11(7).

[2] C.S. Kruse, and K. Goetz, Summary and frequency of barriers to adoption of CPOE in the U.S., *J Med Syst* 2015 Feb; **39**(2):15

[3] E.W. Duggan, K. Carlson, and G.E. Umpierrez, Perioperative Hyperglycemia Management: An Update, *Anesthesiology,* 2017 Mar; **126**(3):547-560.

[4] X. Zhou, S. Murugesan, H. Bhullar, Q. Liu, B. Cai, C. Wentworth, and A. Bate, Anevaluation of the THIN database in the OMOP Common Data Model for active drug safety surveillance, *Drug Saf*. 2013 Feb; **36**(2):119-34.

[5] M.J. Schuemie, Methods for drug safety signal detection in longitudinal observational databases: LGPS and LEOPARD, *Pharmacoepidemiol Drug Saf* 2011Mar; **20**(3):292-9.

Address for correspondence
Division of Mdeical Informatics, Tottori University Hospital
36-1 Nishi-cho, Yonago, Tottori, 683-8504, Japan,kei-tera@umin.net

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190386

Extracting Alcohol and Substance Abuse Status from Clinical Notes: The Added Value of Nursing Data

Maxim Topaz [a,b], Ludmila Murga [c], Ofrit Bar-Bachar [c], Kenrick Cato [a], Sarah Collins [a,d]

[a] *School of Nursing, Columbia University, New York City, NY, USA*
[b] *Data Science Institute, Columbia University, New York City, NY, USA*
[c] *Cheryl Spencer Department of Nursing, University of Haifa, Haifa, Israel*
[d] *Department of Biomedical Informatics, Columbia University, New York City, NY, USA*

Abstract

We applied an open source natural language processing (NLP) system "NimbleMiner" to identify clinical notes with mentions of alcohol and substance abuse. NimbleMiner allows users to rapidly discover clinical vocabularies (using word embedding model) and then implement machine learning for text classification. We used a large inpatient dataset with over 50,000 intensive care unit admissions (MIMIC II). Clinical notes included physician-written discharge summaries (n = 51,201) and nursing notes (n = 412,343). We first used physician-written discharge summaries to train the system's algorithm and then added nursing notes to the physician-written discharge summaries and evaluated algorithms prediction accuracy. Adding nursing notes to the physician-written discharge summaries resulted in almost two-fold vocabulary expansion. NimbleMiner slightly outperformed other state-of-the-art NLP systems (average F-score = .84), while requiring significantly less time for the algorithms development.: Our findings underline the importance of nursing data for the analysis of electronic patient records.

Keywords:
Substance-Related disorders, Alcoholism, Nursing informatics.

Introduction

Over the past decade, widespread adoption of health information technology, such as the electronic health record, resulted in rapidly growing volumes of clinical data. Most of these data are stored in an unstructured format, such as text narratives, within electronic health records across diverse healthcare settings. The narrative data require innovative tools that help clinicians and researchers to extract meaning from large databases of text.

Natural language processing (NLP) is a set of powerful techniques that can help in processing and deriving meaning from clinical narratives. NLP has the potential to identify clinicians' detailed documentation of wound information, mentions of gender identity, documentation of poor self-management status and other concerns about a patient expressed as free text [1–5].

Although significant progress has been made within the medical domain, for nursing and allied health professions NLP development and implementation remains relatively challenging and scarce[6,7]. Some of the challenges include lack of large datasets of labeled clinical notes from nursing or allied health professions needed to create machine learning and text mining algorithms. On the other hand, rule-based NLP systems are hard to construct because large curated vocabularies of nursing or allied health professions-specific lexicons are rarely available. Moreover, the importance of nursing data and its contribution to general tasks such as text classification, remains generally under-studied. We found only one study that used information extracted from nursing notes as an indicator of out-of-hospital mortality[8]. Our main contribution is further assessment of the importance of nursing data.

This study focuses on an important domain-identification of alcohol and substance abuse from clinical data. Prior research suggests that alcohol and substance abuse information is often documented as free text[9]. However, previous NLP studies used physician notes to extract alcohol and substance abuse information[10,11]. In this study, we explore the potential added value of nursing data.

In this study we applied our open source NLP application called "NimbleMiner" [12,13] to generate an algorithm that can automatically identify clinical notes with mentions of alcohol and substance abuse. Our system allows users to rapidly specify clinical vocabularies for a certain domain, apply weakly supervised rapid labeling, and then implement machine learning for text classification. In this study, we first used physician-written discharge summaries to generate the NLP algorithm and evaluate the algorithm's prediction accuracy. We then added nursing notes to the physician-written discharge summaries and evaluated the algorithm's prediction accuracy. We compared the performance of the two NLP algorithms to understand whether adding nursing notes resulted in better prediction accuracy when identifying clinical notes with mentions of alcohol and substance abuse. We also compared our system's performance to other state-of-the art NLP systems that were applied on the same documents.

Methods

Dataset

This study used the large, publically available, de-identified dataset MIMIC II which is comprised of data from adult patients admitted to the intensive care units (ICUs) at the Beth Israel Deaconess Medical Center from 2001 to 2012. The data included over 50,000 hospital ICU admissions. The dataset contained several types of clinical notes, including physician-written discharge summaries (n = 51,201) and nursing notes (n

= 412,343). Nursing notes included nursing admission notes, daily progress and status update notes, case management notes, etc. This study received an Institutional Review Board approval from the University of Haifa, Israel.

NLP System Description

Our NLP system NimbleMiner is an open source system developed by our team[12,13]. User manual and download options can be accessed at: http://github.com/mtopaz/NimbleMiner. Other research or clinical teams can use the system under the GNU General Public License v3.0. NimbleMiner includes several methodological stages of clinical note processing that are briefly described below and presented in Figure 1.

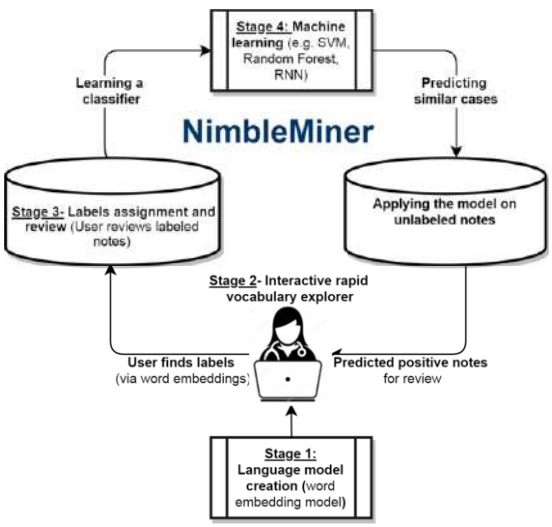


Figure 1– NimbleMiner system process stages.

Stage 1- Language Model Creation:

The user selects a large corpus of clinical notes and defines language model characteristics. We use a word embedding model for language model generation and users can change model settings based on their preferences.

Stage 2- Interactive Rapid Vocabulary

Explorer: The user enters a query term of interest, and the system returns a list of similar terms it identified as relevant. The list of suggested similar terms is based on the cosine term similarity metric extracted from the word embedding model. The user selects and saves the relevant terms. Negated or other irrelevant terms that are not selected by the user are also saved in the system for further tasks, such as negation detection. Figure 2 describes the steps of the vocabulary explorer stage.

Stage 3- Labels Assignment and Review:

The system uses the stage 2 discovered similar terms to assign labels to clinical notes (while excluding notes with negations and other irrelevant terms). Assigning a positive label means that a concept of interest is present in the clinical note. When needed, the user reviews and updates lists of similar terms and negated similar terms. The user reviews the clinical notes with assigned labels for accuracy. This weakly supervised rapid labeling approach is based on a postive labels learning framework validated in previous research [14,15].

Stage 4- Machine Learning:

The user choses a machine learning algorithm to be applied to create a predictive model. The model is then applied to predict what clinical notes might have the concept of interest. The user reviews the predicted notes and can go through stages 2-4 again to add new labels.

NLP System Settings

NimbleMiner's user interface is implemented in R statistical package. To create a word embedding model, we used a skip-gram model implementation called word2vec and phrase2vec in R[16,17]. Parameters of the word embedding model were held constant based on parameters suggested in other studies of word embedding[18]. Specifically, we used a model with window width size = 10, vector dimension = 100, minimum word count = 5, negative sample size = 5, and sub-sampling = 1e-3. For each similar term entered by the user, the system presented 50 potentially similar terms based on the cosine similarity. Our previous experiments [12] showed that the random forest algorithm outperforms other approaches (e.g., J48 Decision trees, Support Vector Machines), hence we used this algorithm in the machine learning stage of this study. The algorithm was trained using only discharge summaries, as was the case in other studies. The Random Forest algorithm was used with default settings (number of iterations = 100, minimum number of instances = 1, minimum variance for split = 1e-3, depth = unlimited).

NLP System Performance Evaluation

The system performance evaluation was performed based on the publically available gold-standard testing dataset generated by Gerhman et al. [19]. To create the gold standard dataset, Gerhman et al. annotated 1,610 discharge summaries for presence of several patient phenotypes, including alcohol and substance abuse. Each discharge summary was labeled by at least two experts in medicine and health informatics [19]. Full agreement about each of the labels was achieved on all the cases. Overall, the annotated dataset included 155 instances of substance abuse and 196 instances of alcohol abuse. For both domains, a relatively high inter-rater agreement was achieved (Cohen's Kappa inter-rater agreement = .86). We applied our NLP algorithms on this dataset to predict alcohol or substance abuse for each of the discharge summaries. We calculated precision (defined as the number of true positives out of the total number of predicted positives), recall (defined as the number of true positives out of the actual number of positives) and F-score (F1, weighted harmonic mean of the precision and recall) to evaluate the performance of our system.

Comparison with Other NLP Systems

NimbleMiner's performance was compared to the results reported by Gerhman et al. who developed a convolutional neural network (CNN) for text classification[19]. CNN is a subtype of machine learning models called neural networks. CNN is built of one or more convolutional layers and then followed by one or more fully connected layers, as in a standard multilayer neural network[20]. CNNs were found to outperform other machine learning methods in several classification tasks, such as image recognition[21].

In addition, NimbleMiner's performance was compared to another open-access rule-based NLP system called the clinical Text Analysis and Knowledge Extraction System (cTAKES)[22]. cTAKES was applied by Gerhman et al. to identify alcohol and substance abuse instances in each of the notes in the testing set[19]. To pre-process the clinical notes,

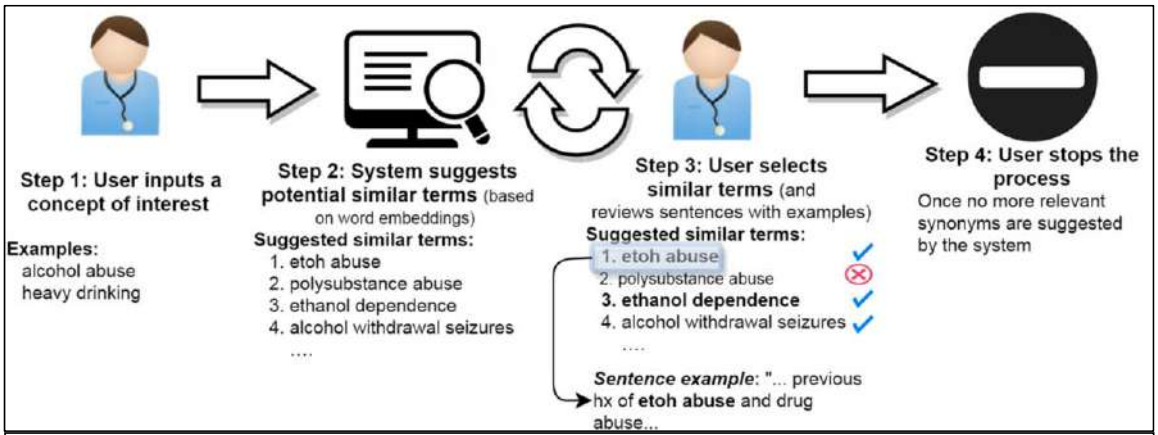


Figure 2– NimbleMiner Stage 2 -Interactive rapid vocabulary explorer process overview.

***Legend**: Interactive Rapid Vocabulary Explorer: The user enters a query term of interest (step 1), and the system returns a list of similar terms it identified as relevant (step 2). The list of suggested similar terms is based on the cosine term similarity metric extracted from the word embedding model. The user selects and saves the relevant terms while negated or other irrelevant terms that were not selected by the user are also saved in the system for further tasks (step 3). The system iteratively identifies new potential similar terms and presents them to the user for review (steps 2-3) and the process continues until there no new similar terms to suggest (step 4).*

cTAKES splits sentences and phrases into individual words (tokenization), normalizes them (e.g., removes plurals) and tags part-of-speech (e.g. noun, verb). Then, the named-entity recognition algorithm is implemented to detect named entities for which a concept unique identifier (CUI) exists in the Unified Medical Language System (UMLS).

Results

Applying NimbleMiner on discharge summaries alone resulted in identifying 73 alcohol abuse-related words and expressions (e.g., "heavy drinker" or "alcohol addiction"). Adding nursing notes resulted in identifying 71 additional unique words and expressions, resulting in a 97% lexicon expansion (total alcohol abuse expressions n = 144). Similarly, 76 substance abuse-related words and expressions (e.g., "illegal drug abuse" or "crack/cocaine abuse") were identified from the discharge summaries alone and 60 additional expressions were identified when nursing notes were added (total substance abuse expressions n = 136). Adding nursing notes enabled 79% lexicon expansion for the substance abuse domain. The lexicon discovery phase for each domain was conducted by one of the study co-authors and took about 4 hours in total to implement (2 hours for each domain) with discharge summaries only and about 6 hours when nursing notes were added (since more potentially relevant words needed to be reviewed). Two more hours were spend on system refinement (e.g., reviewing labeled notes and adding negations) and machine learning algorithm implementation for each domain. Overall, a maximum of 10 hours were spent on vocabulary development and algorithm implementation. Table 1 presents examples of alcohol and substance abuse words and expressions.

In addition, 1,736 terms for substance abuse and 2,255 terms for alcohol abuse were not selected by the user during the interactive rapid vocabulary explorer process (stage 2). These irrelevant terms included a diverse range of negations (e.g., "illicits none" or "not addicted to drugs"), family history expressions (e.g., "family history of alcoholism" or "father used cocaine") and other irrelevant terms that appear in the same context, for example other habits or related diseases (e.g., "tobacco smoker" or "hepatic cirrhosis"). Examples of negated or other irrelevant terms are shown in Table 1.

Table 1– Examples of alcohol and substance abuse expressions

Domain	Examples of relevant terms
Alcohol abuse	"heavy drinker" "alcohol addiction" "etoh [ethyl alcohol] abuse" "ciwa [The Clinical Institute Withdrawal Assessment for Alcohol scale] high" "korsakoff syndrome" [chronic memory disorder common in alcoholics] "alcoholic cirrosis [misspelling of cirrhosis]"
Substance abuse	"ho [history of] ivdu" "heroin withdrawal" "recent crack cocaine" "narcotic overdose" "iv heron [misspelling of heroin]" "cocaine in urine"
	Examples of negated or other irrelevant terms
Alcohol abuse	"drinks alcohol rarely" "occasionally drinks alcohol" "father was alcoholic" "family history of alcoholism" "alcohol abstinence" "social alcohol use"
Substance abuse	"ilicit drug use denies" "denies recreational drug use " "father used cocaine" "illicits none" "drugs none" "not addicted to drugs"

NLP System Performance Evaluation

NimbleMiner's predictive algorithm (Random Forest classifier) slightly outperformed other state-of-the-art NLP approaches in terms of the F-score. The best performing NimbleMiner algorithm was based on words and expressions learned from a corpus of discharge summaries and nursing notes compared to discharge summaries alone. See Table 2 for details on system performance measures

*Table 2– Comparison of system metrics across NLP systems**

Metric	NimbleMiner (discharge summaries+ nursing notes)	NimbleMiner (discharge summaries only)	Best performing Convolutional Neural Network	Text Analysis & Knowledge Extraction System (cTAKES)
Alcohol abuse				
Precision	83	79	85	**88**
Recall	**84**	78	79	79
F1	**84**	78	81	83
Substance abuse				
Precision	87	78	83	**93**
Recall	80	77	**83**	47
F1	**84**	77	83	62

*All systems were tested on discharge summaries only.

Discussion

This study applied a rapid NLP system called NimbleMiner to identify clinical notes with mentions of alcohol and substance abuse. Our approach showed promising results and it performed similarly or outperformed other systems developed for this domain[19]. For traditional machine learning and text mining, large datasets of labeled data are needed to train the system. For example, Gerhmann et al. annotated 1,610 discharge summaries and 70% of these notes were used for system training[19]. This approach required two or more clinicians to read and label the corpus of discharge summaries, resulting in significant time investment of at least 260 hours for data labeling (5 min per note * 1,610 discharge summaries * 2 clinicians= ~260 hours). On the other hand, the strength of rule-based systems lies in carefully curated and developed terminologies and rich hierarchical relationships between the concepts. However, rule-based systems like cTAKES require significant time and expertise investments when creating vocabularies and rules, defining negations, identifying family history, etc.

Our approach offers a hybrid solution where the human expert is interacting with the machine to rapidly create weakly supervised large labeled datasets for further machine learning. NimbleMiner also allows users to create large corpora of negated or other irrelevant terms (such as family history) that should be excluded from the labeled clinical notes processed through machine learning. Using the NimbleMiner, it took less time (4 hours) compared to other systems for algorithm development and implementation with comparable or better results. These findings suggest that for some text mining domains, such as identifying socio- behavioral determinants of health, NLP approaches can be implemented in a rapid, clinician-driven manner.

Our results also demonstrate the added value of extending data sources [23]. to include nursing data for vocabulary exploration. In our approach, adding nursing clinical notes resulted in an almost two-fold vocabulary expansion, leading to better classification performance. Other recent studies confirm our results about the importance of nursing data. For example, a NLP study conducted with the same dataset (MIMIC) has recently found that sentiment presented in nursing notes was significantly associated with a 30-day intensive care patient mortality[8]. Thus, the presence of negative-sentiment related expressions in nursing notes predicted a 30-day patient mortality. These findings underline the importance of nursing data for the analysis of electronic patient records. Other studies conducted in diverse clinical domains should consider using the valuable information from nursing notes to improve text and data mining algorithms.

Limitations

Our study has several limitations. First, our approach might not be applicable to solve more complex NLP challenges, such as word sense disambiguation or relation extraction. Comparing our results with rule-based systems like cTAKES would likely result in biased results in favor of more domain-specific systems like NimbleMiner. In addition, different NimbleMiner users can discover different vocabularies, which might influence the system's performance. Also, the vocabulary discovery phase was implemented by one co-investigator and adding more reviewers to this phase could have resulted in a different vocabulary for alcohol and substance abuse terms. A larger study with more domains is needed to validate the generalizability of our approach.

Conclusions

We applied our open-source NLP system NimbleMiner to conduct clinician-driven concept discovery in clinical narratives. Our results suggest that NimbleMiner can be applied to rapidly discover similar clinical terms and create large datasets of labeled clinical notes required for machine learning. Our system slightly outperformed other state-of-the art NLP systems while requiring significantly less time for the algorithm's development. We believe that NimbleMiner can be used by almost any clinician without special informatics training to create accurate NLP algorithms.

References

[1] S.A. Collins, and D.K. Vawdrey, Reading between the lines; of flow sheet data: nurses' optional documentation associated with cardiac arrest outcomes., *Appl. Nurs. Res.* **25** (2012) 251–7. doi:10.1016/j.apnr.2011.06.002.

[2] S.A. Collins, M. Fred, L. Wilcox, and D.K. Vawdrey, Workarounds used by nurses to overcome design constraints of electronic health records., *NI 2012 11th Int. Congr. Nurs. Informatics, June 23-27, 2012, Montr. Canada. Int. Congr. Nurs. Informatics (11th 2012 Montr. Quebec)*. **2012** (2012) 93.

[3] S.A. Collins, K. Cato, D. Albers, K. Scott, P.D. Stetson, S. Bakken, and D.K. Vawdrey, Relationship between nursing documentation and patients' mortality., *Am. J. Crit. Care*. **22** (2013) 306–13. doi:10.4037/ajcc2013426.

[4] K. Cato, J. Romano, R. Vanguri, and Ni. Tatonetti, The Use of Informatics to Reduce Disparities in Transgender Health, in: MIA 2017 Annu. Symp. - Sess. Details, Washington D.C., 2017.

[5] M. Topaz, K. Lai, D. Dowding, V.J. Lei, A. Zisberg, K.H. Bowles, and L. Zhou, Automated identification of wound information in clinical notes of patients with heart diseases: Developing and validating a natural language processing application, *Int. J. Nurs. Stud.* (2016). doi:10.1016/j.ijnurstu.2016.09.013.
[6] M. Topaz, and L. Pruinelli, Big Data and Nursing: Implications for the Future, in: Stud. Health Technol. Inform., 2017: pp. 165–171. doi:10.3233/978-1-61499-738-2-165.
[7] G. Gonzalez-Hernandez, A. Sarker, K. O'Connor, and G. Savova, Capturing the Patient's Perspective: a Review of Advances in Natural Language Processing of Health-Related Text, *Yearb. Med. Inform.* **26** (2017) 214–227. doi:10.15265/IY-2017-029.
[8] I.E.R. Waudby-Smith, N. Tran, J.A. Dubin, and J. Lee, Sentiment in nursing notes as an indicator of out-of-hospital mortality in intensive care patients, *PLoS One*. **13** (2018) e0198687. doi:10.1371/journal.pone.0198687.
[9] B. Tai, and A.T. McLellan, Integrating information on substance use disorders into electronic health record systems., *J. Subst. Abuse Treat.* **43** (2012) 12–9. doi:10.1016/j.jsat.2011.10.010.
[10] E. Chen, and M. Garcia-Webb, An analysis of free-text alcohol use documentation in the electronic health record: early findings and implications., *Appl. Clin. Inform.* **5** (2014) 402–15. doi:10.4338/ACI-2013-12-RA-0101.
[11] Y. Wang, E.S. Chen, S. Pakhomov, E. Arsoniadis, E.W. Carter, E. Lindemann, I.N. Sarkar, and G.B. Melton, Automated Extraction of Substance Use Information from Clinical Texts., *AMIA ... Annu. Symp. Proceedings. AMIA Symp.* **2015** (2015) 2121–30. http://www.ncbi.nlm.nih.gov/pubmed/26958312 (accessed March 18, 2019).
[12] M. Topaz, K. Gaddes, M. McDonald, and K. Bowles, Development and Validation of a Novel Rapid Clinical Text Mining Approach Based on Word Embeddings (NimbleMiner)., *Stud. Health Technol. Inform.* **244** (2017) 88.
[13] M. Topaz, L. Murga, K.M. Gaddis, M. V. McDonald, O. Bar-Bachar, Y. Goldberg, and K. Bowles, Mining fall-related information in clinical notes: comparison of rule-based and novel word embedding-based machine learning approaches, *J. Biomed. Inform.* (2019) 103103. doi:10.1016/J.JBI.2019.103103.
[14] Y. Halpern, S. Horng, Y. Choi, and D. Sontag, Electronic medical record phenotyping using the anchor and learn framework., *J. Am. Med. Inform. Assoc.* **23** (2016) 731–40. doi:10.1093/jamia/ocw011.
[15] C. Elkan, and K. Noto, Learning classifiers from only positive and unlabeled data, in: Proceeding 14th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. - KDD 08, 2008: p. 213. doi:10.1145/1401890.1401920.
[16] T. Mikolov, K. Chen, G. Corrado, and J. Dean, Efficient estimation of word representations in vector space, in: ICLR 2013, 2013: pp. 1–9. doi:10.1162/153244303322533223.
[17] R Core Team, R: A language and environment for statistical computing., (2013). http://www.r-project.org/. (accessed April 13, 2018).
[18] B. Chiu, G. Crichton, A. Korhonen, and S. Pyysalo, How to train good word embeddings for biomedical NLP, in: Proc. 15th Work. Biomed. Nat. Lang. Process., 2016: pp. 166–174. doi:10.18653/v1/W16-2922.
[19] S. Gehrmann, F. Dernoncourt, Y. Li, E.T. Carlson, J.T. Wu, J. Welt, J.F. Jr., E.T. Moseley, D.W. Grant, P.D. Tyler, and L.A. Celi, Comparing deep learning and concept extraction based methods for patient phenotyping from clinical narratives, *PLoS One*. **13** (2018) e0192360. doi:10.1371/JOURNAL.PONE.0192360.
[20] A. Ng, J. Ngiam, C.Y. Foo, Y. Mai, C. Suen, A. Coates, A. Maas, A. Hannun, B. Huval, T. Wang, and S. Tandon, Unsupervised Feature Learning and Deep Learning Tutorial, (n.d.).
[21] A.L. Beam, and I.S. Kohane, Big Data and Machine Learning in Health Care, *JAMA*. **319** (2018) 1317. doi:10.1001/jama.2017.18391.
[22] G.K. Savova, J.J. Masanz, P. V Ogren, J. Zheng, S. Sohn, K.C. Kipper-Schuler, and C.G. Chute, Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications., *J. Am. Med. Inform. Assoc.* **17** (2010) 507–13. doi:10.1136/jamia.2009.001560.
[23] P.F. Brennan, and S. Bakken, Nursing Needs Big Data and Big Data Needs Nursing, *J. Nurs. Scholarsh.* **47** (2015) 477–484. doi:10.1111/jnu.12159.

Address for Correspondence

Max Topaz, mt3315@cumc.columbia.edu.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190387

Reduction of Overwork Time of Nurses by Innovation of Nursing Records Using Structured Clinical Knowledge

Satoko Tsuru[a], Akihiro Nakao[a], Naohisa Yahagi[a], Kouichi Tanizaki[a]
Kumiko Sudo[b] , Shizuka Morimatsu[c], Tomomi Takaki[d], Nobuko Takakusaki[e]
Tomoko Higashi[f], Keiko Nakashigeg[g], Miyuki Takahashi[h]

[a] *The University of Tokyo,*
[b] *Iizuka Hospital,*
[c] *Oguno Hospital*
[d] *Seirei Sakura Citizen Hospital,*
[e] *Tokyo Rinkai Hospital,*
[f] *Kumamoto sekijuji Hospital*
[g] *Jiaikai Imamura sogo Hospital,*
[h] *Nara Medical University Hospital*

Abstract

Hospitalization expenses account for a high rate of national medical care expenditure in Japan. The Japanese national medical care expenditure was 42 trillion 364.4 billion yen in 2015, in which hospitalization expenses were 15 trillion 575.2 billion yen (36. 8%). Therefore, it is necessary to take measures to reduce hospitalization expenses. The total ratio of the labor cost of physicians and nurses accounted for about 1/3 of all expenditures of general hospitals in 2015. Moreover, the personnel cost of nurses accounted for about 1/5 of all expenditure, showing that the personnel cost of nurses is an element with a large influence on hospital management. The objective of this study was to develop a methodology to reduce the overtime work of nurses accounting for a large rate of personnel expenses by focusing on overtime work, a personnel expense-increasing factor, aiming at hospital cost reduction. First, the cause of overtime work, planning, and recording by nurses were analyzed and an IT application increasing the quality and efficiency of the work was developed. Then, fees for the use and maintenance of the IT system meeting the following conditions were set as a strategy to introduce the system: (1) 50% reduction of the overtime work of nurses and (2) fees 50% or lower than the reduced payment for overtime work. This IT application was introduced to the heads and directors of nursing of 5 hospitals and the strategy was proposed. All heads and directors highly evaluated the system and responded to initiate the process for the introduction. It was suggested that the methodology to reduce the overtime work of nurses proposed by this study is useful and feasible.

Keywords:
Hospitalization, Health expenditures

Introduction

Problems with health care are discussed from viewpoints of safety, quality (in a narrow sense), and cost, and all advanced countries have immense tasks in each issue. Aging spurs an increase in national medical care expenditure to which advanced countries inject large public finance.

In Japan, hospitalization accounts for a high proportion of national medical care expenditure. The national medical care expenditure in Japan was 42 trillion 364.4 billion yen in 2015, in which hospitalization expenses was 15 trillion 575.2 billion yen (36. 8%) [1]. Therefore, it is necessary to take measures to reduce hospitalization expenses, which account for 36.8% of the total medical expenses. For 'reduction of the hospitalization expenses' in medical care, which is an important socio-technology to be possessed by a country, it is necessary to aim at 'improvement of efficiency while securing quality', i.e., improvement of added values for customers while optimizing devoted resources.

Table 1 Ratios of personnel expenses in general hospitals in Japan [3]

	Medical corporation	Local government	Former social insurance-related organization
Ratio of personnel expenses (%)	53.3	63.1	51.8
Ratio of personnel expenses for physicians (%)	13.8	14.2	15.9
Ratio of personnel expenses for nurses (%)	17.7	22.2	18.6
Ratio of personnel expenses for physicians and nurses (%)	31.5	36.4	34.5
Ratio of personnel expenses for others (%)	21.8	26.7	17.3

In data analysis of 773 general hospitals (2015 account settlement) [2], the numbers of workers per 100 patients were 9.9 full-time physicians, 2.4 part-time physicians, 65.7 nurses, and 52.7 other workers. Regarding the rate of cost to the medical revenue, the ratio of personnel expenses was 52.4%. The annual medical revenue per worker was 11,953,000 yen

(about 12 million yen) and the personnel expenses per worker were 6,267,000 yen.

On the other hand, in data analysis of 520 sanatorium-type hospitals (2015 account settlement) [2], the numbers of full-time and part-time physicians, nurses, and other works were 3.3, 1.7, 59.4, and 39.9, respectively. The proportion of personnel expenses was 58.5%, the annual medical revenue per worker was 8,999,000 yen (about 9 million), and the personnel expenses per worker was 5,266,000 yen.

Regarding the personnel expenses of general hospitals by hospitals established in 2015[3] are categorized as follows: 1) medical corporation, 2) local government, and 3) former social insurance-related organization, the overall ratio of personnel expenses was: 1) 53.3%, 2) 63.1%, 3) 51.8%, the ratio of personnel expenses for physicians was: 1) 13.8%, 2) 14.2%, 3) 15.9%), and the ratio of personnel expenses for nurses was: 1) 17.7%, 2) 22.2%, 3) 18.6%, showing that the total ratio of the labor cost of physicians and nurses accounts for about 1/3 of all expenditures. In addition, about 1/5 of all expenditures is the personal cost of nurses, showing that the personnel cost of nurses is an element with a large influence on hospital management.

To realize high-quality excellent services, 'co-creation' [4] by customers and service providers is attracting attention, i.e., greater importance is attached to designing for points of contacting customers. For hospitals, it is necessary to focus on ways of working of physicians and nurses who have many important contact points with patients. Their abilities and numbers influence the operational efficiency of elements other than the humans described above. Therefore, 'the ratios of personnel expenses for physicians and nurses' and 'way of working to create value' may be important.

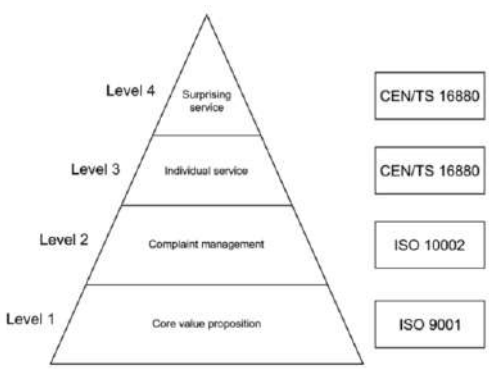


Fig. 1 Service excellence pyramid [4]

This study aims at developing a management method to reduce the hospital cost for the latter. Focusing on a personnel cost-increasing factor, overtime work, reduction of overtime work of nurses, which accounts for a large ratio of personnel expenses, is targeted. In the present study, we proposed the design of digitalization as an innovation of nursing records and investigated its usefulness and feasibility aiming at solving the problem with overtime work in health care.

Methods

We took the following steps for the purpose of a high-quality nurse's record and reduction of the overtime work.

Step1:

We suggested the rate of recording nursing observation of patient condition necessary to deal with individual patients does not reach 50% through a survey to evaluate the quality of nursing records in the current state. Failures of its inclusion in the planning, implementation, and recording have been identified as causes and 'volatizing situation' was suggested [5-7]. Moreover, it was clarified that data production and input depend on individual nurses and many documents are descriptive electronic records, being a long way from re-utilization.

Step2:

To solve failure to include it in a plan, preparation and utilization of many high-quality models of treatment and nursing care plans are useful. Thus, we have developed clinical knowledge content by incorporating NursingNAVI™ (Navigator for Thinking Process in Nursing) into PCAPS™ (Patient Condition Adaptive Path System) as a structured clinical knowledge content [8-12]. In the process of this development, using the Standard Terminology for Nursing Observation and Action (Ministry of Health, Labour Standards since 2016) for electronic medical records, the content of structured nursing care plan was developed.

Step3:

The result of two hospitals carried out record reform for problem-solving was shown five hospitals.

Case1: An acute hospital

A system application to support recording the structured nursing care plan was designed and implemented in a large-scale acute-care hospital with 1,000 beds. The proportion of nurses who took 30 minutes or more to identify the condition of patients in whom they were in charge of decreased from 45.4 to 5% after 3 months. It was suggested that utilization of the high-quality structured nursing care plan, which is clinical nursing knowledge, supports the thinking process of nurses and shortens the work time. After 3 years, 40% of patients applied clinical path, 60% of patients applied NursingNAVI®. It became zero in the overtime for records. The overtime work markedly decreased in the hospital and overtime work for recording was mostly resolved.

Case2: A non-acute hospital

We analyzed the nurse's record. Physicians satisfied the nursing record of observation. Only the PCAPS introduction ward was able to maintain a decrease of the overtimes.

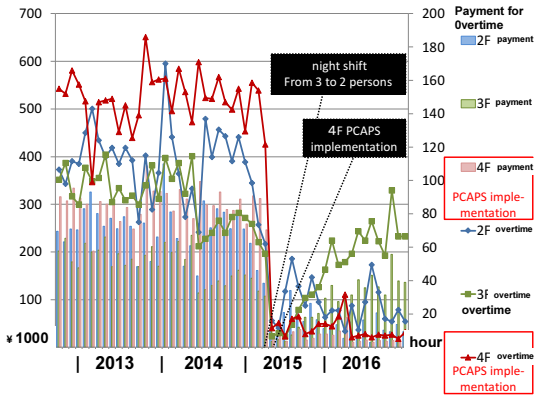


Fig. 2 Overtime and the payment every month

Step4:

These five hospitals carried out follows two surveys. The aim of Survey 1 is to remark the problem of quality of nursing record using our survey sheet[7]. It is set elements of nursing observation in the acute phase after colectomy. We used the 'Master File of Standardized Nursing Practice Terminology'. It is standard terminology for nursing practice by the Ministry of Health, Labour and Welfare in Japan. The set of elements was developed after reviews using the framework of NursingNAVI® [6-8] by multiple clinical nursing specialists. The availability of the survey sheet was verified by comparing surveys among four and fifteen hospitals. The aim of Survey 2 is to remark the problem of the overtime work. It was conducted in each hospital. We prepared output-table of each month and each nursing unit. Senior nursing officers of five hospitals gathered data in each hospital, and input the output-table sum of overtime and sum of overtime payment.

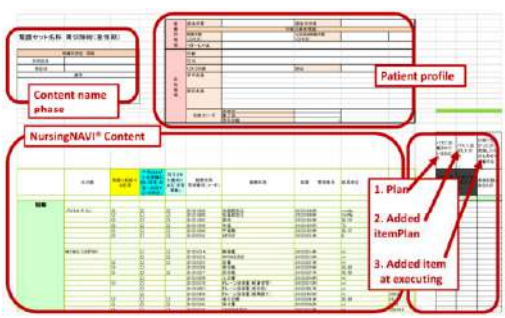


Fig. 3 Survey 1: Surve sheet of nursing documents using NursingNAVI® Content for quality evaluation of nursing observation

Results

Survey 1: Quality of nursing record using our survey sheet

Vital signs, in/out, was relatively recorded, but the observation to affect the monitoring of a symptom, complications of the disease was less than 50%. A certain hospital investigated record time and the number of record letters using log data of HIS. As a result, we understood that they recorded it until the night after daily service duties of 2 change were over. In addition, the actual situation that wrote an enormous amount of description record became clear.

Survey 2: the overtime work

Each hospital surveyed the payment amount and time of overtime work by month and ward for one year. The annual payment was 20-90 million yen in acute-care general hospitals with about 400 beds, 150 million yen in acute-care general hospitals with about 490 beds and 250 million yen or more in university hospitals with about 1,000 beds. In contrast, the payment for overtime work in the large-scale hospital with 1,000 beds described above which introduced the support system for the structured nursing care plan recording was about 25 million yen, confirming that recording work necessary for medical treatment is performed in overtime work.

Five hospitals started activity for the innovation of the nursing record system after having confirmed the results of this investigation.

Discussion

Standardized care plan using structured nursing knowledge is effective and efficient for nursing practice. Nurses can observe systematically using the structure. They can gather data and information for analyzing own nursing. We could introduce NursingNAVI® and PCAPS in the hospital information system. It was suggested that hospital nursing can use two types of systems as structured nursing knowledge.

Even though an excellent system is designed and developed an actual practice sites request its introduction, if the contribution to management cannot be claimed, the realization is difficult. For a strategy to introduce the system, we investigated a business model (1) targeting 50% reduction of overtime work time of nurses and (2) setting fees for the use and maintenance of the system and content at 50% or lower of the reduced payment for overtime work.

It is possible to set a business model in which this system can be introduced into acute care hospitals with 400 beds paying 20 million yen for overtime work a year, which is a relatively small payment, and the hospital income can be increased. For large-scale acute care hospitals, contribution to management increases because the payment for overtime work is high.

Hospitals introducing the system are requested to provide information on 'overtime work time of nurses and payment for it in each ward' as Key Performance Indicators (KPI), and changes in improvement are presented as numerical values.

Conclusions

This strategy and scenario were provided to nursing administrators and hospital administrators of five hospitals, and all hospitals responded to initiate the process for introduction, suggesting that the strategy proposed by this study is highly feasible.

Acknowledgements

We are deeply grateful for the collaboration of the project members and organizations.

This work was supported by JSPS KAKENHI (Grant-in-Aid for Scientific Research) Number 17H01608.

References

[1] Ministry of Health, Labour and Welfare: The general condition of the 2015 national medical expenditure http://www.mhlw.go.jp/toukei/saikin/hw/k-iryohi/15/dl/data.pdf

[2] Summary of the nosocomial business analysis reference index (for 2015 financial statements) http://www.wam.go.jp/hp/Portals/0/docs/gyoumu/keiei/pdf/shihyou/h27-byouin.pdf

[3] Ministry of Health, Labour and Welfare: 2015 hospital management index http://www.mhlw.go.jp/file/06-Seisakujouhou-10800000-Iseikyoku/0000164133.pdf

[4] CEN16880:2015 service in excellence

[5] Satoko TSURU,et.al.: Issues Concerning Volatilizing Situation in Nursing: A Survey on the Nursing Observation for the Post-Gastrectomy Patients, The 14th CJKMI2013, Scientific paper 33-38p, 2013

[6] Satoko TSURU, Fumiko WAKO, Miho OMORI and Kumiko Sudo : Problem Solving for Volatilizing Situation in Nursing : Developing Thinking Process Supporting System using NursingNAVI Contents, MIE2015 in Madrid, Scientific paper 5p , 2015

[7] Satoko TSURU, Miho OMORI, Manami INOUE and Fumiko WAKO : Quality Evaluation of Nursing Observations based on a Survey of Nursing Documents using NursingNAVI® Contents in JAPAN, The 13th International Congress on Nursing Informatics 2016. Scientific paper, 5pages , 2016

[8] Satoko Tsuru, Chitose Watanabe - the supervision and compilation: Clinical nursing in IT era -Navigator of thinking process for nursing-, Japan Standards Association (Tokyo), 2011 (Japanese)

[9] Satoko Tsuru, Fumiko Wako, Miho Omori, Chitose Watanabe, Mutsuko Nakanishi, Sawako Kawamura : Developing the Structured Knowledge Model to navigate the Nurses' Thinking Process in their Professional Judgment and Action. The 12th International Congress on Nursing Informatics 2014. Scientific paper, 5pages , 2014

[10] Satoko Tsuru: Recent concrete approach for medical security - Structuring of the Clinical Knowledge for the Patient Safety using PCAPS-, Quality, 39(4) 60-67, 2009 (Japanese)

[11] Satoko Tsuru, Yoshinori Iizuka, Masahiko Munechika : Structuring Clinical Nursing Knowledge using PCAPS :Patient Condition Adaptive path System, Proceedings of NI2009(Connecting Health and Humans IOS-Press ISBN 978-1-60750-024-7), scientific paper 391-395, 2009

[12] Satoko Tsuru, Yoshinori Iizuka , Masahiko Munechika : Structured Clinical Knowledge and its Application as a Socio-technology - PCAPS, Proc. of the 55th European Organization for Quality Congress, scientific paper CD-ROM PP1-8, 2011

[13] Satoko Tsuru, Akira Shindo, Shizuka Morimatsu, Kumiko Sudo, Akihide Masumoto, Miho Omori: Change Management for Productivity Improvement in Hospital using Structured Documentation System for Nursing, Proc. of the 61th EOQ Congress, Scientific paper CD-ROM 13p, 2017.

Address for correspondence

Satoko TSURU, PhD, Professor
Quality and Healthcare Social System Engineering Laboratory,
Department of Chemical System Engineering,
School of Engineering, The University of Tokyo
E-mail : tsuru@tqm.t.u-tokyo.ac.jp
ADDRESS : 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656 JAPAN
TEL : +81-3-5841-7301
FAX : +81-3-5841-7276

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190388

Estimating the Health-Related Quality of Life of Twitter Users Using Semantic Processing

Karthik V. Sarma[a], Brennan M. R. Spiegel[b], Mark W. Reid[b], Shawn Chen[a], Raina M. Merchant[c], Emily Seltzer[c], Corey W. Arnold[a]

[a] *Department of Radiological Sciences, University of California, Los Angeles, Los Angeles, CA, USA*
[b] *Center for Outcomes Research and Education, Cedars-Sinai Medical Center, Los Angeles, CA, USA*
[c] *Penn Medicine Center for Digital Health, University of Pennsylvania, Philadelphia, PA, USA*

Abstract

Social media presents a rich opportunity to gather health information with limited intervention through the analysis of completely unstructured and unlabeled microposts. We sought to estimate the health-related quality of life (HRQOL) of Twitter users using automated semantic processing methods. We collected tweets from 878 Twitter users recruited through online solicitation and in-person contact with patients. All participants completed the four-item Centers for Disease Control Healthy Days Questionnaire at the time of enrollment and 30 days later to measure "ground truth" HRQOL. We used a combination of document frequency analysis, sentiment analysis, topic analysis, and concept mapping to extract features from tweets, which we then used to estimate dichotomized HRQOL ("high" vs. "low") using logistic regression. Binary HRQOL status was estimated with moderate performance (AUC=0.64). This result indicates that free-range social media data only offers a window into HRQOL, but does not afford direct access to current health status.

Keywords:

Quality of Life, Social Media, Natural Language Processing

Introduction

Social media platforms are increasingly used for clinical research. Investigators now harness social media data for many purposes, including recruiting patients into clinical trials [1], measuring mood and sentiments [2], educating patients about social and medical concerns [3], and many others [4–9]. Although these applications expand the ability to perform health outcomes research on a broad scale, there is currently no reliable technique to measure patient reported outcomes (PROs), such as health related quality of life (HRQOL), using social media data.

The availability of a psychometrically valid, automatically generated social media PRO instrument would offer researchers and clinicians a new method to measure the "real life" effectiveness of interventions, behaviors, and therapies outside the confines of traditional research. Ideally, such a PRO would not require explicitly prompting users to speak specifically about health-related issues. Instead, the PRO would rely on existing social media posts and then estimate HRQOL using techniques such as natural language processing (NLP) and sentiment analysis.

An important step toward using social media for HRQOL estimation is to develop a method for translating diverse, unprompted, unstructured social media posts into information useable by clinicians and researchers. Given concerns about the representativeness and validity of social media data, there are limitations to measuring "true" HRQOL using social media analytics. Specifically, social media users may be systematically different from non-users, and even among users, many likely display a version of themselves that does not reflect their true functional state. Nevertheless, it would be useful to develop a method of extracting meaningful information from social media posts that operates within these limitations.

Several previous studies have looked at whether information about various components of an individual's health status could be inferred from unprompted social media use. For example, one study found that language used in Facebook posts could predict depression [10]. Another study found that aggregated Twitter data could predict county-level mortality from heart disease [11]. Others still have looked at PTSD [12] and county-level life satisfaction [13] with positive results. To the best of our knowledge, however, this is the first study that has tried to directly predict HRQOL measured by a validated PRO instrument using unprompted social media data.

In the present study, we collected social media posts from users of Twitter (http://www.twitter.com/), a web- and smartphone-based tool for users to express comments in short statements ("Tweets"), often appended with images or embedded with links to websites. We contacted users and administered a brief, empirically validated questionnaire that assesses HRQOL. We then used multiple analytic methods to develop predictors of their ground-truth HRQOL scores using data solely from their "Tweets". We hypothesized that, despite the limitations of free-range social media data, HRQOL could be classified with a degree of accuracy greater than chance.

Methods

Study Overview

We recruited a cohort of 1,831 Twitter users between May and December of 2015 through a combination of online solicitation and in-person contact. All participants were asked to complete the four-item Centers for Disease Control Healthy Days Questionnaire (CDC-4) questionnaire [14] at the time of enrollment and 30 days later to measure "ground truth" HRQOL. We then collected their tweets over a period of 60 days: 30 days before and after enrollment. Next, we collected and analyzed the tweets to determine characteristics that distinguish users with low or high HRQOL, using the CDC-4 as a gold standard. Of the 1,831 participants recruited, 878 had at least one accessible tweet during the analysis period, and the remaining participants were not used for analysis.

Participant Recruitment

We recruited participants through three mechanisms. First, we recruited general Twitter users by tweeting announcements through our institutional Twitter account (@CedarsSinai). Second, we recruited a general population sample through Cint, a survey research firm that recruited a sample of Twitter users from its survey cohort. Third, we supplemented our cohort of general Twitter users with a defined group of patients seeking care by recruiting individuals seen at the University of Pennsylvania Emergency Department for non-life-threatening, ambulatory conditions. Subjects in all cohorts were required to be 18 years of age or older and able to read English. Participants were entered into a drawing for one of two randomly selected $500 USD prizes. Participants with Twitter accounts who consented to participate completed the CDC-4 and provided their Twitter handle. Additional information about the University of Pennsylvania cohort has been published previously [15].

Study Measure

The CDC-4 Healthy Days Questionnaire [14] is a brief measure that assesses a number of factors, including general HRQOL on a five-point scale. This measure has been rigorously validated, demonstrating criterion validity with the Short Form 36-item health survey (SF-36) [16] as well as content, construct, and predictive validity, internal consistency, and test-retest reliability across diverse populations. The measure was chosen for its simplicity, reducing respondent burden to a minimum.

Acquisition and Preprocessing of Twitter Data

All tweets were acquired from the Twitter website. Tweets collected from each participant were segregated into two "collection waves": tweets from the 30 days prior to enrollment, and tweets from the 30 days including and subsequent to enrollment.

In order to reduce the proportion of marketing-related tweets in our dataset, we filtered out tweets containing any of a set of key phrases (any URLs, or the strings "sponsored", "my echo", "#mpoints", "giveaway", "@youtube", "giftcard", and "gift card.") that were determined to be associated in our dataset with marketing tweets by manual review of a random tweet sample. In addition, any users who tweeted more than 300 times in a given 30-day collection wave were removed from analysis in that period. This threshold was determined manually in order to remove accounts that were predominated by spam.

We then applied an initial preprocessing pipeline consisting of custom filters as well as Natural Language Tool Kit (NLTK) [17] tools to all tweets. The pipeline consisted of the following steps:

1. Unicode normalization form C (NFC) was applied
2. The Penn Treebank tokenizer was applied to generate tokens
3. Mention and reply tokens (e.g., @username) were removed
4. All tokens were converted to lower case
5. Repeated characters were removed iteratively until either no additional repeat characters existed or the token was transformed into a word in WordNet [18]
6. A Lancaster [19] stemmer was applied to each token

After preprocessing, we created “tweet sets” consisting of the concatenation of all of the tweets from each of the two collection waves (thus creating two tweet sets per participant). We then applied a term frequency-inverse document frequency (TF-IDF) transformer to each tweet set. For each of these tweet sets, we also computed several additional features. We fed each acquired tweet in each document set into the SentiStrength sentiment analysis tool, a validated opinion mining toolkit that computes positive and negative sentiment scores [20–22]. Then, we computed the mean, median, standard deviation, maximum, and minimum of the SentiStrength scores across all tweets in the set, and added them to the set’s data vector. We also added the number of tweets in the set to the vector.

In order to conduct semantic processing on the large volume of tweets, we utilized Latent Dirichlet Allocation (LDA), an automated method for discovering semantic themes in unstructured text [23]. For each concatenated tweet set, a vector representing the topics contained in the set was generated and added to the set’s data vector. Selected examples for demonstration are presented in Table 1.

Finally, we used the cTAKES processing engine [24] to map words in each tweet to concepts found in the Unified Medical Language System (UMLS) Metathesaurus and to detect concept negation. Instances in which the concepts were negated were considered to be separate concepts from non-negated versions. After determining the list of all concepts present in our dataset, for each 30-day tweet set, we created a data vector representing whether or not each concept was present or absent in the set. We then added this vector to the set’s data vector.

Statistical Analysis

We first dichotomized survey responses to HRQOL question 1a – a single 5-level global measure of HRQOL – into a “high” (>3) or “low” (<3) HRQOL binary variable; responses of exactly 3 were discarded. Next, we applied several machine learning techniques to produce a model that predicted this variable using the input vectors described above and 5-fold cross-validation. We employed the following techniques: standard logistic regression, random forest classifiers, and bagging generalized linear models (GLMs) trained using stochastic gradient descent (SGD).

We also applied a range of explicit and implicit feature selection techniques. Explicit methods included the use of a variance threshold for each column, principle component analysis (PCA), selecting the K best columns based on a univariate t-test, a unit origin transformer, and class balance subsampling. The random forest and bagging GLMs also used implicit feature selection, in which random subsets of the feature space were used to train classifiers which were then combined to produce an aggregate prediction.

We analyzed overall model performance in predicting our dichotomized HRQOL measure by conducting receiver operating characteristic (ROC) curve analysis. We evaluated resulting models using a held-out test set of 20% of the dataset. We generated accuracy scores, confusion matrices, and ROC curves for each test. Our main analysis was performed using data from subjects recruited from Cint or Cedars-Sinai. A follow-up analysis was also performed combining this data with data obtained from the University of Pennsylvania.

Results

Participant Characteristics

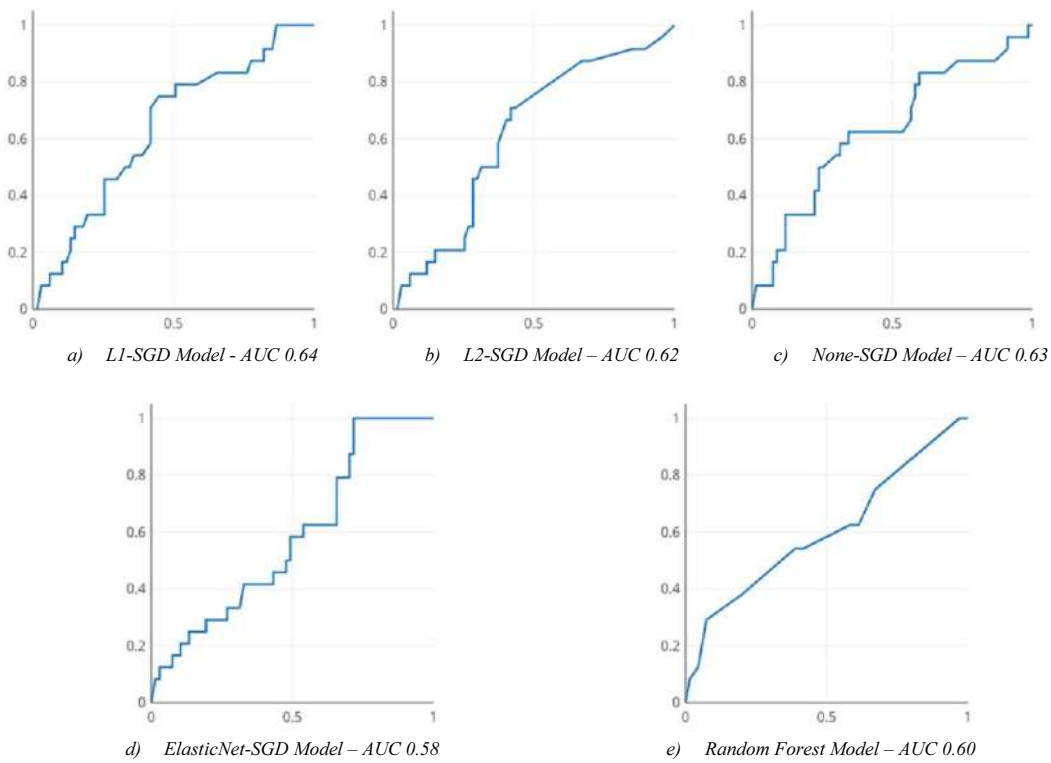
Table 2 presents demographic information of the participants recruited from Cedars-Sinai, Cint, and the University of

Pennsylvania. Because some users signed up multiple times through multiple accounts, we removed duplicate entries from the study. In addition, we removed participants who had protected (i.e., non-public) Twitter accounts that could not be analyzed, or who had zero tweets in the analysis period. In the end, recruitment efforts yielded 1,831 users who completed the survey at enrollment.. We could retrieve at least one tweet from 835 of the original 1,831 subjects. The remaining 996 subjects either did not tweet during the 60-day period or had a protected or deleted twitter account when tweet collection was being performed. Of the 835 subjects with tweets, 581 of them completed both surveys, and the remaining 254 only completed the first survey. The latter subjects were discarded from the analysis. In the follow-up analysis, data obtained from the University of Pennsylvania for 43 patients (with accessible Tweets) was added to the data from Cedars-Sinai and Cint.

Over each 30-day period, participants posted an average of 99 messages (range 0 to 4880). The rate at which patients posted did not significantly differ between the pre-enrollment period and post-enrollment period (p=0.19). The initial tweet corpus had neutral sentiment overall, as measured by the difference of the average positive and negative SentiStrength score across all Tweets in this analysis (M(Pos) = 1.58, SD(Pos) = 0.81, M(Neg) = 1.43, SD(Neg) = 0.84).

Predictive Model Accuracy

To develop our model, we employed a grid search to test various combinations of model designs and subsets of our data. The tested model designs included generalized linear models with varying forms of regularization, as well as a random forest model. We also tested subsetting our data to include or exclude each of the three data sources (Cedars-Sinai, Cint, and UPenn). We also tested different methods of handling subjects who completed both surveys, including using only the first collection wave and survey, and using both collection waves and surveys. The resulting models achieved AUCs ranging from 0.58 to 0.64 for predicting the ground truth binary HRQOL status (Figure 1). The highest AUC of 0.64 was achieved using a bagging GLM model with L1 regularization, using all three datasets and all available collection waves. The model used 100 estimators with a maximum sample and feature proportion of 0.75 with bootstrapping, and was trained with 500 iterations of SGD. Further optimization of logistic regression parameters, model design and feature selection, or tweet subsets was unable to achieve a greater AUC than 0.64.



a) L1-SGD Model - AUC 0.64 *b) L2-SGD Model – AUC 0.62* *c) None-SGD Model – AUC 0.63*

d) ElasticNet-SGD Model – AUC 0.58 *e) Random Forest Model – AUC 0.60*

Figure 1 - Best ROC Curves for Five Classes of Prediction Models. Models evaluated included bagging generalized linear models using a) L1 regularization, b) L2 regularization, c) no regularization, or d) ElasticNet combination regularization, as well as e) a random forest model. The best performance was obtained from the bagging GLM model using L1 regularization, with an AUC of 0.64. This model used 100 estimators with a maximum sample and feature proportion of 0.75 with bootstrapping, and was trained with 500 iterations of SGD. As can be seen from the ROC curves, the three standard logistic regression models outperformed the random forest model and the combination ElasticNet model.

Table 1 – Selected LDA-generated Topics and Highly Related Tweets

Topic Words (First Three)	Example Tweet
data, health, gt	@[redacted] Yes, we needs a digital platform between patients & medical professional to work 4 better outcome with medication / treatment
t, patient, patients	T4 Patients who are actively engaged should get priority but if a patient wants they should go. 5% of attendees! ideally. #hchlitss
cell, sickle, does	@[redacted] Sickle Cell Disease: underfunded, underresourced, underappreciated
lupus, lupuschat, w	@[redacted] I am in DC to advocate for #Lupus Research. From Culver City to Capitol Hill, here to share my journey of 33 years.

Table 2 – Demographic Characteristics of Twitter Users in Study

Demographics	Value (Cint/CSHS)	Value (UPenn)
Number of Subjects	835	43
Age (Mean ± SD)	40.0 ± 12.6	26.9 ± 7.4
Gender (%)		
Male	31.6%	27.9%
Female	68.1%	72.1%
Transgender/Other	0.3%	0%
Race (%)		
American Indian / Alaskan Native	3.0%	0%
Asian	5.6%	0%
Native Hawaiian / Pacific Islander	0.4%	0%
Black or African-American	11.9%	62.8%
White	76.6%	34.9%
Other / Unknown / Declined to answer	2.5%	0%

Discussion

The goal of this study was to develop a PRO instrument that estimates HRQOL from a user's social media posts on Twitter. Further, we studied the feasibility and validity of measuring HRQOL without prompting users to talk specifically about health-related issues. After using a wide range of text processing methods, the best achievable AUC in predicting ground-state HRQOL was 0.64. Although this performance may be considered sub-optimal, it is better than chance and suggests there is at least a correlation between the language used in Tweets and actual HRQOL; a notable finding considering that the data collected was not posted with the intent to provide insight into the HRQOL status of the participants. However, the results also indicate that free-range social media data only offer a window into HRQOL, but do not afford direct access to current health status. Despite the inherent limitations in this pragmatic, naturalistic study, our classification algorithm exhibited moderate performance in estimating true HRQOL status (i.e. high vs. low using CDC-4 as the gold standard) with an AUC of 0.64 by evaluating the language used in 140 character Tweets, the allowable size of microposts during the time this study was conducted. Of note, Twitter currently allows 240 characters, so it is possible that longer posts could include more health-related data and demonstrate a stronger relationship with ground-truth HRQOL.

We hypothesize that the sub-optimal performance of our classifier is due to a poor signal to noise ratio in the dataset. Most people do not routinely post about their health status, including patients undergoing active care (e.g. the ED patients in this sample). In addition, the subjects from the Cint cohort may be more likely than the baseline Twitter user to tweet about marketing-related subjects rather than personal subjects, because they are paid Twitter marketers for Cint. Considering that subjects in this study were obtained from a variety of sources, received no instructions about what to post, included a combination of patient and general population samples, and were limited to a maximum of 140 characters at a time per tweet, the ability to predict HRQOL above chance suggests that this approach – albeit imperfect – still has merit. Despite the inherent limitations in this preliminary study, our classification algorithm exhibited moderate performance in estimating true HRQOL status (i.e. high vs. low using CDC-4 as the gold standard) with an AUC of 0.64 by evaluating the language used in 140-character Tweets.

Other investigators have encountered sub-optimal performance using social media data to estimate health status. For example, Nascimento and colleagues [9] note that their automated processes for identifying "migraine" versus descriptions of an actual patient's migraine experiences were difficult to separate. Indeed, in much of our own dataset, even topics generated using NLP that appeared to be health-related were often about commerce (i.e., other users should consider buying product X, rather than the user attesting to how product X treated their condition). Future research could attempt to more specifically recruit "high-volume" non-commercial Twitter users in order to obtain richer features that maybe more predictive, though this may compromise the generalizability of the method. In addition, future efforts could aim to use a larger patient-specific cohort via targeted recruitment, rather than a general population sample. Nevertheless, the challenge of gaining health insights from unstructured "free-range" social media posts is formidable; additional studies are needed to assess the potential for this data to achieve sufficient accuracy for real world clinical and interventional applications.

Conclusions

In conclusion, we found that analysis of free-range social media posts can predict HRQOL better than chance, but that this technique remains an imperfect method for assessing current health status despite testing a wide range of text processing methods. The best performing model (AUC = 0.64) was a bagging GLM with L1 regularization; this may be helpful when selecting among semantic processing techniques. We hope this study may serve as a template for future research in extracting health data from unstructured social media posts, and believe future studies could improve upon our work by refining recruitment efforts to avoid commercial accounts, expanding the cohort size, and making use of a broader spectrum of social media data from additional platforms.

Acknowledgements

KVS acknowledges support from NIH NCI F30CA210329, NIH NIGMS GM08042, and the UCLA-Caltech Medical Scientist Training Program. BMS acknowledges support from the Robert Wood Johnson Foundation. RMM acknowledges support from NHLBI R01 HL122457 and the Robert Wood

Johnson Pioneer Portfolio. All procedures and recruitment tools were reviewed and approved by the appropriate Institutional Review Board (Cedars-Sinai Pro00038312; University of Pennsylvania PRO 819442). The authors appreciated the opportunity to partially present early components of this work as an abstract in poster form at the 2018 AMIA Annual Symposium and the 2018 NeurIPS Machine Learning for Health workshop.

References

[1] M.S. Tweet, R. Gulati, L.A. Aase, and S.N. Hayes, Spontaneous Coronary Artery Dissection: A Disease-Specific, Social Networking Community-Initiated Study, *Mayo Clin. Proc.* **86** (2011) 845–850. doi:10.4065/mcp.2011.0312.

[2] S.A. Golder, and M.W. Macy, Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures, *Science (80-.).* **333** (2011). doi:10.1126/science.1202775.

[3] S.I. Thaker, A.S. Nowacki, N.B. Mehta, and A.R. Edwards, How U.S. Hospitals Use Social Media, *Ann. Intern. Med.* **154** (2011) 707. doi:10.7326/0003-4819-154-10-201105170-00021.

[4] E.R. Weitzman, B. Adida, S. Kelemen, and K.D. Mandl, Sharing Data for Public Health Research by Members of an International Online Diabetes Social Network, *PLoS One*. **6** (2011) e19256. doi:10.1371/journal.pone.0019256.

[5] P. Wicks, T.E. Vaughan, M.P. Massagli, and J. Heywood, Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm, *Nat. Biotechnol.* **29** (2011) 411–414. doi:10.1038/nbt.1837.

[6] C. St Louis, and G. Zorlu, Can Twitter predict disease outbreaks?, *BMJ*. **344** (2012) e2353. doi:10.1136/bmj.e2353.

[7] L. Sinnenberg, C.L. DiSilvestro, C. Mancheno, K. Dailey, C. Tufts, A.M. Buttenheim, F. Barg, L. Ungar, H. Schwartz, D. Brown, D.A. Asch, and R.M. Merchant, Twitter as a Potential Data Source for Cardiovascular Disease Research, *JAMA Cardiol.* **1** (2016) 1032. doi:10.1001/jamacardio.2016.3029.

[8] M. De Choudhury, S. Counts, and E. Horvitz, Predicting postpartum changes in emotion and behavior via social media, in: Proc. SIGCHI Conf. Hum. Factors Comput. Syst. - CHI '13, ACM Press, New York, New York, USA, 2013: p. 3267. doi:10.1145/2470654.2466447.

[9] T.D. Nascimento, M.F. DosSantos, T. Danciu, M. DeBoer, H. van Holsbeeck, S.R. Lucas, C. Aiello, L. Khatib, M.A. Bender, J.-K. Zubieta, A.F. DaSilva, and A.F. DaSilva, Real-Time Sharing and Expression of Migraine Headache Suffering on Twitter: A Cross-Sectional Infodemiology Study, *J. Med. Internet Res.* **16** (2014) e96. doi:10.2196/jmir.3265.

[10] J.C. Eichstaedt, R.J. Smith, R.M. Merchant, L.H. Ungar, P. Crutchley, D. Preoţiuc-Pietro, D.A. Asch, and H.A. Schwartz, Facebook language predicts depression in medical records., *Proc. Natl. Acad. Sci. U. S. A.* **115** (2018) 11203–11208. doi:10.1073/pnas.1802331115.

[11] J.C. Eichstaedt, H.A. Schwartz, M.L. Kern, G. Park, D.R. Labarthe, R.M. Merchant, S. Jha, M. Agrawal, L.A. Dziurzynski, M. Sap, C. Weeg, E.E. Larson, L.H. Ungar, and M.E.P. Seligman, Psychological Language on Twitter Predicts County-Level Heart Disease Mortality, *Psychol. Sci.* **26** (2015) 159–169. doi:10.1177/0956797614557867.

[12] A.G. Reece, A.J. Reagan, K.L.M. Lix, P.S. Dodds, C.M. Danforth, and E.J. Langer, Forecasting the onset and course of mental illness with Twitter data, *Sci. Rep.* **7** (2017) 13006. doi:10.1038/s41598-017-12961-9.

[13] H.A. Schwartz, J.C. Eichstaedt, M.L. Kern, L. Dziurzynski, R.E. Lucas, M. Agrawal, G.J. Park, S.K. Lakshmikanth, S. Jha, and M.E.P. Seligman, Characterizing geographic variation in well-being using tweets, in: Seventh Int. AAAI Conf. Weblogs Soc. Media, 2013.

[14] Centers for Disease Control and Prevention, Measuring healthy days: Population assessment of health-related quality of life, *Atlanta CDC*. (2000) 4–6.

[15] K.A. Padrez, L. Ungar, H.A. Schwartz, R.J. Smith, S. Hill, T. Antanavicius, D.M. Brown, P. Crutchley, D.A. Asch, and R.M. Merchant, Linking social media and medical record data: a study of adults presenting to an academic, urban emergency department, *BMJ Qual. Saf.* **25** (2016) 414–423. doi:10.1136/bmjqs-2015-004489.

[16] R.D. Hays, C.D. Sherbourne, and R.M. Mazel, The RAND 36-Item Health Survey 1.0., *Health Econ.* **2** (1993) 217–27. http://www.ncbi.nlm.nih.gov/pubmed/8275167 (accessed April 16, 2017).

[17] S. Bird, E. Klein, and E. Loper, Natural language processing with Python: analyzing text with the natural language toolkit, 3rd ed., O'Reilly Media, Inc., 2009.

[18] G.A. Miller, and G. A., WordNet: a lexical database for English, *Commun. ACM*. **38** (1995) 39–41. doi:10.1145/219717.219748.

[19] C.D. Paice, and C. D., Another stemmer, *ACM SIGIR Forum*. **24** (1990) 56–61. doi:10.1145/101306.101310.

[20] M. Thelwall, K. Buckley, and G. Paltoglou, Sentiment strength detection for the social web, *J. Am. Soc. Inf. Sci. Technol.* **63** (2012) 163–173. doi:10.1002/asi.21662.

[21] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, Sentiment in short strength detection informal text, *J. Am. Soc. Inf. Sci. Technol.* **61** (2010) 2544–2558. doi:10.1002/ASI.V61:12.

[22] M. Thelwall, K. Buckley, and G. Paltoglou, Sentiment in Twitter events, *J. Am. Soc. Inf. Sci. Technol.* **62** (2011) 406–418. doi:10.1002/asi.21462.

[23] D.M. Blei, A.Y. Ng, and M.I. Jordan, Latent Dirichlet Allocation, *J. Mach. Learn. Res.* **3** (2003) 993–1022.

[24] G.K. Savova, J.J. Masanz, P. V Ogren, J. Zheng, S. Sohn, K.C. Kipper-Schuler, and C.G. Chute, Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications., *J. Am. Med. Inform. Assoc.* **17** (2010) 507–13. doi:10.1136/jamia.2009.001560.

Address for correspondence

Corey W. Arnold, PhD, UCLA Medical Imaging Informatics Group, 420 Westwood Blvd, Suite 420, Los Angeles, CA 90024-2943. cwarnold@mednet.ucla.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190389

An Evaluation of the Technical Quality Within the Belgian Electronic Prescription: A Cross-Sectional Study

Sven Van Laere, Ronald Buyl

Department of Public Health (GEWE), Vrije Universiteit Brussel, Brussels, Belgium

Abstract

The increased use of eHealth services has shifted focus towards quality and interoperability. This study quantifies the technical quality of the ePrescription in Belgium by validating the construction of the Kind Messages for Electronic Health Records (KMEHR) used for communicating between prescriber and pharmacist, and by validating the digital signature within the ePrescription. A two-way cluster randomized subset of prescriptions of the tarification service Koninklijk Limburgs Apothekers Verbond (KLAV) is used (n=82,952). Here, 38,032 prescriptions (45.85%) were handled electronically, but 180 prescriptions (0.22%) contained an empty KMEHR message at tarification. All 37,852 ePrescriptions (100%) passed the XSD Schema validation, but only 29,428 ePrescriptions (77.74%) passed the Software Development Kit (SDK) XPath validation rule set of the national Recip-e project. 37,622 ePrescriptions (99.39%) passed the validity test for having a correct digital signature. Improvement towards the construction and the digital signature is required in order to fully dematerialize the ePrescription workflow.

Keywords:

Electronic Prescribing, Quality Control, Community Pharmacy Services.

Introduction

In general, an electronic health (eHealth) system on a national basis can help in making healthcare systems economically sustainable. An increased use of eHealth services and in particular the usage of electronic prescriptions (ePrescriptions), has shifted focus towards quality [1] and interoperablity[2].

Electronic prescribing, or "ePrescribing" is the computer-based electronic generation, transmission and filling of a prescription, taking the place of paper and faxed prescriptions [3]. In the flow of ambulatory ePrescribing, there are often three human actors (the prescriber, the patient and the pharmacist) and two system actors (prescribing and dispensing providers).[4] Sometimes, a reference to the 3 Ps will be made in literature when addressing these human actors.[5] In Belgium, a third party payment applies, where the patient only pays the amount to the pharmacist that is not reimbursed by healthcare insurance. That is why in this flow of digital communication, the tarification services –responsible to invoice the remaining amount of medication costs to the healthcare insurance of the patient– might be added.

In Belgium, ePrescribing made its entrance in April 2007, where the Belgian government launched the e-MED project with the intention of developing a coherent action plan for making the ePrescription of medicines in ambulatory care possible.[6] After a national pilot phase in the years 2009-2012, the national project was introduced to the public in 2014. From then on, patients that visited a general practitioner (GP) to prescribe his medication, using the integrated electronic health record (EHR) of the patient, would get a proof of electronic prescription when the GP's software was connected to the national ePrescribing services. Independently, the national project Recip-e is responsible for the temporary storage of the encoded ePrescriptions on a national server. Both the paper-based prescription and the ePrescription were –and are still– accepted as a medium for communication between the prescriber, the patient, the pharmacist and the tarification service.

In a later stage, the national project evaluated the authentication methods used when prescribing. With paper-based prescriptions, the combination of the healthcare party identifier (RIZIV number provided by Rijksinstituut voor Ziekte- en Invaliditeitsverzekering (RIZIV)) and the signature was used to identify the prescriber. When using the ePrescripion, the combination of the RIZIV number with the digital signature, in an encrypted and timestamped message was chosen as evidential value to show that the prescriber made the ePrescription. In 2016, the project opted for using an additional SAML-token (Security Assertion Markup Language token) included in the encrypted message based on the RIZIV nr of the prescriber, to allow healthcare providers to have a single sign-on for a couple of hours.

The RIZIV number for each care provider is provided by the RIZIV organisation. This number serves as a unique healthcare identification number and is constituted of 11 numerical characters. This number can be broke down into 4 groups with each having a function for identification. The structure of the number consists of (1) a professional number, (2) a serial number, (3) a checksum number and (4) a qualification number, e.g. 1-12345-12-123. With the professional number, a distinction can be made between physicians, dentists, hospitals, pharmacists, etc. Since the qualification number (last 3 digits) is, in fact, a reference towards the specialisation and not really identifying as such, one could consider the first 8 numerical characters as identifying for the healthcare provider. However, for reimbursement purpose, it is necessary that the KMEHR message contains the correct RIZIV number consisting out of 11 numeric characters.

As of 2017, an additional change compared to the paper-based prescription came into place, where an extra barcode on top of the electronic proof of prescribing was placed, i.e. the Recip-e ID (RID). This RID barcode is used to uniquely identify an ePrescription and to prevent patients from frauding with their medication prescriptions. In Belgium, people still get a paper-based proof of prescription where this RID is printed on, in case of an ePrescription. It does not fulfill the function of prescription, but can only be used by the community pharmacies in combination with the patient's electronic identity (eID) to get dispensed. The Belgian government is thinking

about dematerialising the prescription process. This means that this paper proof, that is actually used in the transition phase towards dematerialisation, will be removed in this process of ePrescribing.

When digitalising the healthcare sector on a national basis, one needs to adapt to healthcare standards. Healthcare standards often used for ensuring the interoperability are ISO 13606 [7], HL7 [8] and SNOMED CT [9]. In Belgium, the KMEHR standard [10] is used as a communication standard to be able to send messages and the CNK (Code National(e) Kode) standard [11] for identifying the medication. In essence, the KMEHR standard is a structured XML message.

The Belgian government is responsible for the maintenance of both standards and sending updates when necessary, e.g. when a new medication is brought to the market. Proper versioning of these standards among different eHealth initiatives on a national basis is thus required in order to be able to interpret the ePrescription in a uniform way.

In Belgium, a free market for the generation of software for both prescriber and community pharmacist additionally complicates the eHealth story. However, to be able to talk health informatics in a uniform way for ePrescriptions, the Recip-e party developed a software development kit (SDK) for software vendors.[12] The usage of this SDK is not an obligation, but it is a recommendation for software vendors providing software willing to interpret the electronic KMEHR messages in a uniform manner. An incorrect creation of an ePrescription, possibly caused by a software not integrating the SDK, can lead to an incorrect interpretation of the ePrescription KMEHR message at the community pharmacist's side and maybe lead to the pharmacist incorrect dispensing the patient's medication.

Currently, the patient still has a paper proof of prescription that he receives in order for the community pharmacist to read in the RID printed on top of the ePrescription. The patient will then be able to retrieve the prescribed medication. This RID barcode on top facilitates the reading in the software of the community pharmacist. When dematerialising, only the eID of the patient will be used to retrieve ePrescriptions linked to the patient's eID. Therefore, it will be necessary for KMEHR messages to not contain any technical errors. In this study, the soundness of the Belgian ePrescription will be investigated. The focus will not be put on the correct medication prescribing as such, but rather on the technical construction of these messages and the correct use of digital signatures in Belgian ePrescriptions.

Methods

Data of the tarification service Koninklijk Limburgs Apothekers Verbond (KLAV), one of the bigger tarification services in Belgium, was used in order to obtain a randomized subset. Since this tarification service provides services all over Belgium, location was used as a first randomization factor. A second randomization factor used was the amount of prescriptions a community pharmacy sends to the tarification service. These factors are now combined in a two-stage cluster random sampling technique, where randomization was first performed over the location (province) within Belgium and secondly over the community pharmacies within the selected provinces based on their number of prescriptions.

With this sampling technique, a random sample of 50 community pharmacies was selected. The prescriptions retrieved by KLAV from the month June 2018 were used for observing the technical quality of both the KMEHR message and the digital signature within the retrieved ePrescriptions.

Tarification services do not receive these KMEHR messages in plain text. Pharmacists use multiple encryptions to encapsulate the message in different constructs in a time-stamped manner before sending it for tarification. Different elements within different levels of encryption will be required to verify the quality.

Ethical approval for this study was obtained from the ethical committee of UZ Brussels (nr. 2018/218).

Quality of the KMEHR Message

Since the KMEHR message is a structured XML message, the construction of it can first be validated towards an XSD Schema. However, the quality of the KMEHR message was mainly checked by applying the validations that are found in the SDK provided by Recip-e.

These validations exist of XPath validations rules to verify the construction of the KMEHR messages. In these rules a minimum and a maximum number of occurrence for a specific tag or attribute were used, see Rule example below. These rules are normally used in the prescriber software to validate the message before sending it to the national server in an encrypted way. However, since two different KMEHR structures are accepted in this validation of the SDK, i.e. against KMEHR version 1.17.1 and KMEHR version 1.19.0, different validation rule sets applied.

Hereunder, an example of an XPath validation rule is explained.

```
/kmehrmessage/folder/patient;1;1
```

This rule states that within the folder element of the KMEHR message, there can only be one patient element (minimum = 1; maximum = 1). When no or more than one patient element is specified, this XPath expression will fail.

Quality of the Digital Signature

The evidential value of a digital signature in the Belgian eHealth system is two-fold. First of all the certificate within the sealed KMEHR message has to match the holder of key (HOK) certificate within the SAML message. Secondly, also the healthcare identifying RIZIV numbers within the KMEHR and SAML have to match, with an extra requirement of having an 11-character RIZIV number used in the KMEHR message.

One exception applies to hospitals. Within the hospital, you have different specialist physicians. To lower the administrative burden within the national ePrescribing project a "circle of trust" is organised between the organization, which is the hospital, and the employer working for the organization, which is the specialist physician. Since there is a circle of trust relationship, the prescriber will use the certificate of the hospital. When verifying the quality of the digital signature, the certificates should match with that of the hospital organization and in the KMEHR message the hospital should correspond as being the sender and the physician should act as an author. Both are elements in the KMEHR XML structure. In an ePrescription of a healthcare provider not working at a hospital, both the sender and author element should represent the same person.

One should note that June (month of inclusion) is one of the first months in Belgium where also ePrescriptions are used in the hospital context.

Results

In Belgium, the population of community pharmacies exists of in total 4,943 community pharmacies (source Sirius Insights, 2017 - [13]). In the month June 2018, 563 community pharmacies (11.39%) used the KLAV tarification services (see Table 1). In that month, this tarification service collected a total

of 794,724 prescriptions, for which no special tarification rules applied.

Using the two-stage cluster random sampling technique, where a selection of 50 community pharmacies was chosen, a total subset of 82,952 prescriptions was observed. Here all prescriptions of these 50 pharmacies were included (both paper-based prescriptions and ePrescriptions). Within this subset, a total of 38,032 prescriptions treated as ePrescriptions was found (see Table 1). An ePrescription rate of 45.85% was observed within this randomized subset.

However, when analysing the content of these messages that were retrieved at the side of the tarification service, 180 ePrescriptions (0.48%) appeared to have no KMEHR message (see Figure 1). This can be explained by a possible fallacy in the Belgian eHealth system when trying to consult eHealth services of the national eHealth platform when transmitting the data to the tarification service or due to a software error at the side of the pharmacist. These errors are solved by retransmitting the data for tarification in the next month.

In these messages, both the quality of the KMEHR message as the quality of the digital signature was checked (see Figure 1). Quality of the KMEHR message is evaluated in terms of the XPath validation rules provided in the SDK of Recip-e that serves software vendors for prescribers, and in terms of the XSD Schema evaluation. Quality of the digital signature was mainly dependent on the prescriber type (GP, hospital or dentist).

Table 1. Community pharmacy and KLAV's subset of prescriptions characteristics

Community pharmacies		
Level	**N**	**%**
National	4,943	100.00
KLAV (total)	563	11.39
KLAV (random sample)	50	1.01
Prescriptions at KLAV		
Level	**N**	**%**
Total in June 2018	794,724	100.00
Prescriptions within randomized group of community pharmacies	82,952	10.44
ePrescriptions within randomized group of community pharmacies	38,032	4.79
ePrescriptions with no KMEHR message	180	< 0.01

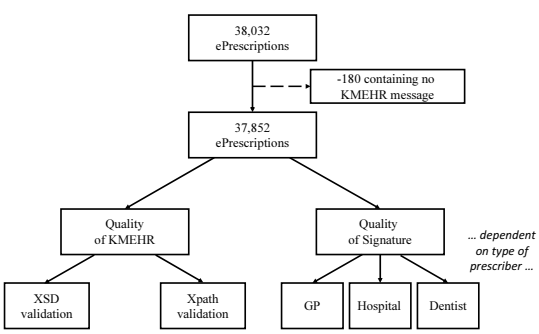


Figure 1. Flow chart of quality validation for the KMEHR and the digital signature

Quality of the KMEHR Message

Currently, KMEHR version 1.19.0 is the latest version that is accepted to communicate with in terms of the ePrescription. Version 1.17.1 is also acceptable in the Belgian context, but preference is given to the later standard. Since the versioning plays an essential role in the XPath validation of the KMEHR message, the distribution of KMEHR standards used over the different software packages at the prescriber's side is represented in Table 2. This versioning is retrieved within the KMEHR message using the SV attribute (scheme version). When no KMEHR message was observed, the versioning was not able to be retrieved.

Table 2. Distribution of different versions of KMEHR communication standard (n = 37,852)

Versioning	**N**	**%**
SV=1.1 – KMEHR version 1.0.0	601	1.59
SV=1.3 – KMEHR version 1.1.0	85	0.22
SV=1.6 – KMEHR version 1.5.0	47	0.12
SV=1.8 – KMEHR version 1.7.0	14	< 0.01
SV=1.18 – KMEHR version 1.17.1	*6,906*	*18.24*
SV=1.19 – KMEHR version 1.18.0	328	0.87
SV=1.20 – KMEHR version 1.19.0	*23,470*	*62.00*
SV=1.23 – KMEHR version 1.22.0	6,401	16.91

Validation against the XSD Schema of KMEHR version 1.19.0, resulted in a 100% success rate, see Table 3. However, when looking at the XPath validation rule set, a lower validation rate of only 77.74% in total was found. This result was split into two parts, according to the version of the KMEHR standard.

Table 3. Prescriber type for ePrescription (n = 37,852)

XSD validation	***SUCCESS***		***FAIL***	
Version	**N**	**%**	**N**	**%**
1.19.0	37,852	100.00	0	0.00
Total	*37,852*	*100.00*	*0*	*0.00*
XPath validation				
Version	**N**	**%**	**N**	**%**
1.17.1	6,710	17.73	31,142	82.27
1.19.0	22,718	60.02	15,134	38.98
Total	*29,428*	*77.74*	*8,424*	*22.26*

Quality of the Digital Signature

Verifying the evidential value of a digital signature on an ePrescription requires to know the prescriber type of it, since physicians working in a hospital are treated as working within the circle of trust of the hospital. Table 4 provides descriptive statistics about the prescriber type within the ePrescriptions in the retrieved subset. When no KMEHR message was observed, the prescriber type was not able to be retrieved.

The majority of ePrescriptions was generated by a general practitioner (GP) (n=34,455; 91.03%). The physicians working in a hospital prescribed 3,061 ePrescriptions (8.09%). Dentists prescribed only 336 prescriptions (0.89%) electronically.

The evidential value of a digital signature in Belgium consists of correct use of certificates and correct use of the identifying RIZIV number of the prescribing healthcare party. In Table 4, the matching percentages for the RIZIV number within the SAML versus the RIZIV number within the KMEHR message and the HOK certificate within the SAML versus the certificate in the sealed KMEHR message are represented.

When observing the use of the certificates (see Table 5), a 100% matching rate was observed. When looking at the identifying RIZIV number of the prescriber, a matching rate of 99.39% between the SAML and the KMEHR message was observed. These rates varied across different types of prescribers.

The mismatch between RIZIV identifiers, where an 11 numeric character string is required in the KMEHR message (see Table 5), occurred most frequently when prescribed by dentists (1.79%), followed by an ePrescription generated by a GP (0.65%). No mismatch between RIZIV identifiers occurred within the hospital setting.

Table 4. Prescriber type for ePrescription (n = 37,852)

Prescriber type	**N**	**%**
General practitioner (GP)	34,455	91.03
Hospital	3,061	8.09
Dentist	336	0.89

Table 5. Evidential value of digital signature (n = 37,852)

Certificate match	***SUCCESS***		***FAIL***	
Prescriber type	**N**	**%**	**N**	**%**
GP	34,455	100.00	0	0.00
Hospital	3,061	100.00	0	0.00
Dentist	336	100.00	0	0.00
Total	*37,852*	*100.00*	*0*	*0.00*
RIZIV number match with 11 characters in KMEHR				
Prescriber type	**N**	**%**	**N**	**%**
GP	34,231	99.35	224	0.65
Hospital	3,061	100.00	0	0.00
Dentist	330	98.21	6	1.79
Total	*37,622*	*99.39*	*230*	*0.61*

Discussion

A two-stage cluster randomized subset of prescriptions was used from one of the biggest tarification services in Belgium, i.e. KLAV. In total, 45.85% of the prescriptions (38,032 out of 82,952 prescriptions) were handled as ePrescription by the community pharmacist. 180 prescriptions were excluded, since not containing a KMEHR message at the side of the tarification service. This study shows that the construction of the Belgian ePrescription is 100% valid towards the XSD Schema, but is only valid in 77.74% (29,428 out of 37,852 ePrescriptions) of the cases when comparing the KMEHR messaging standard used to the national ePrescription project validation rule set. Secondly, a higher validation rate of 99.39% (37,622 out of 37,852 ePrescriptions) was observed when observing the digital signature.

The low validation rate of 77.74% for the KMEHR standard can partly be explained by the free market of software vendors in Belgium for both the prescriber and community pharmacist. Using different versions of the KMEHR standard over the different vendors might cause problems interpreting ePrescriptions in different software at the pharmacist side.

However, since the XSD validates all prescriptions correctly, one may state that the Belgian KMEHR standard is used quite loosely when validated towards its schema. A technique like Java Architecture for XML Binding or JAXB [14], might help to capture data in a uniform way using the XML of the KMEHR message in the back-end of the software.

However, since the validation rules do not always apply to the KMEHR message and most software programmers use only a limited subset of test case ePrescriptions to work with when writing the software, the Belgian ePrescription might give problems in reading and interpreting them correctly once entered in the working field.

A study by Anderson, reveals a possible reason why this discrepancy might be observed between the actual ePrescription (KMEHR) and its validation rules. In his work [15] about barriers to eHealth adoption in the practice of the physician, 79.3% of the physicians named the vendor's inability to deliver acceptable products as a significant barrier to implementation of IT in their practice. In Belgium, a prescriber might think that by using his EHR software, he produces standard ePrescriptions in the front-end and standard KMEHR messages in the back-end. However, this study demonstrates this is not always the case.

The benefits of putting a digital signature are first of all that it ensures the identity of the sender. Secondly, compared to the manual signature on the paper-based prescription, it saves time since the use of a SAML-token lasts for a couple of hours. Thirdly, the digital signature will ensure the validity of the content of the prescription. The slightest change to the digital content will revoke the validity of the digital signature.

Plans of dematerialisation for the Belgian ePrescription require the validity of the digital signature in order to guarantee the authenticity of the sender. In Belgium, the digital signature was introduced at the beginning of the existence of the ePrescription. However, not all prescribers support the idea of using a digital signature using an ePrescription. For example, in Turkey [16] 78.23% (194 of 248 prescribers) reported that they did not use an electronic signature to sign ePrescriptions and 52.42% (130 of the 248 prescribers) doubts or does not support the use of an electronic signature.

Sometimes an invalid signature might lead to errors in the workflow of the community pharmacist. In a study by Reed-Kane et al. [17] over a four-week study period for pharmacy compounding, they observed the most frequent error was an invalid signature (31.25%; 10 of the 32 errors identified within 138 ePrescriptions). This potentially leads to an increase in working time handling the ePrescription. In that same study, the most frequent used error resolution method was a phone call to the physician (59.38%; 19 of the 32 errors identified within 138 ePrescriptions).

To the best of our knowledge, this is a first study that evaluates the technical construction of this kind of messages. Lots of studies in the field of ePrescriptions evaluate the number of prescribing errors and its causes, in different settings [18-21]. However, this study also has some limitations. Due to the cross-sectional nature of this study, it is not able to sort out the exact reason for this rather low validity. For example, the author of the prescription is responsible for the placing of a valid signature using the right certificate, but the vendor is responsible for a good implementation of it when using the identifying RIZIV numbers. Multiple aspects are thus expected to be at the basis of these validation rates. A second limitation is that this study uses prescriptions of community pharmacies affiliated with one tarification service in Belgium. However, due to the fact that KLAV delivers services to all provinces in Belgium and since a two-stage cluster randomized subset was used, the conclusions may be generalized to the population of ePrescriptions in Belgium. Another limitation is the low number of ePrescriptions coming from hospitals, limiting the generalizability of the study.

Future research should focus on a timely review of these validations in order to be able to continue towards a national dematerialisation. This study only focuses on the construction of the KMEHR message and the validity of the digital signature within the Belgian ePrescription. An important part of the ePrescription lies in the validity of the medication section of an ePrescription. This part of the prescription should undergo a similar evaluation towards the CNK standard, which is the identifying code in Belgium for medication products. Currently, pharmacy compounding prescribing within the Belgian ePrescription occurs in plain text in the KMEHR standard. A future extension of the KMEHR standard to integrate this in a structured way is required. Validation towards

this future standard is also required before removing all paper (proof of) prescriptions.

Conclusions

This study evaluates the quality of the Belgian ePrescription by looking at the validity of the construction of the KMEHR message and by looking at the validity of the digital signature of the prescriber. The construction of the KMEHR message showed a 100% validity towards the provided XSD Schema, but only 77.74% succeeded to validate against the Recip-e project validation rule set. When looking at the validity of the digital signature, a rather successful validation rate of 99.39% was observed.

The incorrect construction of the underlying KMEHR message might lead to an incorrect interpretation of certain fields represented in the software of the pharmacist, whereas incorrect signatures might lead to errors and an increased workload in the pharmacist's workflow. Timely review of these validations is required in order to dematerialize the ePrescription in a safe way.

Acknowledgements

We would like to thank Koninklijk Limburgs Apothekers Verbond (KLAV) for the provided electronic prescriptions of community pharmacists, using their tarification services.

References

[1] B. Astrand, E. Montelius, G. Petersson, and A. Ekedahl, Assessment of ePrescription quality: an observational study at three mail-order pharmacies, *BMC Med Inform Decis Mak* **9** (2008), 8.

[2] S. E. Ohlund, B. Astrand, and G. Petersson, Improving Interoperability in ePrescribing, *Interact J Med Res* **1(2)** (2012), e17.

[3] eHealth Initiative (eHI), and the Center for Improving Medication Management (CIMM), A clinician's guide to electronic prescribing (2008).

[4] M. Samadbeik, M. Ahmadi, F. Sadoughi, and A. Garavand, A Comparative Review of Electronic Prescription Systems: Lessons Learned from Developed Countries, *J Res Pharm Pract* **6(1)** (2017), 3-11.

[5] P. Esmaeil Zadeh and M. C. Tremblay, A review of the literature and proposed classification on e-prescribing: Functions, assimilation stages, benefits, concerns, and risks, *Res Social Adm Pharm* **12(1)** (2016), 1-19.

[6] J. Van Calster and M.-L. Bouffioux, *E-MED of het project rond het elektronisch voorschrijven van geneesmiddelen en de "authentieke bron" geneesmiddelen van het FAGG* (2009), Available: https://www.fagg-afmps.be/sites/default/files/downloads/newsletter_actua_01012009_nl.pdf [Accessed on March 18, 2019].

[7] International Organization for Standardization (ISO), ISO 13606 Health informatics - Electronic health record communication (2008).

[8] HL7, Health Level Seven - HL7 (1987).

[9] SNOMED International, Systematized Nomenclature of Medicine, Clinical Terms (SNOMED CT) (2009).

[10] eHealth Platform, *Kind Messages for Electronic Healthcare Record - KMEHR standard* (2008). Available: https://www.ehealth.fgov.be/ standards/ kmehr/en [Accessed on March 18, 2019].

[11] Algemene Pharmaceutische Bond. *De CNK code* (2015). Available: https://www.apb.be/nl/corp/De-Algemene-Pharmaceutische-Bond/Ontdek-onze-diensten/De-CNK-code/Pages/default.aspx [Accessed on March 18, 2019].

[12] Recip-e. *Ontwikkelaars binnen Recip-e (documentatie, SDK, ...)* (2018). Available: https://recip-e.be/profielen/ontwikkelaars/?lang=nl [Accessed on March 18, 2019]

[13] I. Sirius. *Aantal apotheken in België* (2017). Available: http://siriusinsight.be/nl/whitepaper-nl/spreiding-apotheken/ [Accessed on March 18, 2019].

[14] Oracle. *Java Architecture for XML Binding (The Java Tutorials) (1995)*. Available: https://docs.oracle.com/javase/tutorial/jaxb/ [Accessed on March 18, 2019].

[15] J. G. Anderson, Social, ethical and legal barriers to e-health, *Int J Med Inform* **76(5-6)** (2007), 480-483.

[16] Ö. Gider, S. Ocak, and M. Top, Evaluation of electronic prescription implications in Turkey: an investigation of the perceptions of physicians, *Worldviews Evid Based Nurs* **12(2)** (2015), 88-97.

[17] D. Reed-Kane, K. Vasquez, A. Pavlik, J. Peragine, and M. Sandberg, E-prescription errors and their resolution in a community compounding pharmacy, *Int J Pharm Compd* **18(2) (**2014), 159-161.

[18] M. P. Tully, Prescribing errors in hospital practice, *Br J Clin Pharmacol* **74(4)** (2012), 668-675.

[19] N. A. Caldwell and B. Power, The pros and cons of electronic prescribing for children, *Arch Dis Child* **97(2)** (2012), 124-128.

[20] M. P. Tully, D. M. Ashcroft, T. Dornan, P. J. Lewis, D. Taylor, and V. Wass, The causes of and factors associated with prescribing errors in hospital inpatients: a systematic review, *Drug Saf* **32(10)** (2009), 819-836.

[21] E. Ammenwerth, P. Schnell-Inderst, C. Machan, and U. Siebert, The effect of electronic prescribing on medication errors and adverse drug events: a systematic review, *J Am Med Inform Assoc* **15(5)** (2008), 585-600.

Address for Correspondence

Sven Van Laere, Department of Public Health (GEWE), Vrije Universiteit Brussel, Laarbeeklaan 103, 1090 Brussel, Belgium.
Email: Sven.Van.Laere@vub.be – Phone: (0032) 2 477 44 24

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190390

Development and Progression in Danish eHealth Policies: Towards Evidence-Based Policy Making

Sidsel Villumsen[a], Arild Faxvaag[b], Christian Nøhr[a]

[a] Maersk Mc-Kinney Møller Institute, University of Southern Denmark, Odense, Denmark,
[b] Faculty of medicine and health sciences, NTNU, Trondheim, Norway

Abstract

In order to realise the potential benefits of eHealth, governments develop eHealth policies to define and prioritise initiatives, the strategic goals and the resulting benefits. During the 23 years with eHealth policies in Denmark only a few status reports with a systematic and transparent evaluation have been made. This paper advocates a more systematic approach to strategic planning of development and implementation of eHealth systems, by encouraging the concept of evidence-based policy making through analysis of how focus of the Danish eHealth policies have evolved. The Danish eHealth policies have very different framings following the different focus points for the policies. Interestingly, strategies for evaluating the development of eHealth and eHealth policies were very sparcely noted in the policies. For the first time the de-emphasising of evaluations of eHealth policies in Denmark has been empirically demonstrated, thus undermining the objective of obtaining evidence-based eHealth policies.

Keywords:
Policy, Medical Informatics, Learning

Introduction

Health information technologies (HIT) are viewed as a means to easing the everyday life for patients, relatives and health professionals, modernizing and better utilising the resources of the healthcare system [1,2]. The promise of HIT has been profound. In Denmark in 1968, the vision was that within a couple of years, there would be no paper on the doctor's desk, and the patient record could be retrieved on a 'data screen' in a split second [3]. In order to realise the potential benefits, governments develop HIT or eHealth strategies to define and prioritise initiatives, the strategic goals and the resulting benefits [1]. Policies have been a tool for reaching consensus on the prioritisation of eHealth service development and large scale implementations in Denmark since 1995 [4], where the white paper "From Vision to Action: Info-society in Year 2000" described guidelines for national developments towards the information society and pointed out a number of specific priority areas – one of which was health care [5]. The first national eHealth policy, called "Action Plan for Electronic Patient Records" (The HEP-program), was published by the Ministry of Health in 1996 [6]. Since then, five additional national eHealth policies have been published (see Table 1).

During the 23 years with eHealth policies in Denmark only a few systematic and transparent evaluations have been made. In the absence of a firm base for the next policy, the subsequent policies have randomly involved data from the past and have mainly been designed as political documents without references. They seem to reflect the current balance of power between the municipal, regional and central levels.

It is now an established practice within healthcare to review clinical practice, in order to learn and improve [7]. eHealth is no different from other tools for improving healthcare systems and the development, implementation and use hereof needs to be based upon best available knowledge. In order to ensure learning and continuous improvement, evaluations identifying successes and failures and their causes are paramount [7]. Both policy and practice in eHealth should be based on scientifically obtained facts, as they are in healthcare in general [7], and evaluation is the only way to obtain knowledge on the effects of eHealth. Best practice for using evidence to develop a policy include assessing evidence of the likely effectiveness of policy options in order to inform decisions on future policy actions. Moreover, planning for collection of evidence from evaluations of implemented policies to inform decisions 'on whether to continue or how to adjust and improve policies and to contribute to the evidence base to inform future consideration of policy options' [8] is essential.

It has been the ambition of the authors to introduce a more systematic approach to strategic planning of development and implementation of eHealth systems, by encouraging the concept of evidence-based policy making. Figure 1 shows a model for strategic management [9] inspired by the continuous learning cycle of PDSA (Plan-Do-Study-Act) and applying this rationale to the process of strategic management of eHealth policy, development, and implementation. Strategic planning and development starts with formulating strategic goals. These strategic goals are the basis for reaching consensus among the involved interests and hence ensure commitment. As a result, a specific plan is drafted specifying operational and specific goals. These operational goals can point at particular

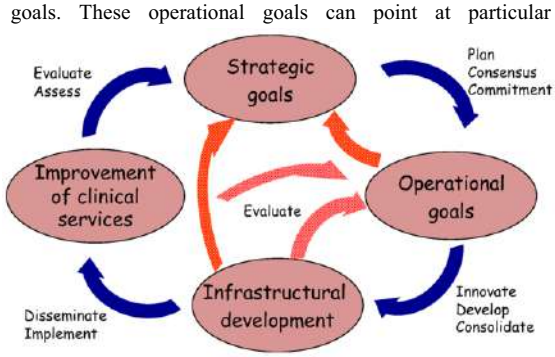


Figure 1 – Strategic management of eHealth policy, development

technology innovations or issues that need to be consolidated to achieve these goals. The third step is to develop infrastructural elements and prepare them to be implemented and disseminated in the healthcare system. The aggregate aim is to improve clinical services which must be assessed and evaluated to determine how well the strategic goals have been achieved.

In order to learn from the development of previous policies, the first step is to analyse how the policies have evolved and how the focus has changes over the years. Such an analysis was published in 2008 analysing the four national Danish policies for IT in the healthcare sector at the time [4]. Since then, ten years have passed, and two more policies have been launched. This paper aims at analysing how focus of the Danish eHealth policies have evolved from the first eHealth policy published in 1996 to the current policy published in 2018. The analysis shall reveal to what degree evaluation of each step in the model of strategic management in figure 1 has been carried out. It is the hope that other nations will adopt a similar approach to strategic management of eHealth development and contribute to a more rational and evidence-based policy making regarding eHealth.

Methods

By using the model of strategic management (figure 1) as a reference, the policy documents were analysed by means of a text analysis tools by two researchers following three steps. Using the text annotation tool NVIVO (NVIVO 12 for Mac), the first researcher (SV) annotated the texts. Sentences and sections that contained statements about the five factors: Strategic goals, Operational goals, Infrastructural development, Improvement of clinical services, and Evaluation were identified and coded. Examples of coded text are presented in the results. As the policies were annotated, the codebook was extended to cover two aspects of evaluation: a) references to previous evaluations or follow-up, and b) references to ongoing, planned or desired evaluation or follow-up. Thereafter, a second researcher (CN) validated the coding based on the codebook. Common consensus on whether the document section deserved a particular code was reached through discussion between the researchers. Coded statements were sorted within the five focus areas and the frequencies counted.

Results

Some difficulties were encountered when coding the policies within the five facets. Strategic goals were most often formulated as either broad/vague initiatives and their supposed benefits (i.e. '*The vision is that IT is easy to use for the staff, gives access to necessary information and facilitates recording and documentation of delivered professional health care*' [10]), or as an imperative need or problem that needs to be addressed or solved (i.e. '*It must not be necessary to give the same information each time you encounter a new instance of care* […]' [10]). A more clearly defined example is found in [10]: '*To some extent, physicians, patients and pharmacies all lack an overview of what medication is prescribed and actually taken by the individual patient* [Strategic goal – need based]. *One way to generate such an overview is to provide safe access to the personal electronic medicine profile through the public health portal* [Operational goal]. *The portal will host information concerning prescribed medicine, dosage, indication, delivery, etc.* [Infrastructural development]. *This will lead to better utilization of drugs with subsequent consequences for the entire health care system and the public finances* [Improved clinical services].'[10]. Ideally, improvements of clinical outcomes should be described as SMART-goals. However, only few of the benefits described in the policies render this granularity (i.e. '*Before the end of 2015, 80% of all applications, reporting, letters and written communication between the healthcare system and the citizen should be digital.*' [11]).

The first national eHealth policy HEP from 1996 was focussed on collecting and sharing data electronically. Connecting local systems to national registers had high priority – creating a health data network. Using national and international standards and terminology was a central point to be explored further [6]. The policy included a national survey of the counties' status on Electronic Patient Records (EPR) and local eHealth strategies. The action plan focused on local pathfinder projects aiming to exploit IT to gain better service and faster and more efficient treatment of patients. Interestingly, this policy had a high proportion of operational goals in relation to strategic goals as well as notations of specific IT/IS functionalities and focus on changes in the organisation, collaborations, workflow and documentation (infrastructural development), which might reflect the document's status as an action plan rather than a policy (figure 2).

The action plan was followed by the first 'National strategy for IT in the Hospital system 2000-2002' published by the Ministry of Health, the National Board of Health, the Counties, and the Capital Area's Hospital Corporation (H:S) [12]. This policy pointed out initiatives to support the national goals for the hospital sector. Great emphasis was put on developing and testing a 'Basic Structure for Electronic Health Records' (BEHR) with the aim of complete EPR coverage on all Danish hospitals within 2005.

Table 1 - Included policy papers

Published	Author	Title
1996	The Danish Ministry of Health	Action plan for Electronic Patient Records (EHR) – strategy report
1999	The Danish Ministry of Health	National strategy for IT in the Hospital system 2000-2002
2003	The Ministry of the Interior and Health	National IT Strategy 2003-2007 for the Danish Health Care Service
2008	Connected Digital Health in Denmark	National Strategy for Digitalisation of the Danish Healthcare Service 2008-2012 - to promote public health as well as prevention and treatment
2013	The Danish Government Local Government Denmark Danish Regions	Making eHealth Work - National Strategy for Digitalisation of the Danish Healthcare Sector 2013-2017
2018	The Ministry of Health The Ministry of Finance Danish Regions Local Government Denmark	A Coherent and Trustworthy Health Network for All - Digital Health Strategy 2018–2022

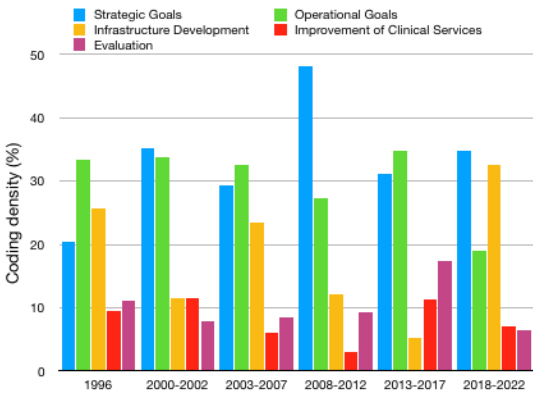


Figure 2. Density of annotations of strategic management factors in the six Danish eHealth policies

The 2000-2002 policy had a strong focus on standards and integrations, but also on the organisational aspects of implementing such as organisational changes and training of staff – formulated as strategic goals and operational goals mainly. The need for research within health informatics was stated explicitly in this policy.

The succeeding policy 'National IT Strategy 2003-2007 for the Danish Health Care Service' broadened the scope to cover the IT initiatives necessary for not only the hospitals but the entire health sector to support the realisation of the national health policy goals [10]. The 2003-2007 policy further supports the development of BEHR as a generic information model that sets the national standard for EHRs. The central vision was '*that citizens, health care professionals, authorities and administrators have access to updated information through channels perceived to be free of any undue obstructions'* [10]. The 2003-2007 policy refers to the independent EPR-Observatory for objective follow-up on the national status on eHealth. In 2005 a status report from the EPR-Observatory concluded that complete EPR coverage would not be obtained before 2013 at best [13], and in 2007 two reports (one by the Public Accounts Committee) raised critique of BEHR and the general level of IT support in the hospital sector [14,15], stating that national use of BEHR would not be feasible within the desired time span [14]. That might be one of the reasons why the 'National Strategy for Digitalisation of the Danish Healthcare Service 2008-2012 - to promote public health as well as prevention and treatment' did not comment further on the BEHR [16].

The 2008-2012 policy was very different from the previous ones, setting a new course for eHealth in Denmark. Focus turned towards consolidating the IT systems to ensure that diverse solutions could act together and exchange or share data. This policy described the ways of working towards joint digital healthcare services rather than describing the specific initiatives [16], which is reflected in the proportion of strategic goals compared to operational goals (figure 2). The 2008-2012 policy presented the national goals for the healthcare system in general (agreed on in other policies) and specified the eHealth contribution of the eHealth policy in attaining these visions. In 2007 the 14 Danish Counties were merged into five Regions. Each Region is responsible for the secondary healthcare services (i.e. hospitals). This implied that the different IT-solutions in the counties now needed to be consolidated within the Regions. The Regions were charged with setting specified goals for use and value of eHealth and to work towards attaining these goals [16]. Consolidating initiatives became a focus point not only for the Regions, but also for the municipalities and General Practitioners. Interestingly, the 2008-20012 policy includes an appendix presenting the conclusions of an external review of the EPR status in Denmark.

'Making eHealth Work - National Strategy for Digitalisation of the Danish Healthcare Sector 2013-2017' focused on exploiting the digital possibilities to the fullest and create better coherence in the digitisation effort [11]. The strategic goals where to ensure that patients and staff would profit from the benefits of ongoing IT projects through an increased focus on full dissemination and use, reflected in the higher proportion of operational goals (figure 2). This policy had a stronger focus on benefits realisation than the previous policies (figure 2 - Improved clinical services), and indicators, follow-up, and evaluation was distinctly more prominent in this policy than in any of the others. This may be due to the strong focus on consolidating and "making eHealth work" through learning and taking actions to improve the dissemination and use of health IT.

The present policy, 'A Coherent and Trustworthy Health Network for All - Digital Health Strategy 2018–2022', reflects two main aims: putting the citizens' needs at the centre and easing the everyday work for healthcare staff [2]. The 2018-2022 policy is framed differently than the others. Focus lies on five specific areas, identifying 27 initiatives. The initiatives are described with respect to the technological and implementational deliveries needed, which is reflected in the high proportion of Infrastructure Developments mentioned (figure 2). The 2018-2022 policy concludes with a section on follow-up, where it is described how the National Board of eHealth will continuously follow the progression of the initiatives and adjust. However, evaluation is not mentioned specifically and does not play a central role in the current eHealth policy.

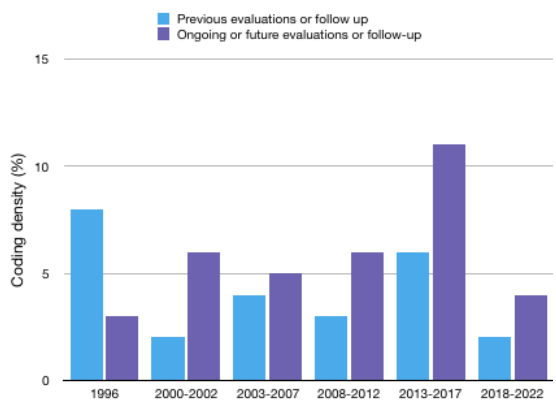


Figure 3. Density of annotations of references to evaluation in the six Danish eHealth policies

A prerequisite for strategic evidence-based management of eHealth development is evaluation [7]. In order to ensure learning and continuous improvements, systematic and transparent evaluations must be performed throughout the stages of the management process described in figure 1. When looking only on statements specifically mentioning evaluation, only a few counts can be made. In this analysis, evaluations have therefore been recognized in broad terms, spanding hearings on the policies, descriptions on follow-up activities, referencing to statistics on EPR dissemination status etc. It is interesting to note that the first Danish national eHealth policy (1996) had a high proportion of references to previous evaluations and follow-up, and planned on presenting yearly

statusreports and adjusted action plans for the Danish Parliament [6] (figure 3). The 2000-2002 policy had only few references to past evaluations but highlighted the need for creating a National Strategy Group for IT in the hospital sector and a National Reference Group for health informatics, specifying the roles and responsibilities for these groups [12]. The following two strategies (2003-2007 and 2008-2012) aimed the evaluation towards the initiatives described in the policies, thus delegating the evaluation to steering groups and centers of excellence. In the two latest strategies (2013-2017 and 2018-2022) the responsibility of follow-up and evaluation has been chartered to the National Board of eHealth primarily. In the 2013-2017 policy, follow-up had a more prominent role than in any of the other policies, with a strong focus on designing and monitoring indicators for eHealth availability and use in order to inform future planning of development and implementation. This strong focus on follow-up and evaluation was not continued in the 2018-2022, where notions on following the status and progress on the initiatives are found in the last chapter .

Is has not been possible to map the statements on evaluation across the policies to the steps in the model of strategic management (figure 1) due to the vast differences in how evaluation and follow-up were described and defined. When looking at ongoing and future planning of follow-up and evaluation, the policies point to boards and groups responsible for this task. The responsibility has changes over the years, but since 2013, the National Board of eHealth has been the coordinating organ (table 2).

Table 2 - Groups and boards responsible for evaluation of eHealth policies

Policy	Responsible groups and boards
1996	The Danish Ministry of Health
2000-2002	National strategy group for IT in the hospital sector (The Danish Ministry of Health and hospital owners). National reference group for health informatics
2003-2007	EPR Observatory National strategy group for IT in the health care service (National Board of eHealth, the hospital owners and the National Association of Local Authorities). Steering groups for initiatives of the IT strategy.
2008-2012	Connected Digital Health in Denmark Centres of expertise
2013-2017	National Board of eHealth Danish Regions
2018-2022	National Board of eHealth Existing boards (e.g. MedCom) New boards

Discussion

In this study, we found that Danish eHealth policies have significantly shifted focus and it has been impossible to deduct any rational reasoning behind the shifting focus. An evidence based approach would have based the visions of a new policy on references to former policies or at least some reported knowledge about former experience or documented needs. The scoring of the content of the policies also show a very low occurrence of evaluation activities – in fact our analysis revealed that evaluation is the least mentioned theme. The Danish policy makers have obviously not laid importance in establishing learning loops which would require evaluation – a situation which is not unique to Denmark, as it is seen in other countries as well [7]. However, to establish evaluation introduces a risk to the policy makers. The purchase and implementation of eHealth technologies e.g. an EPR system, is a massive expense and time commitment. If there are major problems - and there usually are - no one in charge wants to have that publicized. Hence: evaluation might not be desired.

As the analysis shows, there is no clear tendency in the policies over the years. Each policy appears as a unique new policy based on prevailing trends. The strategic goals are formulated as relatively abstract visions with only limited connections to specific changes in clinical services or the workflows that produce them. '*If you don't know where you are going, any road will take you there'* (Alice in wonderland). For the policies to point at a specific road to follow, the destination has to be formulated in continuation of the visions. Nonetheless, the analysis shows that relatively little attention has been payed to explaining how specific clinical services should be changed. In the implementation of such strategies the primary actors need to fully agree on why they find it important to implement [17].

An approach to support the Strategy Management Model in developing and implementing eHealth policies is framing a benefit dependency network (BDN) as described by Ward and Daniel [18]. A BDN visualises the dependencies between the strategic goals (*Why*), the operational goals and intended clinical improvements (*What*) and which changes are warranted in the organisation and workflows in order to reap the benefits of a eHealth system or functionality (*How*) [17]. The operational goals define the roads to take, and hence should translate to the infrastructure development – what is necessary to innovate and design in order to realize the organisational and technological change. And finally, how should this be implemented and disseminated. The policies display a significant variability in how each of the steps in the strategy management model are prioritized. This constitutes an impediment, since progress and improvements require continuity and persistence. Because evaluation activities are prioritized so low it becomes very challenging to identify inadequacies and insufficiencies, and realigning a failed course or refining the approach in order to obtain the goals and visions aimed for.

The strongest focus on monitoring and evaluating eHealth was found in the 2013-2017 policy, where benefit realization and national indicators were central themes. In this period a sub-page on the National Health Data Authority official website was dedicated to indicators of eHealth (figure 4). However, the sub-page was empty – updated last time in 2016 and now it does not exist anymore. The unresolved situation around monitoring indicators was silenced, and indicators were not mentioned in the 2018-2022 policy. However, indicator work has been continued and in March 2019 the first follow-up on the current policy was published [19]. The status report contains data on three of the five areas mentioned in the policy, but does not inform why there are no indicators on the remaining two focus areas.

Evaluation of eHealth can be difficult to quantify and is both nuanced and profoundly complex. Policymakers must understand the powers, problems, and implications of eHealth services in order to evaluate the effects. That is a daunting challenge, but no viable alternatives exist [7]. A method for monitoring the progression and effect of eHealth policies are by viewing the strategic goals as constructs. To every construct a series of indicators reflecting the construct can be developed and monitored.

Figure 4 - National Health Data Authority official website dedicated to indicators of eHealth

If a clear strategic management approach and BDNs have been framed, indentifying indicators reflecting the intention of the strategic goals will be feasible.

Allthough awareness on the value of evidence-based policy making is increasing, evaluation studies on national eHealth policies are not carried out frequently or consistently in the Nordic countries. Since 2012 the Nordic eHealth Research Network, supported by the eHealth group of the Nordic Council of Ministers, has been developing, testing and assessing a common set of indicators for monitoring eHealth in the Nordic Countries. The overall goal is to support national and international policy makers and scientific communities to develop Nordic welfare.

The Strategy Management Model and the insights gained from the analysis of the Danish ehealth policies may serve as an inspiration and example for managing eHealth policies, thus aiming at creating a learning healthcare system [20] and evidence-based eHealth policies that will increase the value of eHealth.

Conclusions

The Danish eHealth policies have very different framings following the different focus points for the policies. Interestingly, strategies for evaluating the development of eHealth and eHealth policies were very sparcely noted in the policies. For the first time the de-emphasising of evaluations of eHealth policies in Denmark has been empirically demonstrated, thus undermining the objective of obtaining evidence-based eHealth policies. In order to realise evidence-based policy making it is neccesary for the eHealth policy makers to focus on continous learning and evaluation of the previous and current eHealth policies. This work may be leveraged by using the Strategy Management Model aimning at supporting learning healthcare systems and optimizing the value of eHealth.

Acknowledgements

The Strategic Management Model presented in figure 1 was developed in cooperation with Søren Vingtoft, MD. Concepts presented in this paper originates from fruitful discussions in the Nordic eHealth Research Network.

References

[1] J. Adler-Milstein, E. Ronchi, G.R. Cohen, L.A.P. Winn, and A.K. Jha, Benchmarking health IT among OECD countries: Better data for better policy, *J. Am. Med. Informatics Assoc.* **21** (2014) 111–116. doi:10.1136/amiajnl-2013-001710.

[2] The Ministry of Health, The Ministry of Finance, Danish Regions, and Local Government Denmark, A Coherent and Trustworthy Health Network for All - Digital Health Strategy 2018–2022, 2018.

[3] Ingeniørens ugeblad, Avanceret databehandlingsanlæg med "fjernsynsskærme" til Rigshospitalet, *Ingeniørens Ugebl.* (1968).

[4] M. Bruun-Rasmussen, K. Bernstein, and S. Vingtoft, Ten years experience with National IT strategies for the Danish Health Care service, *HIC 2008 Aust. Heal. Informatics Conf.* (2008) 1–5.

[5] Fra vision til handling: info-samfundet år 2000. Redegørelse til Folketinget om "Info-samfundet år 2000" og IT-politisk handlingsplan 1995, 1995.

[6] The Danish Ministry of Health, Action plan for Electronic Patient Records (EPR) – strategy report, Sundshedsministeriet, 1996.

[7] M. Rigby, Evaluation: 16 powerful reasons why not to do it - And 6 over-riding imperatives, *Stud. Health Technol. Inform.* **84** (2001) 1198–1202. doi:10.3233/978-1-60750-928-8-1198.

[8] I. Sanderson, Evaluation, Policy Learning and Evidence-Based Policy Making, *Public Adm.* **80** (2002) 1–22. doi:10.1111/1467-9299.00292.

[9] C. Nøhr, S. Koch, V. Vimarlund, H. Gilstad, A. Faxvaag, G.A. Hardardottir, H.K. Andreassen, M. Kangas, J. Reponen, P. Bertelsen, S. Villumsen, and H. Hyppönen, Monitoring and benchmarking ehealth in the nordic countries, 2018. doi:10.3233/978-1-61499-852-5-86.

[10] The Ministry of the Interior and Health, and Sundhedsstyrelsen, National IT Strategy 2003-2007 for the Danish Health Care Service, (2003) 60.

[11] The Danish Government, Local Government Denmark, and Danish Regions, Making eHealth Work - National Strategy for Digitalisation of the Danish Healthcare Sector 2013-2017, 2013.

[12] The Danish Ministry of Health, National strategy for IT in the Hospital system 2000-2002, 1999.

[13] S. Vingtoft, M. Bruun-Rasmussen, K. Bernstein, S.K. Andersen, and C. Nøhr, EPJ-Observatoriet - Statusrapport 2005, 2005.

[14] Deloitte, Strategiske udviklingsveje for EPJ - Eksternt review af det hidtidige EPJ-arbejde, 2007.

[15] Statsrevisorerne, Beretning om it-understøttelsen af sygehusenes opgaver, 2007.

[16] Connected Digital Health in Denmark, National Strategy for Digitalisation of the Danish Healthcare Service 2008-2012 - to promote public health as well as prevention and treatment, 2008.

[17] S. Villumsen, C. Nøhr, and A. Faxvaag, Translating e health visions from strategy to practice - a benefit management approach, 2018. doi:10.3233/978-1-61499-852-5-885.

[18] J. Ward, and E. Daniel, Benefits management: how to increase the business value of your IT projects, Chichester, West Sussex: Wiley, 2012.

[19] Sundhedsdatastyrelsen, Indikatorer for udbredelse og anvendelse af sundheds-it, 2019.

[20] C.P. Friedman, A.K. Wong, and D. Blumenthal, Achieving a nationwide learning health system., *Sci Transl Med.* **2** (2010). doi:10.1126.

Address for correspondence

Sidsel Villumsen, email: sivi@mmmi.sdu.dk

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190391

Establishment of a Comprehensive Information Infrastructure and a Support Organization for Rare Disease Research in Japan (RADDAR-J)

Izumi Yamaguchi[a], Yoshihiko Furusawa[b], Takahisa Kawaguchi[a], Naoko Yagishita[c], Kazumasa Tanzawa[d], Yoshihisa Yamano [c], Fumihiko Matsuda[a]

[a] Center for Genomic Medicine, Graduate School of Medicine, Kyoto University,
[b] Department of Neurology, National Center Hospital, National Center of Neurology and Psychiatry,
[c] Department of Rare Diseases Research, Institute of Medical Science, St. Marianna University School of Medicine,
[d] EPS Corporation

Abstract

There are more than 300 research groups for rare diseases in Japan. Although various clinical and genomic information of patients are being collected by the groups, the information is managed individually by each research group and the current practices for managing and sharing research data are not very efficient. Since "rare diseases" are literally rare, the understanding of the underlying disease mechanisms are incomplete and collecting a sufficient number of patients for clinical trials is difficult. Therefore, there is a need to collect and integrate the data and construct a data integration platform for rare disease research. Funded by the Japan Agency for Medical Research and Development, a national research and development project to establish a standard platform and supporting organization for rare disease registries in Japan is currently under way. In this article, we report the background, purpose, process, results, current status, and future plans of this project.

Keywords:

Rare diseases, registries, data sharing.

Introduction

In Japan, research studies on rare diseases have been conducted for over four decades, and the Ministry of Health, Labour and Welfare of Japan (hereafter, referred to as "MHLW") has been supporting these studies.

The MHLW or the Japan Agency for Medical Research and Development (hereafter called "AMED" and was established in 2015 as a funding agency for medical research in Japan) supports about 300 research groups for rare diseases in Japan. In addition, there are research groups of various sizes that are not funded by MHLW or AMED.

All of these research groups are collecting various kinds of clinical and genomic information from patients for their own purposes. However, in most cases such data are managed individually by each research group, and long-term stable management of the data or effective data sharing among research groups are not sufficiently achieved because of the burden on each research group.

Since the number of the rare disease patients is low, collecting a sufficient number of patients for clinical trials is frequently difficult. As a result, the quantity and quality of the information about rare disease patients which a single research group can collect is limited, and the understanding of the underlying disease mechanisms for rare diseases is largely unknown.

Therefore, there is a strong need for collecting and integrating the data from rare disease patients and constructing a data integration platform for rare disease research in Japan. Such a platform will contribute greatly to the advance and expansion of rare disease research.

With this in mind, a research project was launched by AMED in 2017 to establish a national research infrastructure for rare disease registries in Japan, and our research group was placed in charge of the project [1].

Methods

The purpose and the goal of this research project is to establish a national information infrastructure and management organization to enable: stable longitudinal management; effective sharing among research groups; and integration, analysis, and secondary use of the clinical and genomic information from the research participants.

The information system and the supporting organization was named "RADDAR-J" (RAre Disease DAta Registry of Japan) and the following tasks were carried out:

1. Administering a web-based survey of the rare disease research groups supported by AMED or MHLW
2. Defining the rules and regulations
3. Designing and developing the information infrastructure
4. Establishing a support system for the management of registries

Web-based survey

Modeled on a prior survey in Europe [2], a web-based survey targeting rare disease research groups supported by AMED or MHLW was conducted in order to investigate the current development and management of rare disease registries in Japan, the needs for support, and so on. The survey items were as follows:

- Characteristics of the research project (5 questions)
- Characteristics of the disease registry (25 questions)
- Information system of the registry (24 questions)
- Management and registration scheme (35 questions)
- Collected data (48 questions)
- Management of the data (18 questions)
- Biorepository (16 questions)
- Needs and expectations (2 questions)

Defining the rules and regulations

Rules and regulations necessary for collecting, integrating, sharing, and providing the clinical and genomic information from the research groups included:

- Designing the organization
- Developing the operating policy
- Developing the data-sharing policy

Designing and developing the information infrastructure

The overall design of the information infrastructure necessary for RADDAR-J was created, and the development of the system was started.

Establishing a support system for registry management

A support system for registry management was organized by:

- Establishing a secretariat of RADDAR-J
- Setting up and operating the portal site in both Japanase and English
- Preparing the standard documents for management and operation of the registry
- Holding consultation meetings for research groups about the disease registry

Results

Web-based survey

A total of 303 rare disease research groups supported by AMED or MHLW were surveyed, and 89% (271 groups) answered our web-based questionnaire, which consisted of 173 questions mentioned above. The initial analysis of the answers showed that :

- 74% (200 groups) of the responding groups had some kind of registry.
- 29% (78 groups) were operating some data entry system for registry.
- 18% (49 groups) had a problem in collecting research participants.
- 12% (32 gropus) had a problem in utilizing the collected data effectively.
- 37% (100 groups) were short-staffed (especially with regard to specialist staff).

A part of the survey data was released to the public on our portal site, and all the data obtained from the survey are now under analysis in detail.

Defining the rules and regulations

Designing the organization of RADDAR-J

Design of the organization of RADDAR-J was modeled on an international organization in Europe named "RD-Connect" (https://rd-connect.eu/), which aims to accelerate rare disease research by providing a global platform for database, registries, and biobanks [3]. Based on the result of the web-based survey, the roles of RADDAR-J were set as follows:

- Collecting, managing, and providing the information and data shared from each disease research registry
- Integrating and analyzing the data provided from the disease registries
- Supporting the management of each disease registry research
- Returning the analytic results of the integrated data by RADDAR-J to the original research groups

Considering the diversity of the data types and the importance of the patients' personal information management, it was decided that RADDAR-J should consist of three main departments (Personal Information Management department, Clinical Information Integration department, and Genomic Information Integration department) and a RADDAR-J secretariat, which controls the whole organization. A committee was formed for each department.

Developing the operating policy of RADDAR-J

Detailed operating policies of RADDAR-J, such as the organizational structure, the role of each department, the operation of each committee, the necessary specifications of the standard registry system for reseach groups, the rules for the registry system management and the data management, were developed and documented.

1. Personal Information Management department is mainly in charge of the management of patients' personal information including real names. The department performs a “name-based aggregation” of patient information within one research project group as well as between different research project groups, so that patients are not duplicated within projects and can be tracked between projects.

   Since there is no national identification number available for medical research in Japan, personal information including real names need to be used to identify each patient correctly in order to avoid double registration of a patient in one research project or to find the existence of a patient that is registered in multiple disease research projects. After the “Name-based aggregation” process, patients whose real names, dates of birth, and gender match each other will be extracted as “possibly identical patients”.
2. Clinical Information Integration department is mainly in charge of :
   - supporting research groups in the development and the management of registry.
   - collecting, integrating, and managing clinical data provided by each research group.
   - aggregating and analyzing the integrated data.
3. Genomic Information Integration department is mainly in charge of :
   - collecting and managing genomic information and omics information provided by each research group.
   - collecting some clinical information from Clinical Information Integration department necessary for genomic analysis or omics analysis, and performing an integrated analysis of these data and genomic and/or omics data.
   - supporting research groups in genomic analysis such as providing some disease analysis tools, and providing genomic and/or omics information of control groups available for disease analysis.

Developing the data-sharing policy of RADDAR-J

A data-sharing policy of RADDAR-J was developed based on “Data Sharing Policy for Facilitation of Genomic Medicine” of AMED [4]. Three forms of data-sharing are defined in the policy as follows:

1. Group sharing data

   Data will be shared only between RADDAR-J and the research group, the original source of the data.
2. Controlled-Access Data

Data will be provided to third parties based on review and approval of RADDAR-J if the provision is approved in advance by the original source of the data.

3. Open Data

   Data will be published by RADDAR-J publicly if the publication is approved in advance by the original source of the data.

In order to operate these policies, researchers should obtain written informed consent from patients or use an opt-out consent process only when the research is an academic study and opt-in procedures are impractical, as described in the national ethical guidelines.

The basic data sharing policy of RADDAR-J was decided as group sharing between each research group and RADDAR-J in order to make it easier for each reseach group to obtain agreement with patients on data sharing. Each reseach group develops their own data sharing policy with RADDAR-J and its external seconady users.

Designing and developing the information infrastructure

Based on the organizational structure and the role of each department, a system configuration of RADDAR-J was designed, in which each department holds its own dedicated system and all the systems cooperate with each other as shown in Figure 1.

- Personal Information Management System, used in the Personal Information Management department, stores and manages the patients' basic personal information including real names in encrypted form. In addition, it is able to compare the personal information without decryption for privacy protection, so that it can safely perform "name-based aggregation" of patients as necessary [5].
- In order to standardize the clinical information of each rare disease registry as much as possible, a standard registry system was prepared for each research group to use at a reasonable price, reducing the burden of the system development and management. The system is based on an electronic data capture system that one of our research collaborators had developed as a commercial product. It stores and manages various kinds of patient clinical information in de-identified form and some personal information in encrypted form, which are necessary for the operation of each registry secretariat. International standard codes such as ICD-10 and HPO (Human Phenotype Ontology) were implemented on the standard registry system to facilitate international sharing of information.
- Regarding Clinical Information Integration System that will be used in the Clinical Information Integration department, the specification including the system linkage is currently being considered, and the development will start shortly. The system collects the fixed standard clinical information from each registry system in de-identified and standardized form, and receives the result of "name-based aggregation" from the Personal Information Management System as necessary.
- Genomic Information Integration System, used in the Genomic Information Integration department, stores and manages genomic information, omics information, and some clinical information necessary for genomic analysis in de-identified form. In addition, some simple genetic analysis functions are planned to be implemented in the system. This system is currently under development. The basic functions such as storing or managing genomic information have already been developed, and additional functions such as searching for genetic mutations are under development.

Support system for the management of registries

The official website of RADDAR-J was launched in both Japanese and English (https://raddarj.org/en/). An explanation of the RADDAR-J project and the result of the web-based survey about rare disease registries in Japan are available on the site.

The RADDAR-J secretariat held an explanatory meeting about the RADDAR-J project with rare disease research groups funded by AMED or MHLW, and is conducting individual consultations for research groups that request support for developing their disease registries. For an individual consultation, members of the secretariat have an interview with the reseach group to determine their current situation and demand, and consider how to support them.

Various kinds of standard documents for developing and managing rare disease registries are being developed as shown in Table 1. They will be provided at no charge to reseach groups according to their needs.

Discussion

Background of this research project

According to AMED policy, data obtained from research and development supported by public funds is national intellectual property, and AMED places great importance to the appropriate collection, quality assurance, presevation, utilization and application of data. To that end, AMED has established "Data Sharing Policy for the Realization of Genomic Medicine" [4] and promotes data sharing. Based on the policy, AMED launched a research project for the construction of information infrastructure for rare/intractable diseases, and awarded funding to the current authors' researh group.

RADDAR-J aims to contribute to the acceleration of rare disease research by providing the infrastructure and the organization which enables effective sharing and integration of clinical and genomic information that rare disease research groups collect.

Need for support

Although the data from the web-based survey, which was conducted in advance of designing the organization and the information system of RADDAR-J, are currently under detailed analysis, the survey showed that about three quarters of the responding research groups hold some kind of registry, but that less than half of them operate some data entry system for registry. In addition, slightly more than one-third of the responding research groups are understaffed (especially lacking in specialist staff).

From these results, we concluded that support by RADDAR-J for the information system and the operation of registry would help to advance the progress of rare disease research.

What we have done so far

After the whole picture for RADDAR-J including the necessary systems and regimes were designed, the development of the information system and the supporting organization for the registry systems was prioritized. The rare disease research groups could thus smoothly launch and start using their own registries because this part is, as it were, a gate for the data into RADDAR-J.

In order to secure data provenance, a standard registry system with operational log function and user authentification function, as well as standard documents for organizational operation and procedures are offered by RADDAR-J. Using these, the research groups can manage their registries and the

information appropriately and undergo monitoring when necessary.

Next, an explanatory meeting about the RADDAR-J project was held with rare disease research groups funded by AMED or MHLW, and individual consultations started to be conducted for research groups that request support for developing their disease registries. Being able to exchange frank opinions which can improve future support, these consultations were found to be good opportunities to get to know deeply the actual circumstances of the research groups.

What we are doing

Currently, the information system and the organization for integration, analysis, and secondary use of the patient data collected by the registry system is being developed. Since multiple systems of different origins and by different software manufacturers need to link to each other mainly due to budget, there are many problems to be solved such as the specification of system linkage, the specification of data integration (e.g., patient ID, project ID, data format), the system operations, and so on. Therefore, the detailed specification development has been challenging and time consuming.

There is no national identification number available for medical practice in Japan, and so harmonizing information between registries and clinical record systems is a big challenge for this project. A new system of national identification number for medical practice is recently being designed, which may enable such harmonization in Japan as in the Nordic countries [6].

It can be said that this research project is still in the midst of development and there are numerous issues that remain to be resolved. However, when complete, this infrastructure and organization will greatly fascilitate effective data sharing and promote the progress of rare disease research. International cooperation with some similar organizations in Europe is currently under cosideration.

Conclusions

Data sharing between research groups will play an important role in the field of rare disease research due to the rarity of patients. In this research project (RADDAR-J), an information infrastructure and a supporting organization are being developed for building rare disease registries and sharing patient data, which are expected to greatly contribute to the progression of rare disease research.

Table 1– Examples of the standard documents to be provided

Name of Documents
Sample of Protocol
Sample of Informed Consent Form
Sample of Informed Assent Form
Guide to modification of an already existing protocol
Guide to modification of an already existing informed consent form
Sample of information disclosure document for opt-out
Guideline for data sharing
Standard Operating Procedure for research
Operating Procedure for data management
Operating Procedure for configuration and management of EDC (Electric Data Capture)
Operating Procedure for monitoring and audit
Manual for data collecting
Manual for data entry to EDC
Database structure definition
Case Report Form
Statistical analysis plan

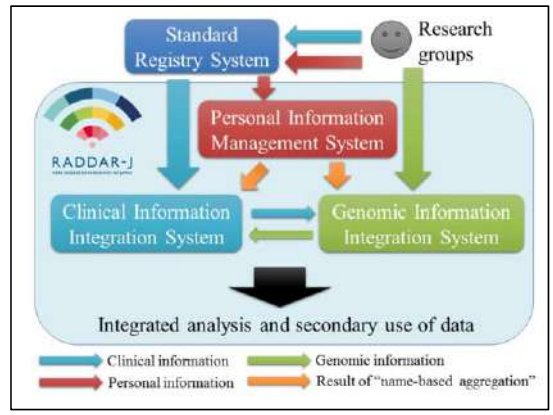


Figure 1– System configuration and data flow of RADDAR-J

Acknowledgements

RADDAR-J is a three-year research project which started in February 2017 and is supported by AMED under Grant Number JP17ek0109196 and JP18ek0109348.

References

[1] Furusawa Y, Yamaguchi I, Yagishita N, et al. National platform for Rare Diseases Data Registry of Japan. *Learn Health Systems*. 2019;e10080.

[2] Taruscio D. et al. The Current Situation and Needs of Rare Disease Registries in Europe. *Public Health Genomics* 16 (2013), 288–298.

[3] Thompson R. et al. RD-Connect: An Integrated Platform for Rare Disease Research. *Journal of General Internal Medicine*, 2014 Aug;29 Supple 3:S780-7.

[4] Japan Genomic Medicine Project Data sharing Policy for the Realization of Genomic Medicine. https://www.amed.go.jp/content/000017356.pdf

[5] Yamaguchi I. et al. Development of a real name registration and verification system for epidemiologic studies in order to avoid overlapping of patient registration. *Japan Journal of Medical Informatics*, 2017; 37(Suppl.): 939-942.

[6] Andreassen OA. eHealth provides a novel opportunity to exploit the advantages of the Nordic countries in psychiatric genetic research, building on the public health care system, biobanks, and registries. *American Journal of Medical Genetics Part B*. 2018;177B:625–629

Ethics approval

The positive ethics vote (No. R1399-2) from the ethics committee of Kyoto University Graduate School and Faculty of Medicine, Kyoto, Japan was obtained (chairman: Atsushi Asai).

Address for correspondence
Izumi Yamaguchi, M.D.
Center for Genomic Medicine,
Graduate School of Medicine, Kyoto University
Nanbusogo-Kenkyu-To-1, 5F
53 Shogoinkawahara-cho, Sakyo-ku, Kyoto 606-8507, JAPAN
Tel: (+81)-75-751-4157
FAX: (+81)-75-751-4167
E-mail: yamaguchi@genome.med.kyoto-u.ac.jp

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190392

Identifying RNA Biomarkers for Oesophageal Squamous Cell Carcinoma

Si Zheng[a], Jiagen Li[b], Li Hou[a], Jiao Li[a]

[a] *Institute of Medical Information & Library, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China,*
[b] *Cancer Institute and Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China*

Abstract

It is challenging to predict patient's prognosis in oesophageal squamous cell carcinoma (abbreviated to OSCC) or perform early intervention and prevention for this disease. With the development of clinical genomics, genetic heterogeneity in tumors was found to be relevant to OSCC tumorigenesis in precision medicine studies. Over the past decade, global microarray expression profiling has been used to investigate biomarkers for human diseases. In this study, we proposed two computational strategies for identifying RNA biomarkers from microarray expression profiles of OSCC. Firstly, the logistic regression model with a least absolute shrinkage and selection operator (LASSO) regularization was used to analyze RNA expression profiles from OSCC tissues. Secondly, differential expression analysis for lncRNA profiles from both tumor and the corresponding para-cancerous tissues was performed to identify biomarkers related with lymph node metastasis, which might contribute to the early diagnosis and treatment of the disease.

Keywords:

Esophageal Squamous Cell Carcinoma, RNA, Biomarkers

Introduction

Oesophageal cancer is one of the most malignant forms of tumor worldwide regarding its prognosis and mortality, with its incidence expected to increase in the following decade [1]. According to the statistics, esophagus cancer was the eighth most common cancer worldwide with an estimated 456,000 new cases (3.2% of the total), as well as an estimated 400,000 deaths (4.9% of the total) in 2012 [2]. Nevertheless, despite the advances in diagnosis and treatment strategies, therapeutic control of this disease remains limited, with the 5-year survival rate ranging from 15% to 25% [3]. There are two major subtypes of oesophageal cancer, and one of the most common histological subtypes in developing countries, especially in the Asian countries, is the oesophageal squamous cell carcinoma (abbreviated to OSCC) [4; 5].

Generally, OSCC is difficult to diagnose in the early stage, as it has no significant symptoms or signs. However, in patients with early oesophageal cancer who have undergone surgical or minimally invasive endoscopic treatment, the 5-year survival rate can reach > 90%. Lymph node metastasis is a pathognomonic feature of spreading tumors and is widely believed to be one of the most powerful prognostic factors influencing the outcomes of many solid tumors, including oesophageal cancers [6]. Clarifying the molecular mechanisms underlying the development and progression of OSCC and identifying biological markers for screening high-risk populations are therefore of great clinical importance for early diagnosis and treatment.

Studies based on genomic information have confirmed that mutations and aberrant expressions can contribute to the pathogenesis of OSCC [7-9]. For example, Togashi Y *et al*. identified that frequent amplification of ORAOV1 gene in oesophageal cancer can promote the tumorigenesis [10]. High expression of XRCC1 and JWA were found to be correlated with longer median overall survival, and these genes are useful biomarkers in oesophageal cancer [9]. Meanwhile, an increasing number of studies have confirmed that non-coding RNAs also play an important role in tumor biological process, and might be potential biomarkers for tumorigenesis [11-14]. For instance, Hou *et al*. showed non-coding RNAs could be new biomarkers and therapeutic targets for esophageal cancer [15]. Chen and Li *et al*. analyzed the microRNA (abbreviated to miRNA) and long non-coding RNA (abbreviated to lncRNA) expression profiles of 119 paired OSCC samples separately by microarray technology and identified four-miRNA signatures and three-lncRNA signatures that can predict patient survival [16; 17]. Yu *et al.* confirmed that miR-130b plays an oncogenic role in OSCC cells [18]. Actually, lncRNAs might involve in governing critical biological processes in cancer by regulating target genes expression at multiple stages of cancer development. Although various regulatory mechanisms of non-coding RNAs, mRNAs, as well as their interaction mechanisms for disease pathogenesis, have been uncovered. Still, systematic and integrative strategies for analyzing RNA expression profiles are still worth exploring.

On the other hand, previous studies indicated that genome-wide alterations in tumor compared with adjacent tissues may be predictive biomarkers for cancer [19-21]. Thus, identification of the molecular changes associated with cancer risk factors like lymph node metastasis in para-cancerous tissues (adjacent normal tissues) may aid in determining the earliest events of carcinogenesis and informing cancer prevention and treatment strategies.

In this research, we proposed two strategies for identifying RNA biomarkers from microarray expression profiles of OSCC. Followed with a literature search for validating the significance of the proposed biomarkers.

Methods

Strategy_1: Combination Analysis of miRNA, lncRNA and mRNA with Logistic Regression

RNA expression profiles (mRNA, miRNA and lncRNA) for cancer and adjacent normal tissues (para-cancerous tissues) of 119 OSCC patients were obtained from the study by Chen and Li *et al* [16; 17]. Followed with data filtering and normalization to reduce the number of variables. Here we used logistic regression to check whether the combination of different kinds of RNA molecules might be effective prognostic biomarkers for OSCC (Figure 1).

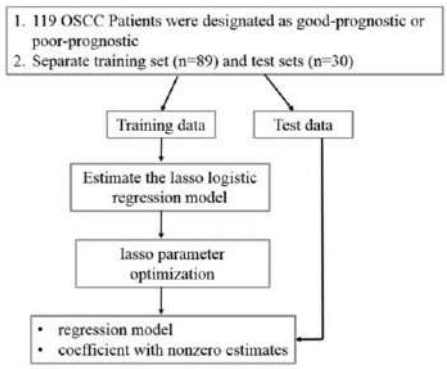


Figure 1. Workflow for strategy 1

- Step 1. Patients were designated as good-prognostic (47 patients) or poor-prognostic (72 patients), based on whether the survival time was longer than 5 years or not (Figure 1). These can be the categorical variable indicating the disease prognosis. Then, split 75% of the 119 OSCC patients selected randomly into training dataset (n=89) and the remaining 25% (n=30) sample into test dataset.
- Step 2. After data filtering (lncRNAs with significant variable expression, differentially expressed between normal and tumor, etc.) and normalization, expression profiles of 286 miRNA, 149 lncRNA and 175mRNA in tumor tissue of OSCC patients were obtained.
- Step 3. Adopt the logistic regression model as follows:
- Logit(P(Y=1|Z))=a+BZ′, Y is the categorical variable with 1 denotes good-prognostic, while 0 denotes poor-prognostic, and Z is the matrix constituted by expression profiles of the mRNA, miRNA and lncRNA selected in Step 2. Here, lasso regularization [22-24], which optimizes the regression output and select features by enforcing sparsity, was used in the regression model to extract RNA biomarkers that can significantly affect the prognosis. It was confirmed to be an efficient algorithm and is suitable for high dimensional problems. We employed the 10-cross validation to optimize the parameter λ that controls the regularization penalty, the one that gives minimum mean cross-validated error was defined as the final penalty parameter.
- Step 4. RNA molecules with nonzero estimates in the final regression model could be the potential biomarkers. Finally, check the efficiency of the classification model composed of the selected RNA biomarkers in the test dataset.

Strategy_2: Extracting RNA Biomarkers from Para-Cancerous Tissues with Differential Expression Analysis

Microarray expression profiles of 8900 human lncRNAs in tumor and adjacent normal tissues of OSCC were obtained [17]. The hypothesis of this strategy is that genetic alterations in normal tissue adjacent to established carcinomas can facilitate tumorigenesis. Here we focused on researching the relationships between lncRNA expression and lymph node metastasis in para-cancerous tissues of OSCC. The workflow of this strategy is shown in Figure 2.

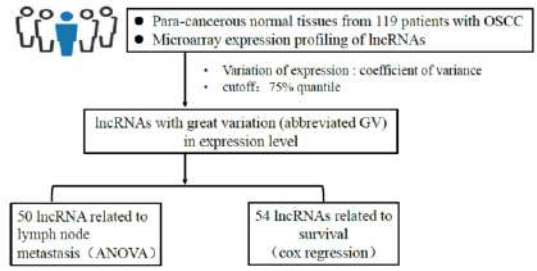


Figure 2. Workflow for strategy 2

- Step 1. Coefficient of variance (abbreviated to CV) was used to measure the specific patterns of expression variation across para-cancerous tissue samples for each lncRNA molecular. For dimensionality reduction, 75% quantile of the CV was used to select lncRNAs with great variation in expression level (abbreviated to GV_lncRNAs).
- Step2. Number of lymph nodes was used to measure the metastasis. Para-cancerous samples from the 119 OSCC patients were designated as subgroup-1 (lymph node: N0) (n=54) or subgroup-2 (lymph node: N1, N2, N3) (n=65), based on the number of lymph nodes (N stage, come from the TNM classification system) of the patients. This can be the categorical variable indicating the disease prognosis [25].
- Step 3. Among the GV_lncRNAs, identify lncRNAs significantly differentiated between para-cancerous tissues of OSCC patients from subgroup-1 and subgroup-2 using ANOVA. Do survival analysis to define lncRNAs whose expression level in para-cancerous tissue was associated with the survival time of patients were defined (Cox regression, $p < 0.05$). These could be potential biomarkers associated with disease prognosis.

Results

Combination of miRNA, lncRNA and mRNA Emerged to be Powerful Biomarkers for OSCC

Followed with the method mentioned in Strategy_1, we defined RNA biomarkers constituted by 10 mRNA, 4 miRNAs and 7 lncRNAs, which achieved high predicting performance (with 20% misclassification error) in our test dataset.

Among the predicted RNA biomarkers, several have been reported to be oesophageal tumour-related items. For instance, DKK1 is the inhibitor of the canonical Wnt signaling pathway, and aberrant expression of this gene was implicated as a novel predictor of poor prognosis in OSCC [26]. WDR66 was reported to be a novel marker for risk stratification and involved in epithelial-mesenchymal transition of esophageal squamous cell carcinoma [27]. Besides, the miRNA biomarker microRNA-1290 and its target gene may play crucial roles in esophageal carcinogenesis and progression and oral squamous cell carcinoma [28].

Expression Level of lncRNA in Para-Cancerous Tissue might Reflect Tumorgenesis

Firstly, variation of expression level for each lncRNA in tumor and para-cancerous tissues showed no significant difference, similar results were observed in those lncRNAs related to lymph node metastasis and survival (Figure 3). Therefore, we

further focused on those lncRNA molecules with great variation in expression level in para-cancerous tissues.

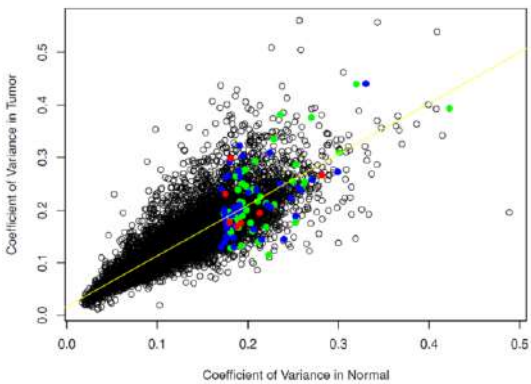


Figure 3. Comparing the variation of expression level for each lncRNA in tumor and para-cancerous tissues. Red dot, lncRNAs whose expression level in normal tissue is associated with patients' survival time and differentially expressed between the para-cancerous tissues of patients from subgroup-1 and subgroup-2; Green dot, lncRNAs whose expression level in para-cancerous tissues is associated with patients' survival; Blue dot, lncRNAs who differentially expressed in normal tissue between patients from subgroup-1 and subgroup-2.

Among the lncRNAs with great variation in expression level, 50 lncRNAs significantly differentially expressed between para-cancerous tissues of patients from subgroup-1 and subgroup-2, as well as 54 lncRNAs whose expression level in para-cancerous tissue associated with the survival time of patients were defined. Resulting in 7 overlapping lncRNAs (linc_XLOC_008811, linc_ENST00000501008.2, linc_ENST00000443445.1, linc_ENST00000504436.1, linc_NR_029392, linc_XLOC_001499, linc_ENST00000582654.1), these could be the potential biomarkers associated with disease prognosis.

For the 50 lncRNAs differentially expressed in para-cancerous tissue of OSCC patients with different lymph nodes metastasis, we divided these molecules into different categories, based on the mean expression level in tumor and para-cancerous tissue. 7 potential lncRNA biomarkers who highly expressed in para-cancerous tissue of patients with lymph node metastasis were identified (Figure 4, Table 1). It is worth noting that the expression level of linc_ENST00000504436.1 is also correlated with the five-year survival rate of patients, indicating that this might be as a potential biomarker for the early diagnosis and prognosis prediction of the disease.

Similarly, we compared the mean expression level of the 54 lncRNAs associated with survival in tumor and para-cancerous tissues and divided these molecules into different categories based on the mean expression level in tumor and para-cancerous tissue. (Figure 5). Expression level of 5 lncRNAs highly expressed in para-cancerous tissue are also associated with tumor prognosis, these molecules are worthy of further research for predicting tumor prognosis.

Table 1. 7 lncRNAs highly expressed in para-cancerous tissue of patients with lymph node metastasis.

lncRNA	N0_ mean	N1_N3 _mean	diff_ mean	CV _N	CV _T	diff_ CV
linc_ENST00000504436.1	*6.47*	*7.25*	*0.78*	*0.21*	*0.20*	*0.01*
linc_XLOC_004685	*4.28*	*4.96*	*0.68*	*0.25*	*0.19*	*0.06*
linc_ENST00000412241.1	*4.58*	*5.09*	*0.51*	*0.22*	*0.21*	*0.01*
linc_ENST00000447557.1	*4.22*	*4.72*	*0.50*	*0.26*	*0.24*	*0.01*
linc_ENST00000422206.1	*3.79*	*4.25*	*0.46*	*0.30*	*0.27*	*0.03*
linc_XLOC_013446	*4.85*	*5.24*	*0.38*	*0.18*	*0.17*	*0.01*
linc_ENST00000548124.1	*3.12*	*3.41*	*0.29*	*0.18*	*0.15*	*0.03*

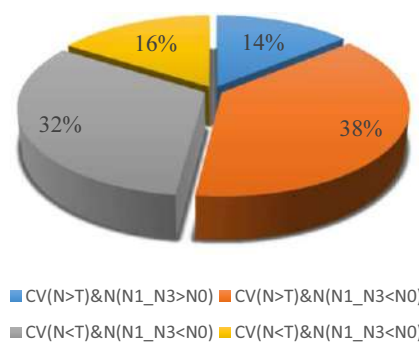


Figure 4. Distribution of 50 lncRNAs based on mean expression level. CV(N>T) represents mean expression level in para-cancerous tissue was higher than in tumor, and vice versa. N(N1_N3>N0) represents mean expression level in para-cancerous tissue of patients designated as N1, N2 or N3 was higher than in normal tissue of those patients designated as N0, and vice versa.

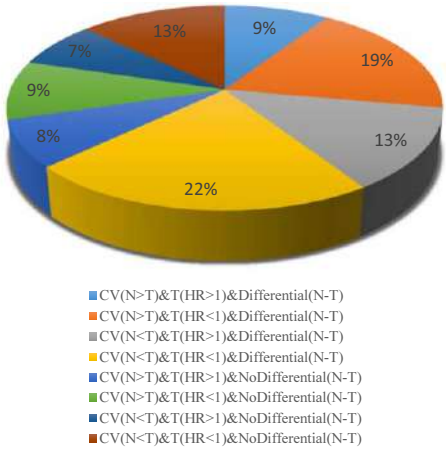


Figure 5. Distribution of 54 lncRNAs based on mean expression level. CV(N>T) represents mean expression level in para-cancerous tissue was higher than in tumor, and vice versa. T(HR>1) represents protective factor, and T(HR>1) represents a risk factor. Differential (N-T) means lncRNA differentially expressed between tumor and para-cancerous tissue, and NoDifferential (N-T) means it is not differentially expressed between tumor and para-cancerous.

Discussion and Conclusions

In this study, we proposed two potential strategies for identifying RNA biomarkers for OSCC. For the integrative analysis strategy, there might be a synergistic effect between mRNAs and non-coding RNAs in tumorigenesis. Generally, based on the extracted biomarkers composed by mRNA, miRNA and lncRNA, patients in test dataset can be classified into good-prognostic and poor-prognostic with low misclassification error. We further searched evidence from published literature and confirmed that several RNA biomarkers have been reported to be oesophageal related risk factors. Actually, selecting prognostic biomarkers from microarray expression profiles is a frequently-used approach, but always limited to dimension of input variables. The lasso approach for logistic regression can limit the number of effective signatures by a proper regularization penalty and extract a minimum number of significant features, and overcome overfitting issues. Application of the lasso approach showed significant power in extracting important prognostic biomarkers for OSCC from high dimensional data.

Nevertheless, further biological effect, as well as the precise regulatory mechanisms between mRNAs and non-coding RNAs in the pathogenesis of oesophageal cancer are essential. Also, due to the sample size limitation, we adopted the cross-validation rather than other independent datasets for penalty parameter optimization, to some extent, might also reduce the model accuracy.

For another strategy, we systematically compared the expression profiles of lncRNA in tumor and adjacent normal tissues from patients with OSCC, to identify potential biomarkers for lymph node metastasis and prognosis prediction. Normal tissue samples are often employed as a control for understanding disease mechanisms, ignoring many of the information that may exist on their own. Actually, exposure to the same biological and physical environment, some genomics mutations occurring in adjacent tissues may also have a certain effect on the development of tumors. Generally, this method can generalized to the analysis of mRNA and miRNA expression profiles. However, the results of informatics research need to be further verified by experimental methods.

Acknowledgements

This work was supported by the National Key research & development Plan of China (Grant No. 2016YFC0901901), the National Natural Science Foundation of China (Grant No.81601573), the National Key research & development Plan of China (Grant No. 2017YFC0907503). The authors declare there's no potential conflict of interest for this study.

References

[1] K.J. Napier, M. Scheerer, and S. Misra, Esophageal cancer: A Review of epidemiology, pathogenesis, staging workup and treatment modalities, *World Journal of Gastrointestinal Oncology* **6** (2014), 112-120.
[2] The Global Cancer Observatory (GCO), in.
[3] A. Pennathur, M.K. Gibson, B.A. Jobe, and J.D. Luketich, Oesophageal carcinoma, *Lancet* **381** (2013), 400-412.
[4] A. Jemal, F. Bray, D. Forman, M. O'Brien, J. Ferlay, M. Center, and D.M. Parkin, Cancer burden in Africa and opportunities for prevention, *Cancer* **118** (2012), 4372-4384.
[5] Y.X. Jiang, D.W. Zhang, Y. Chen, H.H. Sun, S.C. Xu, and H.J. Gao, The characteristics of oesophageal squamous cell carcinoma: an analysis of 1317 cases in southeastern China, *Contemp Oncol (Pozn)* **19** (2015), 137-141.
[6] H. Kashima, K. Noma, T. Ohara, T. Kato, Y. Katsura, S. Komoto, H. Sato, R. Katsube, T. Ninomiya, H. Tazawa, Y. Shirakawa, and T. Fujiwara, Cancer-associated fibroblasts (CAFs) promote the lymph node metastasis of esophageal squamous cell carcinoma, *Int J Cancer* (2018).
[7] M. Matejcic, M. Vogelsang, Y.B. Wang, and M.I. Parker, NAT1 and NAT2 genetic polymorphisms and environmental exposure as risk factors for oesophageal squamous cell carcinoma: a case-control study (vol 15, 150, 2015), *Bmc Cancer* **15** (2015).
[8] B. Vogelstein, N. Papadopoulos, V.E. Velculescu, S. Zhou, L.A. Diaz, Jr., and K.W. Kinzler, Cancer genome landscapes, *Science* **339** (2013), 1546-1558.
[9] B. Wei, Q. Han, L. Xu, X. Zhang, J. Zhu, L. Wan, Y. Jin, Z. Qian, J. Wu, Y. Gao, J. Zhou, and X. Chen, Effects of JWA, XRCC1 and BRCA1 mRNA expression on molecular staging for personalized therapy in patients with advanced esophageal squamous cell carcinoma, *Bmc Cancer* **15** (2015), 331.
[10] Y. Togashi, T. Arao, H. Kato, K. Matsumoto, M. Terashima, H. Hayashi, M.A. de Velasco, Y. Fujita, H. Kimura, T. Yasuda, H. Shiozaki, and K. Nishio, Frequent amplification of ORAOV1 gene in esophageal squamous cell cancer promotes an aggressive phenotype via proline metabolism and ROS production, *Oncotarget* **5** (2014), 2962-2973.
[11] D. Han, M. Wang, N. Ma, Y. Xu, Y. Jiang, and X. Gao, Long noncoding RNAs: novel players in colorectal cancer, *Cancer Lett* **361** (2015), 13-21.
[12] E.N. Kontomanolis and M.I. Koukourakis, MicroRNA: The Potential Regulator of Endometrial Carcinogenesis, *Microrna* **4** (2015), 18-25.

[13] J. Liz and M. Esteller, lncRNAs and microRNAs with a role in cancer development, *Biochim Biophys Acta* **1859** (2016), 169-176.
[14] G.K. Pandey, S. Mitra, S. Subhash, F. Hertwig, M. Kanduri, K. Mishra, S. Fransson, A. Ganeshram, T. Mondal, S. Bandaru, M. Ostensson, L.M. Akyurek, J. Abrahamsson, S. Pfeifer, E. Larsson, L. Shi, Z. Peng, M. Fischer, T. Martinsson, F. Hedborg, P. Kogner, and C. Kanduri, The risk-associated long noncoding RNA NBAT-1 controls neuroblastoma progression by regulating cell proliferation and neuronal differentiation, *Cancer Cell* **26** (2014), 722-737.
[15] X. Hou, J. Wen, Z. Ren, and G. Zhang, Non-coding RNAs: new biomarkers and therapeutic targets for esophageal cancer, *Oncotarget* **8** (2017), 43571-43578.
[16] Z. Chen, J. Li, L. Tian, C. Zhou, Y. Gao, F. Zhou, S. Shi, X. Feng, N. Sun, R. Yao, K. Shao, N. Li, B. Qiu, F. Tan, and J. He, MiRNA expression profile reveals a prognostic signature for esophageal squamous cell carcinoma, *Cancer Lett* **350** (2014), 34-42.
[17] J. Li, Z. Chen, L. Tian, C. Zhou, M.Y. He, Y. Gao, S. Wang, F. Zhou, S. Shi, X. Feng, N. Sun, Z. Liu, G. Skogerboe, J. Dong, R. Yao, Y. Zhao, J. Sun, B. Zhang, Y. Yu, X. Shi, M. Luo, K. Shao, N. Li, B. Qiu, F. Tan, R. Chen, and J. He, LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma, *Gut* **63** (2014), 1700-1710.
[18] T. Yu, R. Cao, S. Li, M. Fu, L. Ren, W. Chen, H. Zhu, Q. Zhan, and R. Shi, MiR-130b plays an oncogenic role by repressing PTEN expression in esophageal squamous cell carcinoma cells, *Bmc Cancer* **15** (2015), 29.
[19] K.C. Johnson, E.A. Houseman, J.E. King, and B.C. Christensen, Normal breast tissue DNA methylation differences at regulatory elements are associated with the cancer risk factor age, *Breast Cancer Research* **19** (2017).
[20] Z. Kaya, M. Akkiprik, S. Karabulut, I. Peker, G.G. Amuran, T. Ozmen, B.M. Gulluoglu, H. Kaya, and A. Ozer, Comparison of telomere length and insulin-like growth factor-binding protein 7 promoter methylation between breast cancer tissues and adjacent normal tissues in Turkish women, *Journal of Clinical Laboratory Analysis* **31** (2017).
[21] L.W. Qu, B. Zhou, G.Z. Wang, Y. Chen, and G.B. Zhou, Genomic variations in paired normal controls for lung adenocarcinomas, *Oncotarget* **8** (2017), 104113-104122.
[22] R. Tibshirani, The lasso method for variable selection in the cox model, *Statistics in Medicine* **16** (1997), 385-395.
[23] R. Tibshirani, Regression shrinkage and selection via the lasso: a retrospective, *Journal of the Royal Statistical Society Series B-Statistical Methodology* **73** (2011), 273-282.
[24] L. Meier, S.A. van de Geer, and P. Buhlmann, The group lasso for logistic regression, *Journal of the Royal Statistical Society Series B-Statistical Methodology* **70** (2008), 53-71.
[25] X.B. D'Journo, Clinical implication of the innovations of the 8(th) edition of the TNM classification for esophageal and esophago-gastric cancer, *J Thorac Dis* **10** (2018), S2671-S2681.
[26] T. Makino, M. Yamasaki, I. Takemasa, A. Takeno, Y. Nakamura, H. Miyata, S. Takiguchi, Y. Fujiwara, N. Matsuura, M. Mori, and Y. Doki, Dickkopf-1 Expression as a Marker for Predicting Clinical Outcome in Esophageal Squamous Cell Carcinoma, *Annals of Surgical Oncology* **16** (2009), 2058-2064.
[27] Q. Wang, C.M. Ma, and W. Kemmner, Wdr66 is a novel marker for risk stratification and involved in epithelial-mesenchymal transition of esophageal squamous cell carcinoma, *Bmc Cancer* **13** (2013).
[28] H. Nakashima, R. Yoshida, A. Hirosue, K. Kawahara, J. Sakata, H. Arita, T. Yamamoto, R. Toya, R. Murakami, A. Hiraki, M. Shinohara, T. Ito, Y. Kuwahara, and H. Nakayama, Circulating miRNA-1290 as a potential biomarker for response to chemoradiotherapy and prognosis of patients with advanced oral squamous cell carcinoma: A single-center retrospective study, *Tumour Biol* **41** (2019), 1010428319826853.

Address for Correspondence

Jiao Li, Ph.D.
Institute of Medical Information & Library,
Chinese Academy of Medical Sciences & Peking Union Medical College
No. 3 Yabao Road, ChaoYang District, Beijing 100020, China
Email: li.jiao@imicams.ac.cn
Phone: +86-(010)-52328740

IV. The Human Element in Medical Informatics

This page intentionally left blank

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190394

Pediatric Asthma Care Assessment for the Emergency Department

Adeola Akinfaderin[a], Kathryn Kovalenko[a], Pamela Newsome[a], Jeritt Thayer[a], Yang Gong[a]

[a] *School of Biomedical Informatics. University of Texas Health Science Center at Houston, Texas, USA*

Abstract

In the United States alone, approximately 6.2 million children have asthma. Managing the care of pediatric patients presenting in the emergency department with asthma requires clinicians to access multiple locations within the electronic health record. This wealth of information makes it vital to develop methods for sifting through the clutter to information that is relevant to the specific task to be completed. Unfortunately, current information displays have not been developed to maximize provider efficiency when integrating information across a number of disparate locations. Using proven data display theories, we aimed to create a consolidated acute pediatric asthma display to aid in the patient's care with the goal of improving the efficiency and consistency by which they are treated - ultimately reducing the length of stay, complications, and morbidity in these patients.

Keywords:
Data Display; Delivery of Health Care; Asthma

Introduction

The amount of information available to the clinician has increased exponentially with the advent of more sophisticated technological tools. Prior to this technology era, the clinician had two sources of information for patient records: the patient's paper chart and the patient. Now, the clinician has access to multiple sources with overwhelming information; the electronic health record (EHR) within their practice, health information exchanges, emergency department (ED) records, and records from other clinicians in the patient's care team, clinical guidelines, and the patient. This wealth of information makes it even more important to develop methods for sifting through the clutter to information that is relevant to the specific task to be completed at a point in time during care delivery.

Background

In the United States alone, approximately 6.2 million children have asthma [2]. Each year, acute asthma events result in 600,000 visits to the ED [2]. In the ED, emergency medicine physicians (occasionally with pediatric specialty) and nurses are the primary health care providers for children presenting with acute asthma exacerbations. There are two major tasks in the treatment of these patients: 1) assess and diagnose the acute asthma event and; 2) determine and monitor asthma severity.

In today's emergency room environment, the data needed to support these tasks are dispersed throughout many areas within the electronic medical record. Information regarding diagnostic imaging results are included in the imaging section, results on arterial blood gas levels would be found in laboratory notes, and additional testing such as spirometry results may be entered in a flow sheet. These multiple sources result in the need for extensive internal data representation and increased cognitive burden to identify trends in the patient's respiratory status and asthma history. For some patients, the declining trend may be represented slowly over days to critical levels if left improperly treated, while others may experience a rapid decline to complete respiratory failure within hours [2].

To further complicate this scenario, it is less common for providers in these departments to have proficiency in managing these severe conditions. In a study conducted by Bratton et al. [3] on the care provided to 13,552 pediatric critical care cases, the results showed a drastic variety in not only the prescribed medications and use of mechanical support, but also the inconsistency in testing and assessment. Sigsbee and Bernat [9] have reviewed potential outcomes ranging from fatigue and job unsatisfaction to increased rates of suicide and depression. Among the four factors in clinician burnout studied were the complexity and time consumption of maneuvering in the electronic health system [9]. For these reasons, the creation of a consolidated acute pediatric asthma display will aid in the care provided to these patients with the goal of improving the efficiency and consistency by which they are treated - ultimately reducing the length of stay, complications, and morbidity in these patients.

Theoretical Frameworks and Methods

Prior to the development of the new Asthma Visualization, current literature on pediatric asthma care in the ED was reviewed. Opportunities for improvement were identified and extracted for evaluation. Our findings suggest that inadequate assessment of patient visit information, history, results, and inaccurate evaluation of asthma severity were the overarching gaps in the current state of pediatric asthma care.

To arrive at an impactful solution to improve and support care delivery to pediatric patients presenting with asthma in the ED, we performed a twofold evaluation specific to the identified gaps in asthma care: 1) the tasks associated with clinical care delivery for pediatric asthma patients in the ED and; 2) the tasks associated with the use of current clinical tools utilized in pediatric asthma care assessment (PACA) in the ED to determine severity. The evaluation process involved a review of the literature, user interviews, and an ethnographic analysis within the users work context.

Care delivery tasks for pediatric asthma patients in the ED

To adequately assess a pediatric patient presenting with asthma symptoms in the ED, the current literature suggests that four clinical tasks should be completed:

Task 1 - A Chest X-Ray must be ordered to help rule out other respiratory problems like pneumothorax and pneumonia.
Task 2 - An arterial blood gas (ABG) test is ordered to evaluate

the level of oxygen in the bloodstream

Task 3 - Pulmonary function tests such as the spirometry and peak flow ratio are performed at regular intervals to monitor the patient's airflow.

Task 4 - Current allergies and medications and reviewed and reconciled.

Tasks for the assessment of pediatric asthma in the ED

Our research findings indicated that current tools used to review patient visit information, screening and diagnostic test results and other pertinent history information were not consolidated, but rather, were in multiple locations within the EHR, in different EHRs, with some even still on paper. We identified the tasks and sub-tasks to be completed for adequate review of the ED visit information and accurate assessment of asthma severity (Table 1).

Table 1 Five Tasks without PACA

5 Tasks without PACA
Task 1 - Review chest X-ray impression
• Subtask 1 – Login to EHR (or source where located)
• Subtask 2 – Find the desired patient
• Subtask 3 – Find X-Ray results for the desired date
• Subtask 4 – Open X-Ray results
○ Subtask 1 – (If image desired) Login to PACs system
○ Subtask 2 – Find Patient
○ Subtask 3 – Find X-Ray image for the desired date
○ Subtask 4 – Click on image to view
○ Subtask 5 – Retain X-Ray results (cognitive task)
Task 2 - Review arterial blood gas results
• Subtask 1 – Login to EHR (or source where located)
• Subtask 2 – Find the desired patient
• Subtask 3 – Find ABG results for the desired date
• Subtask 4 – Open ABG results
• Subtask 5 – Retain ABG results (cognitive task)
Task 3 - Review pulmonary function test results
• Subtask 1 – Login to EHR (or source where located)
• Subtask 2 – Find the desired patient
• Subtask 3 – Find pulmonary function test results for the desired date
• Subtask 4 – Open pulmonary function test results
• Subtask 5 – Retain pulmonary function test results (cognitive task)
Task 4 - Review current medications, allergies, and other pertinent historical information
• Subtask 1 – Login to EHR (or source where located)
• Subtask 2 – Find the desired patient
• Subtask 3 – Find allergies and medications section
• Subtask 4 – Retain allergies and current medications (cognitive task)
Task 5 - Indicate Asthma severity
• Subtask 1 – Login to EHR
• Subtask 2 – Find the desired patient
• Subtask 3 – Find the section in EHR to indicate asthma severity
• Subtask 4 – Recall information about task 1 – 4 from multiple sources (cognitive task)
• Subtask 5 – Indicate asthma severity

We hypothesized that a single display with pertinent recent information on the asthma patient, and a design process incorporating the theories of information display would enable adequate assessment and directly impact the accuracy of asthma severity (Figure 1).

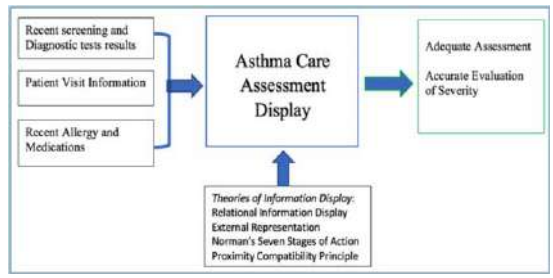


Figure 1 Asthma Care Assessment Display Methodology

To eliminate the inefficiencies associated with current EHRs, we focused on developing an information visualization that would eliminate the need for users to forage through the EHR to find relevant results. Instead, we sought to reduce the above list of tasks to just one (Table 2).

Table 2 One Task with PACA

1 Task with PACA
Task 1 - Review Asthma Assessment Display
• Subtask 1 – Login
• Subtask 2 – Find the desired patient
• Subtask 3 – Review X-Ray results for the desired date
○ Subtask 1 – (If image desired) click or hover over chest X-Ray section
• Subtask 4 – Review Arterial Blood Gas (ABG) test results
• Subtask 5 – Review pulmonary function tests results
• Subtask 6 – Review current medications, allergies, and other pertinent historical information
• Subtask 7 – Select from drop-down asthma severity

The design would incorporate a combination of four theories of information display including, the theory of Relational Information Display, the theory of External Representation, Norman's Seven Stages of Action, as well as elements from the Proximity Compatibility Principle.

Relational Information Displays

According to Zhang [12], Relational Information Displays (RID) represent relations between dimensions. The dimensions are the values to be represented in a display. There are two types of dimensions in a RID, the value to be represented, and how the value is represented. For instance, given a patient weight of 150 pounds, the value to be represented is 150. How we choose to represent it could either be as a name or in relation to other values in order of magnitude. "In order for a RID to be accurate and efficient, its physical dimensions must represent the dimensions of the original domain accurately and efficiently" [12]. Additionally, the potential tasks to be completed using the RID should be at the forefront of the design.

External Representation

The classical view of cognition was that all information processing occurred within the mind and that the external environment (i.e., sight or sound) acted as mere inputs [13]. It wasn't until recently that the cognitive science community began to recognize the interplay of both internal (in the mind) and external (in the world) representations of a task [7]. A common example used to describe this interactivity is the use of multiplication aids. Mental multiplication by itself is not a difficult concept, however, as the numbers grow larger, the partial results of the calculation must be stored in the mind, which tends to exceed the limits of a person's memory. An external structure, writing the partial results on paper, is used to allow a

Figure 2 Asthma Care Assessment Display

person to store this information externally, thus extending their working memory [4]. Other forms of external representations include images, sounds, or even the physical constraints of an object (i.e., being unable to move an object in some desired fashion).

Norman's Seven Stages of Action

This feeling of direct interaction can be broken down into two phases and seven stages according to Norman's model of the Seven Stages of Action [8]. The first phase is composed of stages that comprise the *execution* of a task, and the second phase addressing the stages that enable *evaluation* of the task [8]. The complexity of turning a goal into a path of execution is referred to as the "gulf of execution", where the ability to evaluate the effectiveness of the result is considered the "gulf of evaluation" [6]. Well-designed direct interaction interfaces will attempt to reduce the 'distance' of each gulf by aligning the interface to the way the user would describe a task (semantic distance), and ensuring the forms or images used are accurate to represent the task (articulatory distance) [6]. Reducing these gulfs will support rapid user adoption and efficiency, but rigidness of design may limit users' ability to imagine new ways to evaluate the data [6].

Proximity Compatibility Principle (PCP)

Wickens and Carswell [11] describe PCP as a guideline to use in determining where a display should be located, given its relatedness to other displays. As task complexity increased, it was shown that for integrative tasks, the dimension of the representation should also increase. These results suggest that by applying PCP to data displays, designers can optimize the integration or correlation of displayed information and maximize parallel processing, while also allowing specific elements to be extracted through focused attention. Previous research has concluded that there exists a link between task proximity (i.e., how closely related are individual tasks) and display proximity (i.e., how closely related are the object representations) [11], and displays that adhere to this theory are associated with more efficient and effective information retrieval.

Results

Our Pediatric Asthma Care Assessment visualization contains access to all pertinent information needed to allow the clinician to assess the asthma severity level of a child presenting in the ED with respiratory distress. Our screen design format emphasizes overview-first followed by a click and expand capability when more detail is desired. We employed the use of spatial proximity, enclosure, and colors to separate and group items of similarity. The two main tasks to achieve the clinician's goal were identified, and our screen design aimed to provide the user's ease in executing and evaluating these tasks. We maximized external display to limit cognitive burden and ensure the categories are appropriately represented in a relational information display to avoid over or under-representation of data. There are four sections to the PACA display (Figure 2).

Section 1 The area along the top quarter of the screen contains the patient and care team detail. This section is where one would find patient detail such as gender, age, and name. The right portion notes all of the participating care team members. In the middle of section 1 is an initial asthma severity drop down to allow the ED clinician to review the initial severity level assigned (Figure 4).

Section 2 This area is outlined in yellow to draw attention to the dimensions required to assess and diagnose the acute asthma event. The dimensions accessed in this section are reported triggers, allergies, and medications. Also visible are lab results such as chest X-rays, lung function graphs, and ABG findings. Hovering over the chest x-ray description brings up the actual patient images for review (Figure 3). Color is used to represent the lung function test results. Color is also used to differentiate between normal, caution, and medical alert to speed up information processing.

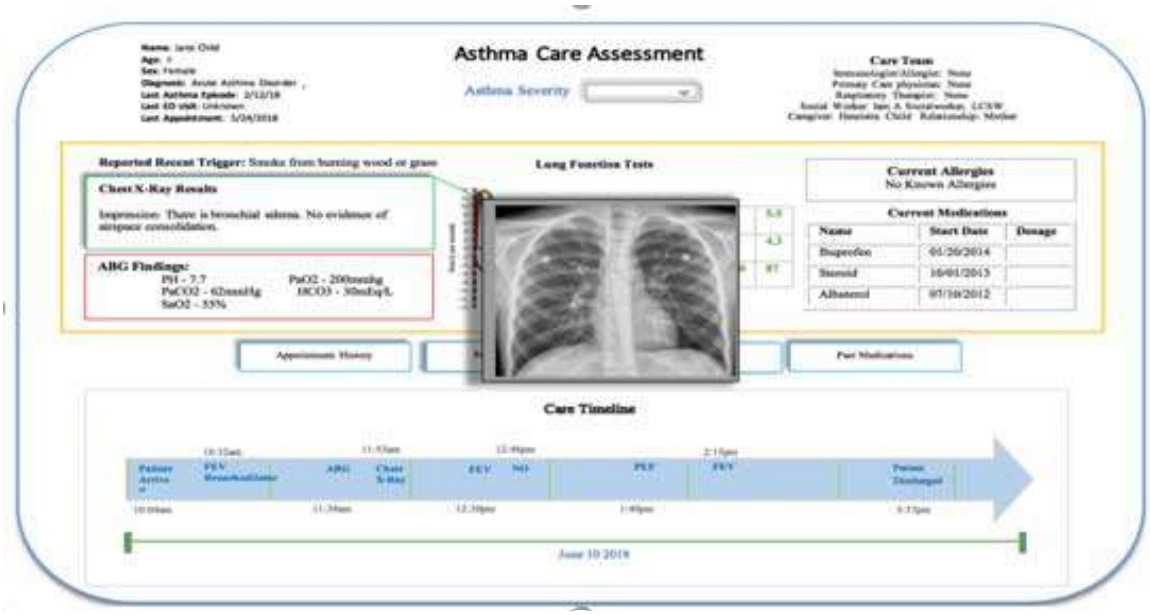


Figure 3 Asthma Care Assessment Display with Imaging View

Section 3 This area will show buttons to expand on further patient information if required by the clinician. Patient information like past surgical history, appointment history, other test results, and past medication is accessible from this section. Our design chooses to represent non-essential data as a button for selecting when desired, rather than cluttering up the screen with extraneous data. The screen is not busy and allows the clinician memory and attention to be uncluttered and focused.

Section 4 This area is outlined in blue to draw attention to the dimensions required to determine and monitor asthma severity. The care timeline shows events such as admit/discharge times, lung function readings, timestamps for each test and when each drug is administered. The section allows the clinician to easily review the care given to the child from ED admission to discharge. All pertinent events for a pediatric ED visit are visible in the line graph.

Discussion

The PACA visualization was developed to reduce the cognitive load on clinicians caring for pediatric patients in the ED. Relevant data that is often located on a number of separate screens in the EHR was represented on a single. Additionally, each data dimension was matched with an appropriate representing dimension to maximize the external representation within the display. For example, the lung function test, which was represented only in tabular format is now also represented graphically. To accomplish this, multiple information visualization principles and theories were referenced. Zhang and Walji [14] provided the TURF framework that informed the initial analysis of the user's work domain. The theory of relational information displays and the theory of external representation provided guidance on mapping the represented and representing dimensions for our given distributed task. The proximity compatibility principle focused our efforts on grouping related objects of interest to reduce the cognitive cost of clinicians integrating information internally.

Our work aligns with other literature on the use of efficient data displays to reduce cognitive load. Teets, Tegarden, and Russell [10] demonstrated that when the PCP was combined with the theory of cognitive fit, there was increased efficiency in information integration during simulated tests within a set of manufacturing facilities. Marino and Mahan (2005) showed that configural displays could increase the speed at which consumers were able to process the relationship between nutrients on a food label. Alonso, Rose, and Plaisant [1] found much faster response times related to a graphical information display in comparison to tabular format for personal history information. Unfortunately, only a few studies have attempted to analyze the impact of data display within healthcare. Even fewer have deployed the displays within a live clinical setting such as the ED. Furthermore, while EHRs have been around for some decades, common data display methods have been habitually ignored. As far back as 30 years ago, a physician named Larry Weed introduced the problem-oriented medical record, but clinicians are still forced to search across multiple screens to find all of the information they require.

Data display continues to be an essential component within healthcare to realize the full potential of information technology. Healthcare is an incredibly complex sociotechnical environment, which creates the opportunity for workarounds or mistakes that can lead to serious adverse events. The tragic story of Genesis Burkett is one example of such an event, where the patient died after a pharmacy technician entered incorrect information onto a screen [5]. As a result, an automated system prepared an IV solution that was more than 60 times the amount ordered by the physician [5]. This story highlights the profound impact health data displays can have on patient safety. While the PACA visualization adheres to proven theories and methods in the design of data displays and usability, further research is needed to ensure the application is used as intended when implemented within a live production EHR.

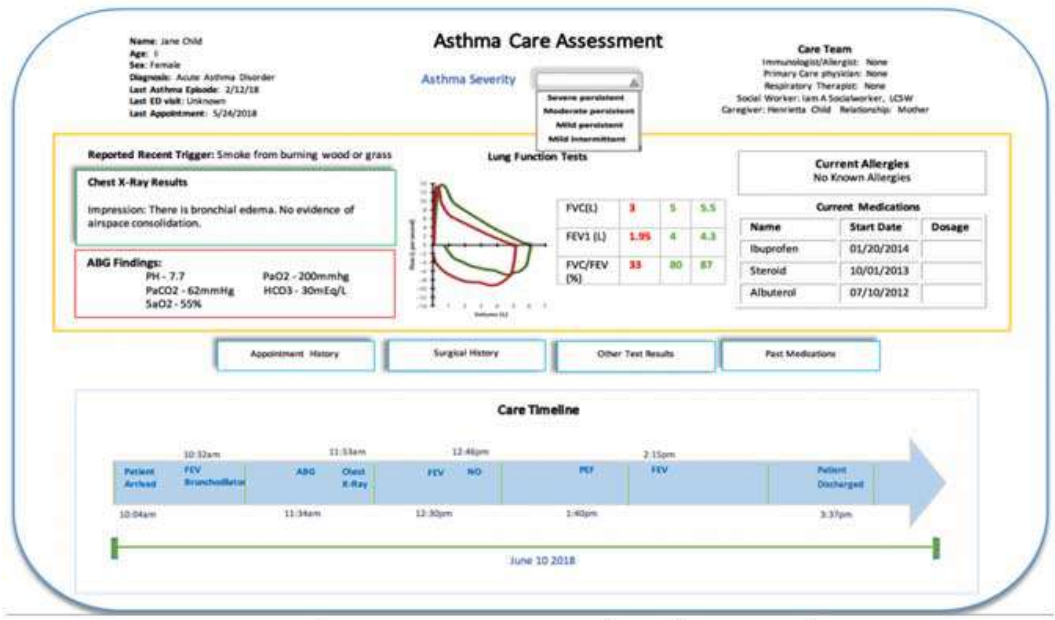


Figure 4 Asthma Care Assessment Display with Severity Selection

Limitations and future work

Although our design is based upon proven framework and theories, we would prefer there be a more formalized expert domain review of tasks to ensure a solid understanding by the developers. This screen would need to be utilized as part of a clinical workflow of an actual ED to prove the efficiency of the tasks to complete the goal.

Conclusion

The increased demand on clinician has produced a critical problem with clinician burnout. Through the application of the visualization frameworks and design principles, the PACA visualization is expected to reduce the cognitive load and burden on the clinician. While the user may benefit and reduce the effects of clinician burnout, the resulting display most importantly allows for the informed, efficient, and consistent treatment of pediatric patients experiencing an acute asthma event.

References

[1] D.L. Alonso, A. Rose, C. Plaisant, and K.L. Norman, Viewing personal history records: A comparison of tabular format and graphical presentation using LifeLines, *Behaviour & Information Technology* **17** (1998), 249-262.

[2] R.A. Batabyal and K. O'Connell, Improving Management of Severe Asthma: BiPAP and Beyond, *Clinical Pediatric Emergency Medicine* **19** (2018), 69-75.

[3] S.L. Bratton, C.J. Newth, A.F. Zuppa, F.W. Moler, K.L. Meert, R.A. Berg, J. Berger, D. Wessel, M. Pollack, and R. Harrison, Critical care for pediatric asthma: wide care variability and challenges for study, *Pediatric Critical Care Medicine* **13** (2012), 407.

[4] M. Card, *Readings in information visualization: using vision to think*, Morgan Kaufmann, 1999.

[5] J. Graham and C. Dizikes, Baby's death spotlights safety risks linked to computerized systems, *Chicago Tribune* **27** (2011).

[6] E.L. Hutchins, J.D. Hollan, and D.A. Norman, Direct manipulation interfaces, *Human-computer interaction* **1** (1985), 311-338.

[7] D.R. Kaufman, T.G. Kannampallil, and V.L. Patel, Cognition and human computer interaction in health and biomedicine, in: *Cognitive informatics for biomedicine*, Springer, 2015, pp. 9-34.

[8] D. Norman, *The design of everyday things: Revised and expanded edition*, Constellation, 2013.

[9] B. Sigsbee and J.L. Bernat, Physician burnout A neurologic crisis, *Neurology* **83** (2014), 2302-2306.

[10] J.M. Teets, D.P. Tegarden, and R.S. Russell, Using cognitive fit theory to evaluate the effectiveness of information visualizations: An example using quality assurance data, *IEEE transactions on visualization and computer graphics* **16** (2010), 841-853.

[11] C.D. Wickens and C.M. Carswell, The proximity compatibility principle: Its psychological foundation and relevance to display design, *Human Factors* **37** (1995), 473-494.

[12] J. Zhang, A representational analysis of relational information displays, *International Journal of Human-Computer Studies* **45** (1996), 59-74.

[13] J. Zhang, The nature of external representations in problem solving, *Cognitive science* **21** (1997), 179-217.

[14] J. Zhang and M.F. Walji, TURF: Toward a unified framework of EHR usability, *J Biomed Inform* **44** (2011), 1056-1067.

Address for correspondence

Yang Gong, MD, PhD

7000 Fannin St Suite 600, Houston, TX 77030, USA

Yang.Gong@uth.tmc.edu

Phone: +1 713 500 3547, Fax: +1 713 500 3907.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190395

User-Centered Design of a Pediatric Vaccination Module for Patients

Sebastian Minoletti[a], Romina Rapisarda[a], Liliana Giraldo[a], María Grande[b], Janine Sommer[a], Fernando Plazzotta[a], Daniel Luna[a]

[a] *Department of Health Informatics, Hospital Italiano de Buenos Aires, Ciudad de Buenos Aires, Argentina*
[b] *Department of Internal Medicine, Hospital Italiano de Buenos Aires, Ciudad de Buenos Aires, Argentina*

Abstract

Immunization information systems and registries are essential for vaccine programs to succeed. Personal Health Record should improve decision making of healthcare providers and consumer outcomes. This article aims to describe the software development of an immunization module using a user-centered design, starting from the analysis of potential users' needs. The design was made through cycles of iterations, adjustments were made validating real scenarios and taking into account the user needs. The main features identified through interviews were: vaccination schedule, further information on vaccines' usefulness, notification, downloadable patient vaccination status and other vaccines outside the official calendar. The final scope is to create a simple, efficient and safe platform for patient immunization management.

Keywords:
Personal Health Record, Vaccines, Software Design

Introduction

Personal Health Records (PHR) can be defined as "*an electronic application through which individuals can access, manage and share their health information, and that of others for whom they are authorized, in a private, secure, and confidential environment*" [1]. In this way, patients and their families can interact with health care systems, becoming involved in the medical decision-making process and thus achieve empowerment [2,3].

PHRs should improve the decision making of caregivers and patient outcomes. However, despite the promise of better health care through patient-centered information empowerment, little progress has been made in terms of immunization information systems.

On the other hand, smartphones have demonstrated the ability to improve immunization practices, and the use of reminder text messages increases vaccination coverage in both, pediatric and adult populations [4–6]. Furthermore, a group of experts in Canada has developed a free smartphone app that currently functions as a digital tool for people to monitor their own and their families' vaccines. This app allows to access vaccination information and see outbreaks of preventable diseases with vaccines close to the user area [7].

Immunization information systems and their registries are critical to the success of immunization programs [8]. The PHR establishes a digital connection between patients and health information systems. This is an opportunity to develop new services and collect new information by patients themselves or their relatives connected into their own clinical records.

Access to vaccination records through health information technology presents an opportunity to empower patients easily with their own health information. Accurate immunization status has a greater value for patients, vaccine providers, public health officials and industry [9].

It is essential to record all vaccines administered to children. For some providers have an "unregistered vaccine is equivalent to non administered vaccine" and, overall, there are safe and centralized computer systems that could allows knowing the immunization status at any time and everywhere. [10].

Currently, caregivers continue to use paper vaccination notes to provide families with essential information about the immunization status of their children. There are many emerging challenges for the adoption of new systems, which include: the frequently release of new vaccines to the market, and the consequent changes in the official vaccination calendar, new recommendations, multiple brand products, the use of multiple immunization providers, under registration, and / or administrations in several locations (included in the intrahospital network, or not) [11].

We propose, through the use of a new information technology development, to overcome these challenges, to explore the needs and barriers of the users, with the scope of designing a user-centered interface for the management of immunizations and evaluating the effectiveness and user satisfaction.

Methods

Setting

The Hospital Italiano de Buenos Aires (HIBA) is a non-profit organization with 165 years of history in Argentina. Its healthcare network includes a university hospital of high complexity that covers health care from outpatient, inpatients, emergencies, critical care, home care, chronic care and medical and surgical specialties. It has its own medical insurance service (Health Insurance Plan), with more than 160,000 patients enrolled, and provides health services to 1,500,000 people with other health insurances. Annually, more than 45,000 patients are admitted to their hospitals, and 45,000 surgical procedures and 3,000,000 outpatient visits are made.

Since 1998, the HIBA has his own health information system (in-house) that includes the management of clinical and administrative information. Its Electronic Health Record (EHR) is an integrated, modular, problem-oriented and patient-centered system, used in the different clinical scenarios (ambulatory, hospitalization, emergency center and home care) [12,13].

As part of the information system, an integrated PHR called POPES is available to all outpatients since 2007. The PHR allows patients receiving medical care in the hospital network to access, verify clinical and administrative information and interact with the health system. Among the main functions, the PHR allows users to update their personal information, share

information, manage schedule appointments, view orders results, have a messaging service with the general practitioner, check, order and buy medication prescribed, and the possibility to consult with the health team by Telemedicine tools, among others. At present, POPES has approximately 400 thousand registered users, of which 15.000 are under the age of 18 years old [14].

The project was carried out in the Health Informatic Department of our center, a university hospital of high complexity in Argentina.

Design

This article uses a mixed methodology that assesses and describes the user centered development of a module to manage the immunization of the patient. The development was divided into 2 stages, the *situational diagnosis* and the *prototype development.*

Stage 1: Situational diagnosis

The first stage consisted of a qualitative analysis through individual interviews with adult patients with at least one child in charge, in order to explore the user needs. For the selection of the sample, a list of the patients who fulfilled the following inclusion criteria was requested:

- Argentine nationality,
- current users of the institutional PHR,
- adult over 18 years, with at least one child (under 18 years old) in charge,
- the patient should have an ambulatory appointment programmed in the hospital during the month that the interviews were planned.

A convenience sampling was conducted, establishing contact via email and to offer an interview before or after the assigned appointment.

Oral informed consent was given before the interviews, which lasted between 30 and 45 minutes, were audio recorded for further analysis, and the data was kept anonymous. The analysis process was carried out simultaneously with the information recollection process, through the codification and categorization of the information, based on a comparison process. The categories were built taking into account both, the central questions of the research team and the findings emerged from the interviews [8].

Stage 2: Prototype development

This stage was divided into two phases, the *prototype design* and 5 rounds of *usability testing*, in order to achieve an efficient design, with a lower error rate and greater value for the user.

Phase 1, prototype design: A new vaccine module was built for the PHR. This module was determined by a list of vaccines segmented by periods of vaccination (newborn, 2 months, 3 months, etc.) similar to the vaccination calendars on paper. Each segment contained the type of vaccine for that age, administration status, number of doses required, previous vaccine received and date of administration, information related to the vaccine and administration data.

The user possible actions were located in the top of the screen (registration, request, reminders settings and download immunization status). The prototype was completed with a link to the current official vaccination calendar and a series of filters for viewing the vaccine status. For the registration interface (new vaccination), the following fields were included: the corresponding age of vaccination, date of administration, institution, vaccine or brand name, batch and voucher. In the "prescribe vaccines form", it was enough to select the ages on the calendar (for example, those corresponding to 2 months of age) to be prescribed electronically by the health professional. For the functionality "download vaccination status", we chose to create a PDF with patient identification data and a description of the vaccines administered and validated by the attending physician. Finally, for the "notification settings" interface, it was decided to replicate what already exists in the PHR, where the patient selects by which way they are used to be notified (SMS, email or through the PHR).

All the prototypes were designed in Illustrator (Ai) and Balsamic, 5 iterations were needed to achieve the final version.

Phase 2, usability test: Consists in the interviews with the patient, showing the developed prototype to 5 users of the PHR (who meet previously mentioned inclusion criteria), to detect usability problems.

The test is an exercise of practice of use. The participant must read aloud the tasks and try to solve them. At the same time, they are requested, as the test goes forward, to express what happens, their interpretation and thoughts. The interviews include 3 participants: the user, the facilitator (who moderates) and the observer (who observe and records on paper). Finally, at the end of the test, a System Usability Scale (SUS) questionnaire was conducted to evaluate satisfaction of use. This entire exercise is recorded with MORAE, with prior informed consent.

Five iterative testing cycles were made, modifying the usability problems detected until reaching a final proposal. Each round of testing was carried out with five different users. In the validation of the prototypes, for each of them, the metrics related to the measurement of effectiveness (percentage of tasks performed correctly), and satisfaction of use (SUS) were established.

Results

About the user perceptions of vaccine registry, we detected through the interviews the following main user needs:

- digitization of vaccination calendar,
- notification system,
- downloading updated vaccination status,
- complementary information about the usefulness of vaccines (mandatory vaccines and extra official calendar).

The Italian Hospital, where the dispensation of vaccines is determined by the medical prescriptions, make as decided to include a functionality to "Request Vaccine Prescription". This functionality already exists for other prescriptions, (e.g. medications) and allows the users to order a delivery at home of vaccine prescribed.

First iteration:

Although 100% complete correctly the registration form (Fig. 2), the users had difficulties to find the main actions due to the amount of elements on the screen. The difficulty was bigger at

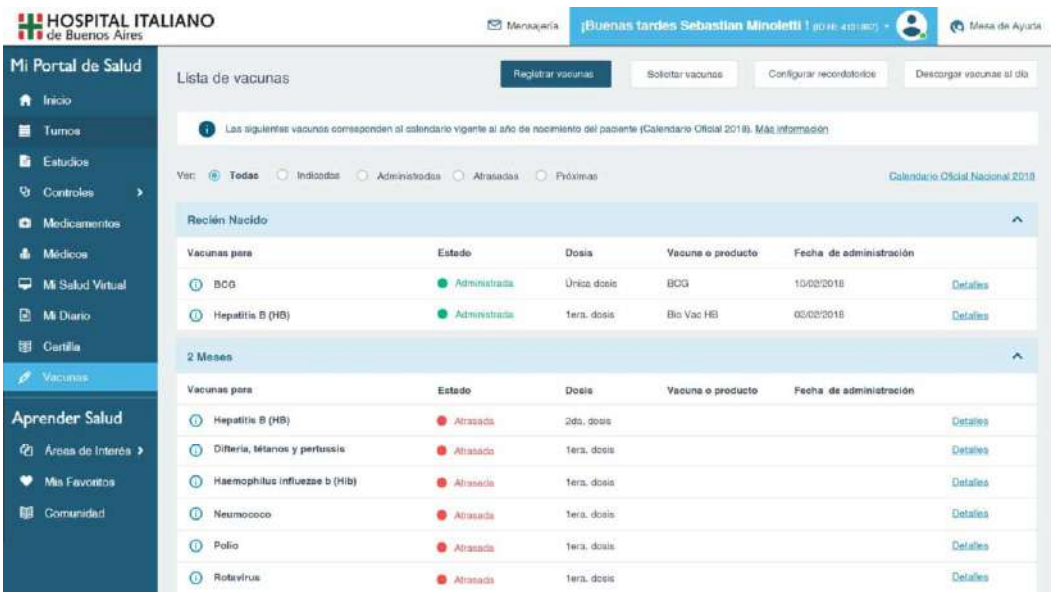


Fig. 1 - Personal Health Record, vaccine module view

displaying the results, then they try to look for a vaccine, they could not recognize which was the correct option because, according to them, the vaccines were similar. (Fig. 1) Another limitation was the expected feedback from the system, this was not accurate or clear in its content. Regarding the Request Vaccine Prescription (fig 3) 60% of users completed the task, although they did not understand why they should request vaccines that are free and obligatory. To obtain the vaccination status, 20% of users find it hard to find the option since the chosen term was not clear. As for satisfaction, the SUS value was 85.

Fig. 2 - Registration form view. To add an administered vaccine.

Fig. 3 - Request vaccine prescription view.

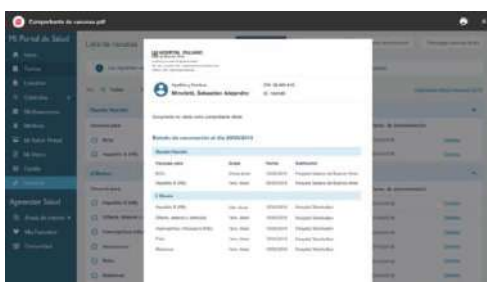


Fig 4 - Vaccination status view.

Second iteration:
The states of the vaccines are clear, but excessively overloaded, making it difficult to recognize the main actions. The lack of understanding of search field results, the status after a new registration and request a vaccine prescription persisted despite the update.

Despite the problems described, 100% of people completed the registration and it was easy after the first record. Only 60% request vaccine prescriptions for the next month. As for the vaccination status, only 20% managed to complete the task, since the name of the button (print state of vaccines) was not clear. Regarding the satisfaction of use, it dropped 1 point in relation to the first iteration, 84.

Third iteration:
A status and age filter was incorporated, to display the vaccines according to the patient calendar, which was accepted by users.

The lack of clarity in the search field results is maintained with the new update. The new changes brought a reduction in the registration form, 80% completed the task. The request task is maintained at 60% and the vaccine status improves to 60% with the new name (voucher download).

The satisfaction of use had a value of 83, one point less than iteration two.

Fourth iteration:

In this stage, we locate the actions "register" and "request" in each row of the list. Although 100% of the users managed to complete the registration, the time to understand the screen and perform the task was higher.
In the registration form, a new visualization of results was proposed, where a brief definition is attached to each vaccine, either generic or brand name, in order to improve understanding.
The rate of request a vaccine was 60%, maintained as well as the voucher download. The SUS scored 87 points.

Fifth iteration:

After 4 iterations, we decided to analyze the functionalities that were understandable and easy to use, and those that were not (Fig 5-7).

Registration: the minimum fields should be the name of the vaccine, the number of doses, date of administration and the voucher.

Request: it was decided not to continue with this functionality since it was not a user need, understood as an action of the health personnel.

List of vaccines: the biggest problem founded was the overloaded information, the user spends a lot of time identifying each element. We decided to simplify the interface, where we eliminated the list of mandatory vaccines to replace it with a button for the registration form. On the right side, a timeline of the current calendar is available as a reminder.

The actions were limited to two: register and download voucher. The 100% of users completed the task of registration and download the voucher, in both cases there were no problems. The SUS value was 84 points.

Table 1 - Effectiveness of each iteration.

Task	Test 1 Correct %	Test 2 Correct %	Test 3 Correct %	Test 4 Correct %	Test 5 Correct %
Find who indicated BCG	Not tested	100	80	100	100
Find the 1st. dose of rotavirus	Not tested	60	80	100	60
Register pentavalent	100	100	80	80	100
Register Hepatitis B	100	100	100	100	100
Register Pneumoc.	100	100	100	100	100
Register Rotavirus	100	100	100	100	100
Register yellow fever	Not tested	Not tested	0	20%	Task not understand able
Request 3rd. month vaccines	80	60	60	60%	Task removed
Vaccination status / voucher	80	20	60	80%	100

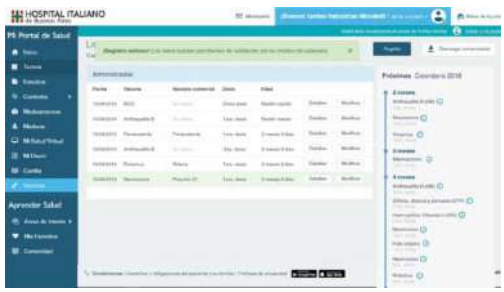


Fig. 5 - Personal Health Record, final vaccine module view.

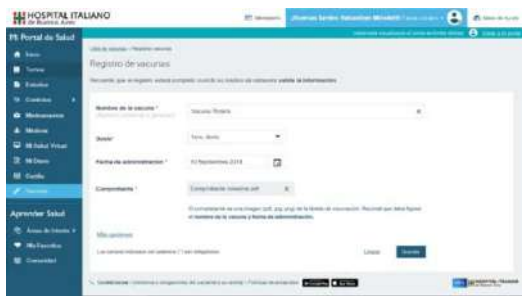


Fig. 6 - Final registration form view. To add an administered vaccine.

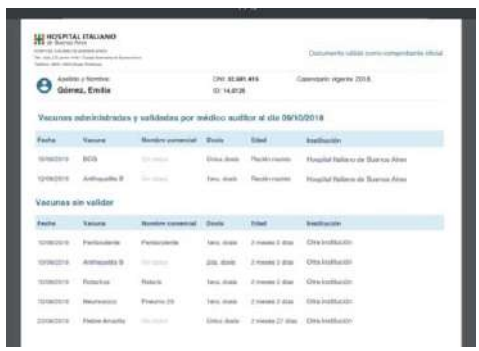


Fig. 7 - Final vaccination status view (PDF).

Discussion

The main strength of this work is that the potential user needs were analyzed at the beginning, and the user was incorporated in all the process of design, which allowed to generate a tool based on user-centered design. Through the iterations, adjustments were made, validating the real scenarios of patients and taking into account their needs. The ultimate goal was to generate a simple and secure platform, being effective and efficient. The SUS value always remained bigger than 80.2, meaning that users love the application, and even recommend the use of it to a partner.

Electronic vaccines management represents an optimal tool to monitor the quality and safety of patient care, through the ability to allow real-time management by patients. In the pediatric population, this challenge becomes even more complex, taking into account the number of vaccines included in the official calendar, the responsibility centralized in the adult in charge and the preventive relevance of immunization in this particular population.

The final version represents an efficient design with great value for the user, adjusted to their needs. Have the potential

to improve communication between health professionals and patients.

Among other features, more information on vaccines was incorporated, which will improve registry and accessibility of patient information.

In the case of yellow fever vaccine, rather than been an error, usability test could not be performed, since being an optional vaccine (not in the official vaccination calendar), it is recorded in a section of the paper vaccination registry where users do not usually read. Also, because it was a test, we could not interfere and induce the user to register that specific vaccine.

We believe that the use of this new technological development will be able to overcome challenges, the complexity of registries and administration of immunizations. It is currently in development. It is planned to carry out a test with a beta version in order to obtain a baseline (starting point), for future measurements and also, based on this experience, we plan to incorporate the adult population.

Conclusions

A tool based on user-centered design was generated, starting from the analysis of potential users needs. Through the iterations, adjustments were made to validate the actual conditions of the vaccination process, the real scenarios of the users, was taken into account. A simple and safe, efficient and satisfactory platform was designed.

Acknowledgements

This study was carried out under the Residency Program in Health Informatics and the Area of Information Technology Oriented to the Community of the HIBA. We want to thank the entire interdisciplinary team (doctors, nurses, designers, developers) for the work you have done.

References

[1] Foundation Markle, The Personal Health Working Group Final Report, 2003.

[2] C. Pagliari, D. Detmer, and P. Singleton, Potential of electronic personal health records, *BMJ : British Medical Journal*. **335** (2007) 330.

[3] J.D. Halamka, K.D. Mandl, and P.C. Tang, Early Experiences with Personal Health Records, *J. Am. Med. Inform. Assoc.* **15** (2008) 1.

[4] M.S. Stockwell, and A.G. Fiks, Utilizing health information technology to improve vaccine communication and coverage. - PubMed - NCBI, (n.d.). https://www.ncbi.nlm.nih.gov/pubmed/23807361 (accessed July 9, 2018).

[5] E. al Leeb A, Using automated text messages to monitor adverse events following immunisation in general practice. - PubMed - NCBI, (n.d.). https://www.ncbi.nlm.nih.gov/pubmed/24794676 (accessed July 9, 2018).

[6] E. al Regan AK, Using SMS technology to verify the safety of seasonal trivalent influenza vaccine for pregnant women in real time. - PubMed - NCBI, (n.d.). https://www.ncbi.nlm.nih.gov/pubmed/24329634 (accessed July 9, 2018).

[7] A. Hoffman. Risk And Reform Of Long-Term Care. https://www.healthaffairs.org/do/10.1377/hblog20141114.042462/full/ (accessed July 9, 2018).

[8] K. Wilson, K.M. Atkinson, S.L. Deeks, and N.S. Crowcroft, Improving vaccine registries through mobile technologies: a vision for mobile enhanced Immunization information systems, *J. Am. Med. Inform. Assoc.* **23** (2016) 207–211.

[9] K. Wilson, K.M. Atkinson, S.L. Deeks, and N.S. Crowcroft, Improving vaccine registries through mobile technologies: a vision for mobile enhanced Immunization information systems, *J. Am. Med. Inform. Assoc.* **23** (2016) 207–211.

[10] T. Bickmore, L. Pfeifer, and B. Jack. Taking the Time to Care: Empowering Low Health Literacy Hospital Patients with Virtual Nurse Agents. http://relationalagents.com/publications/CHI09.VirtualNurse.pdf (accessed July 9, 2018).

[11] E. al Goad JA, Vaccinations administered during off-clinic hours at a national community pharmacy: implications for increasing patient access and convenience. - PubMed - NCBI, (n.d.). https://www.ncbi.nlm.nih.gov/pubmed/24019274 (accessed July 9, 2018).

[12] D. Luna, F. Plazzotta, C. Otero, F. González Bernaldo de Quirós, A. Baum, and S. Benítez, Incorporación de tecnologías de la información y de las comunicaciones en el Hospital Italiano de Buenos Aires, (2012). http://hdl.handle.net/11362/3959 (accessed November 23, 2018).

[13] F. Plazzotta, D. Luna, and F. González Bernaldo de Quirós, [Health information systems: integrating clinical data in different scenarios and users], *Rev. Peru. Med. Exp. Salud Publica*. **32** (2015) 343–351.

[14] M.I. Smith, G. Garcia, M. Simon, L. Bruchanski, J. Frangella, J.A. Sommer, M.V. Giussi, and D.R. Luna, Lessons Learned After Redesigning a Personal Health Record, *Stud. Health Technol. Inform.* **245** (2017) 216–220.

Address for correspondence

Sebastian Minoletti. Visual Communication Designer. Department of Health Informatics, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina.. email: sebastian.minoletti@hospitalitaliano.org.ar

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190396

Developing a Saudi Health Informatics Competency Framework: A Comparative Assessment

Manal Almalki[a], Mowafa Househ[b], Mohammed Alhefzi[c]

[a] *Department of Health Informatics, Faculty of Public Health and Tropical Medicine, Jazan University, Jazan, Jazan Province, Saudi Arabia,*
[b] *Department of Health Informatics, College of Public Health and Health Informatics, King Saud Bin Abdulaziz University for Health Sciences, Riyadh, Riyadh Province, Saudi Arabia,*
[c] *King Faisal Medical City for Southern Regions, Abha, Asir Province, Saudi Arabia*

Abstract

In 2018, the Saudi Commission for Health Specialties (SCFHS) created a national working group composed of key health informatics (HI) professionals, researchers and educators tasked with the development of a draft competency framework for Saudi HI professionals. Over an eight-month period, the research group collected data obtained from literature sources (both academic and grey), international competency standards, participant surveys, focus groups, and expert panel reviews. Through multiple rounds of discussions and graphic visualisation of the information collected using Microsoft PowerPoint and flip charts, the data were summarised and a visual representation of the proposed SHICF was developed. The result of this effort was the development of the first Saudi Health Informatics Competency Framework (SHICF). This paper provides a comparative assessment between the Saudi HI competency framework development and that of other internationally recognised HI competency development frameworks. Challenges related to the development of the SHICF are also discussed.

Keywords:

Health Informatics, Professional Competency, Saudi Certification.

Introduction

Obtaining formal recognition for Health Informatics (HI) as a licensed and credentialed profession was the aim of many leading HI professional organisations such as Canada's Health Informatics Association and HISA (Health Informatics Society of Australia). Projects were carried out to identify the core HI domains and competencies that would describe the performance and set pathways of a health informatics professional. Due to the evolving and multidisciplinary nature of health informatics, it has been a challenge for HI professional organisations to define the competencies and core domains of HI in a manner that would satisfy the variety of stakeholders practicing the HI profession. Nonetheless, this has not stopped HI professional organisations from trying to develop core domains and competencies to define the HI field.

In March 2018, the Saudi Commission for Health Specialties (SCFHS) invited key health informatics educators and researchers to participate in a national working group which, over an eight-month period produced the first competency framework for Health Informatics professionals in Saudi Arabia. The purpose of this paper is to compare the competency development project that was carried out in Saudi Arabia with similar projects in five internationally recognised HI professional organisations and identify the challenges that the Saudi national competency working group faced in the development of SHICF. We anticipate that our work will provide insights for future countries and/or regions when conducting similar work.

Methods

As part of the Saudi HI competency framework development project, the working group searched the literature and identified five primary organisations that previously created HI competency frameworks from Europe, North America and Australia: Digital Health Canada, formerly known as COACH [1], The International Medical Informatics Association (IMIA) [2], EU-US eHealth Collaboration Initiative [3], the Certified Health Informatician Australasia (CHIA) [4], American Medical Informatics Association (AMIA) and CAHIIM [5]. The HI competency frameworks for each country were reviewed and studied through a comparative analysis between the work the different organisations conducted in the creation of their own HI competency framework. An online discussion also took place between the key research members involved in this study between October 1 and November 15, 2018 to identify the challenges and issues relating to the development of the Saudi HI competency and domain-based framework. Comparisons relating to project's objectives, approaches of competency development, key outputs (framework), and framework applications were conducted between the different HI competency and domain-based HI professional frameworks. Figure 1 summarises the methodology followed.

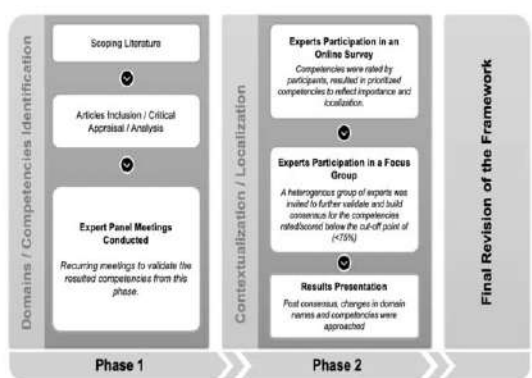


Figure 1– Methodology Followed in Developing the Saudi Health Informatics Competency Framework (SHICF).

Table 1 – Summary of the Comparative Analysis of Competencies Development Projects of Five Leading Organisations in HI with the Saudi Project

Country	Canada	Switzerland	Europe	Australia	USA	Saudi Arabia
Body/ organisation/ initiative description	**Digital Health Canada, formerly known as COACH:** Canada's Health Informatics Association is the organisation focused on advancing health informatics (HI) practices and professionalism in Canada.	**The International Medical Informatics Association (IMIA)** is an independent organisation established under Swiss law in 1989. IMIA plays a major global role in the application of information science and technology in the fields of healthcare and research in medical, and health informatics.	**EU-US eHealth Collaboration Initiative (Workforce Development Group):** Memorandum of Understanding (MoU) was signed between the European Union and the United States (The Office of the National Coordinator for Health Information Technology (ONC). Two work groups were launched: 1. Interoperability. 2. Workforce Development.	**CHIA (Certified Health Informatician Australasia):** It was launched by the Health Informatics Society of Australia (HISA), the Australasian College of Health Informatics (ACHI), and the Health Information Management Association of Australia (HIMAA).	**Health Informatics Accreditation Council (HIAC):** AMIA joined the CAHIIM and established HIAC. It was charged with revising the existing CAHIIM Curriculum Requirements document and the "Accreditation Standards for Masters' Degree Programs in Health Informatics".	**Saudi Commission for Health Specialties (SCFHS)**: National Health Informatics Competency Working Group was established as part of the Saudi National Health Informatics Advisory Committee.
Project's foundation year	2007	2010	2010	2013	2015	2018
Project's main objective/s	To set out a common core or shared set of skills, knowledge, attitudes, and capabilities necessary to effectively perform as a Health Informatics Professional.	To provide a framework for individual curriculum development in the field of biomedical and health informatics (BMHI).	To define standards, develop competencies and produce useful tools that support this work.	To develop a new competency framework based on the existing frameworks but with a focus on the Australian healthcare system.	To build CAHIIM Accreditation Standards to reflect the emergent knowledge, skills, and attitudes reflecting the foundational domains set forth in the AMIA White Paper of 2012.	To develop a new competency framework but with a focus on the Saudi healthcare system.
Approach for competency development	- LR (Journal articles, textbooks and the Internet were reviewed). - Existing provincial, territorial, national and international competency frameworks pertaining to HI were reviewed. - Workshops were carried out to revise list of competencies obtained from LR. Participants were leaders in the field,	- Establishing a Working Group on Health and Medical Informatics Education - Reviewing the publications by organisations on the development of competencies - Revising the existing international recommendations in health informatics /medical informatics education.	Members met every week for 20 months: 1. Narrowing down the focus to one setting where roles could be mapped to and evaluated against Health IT competencies 2. Collating and mapping over 250 job roles and careers in the Acute Care setting in the EU and US 3. LR (Gathering competencies from 15+ organisations from both the EU and US)	Three organisations (AMIA, IMIA, COACH) were reviewed. Any repetitions, overlaps, and redundancies were removed, and the new competencies were restructured, ultimately producing the final competency framework.	Through an iterative process of review and revision by the AMIA Accreditation Committee (AAC) of organisations, papers, and recommendatio ns pertaining HI.	- LR (published and grey literature were reviewed). - Existing international competency frameworks pertaining to HI were reviewed. - Seventeen expert panel meetings - Online survey followed by a focus group (workshop) were carried out to revise

	representative of the HI profession from across Canada.		4. Organising, levelling and categorising the competencies. 5. Reviewing and reworking the competencies.			list of competencies obtained from LR. Participants were leaders in the field, representative of the HI profession from across Saudi.
Key Output	Health Informatics Professional Core Competencies Framework (version 3.0), released 2012. It includes - 3 domains, - 7 areas of competencies, and - 51 competencies.	A three-dimensional framework: - 1) professionals in healthcare, - 2) type of specialisation in BMHI, and - 3) stage of career progression.	HitComp tool (http://hitcomp.org), released 2015. It includes - 5 domains, - 33 areas of skills, and - 1000 competencies.	CHIA Health Informatics Competencies Framework (edition 1), released 2013. It includes 6 domains and 52 competencies	Health Informatics Competencies Framework, released 2016. it includes 10 domains and 22 competencies.	Saudi Health Informatics Competencies Framework (version 1.0), released 2018. It includes 6 domains, 22 subdomains or areas of competencies, and 92 competencies.
Visual illustration of competencies?	Yes	No	No	Yes	Yes	Yes
Framework Applications	- Establishing Professional certification based on the core HI competencies required in the Canadian health system. - Establishing a Career Matrix and the Role Profiles. - Setting the minimum requirements in terms of skills, knowledge, understanding and capabilities that will enable a candidate to perform in a professional environment. - Defining more clearly the body of knowledge underpinning this discipline.	Accreditation of BMHI educational programs. Curricula guidelines for academic institutions that define minimum-level competencies required for each level and knowledge/skill domain. Recommendations for Bachelor, Master and Doctoral Programs in Biomedical and Health Informatics In future, IMIA will also develop teaching credentialing criteria to serve as a guide for teachers wishing to participate in BMHI education.	- Hiring Manager – needs all candidates to have eHealth skills - Technology College – developing new eHealth certificate program in management, quality and improved outcomes - ICT/IT Specialist – wants to transition to eHealth	- Providing the context in which the questions for the exam have been developed. - Setting guidelines for recruiting purposes, definitions of career pathways, or the design of educational and training activities. - Setting the minimum requirements in terms of skills, knowledge, understandings and capabilities that will enable a candidate to perform in a professional environment. - Defining more clearly the body of knowledge underpinning this discipline.	- Curriculum development - Accreditation quality assessment for graduate (Master's level) education in applied health informatics. - Setting the minimum requirements in terms of skills, knowledge, understandings and capabilities that will enable a candidate to perform in a professional environment. - Defining more clearly the body of knowledge underpinning this discipline.	- Establishing a professional credential certification based on the core HI competencies required in the Saudi healthcare system. - Setting the minimum requirements in terms of skills, knowledge, understanding and capabilities that will enable a candidate to perform in a professional environment. - Defining more clearly the body of knowledge underpinning this discipline.

Results

Table 1 presents a high-level summary of the competency development projects in Europe, Canada, Australia, United States and Saudi Arabia. The following sub-sections provide a summary of the body/organisation initiative, foundation year of the project, the project's main objectives, approach for competency development, key outputs, and visual illustrations of the competency framework. It also illustrates challenges and issues relating to the development of the Saudi HI competency framework.

Foundation Year of the Projects

When compared to other countries, Saudi Arabia is the latest country to join in developing a HI competency framework for HI professionals. Canada was the first to begin this task in 2007, followed by Europe in 2010, Australia in 2013, and the United States in 2015. Saudi Arabia's delay in the development of a HI competency framework can be attributed to the novelty of HI education within the country, undefined roles of HI professionals working within the Kingdom, and recognition of HI as a health profession vs a computing science profession.

Project's Main Objectives

Digital Health Canada and CHIA aimed to develop new competency frameworks that set out a core list of knowledge and capabilities necessary to effectively perform as a HI professional. IMIA and HIAC however built accreditation standards and guidelines for curriculum development that reflect the key knowledge and skills in HI as a discipline. On the other hand, EU-US eHealth Collaboration Initiative established Workforce Development that aimed at defining a variety of healthcare roles and careers, levels and areas of knowledge.

In Saudi Arabia, at the beginning of the project, there was a debate among the panel's members on whether we should adapt an existing competency framework or build a new one. Decision was made to build our own framework for the following reasons:

- EU-US eHealth Collaboration Initiative project was not designed to establish a professional credential certification which was an essential aim for our work in Saudi. A total number of 1000 of competencies were identified which means that they did not set the minimum requirements in terms of skills and knowledge to perform in a professional environment.
- HI is a rapidly evolving field. Some competencies were outdated; e.g., IMIA's framework was developed from 2010.
- The Australasian framework was built with a focus on the Australian healthcare system.
- AMIA and CAHIIM competencies were designed for curriculum development, not for providing professional credential certification.

Approaches for Competency Development

The comparative analysis of competencies development projects showed that most of the identified leading HI professional organisations (e.g., COACH, IMIA, etc.) had different project objectives but similar approaches for competency development. For example, review of published and grey literature, review of existing national and international competency frameworks pertaining to HI, and workshops and/or panel meetings were carried out by all HI professional organisations around the globe.

The Saudi approach was different in a number of ways than other approaches (see Table 1) primarily due to the complexity, diversity, and novelty of health informatics education and professional practice within the country. Our approach focused on building consensus among the diverse academic and professional perspectives by generating a list of key professional and educational stakeholders and inviting them to participate in both an online survey and a focus group session to garner their feedback on the proposed Saudi HI competency framework.

The online survey asked the participants to rank the competencies in order of importance. Once the survey and analysis were complete, the stakeholders were invited to a focus group session to discuss the findings and garner consensus for selecting the key elements of the Saudi HI competency framework. The findings were presented to the group via a PowerPoint presentation and each element of the proposed Saudi HI framework was discussed until a consensus was reached either to include, exclude or postpone discussion at a later date for the relevant competency discussed.

Key Outputs

EU-US eHealth Collaboration Initiative developed HitCopm tool (http://hitcomp.org). HitComp is a searchable database designed for interested parties in healthcare information technology and eHealth such as educators, workforce developers, students, eHealth managers, etc. It defines 5 domains (Administration, Direct Patient Care, Engineering/Information Systems/ICT, Informatics, and Research/Biomedicine), 33 areas of skills, and 1000 competencies.

Digital Health Canada built the Health Informatics Professional Core Competencies Framework (version 3.0). It includes 3 domains (Management Science, Health Science, and Information Science), 7 areas of competencies, and 51 competencies. Similarly, HISA developed CHIA Health Informatics Competencies Framework (edition 1). It has 6 domains (Management Science, Health Science, Information Science, Information and Communication Technology ICT, Scientific skills of Health Informatics, Human and Social Context) and 52 competencies.

AMIA and CAHIIM established the Health Informatics Competencies Framework for curriculum development that includes 10 domains and 22 competencies. This framework has six major domains: Health, Information Science and Technology, and Social and Behavioural Science, Professionalism, Interprofessional Collaborative Practice, and Leadership. The first three domains intermingle and provide four co-mingled domains: Health Information Science and Technology, Human Factors and Sociotechnical Systems, and Social and Behavioural Aspects of Health. On the other hand, IMIA built a three-dimensional framework. These dimensions are: professionals in healthcare (e.g. physicians, nurses, BMHI professionals), type of specialisation in BMHI (IT users, BMHI specialists), and stage of career progression (bachelor, master, doctorate).

The Saudi Health Informatics Competency Working Group developed the Saudi Health Informatics Competencies Framework (version 1.0). It has 6 domains (Core Principles in HI, ICT, Health Sciences, Data Science, Education and Research, and Leadership and Management), 22 sub-domains or areas of competencies, and 92 competencies.

All of the above projects were established by independent organisations/bodies that aim to foster the HI profession. However, in Saudi Arabia, SCFHS provided support for the Saudi project although it is not devoted for the HI profession. SCFHS is a professional body that regulates health care-related practices and accreditation at all levels in Saudi Arabia. Lack of independent body for taking charge of the HI profession may obstacle the process of keeping the framework updated. To achieve this state of sustainable professional presence, HI workforce development play a significant role, which can be accomplished by an independent professional body to fulfil this

purpose, or SCFHS via establishing a specialised unit to foster the HI profession. A collaborative initiative by SCFHS and the National Health Information Centre may be needed to set together the future career needs to ultimately improve overall HI profession scope and definition in Saudi Arabia.

Visual Illustrations and Framework Applications

The projects of competencies development that were conducted by HI organisations had led to building a variety of frameworks that visually define a set of fundamental competencies for describing the HI profession. Few organisations did not provide a visual illustration of their competencies. For example, EU-US eHealth Collaboration Initiative developed HITComp tool instead, and IMIA presented their framework in the form of guidelines and recommendations to guide the development of biomedical and health informatics bachelor, master, and doctorate programs.

The applications of the developed competencies-based frameworks varied from one organisation to another. For example, Digital Health Canada and CHIA established professional certification based on their frameworks that define the core HI competencies required in the Canadian and Australian health systems respectively. IMIA and HIAC, on the other hand, applied their framework for curriculum development and quality assessment for undergraduate and graduate education in HI.

In Saudi Arabia, SCFHS will use SHICF to guide the development of a professional credential examination. The SCFHS currently interviews candidates without providing the interviewer or interviewee with specific areas of study, leading to a very subjective interview process. Using the Saudi competency framework would provide valuable guidance and needed consistency to the SCFHS evaluation team when interviewing and credentialing HI professionals. In future phases, a written examination is needed to be developed based on the framework. A team of HI experts from academia and industry should work on developing exam questions using the competency framework produced from this project.

SHICF will also be utilised to refine the Health Informatics academic programs of local educational institutions. Since this initial framework has been developed, the framework itself may not be reflective of what is being taught today in HI programs across the country. Therefore, a collaborative initiative by SCFHS and the Ministry of Education may be needed to build guidelines for current university programs to be refined based on the competency framework, and specifically, the competencies for each sub-domain. Meanwhile, the need for building guidelines based on the framework to design and/or update academic programs/degrees in HI is essential to maintain consistency between the proposed framework and the current educational programs and practises.

Discussion

A number of attempts have been made in the last ten years to define the HI profession through the development of an HI competency framework to guide both HI professional practice and curriculum development. Our results show consistency between our work and the work developed by the selected five organisations. Also, our initial work will provide valuable benefits to the SCFHS in supporting planning activities for health informatics professional credentialing within the Kingdom of Saudi Arabia.

The project that was carried out over the past eight months illustrates the difficulties involved in developing a local framework given the discord between current programs being delivered and international standards for HI professionals. The lack of independent body for taking charge of the HI profession in Saudi could be a big challenge that may obstacle the process of keeping the framework updated.

Conclusions

The Saudi Health Informatics Competency Framework (SHICF) can be utilised to define the field of health informatics within the country. SCFHS can use the framework to define the boundaries of the field of health informatics and distinguish it from areas such as Bioinformatics based on the competencies outlined in the framework. As a result, SHICF will provide a clearer definition and formal recognition of the health informatics filed as it is practiced within the country. In addition, the work conducted in our project could be a first step towards providing a fair, objective, and reliable assessment and guidance for HI professionals, employers, and academic institutions within the country.

The health informatics field will continue to evolve and the SCFHS should continually review the proposed competency framework in order to develop in future a context-based HI competency framework that would make HI professionals competitive, competent, and reliable in helping improve overall healthcare quality within the Kingdom of Saudi Arabia.

Acknowledgements

This work has been funded and produced by SCFHS and led by Dr. Amr Jamal and the Saudi Health Informatics Competencies Working Group.

References

[1] Digital Health Canada. Health Informatics Professional Core Competencies. Available from: https://digitalhealthcanada.com/wp-content/uploads/2017/03/Health-Informatics-Core-Competencies.pdf [accessed April 20, 2018]

[2] J. Mantas, E. Ammenwerth, G. Demiris, A. Hasman, R. Haux, W. Hersh, et al., Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics. First Revision, *Methods Inf Med* **49** (2010), 105-120.

[3] EU-US eHealth Collaboration Initiative. Available from: http://hitcomp.org [accessed April 20, 2018]

[4] F. Martin-Sanchez, D. Rowlands, L. Schaper, and D. Hansen, The Australian Health Informatics Competencies Framework and its role in the Certified Health Informatician Australasia (CHIA) Program, *Stud Health Technol Inform* **245** (2017), 783-787.

[5] AMIA Accreditation Committee. Available from: https://www.amia.org/sites/default/files/AMIA-Health-Informatics-Core-Competencies-for-CAHIIM.PDF [accessed April 20, 2018]

Address for Correspondence

Dr. Manal Almalki. The Saudi Health Informatics Competencies Working Group, the National Advisory Committee for Health Informatics. Department of Health Informatics, Jazan University, Jazan, Jazan Province, Saudi Arabia. Email: Almalki@jazanu.edu.sa

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190397

Effects of Adult Patient Portals on Patient Empowerment and Health-Related Outcomes: A Systematic Review

Elske Ammenwerth[a], Alexander Hoerbst[b], Stefanie Lannig[c], Gerhard Mueller[c], Uwe Siebert[d], Petra Schnell-Inderst[d]

[a] *Institute of Medical Informatics, UMIT – University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria*

[b] *Research Group eHealth and Innovation, UMIT, Hall in Tirol, Austria*

[c] *Institute of Nursing Science, UMIT, Hall in Tirol, Austria*

[d] *Institute of Public Health, Medical Decision Making and Health Technology Assessment, UMIT, Hall in Tirol, Austria*

Abstract

Patient portals are offered by health care organizations to facilitate health information sharing and patient empowerment and support patient-centered care. The aim of this systematic review is to assess the effect of patient portals on patient empowerment and health-related outcomes. After a systematic literature search, ten randomized controlled trials (RCTs) were included in this review. Of these, seven RCTs were conducted in the United States., two in Canada, and one in Japan. Study characteristics, risk of bias, and outcomes were extracted. varied in terms of intervention, included patients, and outcome.. Most studies found no or only a small, clinically non-relevant effect of patient portals. The review showed that future research should develop a taxonomy to describe patient portal functionalities to facilitate the aggregation of evidence.

Keywords:

Electronic Health Records, Review, Patient Participation

Introduction

Patient-centered care has gained importance in both medical research and clinical practice. The concept of patient-centeredness is based on patient empowerment, patient participation, and shared decision-making [1–3]. Health information sharing and patient engagement in health care decisions are seen as preconditions for patient-centered care [1].

To facilitate health information sharing and patient involvement in the care process, healthcare organizations are increasingly offering their patients' access to their health data in the institution-based electronic health record (EHR). Patients can access these data and integrate it into any (electronic or paper-based) type of personal health record [4].

The interface that provides EHR access is called patient portal [5]. These portals are typically web-based, allowing patients independent access to their data from anywhere as a primary feature. A patient portal may also offer additional features such as medication refill requests, appointment scheduling, secure messaging, personal health-related reminders, individual therapeutic recommendations, personal diaries, and social networking with other patients.

In addition to being offered by healthcare organizations, EHR access may also be offered to patients on a national level. Some countries, including Austria, Denmark, and Sweden, have already started eHealth projects to make selected health-related data from various healthcare organizations available to their citizens [6].

A uniform theory or clear evidence of how EHR access via patient portals might contribute to patient-centered care and related concepts such as patient empowerment or patient participation or even improved health outcomes, does not exist. Nevertheless, qualitative reviews have shown that patient portals may improve patient empowerment, patient adherence, and clinical outcomes [7,8]. However, systematic reviews on patient portals have found inconclusive results to date [8–12]. As all of these reviews were published before 2015, it is possible that more evidence is now available. In this systematic Cochrane review, we assess the effects of providing access to EHR for adult patients on patient empowerment and health-related outcomes. We summarize characteristics of the identified Randomized Controlled Trials (RCTs) and present preliminary results on the effect of providing access to EHRs for adult patients on patient empowerment and health-related outcomes.

Methods

We used a Cochrane protocol for conducting this review [13]. We systematically reviewed RCTs investigating the effects of providing EHR access to adult patients. First, we developed a patient portal taxonomy describing seven functionalities:

- *Access*: Access to health-related data (e.g., visit notes, test results, medical history).
- *Remind*: Personalized health care reminders (e.g., for mammography or immunization).
- *Request*: Transactional services (e.g., scheduling appointments, prescription request).
- *Communicate*: Bilateral communication (e.g., secure messaging).
- *Share*: Patient self-documentation and sharing (e.g., patient uploads of blood pressure measurements).

- *Manage*: Disease management (e.g., individualized recommendations from guidelines).
- *Educate*: General health-related education (e.g., disease information leaflets).

Primary outcomes for this Cochrane review are the effect of EHR access on patient knowledge and understanding, patient empowerment, patient adherence, patient satisfaction, and adverse events.

Secondary outcomes for this review are health-related outcomes including quality of life, psychosocial health outcomes, health resource consumption, and patient-provider communication.

In our review, all studies offering EHR access to adult patients were included, independent of the medical condition of the patients. Only studies where EHR access was provided via a web-based application were included, thus excluding office-based systems. Only RCTs were included.

We systematically searched electronic libraries including Central, Medline, Embase, PsycInfo, Scopus, and CINAHL; in proceedings of Medinfo, AMIA, and MIE; and in major health informatics journals. We also searched for studies cited in earlier systematic reviews on patient portals and in identified RCTs. Two authors independently screened all titles and abstracts to determine whether they meet the inclusion criteria.

The following information were extracted by two authors using the software Covidence (Version v1062 115d548c):

- Study identification (e.g., country, clinical setting).
- Study methods (e.g., aim of study, intervention and control group, number of arms, study design, funding source).
- Population (e.g., target group, inclusion and exclusion criteria, age, gender).
- Intervention (e.g., name, functionality, usage patterns).
- Risk of bias (e.g., random sequence, allocation, blinding, selective outcome reporting) [14].
- Outcomes (e.g., methods and timing of assessing outcomes, instruments used, methods for follow-up, effect size, mean change in intervention and control group, adverse events, measure of uncertainty).

Primary and secondary outcome results were extracted in systematic evidence tables. More details of data extraction are published in a Cochrane protocol [13].

Results

We identified ten studies that represent distinct RCTs on EHR access for adult patients (seven RCTs and three cluster RCTs). Figure 1 shows the flow chart of this systematic review. The studies included between 78 and 4,500 patients.

Identified studies were very heterogeneous in terms of included patients' diseases (hypertension, diabetes, asthma, glaucoma, congestive heart failure or unspecific) and outcome measures (e.g., frequently patient adherence and health-related outcomes, less often patient satisfaction and patient empowerment), making meta-analysis challenging.

Seven RCTs were conducted in the United States, two in Canada, and one in Japan (Table 1). The majority of studies were published in 2012 or earlier, indicating a scarcity of newer studies.

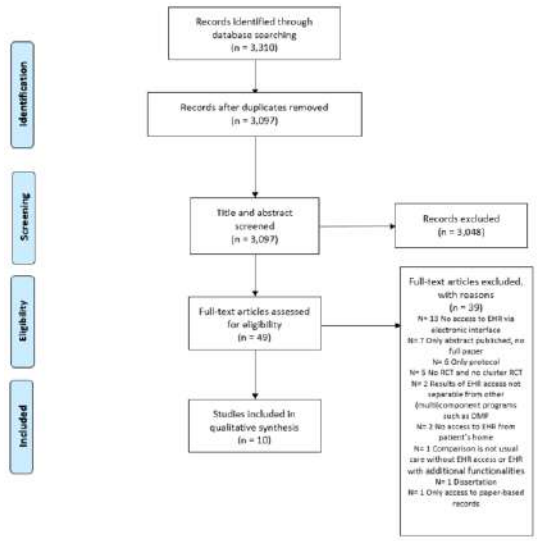


Figure 1 – Flow chart of the systematic review on patient portals.

Table 1 –Identified randomized studies. H-r = health-related outcomes.

Study	Country	Included patients	Outcome
Ahmed (2016) [15]	Canada	Asthma	H-r Outcome
Grant (2008) [16]	United States	Diabetes mellitus	H-r Outcome
Holbrook (2009) [17]	Canada	Diabetes mellitus	Adherence H-r Outcome
Kashiwagi (2014) [18]	Japan	Glaucoma	H-r Outcome
Krist (2012) [19]	United States	Unspecific	Adherence
McCarrier (2009) [20]	United States	Diabetes mellitus	Empowerment H-r Outcome
Ralston (2009) [20]	United States	Diabetes mellitus	H-r Outcome Health Resources
Ross (2004) [21]	United States	Congestive heart failure	Empowerment Adherence Satisfaction

			Adverse Events
			Health Resources
Tang (2013) [22]	United States	Diabetes mellitus	Knowledge
			Adherence
			Satisfaction
			Adverse Events
			H-r Outcome
			Health Recources
Wagner (2012) [23]	United States	Hyper-tension	Empowerment
			Patient Satisfaction
			H-r Outcome

Table 2 – Functionality supported by the patient portals

	Access	Communicate	Share	Manage	Educate	Remind	Request
Ahmed 2016	■	■	■	■	■		
Grant 2008	■		■	■			
Holbrook 2009	■		■	■	■		
Kashiwagi 2014	■						
Krist 2012	■		■		■	■	
McCarrier 2009	■	■	■	■	■		
Ralston 2009	■	■	■	■	■		
Ross 2004	■	■			■		
Tang 2013	■	■	■	■			
Wagner 2012	■	■	■	■	■		

Table 2 shows functionality supported by patient portals, based on our taxonomy. All included studies offer "access to data", as this was an inclusion criteria. No portal offered the functionality "request".

There was substantial heterogeneity across studies regarding instruments used and study outcomes. For example, the three studies measuring patient empowerment used four different questionnaires. For two of these instruments, there was no statistically significant difference in patient empowerment. For the other two instruments, the effect was statistically significant, but the effect size was too small to be clinically relevant.

Eight studies measured changes in health-related outcomes. These include a outcomes such as asthma control and mortality, but mostly risk factors such as hemoglobin A1c (HbA1c), blood pressure, low-density lipoprotein and body mass index. Of the six studies measuring the effect on HbA1c, two found a statistically significant, yet small improvement. Of the four studies measuring the effect on blood pressure, only one found a statistically significant, small improvement.

When looking at usage patterns, the functionality offered by the patient portal was often not used consistently. For example, in one study, the number of logins declined over time [21]. In another study, less than 25% of patients used the portal consistently [20]. In one RCT, 16% of patients never logged in over the three-month study period [15]. Users of the patient portal were more often male, white, commercially insured, and college-educated [16,19].

Sub-group analysis of the intervention group revealed that, in this group, portal users show better outcome than portal non-users in three studies [19,20,22].

Discussion

We identified ten randomized controlled studies that evaluated the effect of patient portals on a range of outcomes. In summary, most studies found no evidence for an effect or only a small, clinically non-relevant effect.

Two studies reported no differences on mortality [22,24], and none of the included RCTs reported other adverse effects of patient portals. In general, EHR access may increase feelings of confusion and anxiety when patients read clinical information that is unclear to them [8], but this effect was not reported in the included studies.

EHR access, like many other digital health solutions, is sometimes said to be created for "people like me", meaning that these digital solutions may only address the needs of "[…] well-educated and well-to-do users rather than the needs of the most disadvantaged in society (the disempowered, disengaged, and disconnected" [25]. Two studies [16,19] found indeed that active portal users were more often white, male, and college-educated.

Among other factors, health literacy may help to reliably interpret content provided in the EHR. The level of health literacy may influence the frequency of use and the potential benefits from accessing EHR. data Besides disease-specific knowledge in one study [22], health literacy was, however, not analyzed in any study.

Three studies found that active users showed better outcomes compared to non-users in the intervention group. The concept of "implementation fidelity" [26] refers to the degree in which an implementation is delivered and used as intended. Low frequency and duration of portal usage may show low fidelity and may explain lack of visible effect. Another likely explanation for the better effects in users may be patient characteristics (e.g., education status) that may be associated with both higher portal use and better health outcome; therefore, the identified difference may be an overestimation of the true effect.

Nine of the ten reviewed RCTs included patients with various chronic diseases (Table 1). While earlier research assumed that chronic patients may benefit from patient portals [27], we did not find clinically relevant effects for this group. We thus do not expect to see effects from portals in other user groups.

Due to the small number of studies, further sub-group analysis such as patient group, portal functionality, or fidelity of implementation are not possible.

We were not certain that we have identified all RCTs on patient portals, due to various terms describing the intervention or the observed effects. So, we also searched in major health informatics proceedings, references of other portal reviews, and references of the identified RCTs.

Conclusions

Preliminary results of this systematic review of RCTs of patient portals did not reveal clear evidence of substantial and consistent positive effects of patient portals on patient empowerment and health-related outcomes.

The number of identified studies, however, is small, quite diverse, and many were published in 2012 or before . Also, in several studies, a part of the intervention group did not use the patient portal consistently.

Future research should develop a taxonomy to describe patient portal functionalities to facilitate the aggregation of evidence in future systematic reviews.

Acknowledgements

The work was partly funded by TWF – Tyrolean Science Foundation.

This abstract is based on a draft pre-peer review version of a Cochrane Review. Upon completion and approval, the final version is expected to be published in the Cochrane Database of Systematic Reviews (www.cochranelibrary.com).

References

[1] I. Scholl, J. Zill, M. Härter, and J. Dirmaier, An integrative model of patient-centeredness - a systematic review and concept analysis, *PLoS One*. **9** (2014) 9(9):e107828.

[2] E. Castro, K. Vanhaecht, W. Sermeus, A. Van Hecke, and T. Van Regenmortel, Patient empowerment, patient participation and patient-centeredness in hospital care: a concept analysis based on a literature review, *Patient Educ. Couns.* **99** (2016) 1923–1939.

[3] T. Irizarry, A. DeVito Dabbs, and C.R. Curran, Patient Portals and Patient Engagement: A State of the Science Review, *J. Med. Internet Res.* **17** (2015) e148. doi:10.2196/jmir.4255.

[4] E. Ammenwerth, P. Schnell-Inderst, C. Machan, and U. Siebert, The Effect of Electronic Prescribing on Medication Errors and Adverse Drug Events: A Systematic Review, *J Am Med Inf. Assoc*. **15** (2008) 585–600.

[5] F. Magrabi, M.S. Ong, W. Runciman, and E. Coiera, Patient safety problems associated with heathcare information technology: an analysis of adverse events reported to the US Food and Drug Administration, *AMIA Annu Symp Proc*. **2011** (2011) 853–857.

[6] A. Moen, W.O. Hackl, J. Hofdijk, L. Van Gemert-Pijnen, E. Ammenwerth, P. Nykänen, and A. Hoerbst, eHealth in Europe - Status and Challenges, *Yearb. Med. Inform.* **8** (2013) 59–63.

[7] T. Otte-Trojel, A. de Bont, T.G. Rundall, and J. van de Klundert, How outcomes are achieved through patient portals: a realist review, *J Am Med Inf. Assoc*. **21** (2014) 751–757.

[8] F. Mold, S. de Lusignan, A. Sheikh, A. Majeed, J.C. Wyatt, T. Quinn, M. Cavill, C. Franco, U. Chauhan, H. Blakey, N. Kataria, T.N. Arvanitis, and B. Ellis, Patients' online access to their electronic health records and linked online services: a systematic review in primary care., *Br. J. Gen. Pract.* **65** (2015) e141-51. http://www.ncbi.nlm.nih.gov/pubmed/25733435.

[9] C.S. Kruse, K. Bolton, and G. Freriks, The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review, *J Med Internet Res*. **17** (2015) e44.

[10] E. Ammenwerth, P. Schnell-Inderst, and A. Hoerbst, The impact of electronic patient portals on patient care: A systematic review of controlled trials, *J. Med. Internet Res.* **14** (2012) e162.

[11] T. Davis Giardina, S. Menon, D. Parrish, D. Sittig, and H. Singh, Patient access to medical records and healthcare outcomes: a systematic review, *J. Am. Med. Informatics Assoc.* **21** (2014) 737–741.

[12] C. Goldzweig, G. Orshansky, N. Paige, A. Towfigh, D. Haggstrom, I. Miake-Lye, J. Beroes, and P. Shekelle, Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review, *Ann. Intern. Med.* **159** (2013) 677–678.

[13] E. Ammenwerth, S. Lannig, A. Hörbst, G. Mueller, and P. Schnell-Inderst, Adult patient access to electronic health records (Protocol), *Cochrane Database Syst. Rev.* **6** (2017) CD01270.

[14] J.P.T. Higgins, and S. Green, Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011], (2011). http://www.cochrane-handbook.org.

[15] S. Ahmed, P. Ernst, S. Bartlett, M. Valois, T. Zaihra, G. Paré, R. Grad, O. Eilyyan, R. Perreault, and R. Tamblyn, The Effectiveness of Web-Based Asthma Self-Management System, My Asthma Portal (MAP): A Pilot Randomized Controlled Trial, *J Med Internet Res*. **18** (2016) e313.

[16] R.W. Grant, J.S. Wald, J.L. Schnipper, T.K. Gandhi, E.G. Poon, E.J. Orav, D.H. Williams, L.A. Volk, and B. Middleton, Practice-linked online personal health records for type 2 diabetes mellitus: a randomized controlled trial, *Arch Intern Med*. **168** (2008) 1776–1782.

[17] A. Holbrook, L. Thabane, K. Keshjavee, L. Dolovich, B. Bernstein, D. Chan, S. Troyan, G. Foster, and H. Gerstein, Individualized electronic decision support and reminders to improve diabetes care in the community: COMPETE II randomized trial, *CMAJ*. **181** (2009) 37–44.

[18] K. Kashiwagi, and S. Tsukahara, Impact of Patient Access to Internet Health Records on Glaucoma Medication: Randomized Controlled Trial, *J Med Internet Res.* **16** (2014) e15. doi:10.2196/jmir.2795.

[19] A. Krist, S. Woolf, S. Rothemich, R. Johnson, J. Peele, T. Cunningham, D. Longo, G. Bello, and G. Matzke, Interactive Preventive Health Record to Enhance Delivery of Recommended Care: A Randomized Trial, *Ann Fam Med*. **10** (2012) 312–319.

[20] K. McCarrier, J. Ralston, I. Hirsch, G. Lewis, D. Martin, F. Zimmermann, and H. Goldberg, Web-Based Collaborative Care for Type 1 Diabetes: A Pilot Randomized Trial, *Diabetes Technol Ther*. **11** (2009) 211–7.

[21] S.E. Ross, L.A. Moore, M.A. Earnest, L. Wittevrongel, and C.T. Lin, Providing a web-based online medical record with electronic communication capabilities to patients with congestive heart failure: randomized trial, *J Med Internet Res*. **6** (2004) e12. doi:10.2196/jmir.6.2.e12.

[22] P.C. Tang, J.M. Overhage, A.S. Chan, N.L. Brown, B. Aghighi, M.P. Entwistle, S.L. Hui, S.M. Hyde, H. Klieman, C.J. Mitchell, A.J. Perkins, L.S. Qureshi, T.A. Waltimyer, L.J. Winters, and C.Y. Young, Online disease management of diabetes : Engaging and Motivating Patients Online With Enhanced Resources-Diabetes (EMPOWER-D), a randomized controlled trial, *J Am Med Inf. Assoc*. **20** (2013) 526–34.

[23] P.J. Wagner, J. Dias, S. Howard, K.W. Kintziger, M.F. Hudson, Y. Seol, and P. Sodomka, Personal health records and hypertension control : a randomized trial, *J Am Med Inf. Assoc*. **19** (2012) 626–34. doi:10.1136/amiajnl-2011-000349.

[24] S.E. Ross, L.A. Moore, M.A. Earnest, L. Wittevrongel, C. Lin, and S.E. Ross, Providing a Web-based Online Medical Record with Electronic Communication Capabilities to Patients With Congestive Heart Failure: Randomized Trial Corresponding Author :, *J Med Internet Res*. **6** (2004) 1–12. doi:10.2196/jmir.6.2.e12.

[25] C. Showell, and C. Turner, Personal health records are designed for people like us, *Stud. Health Technol.*

Inform. **192** (2013) 1037.

[26] C. Carroll, M. Patterson, S. Wood, A. Booth, J. Rick, and S. Balain, A conceptual framework for implementation fidelity, *Implement. Sci.* **2** (2007) 1–9. doi:10.1186/1748-5908-2-40.

[27] L. Tieu, U. Sarkar, D. Schillinger, J.D. Ralston, R. Pasick, C.R. Lyles, and L. Tieu, Barriers and Facilitators to Online Portal Use Among Patients and Caregivers in a Safety Net Health Care System: A Qualitative Study, *J Med Internet Res*. **17** (2015) e275. doi:10.2196/jmir.4847.

Address for correspondence

Elske Ammenwerth, UMIT, Institute of Medical Informatics, Eduard Wallnöfer Zentrum 1, 6060 Hall in Tirol, Austria, elske.ammenwerth@umit.at.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190398

KOTOBAKARI Study: Using Natural Language Processing of Patient Short Narratives to Detect Cancer Related Cognitive Impairment

Eiji ARAMAKI [a], Mai MIYABE[b], Chihiro HONDA[a], Seiko ISOZAKI[a], Shoko WAKAMIYA[a], Akira SATO[c], Isao MIYASHIRO[c]

[a] *Nara Institute of Science and Technology (NAIST), Nara, Japan*
[b] *Suwa University of Science, Nagano, Japan*
[c] *Osaka International Cancer Institute, Osaka, Japan*

Abstract

[Background] Recent reports of some studies have described that the cognitive function of cancer patients often declines by a phenomenon designated as cancer related cognitive impairment (CRCI). For patients' decision-making, detecting CRCI is important. To do so, this study uses language-based CRCI screening to examine participants' language ability. [Objective] This study was conducted to ascertain whether a Natural Language Processing (NLP) based system can detect CRCI, or not. [Materials and Methods] We obtained materials of two types from cancer patients (n=116): (1) speech samples on three topics, and (2) cognitive function level test scores from Hasegawa's Dementia Scale – Revised (HDS-R), a test used in Japan for dementia patients. The test is similar to the Mini-Mental State Examination. [Results and Discussion] Cancer patients with lower HDS-R scores showed a significantly lower Type Token Ratio (TTR). [Conclusion] This result demonstrates the feasibility of the proposed speech–language-based CRCI screening method.

Keywords:
Cancer, Cognitive Dysfunction, Natural Language Processing

Introduction

Rapidly advancing medical technology brings huge numbers of treatment options for patients. Consequently, it is important to choose among various available options to find a suitable treatment. This decision-making process requires careful and thoughtful consideration. Cognitive impairment of patients is an important difficulty that disturbs patient decision making. Recent reports of some studies have described that the cognitive function of cancer patients often declines by a phenomenon designated as cancer related cognitive impairment (CRCI). Although as many as 75% of cancer patients have CRCI, little is known about its mechanisms [1]. Consequently, a large and precise CRCI study must be done.

To date, CRCI studies have adopted various approaches: biomedical studies have investigated genetic factors; neuropsychological examinations have used MRI images; and self-report questionnaire-based surveys have also been used. Although various approaches have presented some results, most have been preliminary studies of small numbers of patients. A difficulty of CRCI studies is obtaining a large number of patients, which in fact motivates the development of a method of rapidly investigating CI. Dementia is diagnosed when cognitive impairment has become sufficiently severe to reduce social functioning. Mild CI (MCI) is intermediate between normal cognition and dementia.

From the perspective of CI (or MCI) measurement, Alzheimer's disease has spurred development of an MCI measurement system. A rapid MCI measurement system is speech-based. A salient benefit of speech-based approaches is that they require only a patient speech sample [2–5] with no intervention. From speech samples, the system estimates the patient vocabulary, which is closely related with cognitive functions.

This study's method was designed to infer CRCI quickly from a speech sample. This study was conducted to evaluate CRCI from a cancer patient's speech narrative. To do so, we recruited cancer patients, and obtained both narrative and the cognitive test results.

The research questions (RQ) of this study are summarized as presented below:

- RQ1: to investigate the ratio of Japanese cancer patients with CRCI.
- RQ2: to investigate whether a system detect the CRCI only from patient speech.

Materials

Participants: Cancer patients have been recruited from the Osaka International Cancer Institute since July 11, 2018. Patient recruiting continues to the present day. In this recruiting, 171 patients in all used the system. Among them, 116 patients finished all tasks (55 (=171-116) were dropout). Participants included 53 men and 61 women, with 2 of unknown gender. Their average age was 64 years old (SD =15.2); the median was 67 years old (min=19, max=90). All participants were Japanese and fluent Japanese language speakers. All were present cancer patients or had a past history of cancer.

Methods

The research design: **Prospective cohort study**.

For this study, we developed a system for cancer patients (Fig. 1). We designate this system "KOTOBAKARI", which means language (/kotoba/) measurement (/bakari/) in Japanese.

System configuration: "KOTOBAKARI" consists of a touch panel computer (Windows 10) and a microphone for speech sampling.

Procedure: The measurement procedure comprises three steps. The measurement time is about 10 min in our preliminary experiments conducted with healthy users.

1. **Login:** a patient enters the booth and touches the screen. The patient then swipes the hospital ID cards to log-in.
2. **HDS-R** [6]. The HDS-R is a short cognitive test that is popular in Japan and which is similar to the Mini Mental State Examination (MMSE).
3. **Narrative Sampling**: The three questions below are given.

> *(1) What is your most joyful event that you experienced in recent days?*
> *(2) How do you come to the hospital?*
> *(3) How did you play during your childhood?*

A patient responds to each question. The system also displays "your talk should be about 3–4 min."

If a patient talks longer than 5 min, then the system displays "the amount of data is sufficient."

For 50% of participants, step 2 and the step 3 were inversed (Step 1 – Step 3 – Step 2) to mitigate order bias.

Inclusion and Exclusion Criteria

Inclusion criteria: cancer patient (any kind of cancer, any stage, any treatment is included).

Measurement Method

The measurement metrics, four NLP metrics, are presented in Table 1.

Table 1– Speech Indicators

Metrics	Description
Type token ratio (TTR)	Vocabulary size
Educational level (EL)	Vocabulary level
Word Commonness index (WCI)	Word commonness
Length (LEN)	Speech length

- **Type Token Ratio (TTR)**: TTR represents the vocabulary size, as derived from the ratio of word type to token (i.e., type/token, where type refers to the number of the use of different words in a text and token refers to the number of words in a text). We have disambiguated orthographic variants using a Japanese morphological analysis system. For example, "*itta* (went; past tense)" was normalized into "*iku* (go; present tense)" using the Japanese morphological analysis system, or JUMAN [7]. This study assessed both the word function and content.
- **Educational Lexicon Level (EL)**: EL represents the average difficulty in lexical choice. This metric was created originally for non-native speakers of Japanese. We used the word scores provided in the Japanese Learner's Dictionary [8]. In this dictionary, the most common 17,928 words are classified into the following three levels: beginner, intermediate, and advanced. To derive the score, we extracted and summed the number of nouns belonging to the intermediate or advanced levels. Then we divided this figure by the sum of intermediate and beginner level nouns. Details are presented in an earlier report [9].

Word Commonness Index (WCI): WCI represents the average use of common words. FPU represents the degree of commonness of words. The manner of calculating the commonness of words is proposed in [10]. We use a dictionary constructed in the work. This measure differed from the frequency of the occurrence of the word because such frequency deals with the number of the usage of the word no matter who the user might be, whereas WCI deals with the number of people who use the word.

- **LEN**: Number of characters of speech.
- **TIME**: Time of speech.

These metrics are calculated automatically from transcripts.

Measuring Cognitive Function Levels: The HDS-R, a screening test used in Japan for dementia patients, is similar to the Mini-Mental State Examination: the tests show a high degree of mutual correlation [6]. HDS-R scores that are 20 or lower (from a possible maximum score of 30) are regarded as indicative of dementia (sensitivity 0.90, specificity 0.82). Normal cognitive function is indicated by an HDS-R score of 27 or higher.

We implemented this HDS-R test into the system. Although most sub scores were judged automatically, several sub-scored items required human judgment, such as "list up as many vegetable names as possible." A clinical psychologist judged such questions.

Ethics Statement: This study protocol was approved by the Medicine Ethics Committee of the Osaka International Institute, 2018.

Figure 1 – "KOTOBAKARI" system. A speech booth in a cancer center (top left and top right) consists of a PC and mouse. Patients can freely enter the booth and use the system (bottom left and bottom right).

Results and Discussion

Results are two-fold, corresponding to two research questions.

RQ1: Does the cancer patient exhibit cognitive impairment, or not?

To answer this question, we investigated the cognitive scores as shown in Figure 3. As described, the cutoff score between healthy and mild cognitive impairment (MCI) is 20 (20 and more indicates healthy; less indicates MCI). The result indicates that not a few patients (n=30) show the MCI level score.

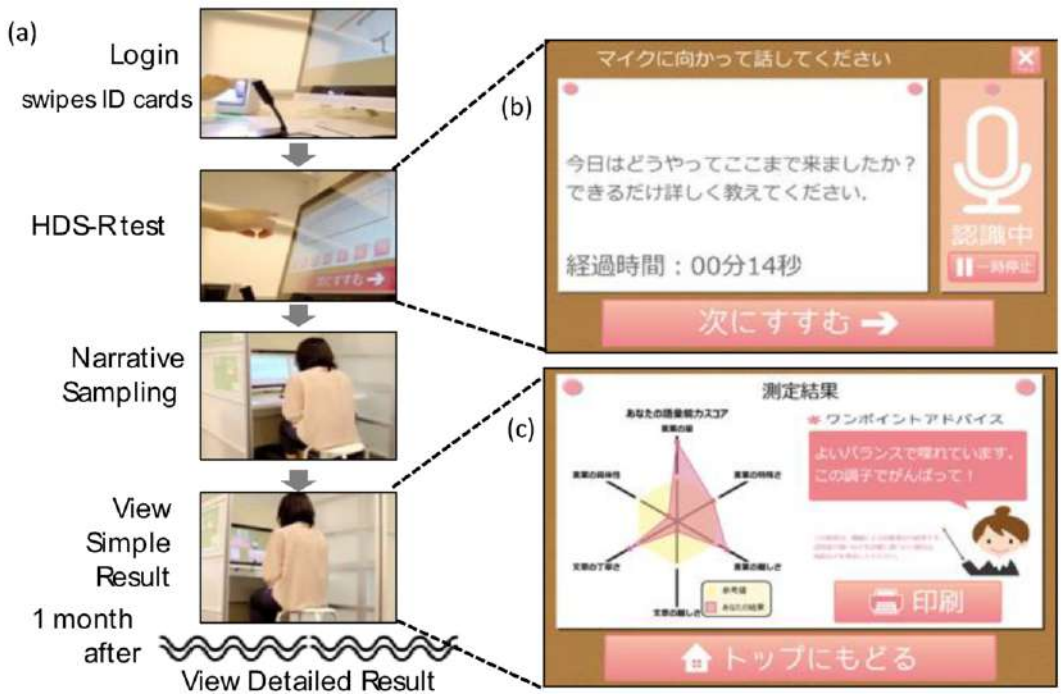


Figure 2 – "KOTOBAKARI" system screenshots. A patient can take tests of two types: (1) cognitive tests (HDS-R) and (2) Speech Tests (proposed method). Speech tests consist of three questions such as "(1) what is the most joyful event that you experienced in recent days?" A patient answers the questions by speaking. After the test, the speech is recognized and the result summary is displayed. This result is a rapid version based on automatic speech recognition. The final result is presented after judgment by a clinical psychologist.

Many factors might bias the results, such as the recruiting method and participant effects. The largest bias might be the patient age. Figure 4 depicts the distribution of HDS-R scores and ages of participants. Results show that age is highly correlated with HDS-R scores. Future studies should pursue more precise analyses that remove these biases. Although various biases exist, cancer patients tend to have cognitive impairment for some reason.

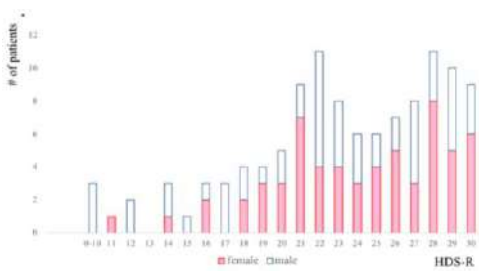


Figure 3 – HDS-R score distribution.

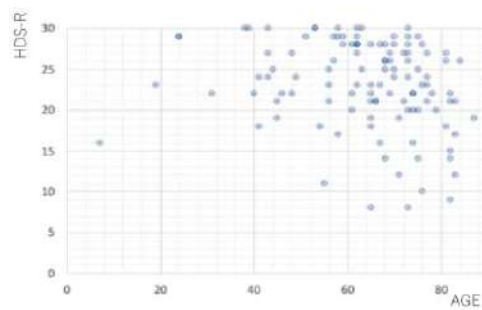


Figure 4 – Age and HDS-R score.

RQ2: Do cancer patient narratives suggest cognitive impairment, or not?

To answer this question, the relation between language abilities and HDS-R scores is shown (Fig. 5). To compare different scale language scores for (1) TTR, (2) EL, (3) WCI, and (4) LEN, scores are shown in order (higher to lower). The X-axis shows the order rank (TTR 10% means the top 10% rank in TTR). The Y-axis shows the HDS-R score.

The correlation ratio between language abilities and HDS-R is TTR (r=-0.34), EL (r=0.15), WCI(r=-0.05), and LEN (r=0.27). Among the four language abilities, TTR shows the highest correlation to the HDS-R score.

Next, we investigated the TTR screening performance. We divided the patients into two groups (higher and lower cognitive

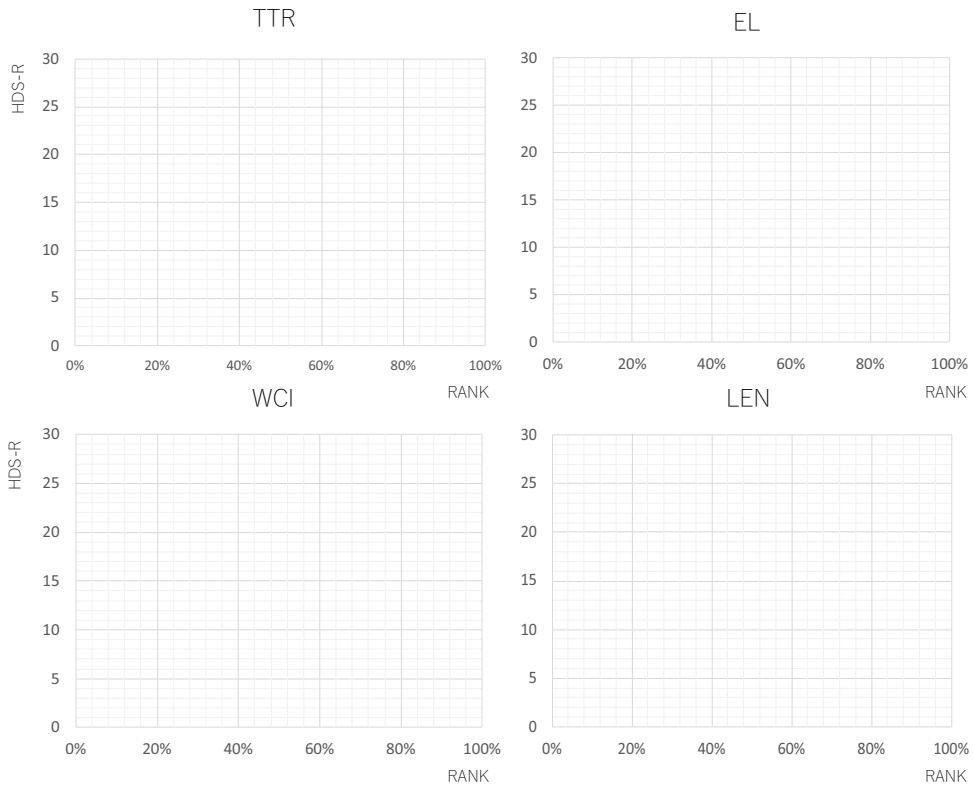


Figure 5 – Language abilities and HDS-R score. X-axis shows the language abilities (ordered by the rank). Y-axis shows the HDS-R score.

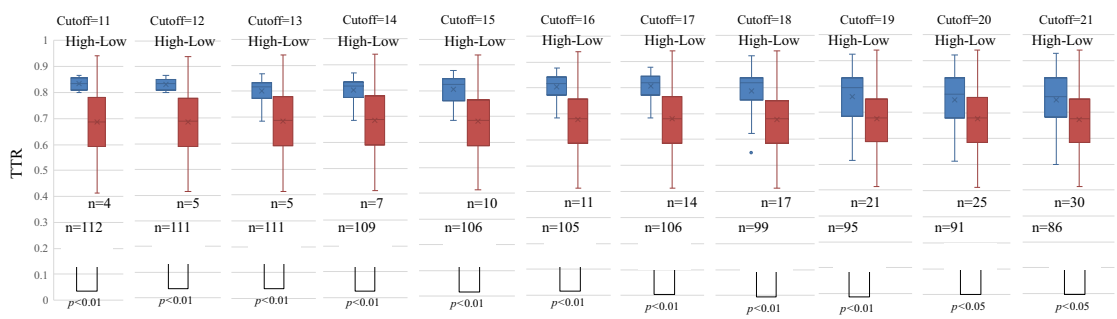


Figure 6 –HDS-R score and Type Taken Ratio (TTR; TTR of higher HDS-R group and lower group). We divided patients into two groups (higher HDS-R group and lower HDS-R group) using various HDS-R cutoff scores. At all cutoff points (from HDS-R 11–21), both groups showed significantly different TTR scores. Especially, HDS-R lower than 19 is significantly different ($p<0.01$), suggesting that serious CRCI is readily distinguished by TTR.

groups) using various HDS-R cutoff scores. The TTR performance is presented in Figure 6. In cutoff scores (HDS-R of 20–21), two groups showed significant TTR difference. Furthermore, in cutoff scores of HDS-R from 11 to 19, many differences were observed.

Results indicate two findings. (1) TTR-based screening of cognitive impairment is feasible. (2) Patients with a low HDS-R score are especially easy to detect.

Discussion

This study assessed the basic feasibility of speech-based cancer related cognitive impairment (CRCI) screening. To date, many studies examining Alzheimer disease patients have shows lower language ability, especially in the Type Token Ratio (TTR) [9, 11–13]. This study first demonstrated that not only Alzheimer's disease, but also CRCI shows lower language abilities. Results also reveal that TTR is the highest performance language indicator. Considering TTR is based on simple metrics (number of type and number of token), TTR-based screening has high potential for practical use.

Nevertheless, the relation between the TTR and diseases affecting cognition such as CRCI and Alzheimer disease remain unknown. Alzheimer disease examinations are classifiable into three types: (1) examinations evaluating cognitive function mainly based on memory, (2) examinations evaluating behavior disorders and mental symptoms, and (3) examinations assessing daily living behavior. Actually, TTR is related with the three functions above. Future studies must investigate the mechanisms of CRCI and language ability.

Limitations

This study has limitations. An important limitation is that the speech recognition accuracy is insufficient. This experiment relied on human transcription: a time-consuming and expensive task. Future clinical trial phases must resolve this problem and either find or produce a fully automated system.

Another limitation is that of screening performance. Results show that group-based differences in TTR (higher cognitive function group and lower cognitive function group showed the significant difference in TTR). However, the individual differences are smaller than group differences: patients sometimes show high TTR but lower HDS-R, or *vice versa*. The speech is highly dependent on the topic and personality, causing the instability. To address this shortcoming, one approach is to control the topic and personality. Moreover, a long time series analysis is highly desired.

Conclusions

Recent reports of some studies have described that the cognitive function of cancer patients is often declining. This study investigated this phenomenon, designated as Cancer Related Cognitive Impairment (CRCI). Using our proposed system, KOTOBAKARI, we collected cancer patients' cognitive function scores and speech samples. Among cancer patients (n=116), 30 showed mild cognitive impairment. Additionally, experiment results revealed that CRCI patients have a significantly lower type token ratio (TTR). This result demonstrated the feasibility of speech-based CRCI screening.

Acknowledgments

This work was supported by JST CREST (JPMJCR16E2), and JSPS (JP16H06395, JP16H06399, JP16K12489) Japan.

References

[1] Janelsins, M.C. et al., Prevalence, mechanisms, and management of cancer-related cognitive impairment. *Int Rev Psychiatry* 2014; **26**(1): 102-13.

[2] Kemper, S. and A. Sumner, The structure of verbal abilities in young and older adults. *Psychol Aging* 2001; **16**(2): 312-22.

[3] Kemper, S., Metalinguistic judgments in normal aging and Alzheimer's disease. *J Gerontol B Psychol Sci Soc Sci* 1997; **52**(3): 147-55.

[4] Orimaye, S.O., J.S.-M. Wong, and K.J. Golden. Learning Predictive Linguistic Features for Alzheimer's disease and related Dementias using Verbal Utterances. in *The Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. 2014.

[5] Nicholas, M. et al., Empty speech in Alzheimer's disease and fluent aphasia. *J Speech Hear Res* 1985; **28**(3): 405-10.

[6] Hosokawa, T. et al., Psychometric equivalence of the Hasegawa Dementia Scale-Revised with the Mini-Mental State Examination in stroke patients. *Percept Mot Skills* 1994; **79**(1 Pt 2): 664-6.

[7] Kurohashi, S. and M. Nagao, Improvements of Japanese Morphological Analyzer JUMAN, in *Treebanks*. 2003, Springer.

[8] Sunakawa, Y. Japanese Learner's Dictionary: Research for the formulation of basic grounds for the construction of a general database for the development of Japanese language learners' dictionaries. [cited 2015 1/8]; Available from: http://dictionary.j-cat.org/JtoE/index.php.

[9] Aramaki, E. et al., Vocabulary Size in Speech May Be an Early Indicator of Cognitive Impairment. *PLoS One*, 2016; **11**(5): e0155195.

[10] Aramaki, E. et al., A Word in a Dictionary is used by Numerous Users. in *International Joint Conference on Natural Language Processing (IJCNLP2013)*. 2013. Japan.

[11] Kempler, D. et al., Sentence comprehension deficits in Alzheimer's disease: a comparison of off-line vs. on-line sentence processing. *Brain Lang* 1998; **64**(3): 297-316.

[12] Weiner, M.F. et al., Language in Alzheimer's disease. *J Clin Psychiatry* 2008; **69**(8): 1223-7.

[13] Wallesch, C.W. and A. Hundsalz, Language function in delirium: a comparison of single word processing in acute confusional states and probable Alzheimer's disease. *Brain Lang* 1994; **46**(4): 592-606.

Address for correspondence:

Eiji ARAMAKI, aramaki@is.naist.jp.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190399

Development of Pictograms for an Interactive Web Application to Help Hispanic Caregivers Learn About the Functional Stages of Dementia

Adriana Arcia[a], Niurka Suero-Tejeda[a], Suzanne Bakken[b]

[a] School of Nursing, Columbia University, New York, NY, USA,
[b] School of Nursing and Department of Biomedical Informatics, Columbia University, New York, NY, USA

Abstract

Caregivers of persons with dementia need anticipatory guidance about the stages of dementia in order to prepare for the caregiving situations they will face. The study objective was to develop a set of pictograms representing the functional stages of dementia for eventual inclusion in a tailored, educational web application. We used a hybrid iterative participatory design process. A graphic designer prepared prototypes in a flat, minimalistic style. These were then culled and refined based on feedback from 16 Hispanic caregivers in six design sessions in English and Spanish. The resulting 19 pictograms representing the functional stages and substages of dementia were acceptable to and easily comprehended by participants. Short, plain-language captions support comprehension and aid discrimination between similar scenarios. Our participants preferred candid depictions of all aspects of dementia, including bodily functions, but acceptability may vary by population so further testing is warranted prior to deployment with a new population.

Keywords:

Consumer Health Information, Health Literacy, Patient Participation

Introduction

Caregivers require timely anticipatory guidance to prepare them for the complex and changing care needs of persons with dementia. Caregivers may need to acquire new skills or equipment, connect with community resources, or recruit additional caregiving assistance as the care recipient's condition progresses. An important way to prepare for the demands of caregiving is to understand the abilities or deficits associated with each of the 16 functional stages and substages of dementia [14; 15]. Numerous digital information resources about dementia exist, such as websites and mobile applications, but barriers like Low Health Literacy (LHL) and Limited English Proficiency (LEP) may limit the comprehensibility of these text-heavy resources for some caregivers in the United States [3].

Hispanics suffer disproportionately from dementia, including Alzheimer's dementia, and the associated caregiving burden, compared to non-Hispanic Whites [13]. Hispanics are also over-represented among individuals with LHL and LEP [16]. As a result, there is a need for culturally-competent resources that are designed specifically to meet the information needs of Hispanic caregivers.

One promising approach to supporting comprehension of health information among people with LHL and/or LEP is the use of visual enhancements, such as infographics [1] or photos. A photographic guide to caregiving for dementia is commercially available, but it is targeted to professional caregivers rather than family members and is only available in English [10]. Several studies report on the use of *fotonovelas*—soap opera-type stories in a photographic comic book format—to help Hispanic caregivers gain a general understanding of dementia [18] or learn strategies to reduce depression and stress [7]. However, our search did not uncover any research on visually-enhanced resources intended to make information about the functional stages of dementia easy for Hispanic caregivers to understand and use.

To meet this need, our objective is to develop an interactive web application, entitled **I**nteractive **F**unctional **A**ssessment **St**aging **N**avigator (I-FASTN), to inform caregivers about the functional stages of dementia. The application will be tailored to the user in that it will highlight the specific functional stage of the care recipient while also allowing the user to review the other stages. Screens for each functional stage will offer links to additional caregiving tips and strategies relevant to that stage. A long-term objective is to make I-FASTN available within patient portals through the Fast Health Interoperability Resource (FHIR) Health Level Seven standard and the Substitutable Medical Applications and Reusable Technologies (SMART) platform that enables systems to behave as 'iPhone-like platforms' through an application programming interface (API) and a set of core services that support easy addition and deletion of third party apps, i.e., the core system is stable and the apps are substitutable [12].

The first step in the creation of I-FASTN was content development. Therefore, the objective of this study was to develop a set of pictograms to represent the functional stages of dementia. For specific examples of functional deficits, we drew from both Functional Assessment Staging (FAST) [15] and the Global Deterioration Scale (GDS) [14]. We accomplished our objective through a hybrid iterative participatory design process in which experts create a starting set of prototype images which are then culled and refined based on participant feedback.

Methods

The Institutional Review Board of Columbia University Irving Medical Center approved this study. Written informed consent was obtained from all participants.

Participants

We recruited participants from the **N**ew York City **His**panic dementia caregiver **R**esearch **P**rogram (NHiRP), a study of family caregivers of persons with dementia. Participants were eligible for the present study if they met the NHiRP inclusion criteria: self-identified as Hispanic, 18-90 years old, related to the care recipient, physically able to provide care, not have a diagnosis of a major psychiatric disorder apart from depression, not have depression with psychotic features or suicidal ideation within the preceding 5 years, and expected to live in New York City for at least 12 months.

Procedures

Prior to starting design sessions, we prepared multiple prototypes to illustrate each functional stage and substage of dementia. In order to maximize our options for I-FASTN content, we aimed to identify as many prototypes as possible that participants indicated were clear representations of the intended meaning. We planned to undertake revisions until reaching design saturation, which is the point at which participant feedback yields no further substantive changes.

Iterative participatory design sessions were held in November and December of 2016 in New York City in English and Spanish. We collected basic demographic information from participants. Two investigators experienced in participatory design (AA, NST) led the sessions which were also attended by the experienced graphic designer (NSG) who had created the initial prototypes. In the sessions, we informed participants of the intended meaning (e.g., "Here we are trying to show difficulty managing complex tasks") and showed the captioned prototypes printed on 8.5" x 11" (22 cm x 28 cm) card stock. We solicited preferences among the prototypes by voice or hand vote and used open-ended questions to elicit suggestions for improvement. Sessions were audio-recorded and study staff took notes.

Analysis was concurrent with data collection: at the end of each session, study staff met to 1) synthesize the findings from the session, 2) identify poorly-performing prototypes to be removed from further consideration, and 3) come to consensus about revisions to the favored prototypes based on participant feedback. We then carried out any needed revisions between sessions. In later sessions after the set of prototypes had been narrowed down, we showed the pictograms in the context of preliminary layouts for the I-FASTN interface. After all the sessions were completed the study staff reviewed the notes, audio recordings, and resultant transcripts to audit the results and ensure important insights had not been missed.

Results

Thirteen women and three men (N = 16) participated in three English (n = 6) and three Spanish (n = 10) sessions. They ranged in age from 49 to 86 (M = 61.8) and had been caregiving for an average of almost 9 years. With respect to educational attainment, 12% (n = 2) had some high school, 44% (n = 7) were high school graduates, 25% (n = 4) had an associate's degree, and 19% (n = 3) had a bachelor's degree. Participants were predominantly born in or had family roots in Caribbean countries, rather than Spain or Central or South America.

We tested a total of 54 prototypes, which we subsequently reduced to a final set of 19 pictograms for the 16 stages and substages of dementia, as shown in Table 1. Stages 1, 4, and 5 each yielded two viable pictograms, so these were retained. Just over half of the pictograms (n = 10) underwent one revision, one went through two revisions and the remainder (n = 8) were unchanged from the original prototypes.

Participants' judgments about the extent to which the prototypes accurately represented a stage or substage were often unanimous and unequivocal, with clear favorites emerging early on. Participants narrowed down to the 19 final pictograms so quickly that by the fourth session we began to combine the pictograms into preliminary layouts for the I-FASTN interface to gather preliminary data for developing the interface. Participants in sessions four through six had no additional suggestions for improvement of the individual pictograms and thus we achieved design saturation. Participants displayed ease of comprehension with the final set of pictograms. For example, in sessions four and five participants were asked to review a preliminary I-FASTN layout for Stage 5 that had thumbnail images—less than 0.5" (1.27 cm) high—without captions of Stages 6a-6e printed in one corner of the page. In each of the two sessions, a participant spontaneously and correctly interpreted the meaning of the thumbnails. One man in session 5 explained: "Either they can't dress themselves, they are not bathing themselves, they need help going to the bathroom. Let me see this… they're having accidents."

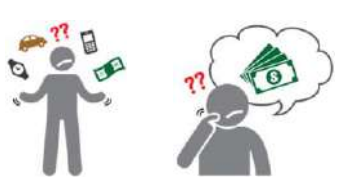
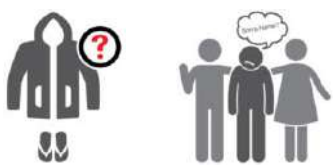
Despite the ease with which participants interpreted uncaptioned pictograms, we concluded that they functioned best when accompanied by the captions included in Table 1, particularly to help differentiate between stages. For example, without captions, the distinction between Stage 5, "difficulty selecting appropriate clothing," and Stage 6a, "difficulty dressing without help," is an important one that might be lost on casual observation.

We noticed a general preference among participants for pictograms that featured people rather than objects. For example, the pictogram for Stage 3 originally had just the street signs; the person shown in the final version was added at participants' request. Participants also preferred imagery that was candid and unflinching. For instance, the initial prototype for 6e, fecal incontinence, showed the figure from the front with only the wavy lines to indicate the meaning. The more explicit depiction of sagging, soiled briefs was made per participants' specific request. In general, we noted that prototypes showing the presence of something were far more readily understood than those that attempted to show the same idea from the perspective of absence. For example, the pictogram for 7e showing a face with a drooping mouth was more readily understood as representing "cannot smile" than a smiling face marked with an X. Similarly, the final pictogram for Stage 4 in which a person appears confused as they think about money was preferred over a prototype showing a stack of cash with a slash through it.

Discussion

In this study, we worked with Hispanic caregivers to develop a set of pictograms that effectively represents the functional stages of dementia in a culturally acceptable manner. The pictograms are an important contribution to consumer health informatics and health communication because they facilitate access to necessary health information for a population at risk for health disparities. It is important to note that this work is distinct from research aimed at developing visualizations for non-literate populations [2] because we find that our population has little-to-no difficulty reading short passages of plain-language text. Rather, we have observed that comprehension deficits are related to uncertainty about having arrived at correct conclusions and limitations in understanding the basic

Table 1– Pictograms of the Functional Stages and Substages of Dementia.

Functional stages and sub-stages ***Captions***	**Related Pictogram(s)**	**Functional stages and sub-stages** ***Captions***	**Related Pictogram**
1. No objective or subjective functional decrement *Able to work; normal functioning*		6d. Urinary incontinence *Urinary incontinence*	
2. Subjective deficit only *Forgets where they placed familiar objects*		6e. Fecal incontinence *Fecal incontinence*	
3. Deficits noted in demanding occupational and social settings *Gets lost when traveling to new locations*		7a. Speech limited to about six words in the course of an average day *Speech limited to a few words*	
4. Deficits in performance of complex tasks of daily life *Difficulty managing complex tasks; unable to manage finances*		7b. Intelligible vocabulary limited to generally a single word in the course of an average day *Speech limited to one word*	
5. Deficient performance in choosing proper attire *Difficulty choosing appropriate clothing; can't remember the names of close family members*		7c. Ambulatory ability lost *Needs help to walk*	
6a. Requires actual physical assistance in putting on clothing properly *Difficulty dressing without help*		7d. Ability to sit up lost *Cannot sit up without support*	
6b. Requires assistance bathing properly *Unable to bathe properly*		7e. Ability to smile lost *Cannot smile*	
6c. Requires assistance with mechanics of toileting *Unable to manage the details of using the bathroom*		7f. Ability to hold head up lost *Cannot lift head by themselves*	

implications of a text. As such, pictograms support comprehension by using familiar imagery to reinforce the meaning of a text. For example, if a reader is uncertain about precisely what is meant in Stage 5 by "appropriate clothing," the pictogram gives a visual example of the incongruous combination of jacket and flip-flops.

The validity of these pictograms as representations of the stages of dementia is underscored by participants' quick, definitive judgments and strong interpersonal consensus when selecting from among prototypes. Participants' spontaneous interpretations of uncaptioned images, even when reproduced in thumbnail size, is robust preliminary evidence of the pictograms' comprehensibility. Group dynamics limit the ability to conclusively establish every participant's level of comprehension, but further testing may be undertaken by adapting methods outlined by the International Organization for Standardization in ISO 9186-1 for testing the comprehensibility of graphical symbols [5; 9].

For this project, we chose to use pictograms in a flat, minimalistic style rather than more detailed illustrations for two reasons. One, prior research has shown that viewers, particularly those with LHL, are at risk of misconstruing the intended meaning of complex illustrations because of a tendency to focus on irrelevant details [8]. Two, we wanted to minimize details within the images that might prevent a diverse population of users from identifying with the figures in the pictograms. Implementation of the chosen style was facilitated by the inexpensive commercial availability of royalty-free stock images which we adapted as needed. This style is consistent with other studies that have successfully used what have been described as "restroom icons" to convey health-related risks [20] and "stick figures" to develop easily comprehensible symbols for wayfinding in health care settings [4; 11].

In addition to style, other aspects of the final set of pictograms are consistent with prior research. When Weiner and colleagues [19] set out to design an illustrated patient satisfaction questionnaire for low-literacy populations, they found that "a few words go a long way" which is in keeping with our conclusions about the value of short, plain-language captions. As in our study, their illustrations successfully used question marks to depict confusion but their attempts to use slashes to indicate the negation of an idea (e.g., illustration of the response option "not applicable") were as unsuccessful as ours. Foster and Afzalnia tested the comprehensibility of various symbols to represent a cash machine and found that images that depicted human interaction with an object (i.e., a hand holding a stack of bills) were more easily comprehended than images of the object alone [6]. In a similar vein, we found that participants generally preferred the images that contained people.

Although these pictograms were developed to suit the needs of Hispanic caregivers in the United States, we believe the images will be easily comprehensible to a broader range of users. However, not all groups may consider all of the pictograms to be culturally acceptable, particularly those that depict bodily functions [17]. Given that our sample predominantly has roots in Caribbean countries, our findings may or may not extend to Hispanics from other regions. Explicitly testing for cultural acceptability and comprehensibility is warranted before pictograms are deployed with a new population.

The next step for our project is to design and test the I-FASTN interface (i.e., heuristic evaluation with experts followed by usability testing with end users) using the pictograms presented in this report. Spanish-language captions and vector files of the pictograms are freely available for non-commercial use upon request from the corresponding author.

Conclusions

Application of hybrid participatory design methods resulted in pictograms that were acceptable and comprehensible to Hispanic caregivers as an essential first step in developing tools to meet their needs. Well-designed pictograms are a useful adjunct to text for meeting information needs in consumer health informatics tools.

Acknowledgements

We thank our graphic designer, Nicole Spiegel-Gotsch. This publication was supported by the **N**ew York City **Hi**spanic dementia caregiver **R**esearch **P**rogram (R01 NR014450-03S1), **Pr**ecision **i**n **S**ymptom **S**elf-**M**anagement (PriSSM) Center (P30 NR016587), and the National Center for Advancing Translational Sciences, National Institutes of Health, (UL1 TR001873). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] A. Arcia, N. Suero-Tejeda, M.E. Bales, J.A. Merrill, S. Yoon, J. Woollen, and S. Bakken, Sometimes more is more: Iterative participatory design of infographics for engagement of community members with varying levels of health literacy, *Journal of the American Medical Informatics Association* **23** (2016), 174-183.

[2] P.A. Barclay and C.A. Bowers, Design for the Illiterate: A Scoping Review of Tools for Improving the Health Literacy of Electronic Health Resources, in: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, SAGE Publications Sage CA: Los Angeles, CA, 2017, pp. 545-549.

[3] N.D. Berkman, S.L. Sheridan, K.E. Donahue, D.J. Halpern, and K. Crotty, Low health literacy and health outcomes: An updated systematic review, *Annals of Internal Medicine* **155** (2011), 97-107.

[4] J. Bolek and J. Cowgill, Development of a symbol system for use in the health care industry, in: *Proceedings of the 2005 Conference on Designing for User eXperience*, AIGA: American Institute of Graphic Arts, 2005, p. 23.

[5] K.A. Brantley and M.S. Wogalter, Oral and written symbol comprehension testing: The benefit of cognitive interview probing, in: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Sage Publications Sage CA: Los Angeles, CA, 1999, pp. 1060-1064.

[6] J.J. Foster and M. Afzalnia, International assessment of judged symbol comprehensibility, *International Journal of Psychology* **40** (2005), 169-175.

[7] D. Gallagher-Thompson, M. Tzuang, L. Hinton, P. Alvarez, J. Rengifo, I. Valverde, N. Chen, T. Emrani, and L.W. Thompson, Effectiveness of a fotonovela for reducing depression and stress in Latino dementia family caregivers, *Alzheimer Disease and Associated Disorders* **29** (2015), 146.

[8] P.S. Houts, C.C. Doak, L.G. Doak, and M.J. Loscalzo, The role of pictures in improving health communication: A review of research on attention, comprehension, recall, and adherence, *Patient Education and Counseling* **61** (2006), 173-190.

[9] ISO, *ISO 9186-1:2014 Graphical Symbols -- Test Methods -- Part 1: Method for Testing Comprehensibility*, International Organization for Standardization, Geneva, 2014.
[10] Grealy J, McMullen H, Grealy J. *Dementia Care: A Practical Photographic Guide.* Chichester, United Kingdom: John Wiley & Sons, Incorporated; 2008.
[11] S. Lee, S.S. Dazkir, H.S. Paik, and A. Coskun, Comprehensibility of universal healthcare symbols for wayfinding in healthcare facilities, *Applied Ergonomics* **45** (2014), 878-885.
[12] J.C. Mandel, D.A. Kreda, K.D. Mandl, I.S. Kohane, and R.B. Ramoni, SMART on FHIR: a standards-based, interoperable apps platform for electronic health records, *Journal of the American Medical Informatics Association* **23** (2016), 899-908.
[13] K.M. Mehta and G.W. Yeo, Systematic review of dementia prevalence and incidence in United States race/ethnic populations, *Alzheimer's & Dementia* **13** (2017), 72-83.
[14] B. Reisberg, S.H. Ferris, M.J. de Leon, and T. Crook, The Global Deterioration Scale for assessment of primary degenerative dementia, *American Journal of Psychiatry* **139** (1982), 1136-1139.
[15] S.G. Sclan and B. Reisberg, Functional assessment staging (FAST) in Alzheimer's disease: reliability, validity, and ordinality, *International Psychogeriatrics* **4** (1992), 55-69.
[16] T. Sentell and K.L. Braun, Low health literacy, limited English proficiency, and health status in Asians, Latinos, and other racial/ethnic groups in California, *Journal of Health Communication* **17** (2012), 82-99.
[17] C.G. Spinillo, Graphic and cultural aspects of pictograms: an information ergonomics viewpoint, *Work* **41** (2012), 3398-3403.
[18] R. Valle, A.-M. Yamada, and A.C. Matiella, Fotonovelas: A health literacy tool for educating Latino older adults about dementia, *Clinical Gerontologist* **30** (2006), 71-88.
[19] J. Weiner, A. Aguirre, K. Ravenell, K. Kovath, L. McDevit, J. Murphy, D.A. Asch, and J.A. Shea, Designing an illustrated patient satisfaction instrument for low-literacy populations, *The American Journal of Managed Care* **10** (2004), 853-860.
[20] B.J. Zikmund-Fisher, H.O. Witteman, M. Dickson, A. Fuhrel-Forbis, V.C. Kahn, N.L. Exe, M. Valerio, L.G. Holtzman, L.D. Scherer, and A. Fagerlin, Blocks, ovals, or people? Icon type affects risk perceptions and recall of pictographs, *Medical Decision Making* **34** (2014), 443-453.

Address for correspondence

Adriana Arcia
School of Nursing, Columbia University, New York, NY, USA
Email: adrianaxarcia@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190400

Resident-Physician Preferences for Electronic Handoff Note Content: Implications for Implementation of a System-Wide Electronic Health Record-Integrated Handoff Tool

Elliot G Arsoniadis[a,b], Rohini Khatri Olsen[c], Steven Skube[d], Jenna Marquard[e], Genevieve B Melton[b,f]

[a] *Department of Colorectal Surgery, Cleveland Clinic Florida, Weston, FL, USA*
[b] *Institute for Health Informatics, University of Minnesota, Minneapolis, MN, USA*
[c] *Department of Pediatric Surgery, Long Island Jewish Medical Center, Queens, NY, USA*
[d] *Department of Surgery, University of Minnesota, Minneapolis, MN, USA*
[e] *College of Engineering, University of Massachusetts, Amherst, MA, USA*
[f] *Division of Colorectal Surgery, Department of Surgery, University of Minnesota, Minneapolis, MN, USA*

Abstract

Increasing attention is paid to the Handoff Process and EHR-integrated tools to enhance the Handoff Process and aid in creating Handoff Notes are becoming more prevalent. In this study, we attempted to determine the ideal content of the Handoff Note based on the preferences of the resident physicians for whom the tool was being constructed. This commenced with an initial semi-structured interview and culminated in a large-scale survey. Overall, 315 resident physicians completed the survey. Plan of Care, Illness Severity, and Patient Summary were the most important content elements to resident physicians. The importance and trustworthiness of other content elements, as well as their preferred display and level of granularity within the Handoff Note, varied considerably. Subjective assessment by a colleague of a patient's hospital course and plan of care, rather than any single piece of objective data, are preferred as ideal content for Handoff Note composition by resident physicians.

Keywords:

Patient Handoff, Electronic Health Records

Introduction

The transfer of patient care between physicians in the acute care setting commonly referred to as the "Handoff Process", has become more frequent with the advent of American Council for Graduate Medical Education (ACGME) restrictions on resident physician work hours. Numerous studies have identified inpatient Handoff as a source of preventable medical errors [1; 2]. The ACGME, Joint Commission, and Agency for Healthcare Research and Quality have all identified inpatient Handoff as a priority for improving patient care [3; 4]. Beginning in 2003, numerous reports have been published on the benefits of Electronic Health Record (EHR)-integrated Handoff Notes [5; 6]. No true standard exists for what content elements should be contained within EHR-integrated Handoff Notes [7; 8].

Methods

In order to investigate the preferred content of Handoff Notes, we first conducted a series of semi-structured interviews with resident physicians engaged in active clinical service on inpatient units within our hospitals. We then utilized the major themes discovered within the semi-structured interviews to develop an online survey containing discrete, multiple choice questions that could provide quantifiable data regarding the preferred content of Handoff Notes by the resident physicians for whom this electronic tool was being designed.

Semi-structured Interview

In order to gain an initial sense of resident physicians' views of current Handoff Note-generation techniques, content, and satisfaction, we conducted semi-structured interviews with resident physicians on inpatient clinical service teams at either our tertiary, acute care adult hospital or our tertiary, acute care pediatric hospital. Resident physicians functioning as junior level residents, who were actively involved in the process of Handoff Note creation as part of their daily work or who functioned as inpatient cross-cover residents who were actively involved in the process of Handoff and receiving Handoff Notes to guide their care of patients, were interviewed. Resident physicians came from Pediatrics, Internal Medicine, and Surgery teams. Pediatrics and Internal Medicine teams were from general pediatrics and general medicine teams. Surgery teams were either General Surgery or a surgical subspecialty, including Colorectal, Thoracic, Vascular, and Surgical Oncology.

Resident physicians were first presented with a sample Handoff Note typical for that specialty. At this time in our institution, different specialties were utilizing different techniques and styles for generating Handoff Notes. Pediatrics and Internal Medicine teams were utilizing quasi-EHR integrated Handoff Note generating techniques while Surgery teams utilized Microsoft Word-based Handoff Notes stored on local computers at specific workstations throughout the hospital. Therefore, when initially asked to comment on the quality and utility of Handoff Notes, resident physicians were presented with examples as similar to their specialty's current practice as possible. During the semi-structured interview, resident physicians were presented with a common clinical scenario encountered during cross-cover periods (a patient with hypotension) and then were asked what data in the Handoff Note was useful for clinical decision making. They were then asked to comment on the trustworthiness of the data elements found within the Handoff Note. Finally, resident physicians were asked to comment on their beliefs for what would be required of an EHR-integrated Handoff Note that would contain optimal clinical information and be easily constructed and shared with colleagues. Interviews were conducted by one of two study investigators (EGA and RKO) and voice recordings were performed for each of the interviews. Narrative content of the interviews was then analyzed for major themes and headings.

Institution-Wide Online Survey

We then utilized the major themes and headings discovered within the narrative content of the semi-structured interviews, as well as previous review of the literature on content of EHR-integrated Handoff Notes (Arsoniadis et al, unpublished), to develop an online survey for distribution among resident physicians enrolled in American Council for Graduate Medical Education (ACGME) – accredited residency and fellowship programs at our institution. The goals of the survey were to assess the importance of different data elements within the Handoff Note, assess the trustworthiness of certain data elements within the Handoff Note, and determine the preferred format and level of granularity at which certain Handoff Data is displayed within an EHR-integrated Handoff Note functionality

The survey was developed by two physicians (EGA and RKO) with experience in health informatics and an expert in human factors and usability (JM). Resident physicians were asked to identify their specialty and level of training, as well as experience using the EHR. Resident-physicians were asked to rate the importance of different content elements within the Handoff Note according to a Likert-style scale, with 1 being "Not Important" and 5 being "Very Important" for accomplishing clinical care. An additional option for free text response was included at the end of the survey. Content elements included in the survey were extracted from either previous review of the literature on standard or suggested content for electronic Handoffs [7; 8], from the prior semi-structured interviews with resident-physicians, or both.

A Likert-style scale was also utilized to determine the level of trustworthiness of certain content elements to resident physicians, including Code Status, Patient Summary, Post-Operative Day, To-Do List, and Anticipatory Guidance. The trustworthiness of these data elements was of special interest, given the varied opinions voiced during the semi-structured interviews performed previously. Options ranged from 1, "I do not trust at all" to 5 "I trust completely".

The preferred format and granularity at which certain Handoff Note data is displayed was also queried through a series of multiple-choice questions in our survey. These questions included content elements such as patient problem list, past medical/surgical history, clinical data (vital signs, imaging and laboratory studies, fluid balance), and medications, among others. Again, these multiple choice questions were formulated based on themes generated from our initial semi-structured interview.

The final content of the survey was reviewed by five resident physicians currently involved in active duty on inpatient clinical service teams and minor revisions made based on their input. It was released to all physicians in ACGME-accredited residency and fellowship programs involved in inpatient acute care (anesthesiology, radiology, ophthalmology, pathology, and dermatology excluded). The survey was approved by the Dean for Graduate Medical Education as well as the Institutional Review Board of the University of Minnesota. Participating resident physicians gave informed consent, and were entered into a lottery for one of two iPad Airs for their participation.

Results

Semi-structured Interviews

Sixteen resident physicians (first through third years of training) from three specialties (Internal Medicine, Pediatrics, and Surgery) were interviewed. They were asked to comment on the clinical utility and trustworthiness of various content elements on an example Handoff Note currently in use by that specialty.

Table 1 – Survey Participants by Training Level and Program

Training Level	Surgery	Pediatrics	Internal Medicine	Total
PGY 1-2	44 (39%)	33 (42%)	54 (43%)	131
PGY 3-4	38 (34%)	29 (37%)	46 (37%)	113
PGY ≥5	22 (20%)	16 (21%)	25 (20%)	63
Total	112	78	125	313

Clinical Utility and Trustworthiness

The utility of content elements in Handoff Notes for clinical decision making during cross-cover periods varied between resident physicians. Summary statements about the patients were nearly universally utilized in clinical decision making. Two-thirds of resident physicians interviewed felt the Medications to also be useful. Some content was specialty specific. For instance, Procedure Name and Date were regarded as useful content by four of five Surgery resident physicians. However, these content elements were not even present in the versions of the Handoff Note being utilized by Internal Medicine and Pediatrics resident physicians. Similarly, the trustworthiness of different content elements varied, with some resident physicians implicitly trusting all or most of the content elements and others trusting almost none of them. The trustworthiness of content elements did not seem to be affected by the source of the data, whether manually entered or automatically generated from the EHR. Interestingly, mistrust of data in the Handoff Note was most prevalent in the comments of Internal Medicine resident physicians, who contributed 18/21 comments (86%) that called the trustworthiness of various content elements into question.

Table 2 – Most Important Handoff Content Elements

Content Element	% Ranking Important (4 or 5)
Plan of Care	95%
Name & Age	89%
MRN & Room Number	83%
Illness Severity	78%
Code Status	77%
Patient Summary	77%

Table 3 – Least Important Handoff Content Elements

Content Element	% Ranking Not Important (1 or 2)
Primary Care Physician	84%
Fluid Balance	50%

Suggestions for Improvement

Transcripts for interviews were coded by two authors (EGA and RKO) and major themes emerged regarding improvements resident physicians felt could be made to EHR-integrated Handoff Notes. Sixty three percent of comments related to improvements focused on the importance of including information from the primary service team regarding their subjective assessment of the patient, including a brief summary of the patient and their hospital course, assessing patients' illness severity, providing anticipatory guidance for addressing possible issues that may arise during the cross-cover period, and providing a To-Do list for tasks that needed to be accomplished in the cross-cover period. The structure and format of the Handoff Note were addressed by several resident physicians. Some favored a more standardized approach with decreased content and "white noise". Others favored a Handoff Note that was more customizable, with resident physicians being able to

control the content elements included in the Handoff Note. Another theme included the level of granularity with which certain data were displayed. Some suggestions included specifying the infusion rate of intravenous fluids and a patient's diet. The display and organization of Medications was a very frequent topic of commentary, with many resident physicians voicing the desire for a more concise medication list, organized in a systems-based format (e.g. pain medications, antibiotics, anticoagulants), while others wanted only scheduled medications reported and PRN (as needed) medications removed.

*Table 4 – Somewhat Important Handoff Content Elements**

Content Element	4-5	3	1-2
Medications	48%	31%	21%
Procedures Performed	46%	31%	23%
Primary Language	44%	28%	28%
Attending & Team	44%	30%	26%
Problem List	43%	34%	23%
Hospital Day	42%	38%	21%
IV Access/Tubes/Drain	40%	32%	28%
Allergies	36%	32%	33%
Labs & Imaging	34%	28%	38%
Vital Signs	33%	28%	39%
Past Medical History	24%	41%	35%
Diet	22%	32%	29%
Prophylaxis	22%	31%	47%
Family Contact	21%	28%	46%
Psychosocial Concerns	21%	35%	44%
Consulting Services	21%	32%	47%

**Percentage of Likert-style scores per Content Element; 4-5(Important), 3(Somewhat Important), 1-2(Not Important)*

Survey

Given the diverse (and at times competing) ideas for ideal EHR-integrated Handoff Note content and structure, we utilized an online survey sent to all resident physicians in our institution practicing in the acute, inpatient setting. The intent was to utilize the themes that emerged from our convenience sample of interviewed resident physicians and create a survey through which more generalizable preferences for ideal EHR-integrated Handoff Note content and functionality could be ascertained. Three hundred thirteen of 715 eligible resident physicians completed the survey (44%). Forty-two percent were in their first or second year of training (Table 1). Thirty-six percent were in their third or fourth year of training, and 20% were in their fifth year or greater of training. Thirty-six percent were enrolled in a surgical specialty, while 25% were from Pediatrics or a pediatric subspecialty and 40% were from Internal Medicine or a medical subspecialty (e.g. Cardiology, Gastroenterology). The vast majority (91%) of resident physicians reported at least three years (including medical school) of experience working with electronic health records and 75% reported at least three years of working with the Epic EHR specifically. Seventy-five percent reported that at least half of their clinical time was spent caring for patients in acute care, inpatient setting (as opposed to outpatient clinics). Seventy-eight percent reported serving as a cross-cover physician an average of at least three times per month or more during their training.

Clinical Utility

When ranking the importance of various content elements using a Likert-style rating scale there were several content elements that resident physicians found most important, with an overwhelming majority (>50%) ranking as Important or Very Important (ranking 4 or 5) (Table 2). These included content elements such as Plan of Care (95%), Name and Age (89%), Medical Record Number and Room Number (83%), Illness Severity (78%), Code Status (77%) and Patient Summary (77%). There were also content elements that the majority (>50%) of resident physicians ranked as not important (1 or 2) (Table 3). These were Primary Care Physician and Fluid Balance. The remaining content elements were more variable in the importance placed in them by resident physicians (Table 4). Some of them, such as Medications and Procedures Performed, had nearly 50% of resident physicians ranking them as important, while others, like Psychosocial Concerns or Consulting Services, were ranked as not important by nearly one-half of resident physicians. Other content elements, such as Vital Signs, and Labs and Imaging, had nearly even distributions, with one third ranking them as important, one third ranking them as somewhat important, and the remaining third ranking them as not important.

*Table 5 – Trustworthiness of Various Handoff Content Elements**

Content Element	4-5	3	1-2
Code Status –			
EHR Generated	41%	43%	16%
Manually Entered	82%	16%	2%
Postoperative Day			
EHR Generated	64%	26%	10%
Manually Entered	46%	36%	18%
Illness Severity**	86%	12%	2%
Patient Summary**	82%	16%	2%
"To Do" List**	91%	9%	0%
Anticipatory Guidance**	91%	8%	1%

**Percentage of Likert-style scores per Content Element; 4-5(Trustworthy), 3(Somewhat Trustworthy), 1-2(Not Trustworthy)*

***Only manually-entered versions of these content elements are available*

Trustworthiness

Resident physicians also ranked various content elements using a Likert-style scale as trustworthy (5) or not trustworthy (1) (Table 5). Some of these content elements could be either manually entered into Handoff Notes or automatically derived from the EHR. The source of the content affected the trustworthiness of the data. For instance, 82% of resident physicians ranked Code Status manually entered into the Handoff Note as trustworthy (4 or 5) while only 41% ranked EHR generated Code Status as trustworthy. This is in contrast to Postoperative Day (the number of days that have passed since a patient's surgery occurred), where the majority (64%) of resident physicians ranked EHR-generated Post Operative Day as trustworthy while only 46% ranked manually entered Postoperative Day as trustworthy.

Other content elements, such as To-Do List and Anticipatory Guidance, could only be manually entered into the Handoff Note. These content elements were ranked as trustworthy by the majority of resident physicians (91% ranking both as trustworthy) (Table 5).

Data Display and Content Source Preferences

Resident physicians also showed some clear preferences for the level of granularity with which certain content elements were displayed in Handoff Notes and the source of such content elements (Table 6). For instance, 65% of resident physicians preferred an Admission Diagnosis and Problem List that was manually entered by a colleague as opposed to automatically entered by the EHR, or Admission Diagnosis alone. Seventy-seven percent of resident physicians preferred Past Medical and Surgical History to be entered manually. Sixty-eight percent of resident physicians preferred both Scheduled and PRN (as needed) medications to be displayed, but not patients' home medications. Sixty-six percent preferred the attending physician's contact information to be displayed. There was a trend toward favoring display of objective data from the past 24

hours, with 78% and 52% favoring displaying the fluid balance and vital signs over the past 24 hours, respectively, as opposed to over the previous nursing shift (8 or 12 hours), or from the entire hospital stay.

Table 6 – Data Display Preferences

Data	%
Prefer Attending Physician Contact Displayed	
Yes	66%
No	34%
Admission Diagnosis & Problem List	
Only Admission Diagnosis from EHR	3%
Only Admission Diagnosis Manual	21%
Admission Diagnosis & Problems List EHR	12%
Admission Diagnosis & Problem List Manual	65%
Past Medical & Surgical History	
EHR generated automatically	23%
Manually generated	77%
Medications	
Scheduled medications	14%
Scheduled and PRN medications	68%
Scheduled, PRN, and Home medications	16%
Scheduled and Home Medications	3%
Vital Signs	
Most recent	32%
Past shift	15%
Past 24 hours	52%
Entire hospital stay	1%
Diagnostics	
Laboratory data only	37%
Imaging data only	6%
Laboratory and Imaging data	57%
Laboratory Values	
Most recent	52%
Past 24 hours	45%
Entire hospital stay	3%
Fluid Balance	
Past shift	16%
Past 24 hours	78%
Entire hospital stay	6%
Fluid Type	
Total Intake/Output	21%
Total Intake/Output & Urine Output	31%
Total Intake/Output, Urine Output, & All Other Drain* Ouptut	26%
Total Intake/Output, Urine Output, Other Drain Output, and Oral Intake	23%
Consulting Service	
Physician Services	54%
Physician and Ancillary** Services	12%
No Consulting Services	34%

**e.g. Nasogastric tube, Surgical drains, Ventriculostomy, etc.*

***e.g. Physical Therapy, Nutrition, Social Work*

Preferences were less clear for the display of other content elements. Fifty-two percent of resident physicians favored the most recent set of Laboratory Values to be displayed while 45% wanted Laboratory Values from the past 24 hours to be displayed. Fifty-seven percent wanted to display both Laboratory and Imaging Data, while 37% wanted to display Laboratory Data alone. The display of Fluid Balance was most variable, with a near-even distribution of resident physicians favoring the display of only Total Intake and Output as opposed to intake and output being parsed out into Urine Output, Drain Output and Oral Intake (Table 6).

Free Text Commentary

Resident physicians were allowed optional free text commentary for other thoughts on Handoff Note content. The free text commentary was reviewed. Several themes emerged from the free text. Many comments focused on the optimal content of the Handoff Note. Generally, the importance of a patient summary statement and anticipatory guidance on potential issues in the ensuing cross-cover period were highlighted in many comments. Many comments also noted that certain content elements were only important for certain patients and in certain situations. Similarly, there were other comments that discussed the importance of certain content elements varying based on the type of medical team and physician specialty. Many comments alluded to the fact that the Handoff Note need not contain every piece of important data on each patient, and that in most instances the resident physicians would look up patients in the Electronic Health Record for more granular data. Other comments noted a preference for brevity in the Handoff Note. Finally, some comments focused on problems with current functionality for EHR-generated Handoff Notes.

Discussion

The importance of the Handoff Process has been increasingly recognized, and current restrictions on resident physician work hours make the occurrence of Handoff a cultural norm at hospitals that employ physicians in training [1]. Despite this, there is still no standard for the Handoff process, although certain formats, such as the I-PASS mnemonic, have been gaining in popularity [9]. Similarly, there is no standard for content that should be included in the Handoff Note that accompanies the Handoff Process. Prior work highlights the diverse content included in various published reports of EHR-integrated Handoff Note content [7; 8]. Here we report our experience with attempting to determine the ideal content of an EHR-integrated Handoff Note to serve as a standard template for resident physicians at a single, tertiary care, University-based adult and pediatric hospital. We did this utilizing both one-on-one semi-structured interviews with a convenience sample of resident physicians, as well as a large scale survey. From our interviews, we learned that views on ideal Handoff Note content and format are extremely variable and that there were great distrust and dissatisfaction with current modalities (both EHR-based and Microsoft Word-based). In order to gain a more generalizable knowledge of resident physicians' preferences that could be translated into actionable items when designing the Handoff Note template, we created an online survey with both discrete data and free text elements.

Engaging resident physicians in user-centered design of electronic Handoff tools is not a new concept [5], with many utilizing committees that included several resident physician members [10; 11]. One other study has described using a survey to derive optimal content for EHR-integrated Handoff Note functionality [12]. Our institution-wide survey encompassing trainees in multiple disciplines and every year of training is the largest reported in the literature to our knowledge.

The most important content elements to resident physicians at our institution included Plan of Care, Illness Severity, and Patient Summary (along with Name, Location, and Code Status). It is interesting to note that the most important content to resident physicians was not a single piece of objective data, such as laboratory values or vital signs, but rather a free text narrative that contained their colleagues' subjective interpretation of the patient and the plan formulated based on that interpretation. Earlier work by Flanagan and Patterson looking at an EHR-integrated Handoff tool that included only objective data noted that a large portion of resident physicians cited the absence of the patient's Assessment and Plan written by a colleague as a major design flaw of that system [13]. Our similar findings regarding the importance of the Assessment and Plan mirror prior work, and perhaps speak to the fact that

the ultimate purpose of the Handoff Note is to provide the physician with a succinct summary of the patient and plan of care. Should clinical decision-making be required during the cross-cover shift because of a newly arising patient issue, this initial understanding of the patient can be further enriched as needed by objective data found within the EHR. In fact, a large portion of the free text commentary in our survey focused on the fact that should further details be necessary, resident physicians would access the EHR for data to guide decision making. They did not have any expectation that the Handoff Note would (or should) contain the level of detail appropriate for such decisions.

The remaining content varied in its importance to resident physicians. Interestingly, even the content elements that ranked least important still had at least 20% of resident physicians ranking it as Important (score of 4 or 5 – see Table 4). Another major theme discovered in the free text commentary was the dependence of certain content elements on the specific situation. Many resident physicians recognized that while certain content elements were not usually important for a cross-cover shift, in certain situations those data points were actually quite useful. This again argues for the importance of a place for free text data to be included so that resident physicians can enter these data based on their judgement in the appropriate clinical context. While automated entry of certain data directly from the EHR into the Handoff Note might prove helpful, high-quality Handoff Notes still require the skill and sound judgement of the author. Indeed, the ability to recognize important data points and succinctly surmise a complex hospitalization into only a few sentences requires sound clinical judgement and practice. In the words of one of our respondents "this is how learning happens and this is how doctors are trained". EHR-based tools to streamline and standardize the Handoff Process must take these details into account; one cannot automate wise judgement and clinical reasoning.

We recommend to system developers that EHR-integrated Handoff Notes contain, in addition to basic patient identifiers, a free text field for Patient Summary, which should include an assessment of illness severity and a free text field for Plan/To-Do List, which should include appropriate anticipatory guidance. Code Status, automatically derived from the EHR, is a useful addition to the Handoff Note. Further content elements in the Handoff Note, such as Vital Signs, Laboratory Values, or other data, should be customized according to the local needs and culture of the physician specialties and levels of training for which its use is intended. The principles of User-Centered Design should be followed, and active resident-physician participation in EHR-integrated Handoff Note development should be sought early in the design process.

Conclusions

The importance of the Handoff Process and the accompanying Handoff Note are becoming increasingly and deservedly recognized. The ideal content is rarely a specific, objective data point. Rather, succinct summaries of patients' hospital course, current trajectory, and plans of care composed in a thoughtful manner by those caring for them are the most important and useful content elements. Those attempting to construct EHR-integrated tools to guide the Handoff Note composition process must take into account that this process can never be completely automated and that sound clinical judgement and attention to detail are at the very heart of excellent Handoff Notes.

Acknowledgements

The authors wish to thank Fairview Health Services and the Office of the Dean for Graduate Medical Education at the University of Minnesota Medical Center for their support of this work.

References

[1] V. Arora, J. Johnson, D. Lovinger, H.J. Humphrey, and D.O. Meltzer, Communication failures in patient sign-out and suggestions for improvement: a critical incident analysis, *Qual Saf Health Care* **14** (2005), 401-407.

[2] A.J. Starmer, T.C. Sectish, D.W. Simon, C. Keohane, M.E. McSweeney, E.Y. Chung, C.S. Yoon, S.R. Lipsitz, A.J. Wassner, M.B. Harper, and C.P. Landrigan, Rates of medical errors and preventable adverse events among hospitalized children following implementation of a resident handoff bundle, *Jama* **310** (2013), 2262-2270.

[3] T.J. Nasca, S.H. Day, E.S. Amis, Jr., and A.D.H.T. Force, The new recommendations on duty hours from the ACGME Task Force, *N Engl J Med* **363** (2010), e3.

[4] K.M. McDonald, V. Sundaram, D.M. Bravata, R. Lewis, N. Lin, S.A. Kraft, M. McKinnon, H. Paguntalan, and D.K. Owens, AHRQ Technical Reviews, in: *Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies (Vol. 7: Care Coordination)*, Agency for Healthcare Research and Quality (US), Rockville (MD), 2007.

[5] E.G. Van Eaton, K.D. Horvath, W.B. Lober, and C.A. Pellegrini, Organizing the transfer of patient care information: the development of a computerized resident sign-out system, *Surgery* **136** (2004), 5-13.

[6] S.L. Airan-Javia, J.R. Kogan, M. Smith, J. Lapin, J.A. Shea, C.J. Dine, K. Ishida, and J.S. Myers, Effects of education on interns' verbal and electronic handoff documentation skills, *J Grad Med Educ* **4** (2012), 209-214.

[7] S.A. Collins, D.M. Stein, D.K. Vawdrey, P.D. Stetson, and S. Bakken, Content overlap in nurse and physician handoff artifacts and the potential role of electronic health records: a systematic review, *J Biomed Inform* **44** (2011), 704-712.

[8] G. Rosenbluth, J.F. Bale, A.J. Starmer, N.D. Spector, R. Srivastava, D.C. West, T.C. Sectish, C.P. Landrigan, and I.P.S.E.E. Committee, Variation in printed handoff documents: Results and recommendations from a multicenter needs assessment, *J Hosp Med* **10** (2015), 517-524.

[9] A.J. Starmer, N.D. Spector, R. Srivastava, D.C. West, G. Rosenbluth, A.D. Allen, E.L. Noble, L.L. Tse, A.K. Dalal, C.A. Keohane, S.R. Lipsitz, et al, Changes in medical errors after implementation of a handoff program, *N Engl J Med* **371** (2014), 1803-1812.

[10] L.P. Cheah, D.H. Amott, J. Pollard, and D.A. Watters, Electronic medical handover: towards safer medical care, *Med J Aust* **183** (2005), 369-372.

[11] K.M. Kochendorfer, L.E. Morris, R.L. Kruse, B.G. Ge, and D.R. Mehr, Attending and resident physician perceptions of an EMR-generated rounding report for adult inpatient services, *Fam Med* **42** (2010), 343-349.

[12] J. Breaux, R. McLendon, R.B. Stedman, R.G. Amedee, J. Piazza, and R. Wolterman, Developing a Standardized and Sustainable Resident Sign-Out Process: An AIAMC National Initiative IV Project, *Ochsner J* **14** (2014), 563-568.

[13] M.E. Flanagan, E.S. Patterson, R.M. Frankel, and B.N. Doebbeling, Evaluation of a physician informatics tool to improve patient handoffs, *J Am Med Inform Assoc* **16** (2009), 509-515.

Address for Correspondence

Elliot G Arsoniadis, MD; arson001@umn.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190401

Analysis of Voluntary User Feedback of the Swedish National PAEHR Service

Annika Bärkås [a], Isabella Scandurra [b], Maria Hägglund [a, c]

[a] *Health Informatics Centre, Karolinska Institutet, Stockholm, Sweden*
[b] *Informatics, School of Business, Örebro University, Örebro, Sweden*
[c] *Department of Women's and Children's Health, Uppsala university, Uppsala, Sweden*

Abstract

"Journalen" is a patient accessible electronic health record (PAEHR) and the national eHealth service for Sweden's citizens to gain access to their EHR. The Swedish national eHealth organization Inera, responsible for Journalen, created an inbox to receive voluntary user feedback about Journalen in order to improve the service from the user perspective. Based on voluntary user feedback via email. This study explored patients' experiences of using the national eHealth service and identified pros and cons. A mixed method content analysis was performed. In total, 1084 emails from 2016-2017 have been analyzed. 9 categories were identified, the most frequent ones related to questions about why some information was not accessible (due to regional differencies), feedback (including only positive or negative comments as well as constructive improvement suggestions), and emails about errors that user found in their record. These data can be successfully used to continuously improve an already implemented eHealth service.

Keywords:

Patient Portals; Health Record, Personal; Consumer Health Informatics; Electronic Health Records; Evaluation

Introduction

eHealth is an increasing part of healthcare worldwide. It involves a service or a tool, which purpose is to improve prevention, contribute to diagnoses and treatment, as well as monitoring and management, and is built up of information and communication technologies (ICT) [1]. An example of an eHealth solution are patient portals, which give patients access to their own health data online and can give the patient access to a patient-accessible electronic health record (PAEHR) [2]. The US initiative OpenNotes is an example of a movement striving to increase and improve patients' access to clinical notes [3], through e.g. patient portals.

Journalen is the national Swedish PAEHR service used by Sweden's citizens to read their health records online. The purpose of Journalen is to contribute to increased quality of care and cost-effectiveness, but primarily to increase patients' empowerment and engagement in their own care [4]. The service aims to give all residents from 16 years access to all information about themselves documented in tax-funded health and dental care [5]. The national eHealth organization Inera AB is responsible for the maintenance and development of the service [6]. To access Journalen, the user must log on to the national patient portal 1177 Vårdguiden through authentication using a nationally approved BankID/eIdentification [7]. Potential users of Journalen are the 10 million Swedish citizens. Currently over 2.6 million citizens use the service, with close to two million logins to the service every month [8]. Sweden has a decentralized healthcare system, i.e. regional responsibility, resulting in that patients have access to different features in the service depending on in which county council or healthcare region they have received care [5].

As Sweden is a world-leading nation in ICT solutions [9-11], it is important also for the healthcare sector that users of different eHealth services find the solutions useful. Since Sweden offers the citizens the right to access their own health documentation online, such a service should also be as easy to use and easily navigated as possible for all kinds of users. In line with the Swedish Vision for eHealth and eHealth Strategy, that all Swedish citizens should have access to all their health information online by 2020 [12, 13], Inera AB created an email inbox in order to get voluntary user feedback about the service to guide their improvement work [5]. From 2015 to spring 2018 users could provide feedback directly to the service provider after logging into Journalen by using the inbox. However, Inera could not provide emails from 2015, as they had only saved emails from 2016.

Despite a lot of interest and debate around the implementation of PAEHR, both in Sweden and internationally, there is still a lack of research and evidence around the benefits and drawbacks of these types of solutions. [14]. In Sweden, the PAEHR Journalen has been studied for various purposes, e.g. to capture patients and healthcare professionals' opinions [15], to analyse the national regulatory framework for citizens access to their health records [16, 17], to describe challenges during the implementation [7, 18] – all relevant to this study. However, there are few studies that focus on how end-users (patients) themselves experience the use of the service. This study will explore patients' experiences of using Journalen as expressed in voluntary feedback to the service provider. Focus is on the problems and benefits, and how Journalen, according to the real users, can be improved by illustrating the problems they have experienced when using the service.

Methods

A mixed method analysis was conducted, including a qualitative content analysis of the emails, and a quantitative analysis of frequency. An inductive approach has been used throughout the research, which followed an iterative process for the categorizations and analysis of the experience's patients/users expressed in the feedback emails. The data consisted of a large amount of emails, all highlighting the views from a user perspective of Journalen. Therefore, the quantitative perspective of

the study was deemed important, as it was of interest to show the number of emails belonging to each category, hence, a mixed method content analysis [19].

In the PAEHR service, the users were invited to share their experiences of the service, and how they used the information in Journalen [5]. The voluntary user feedback was analysed manually, a main-analysis made by the first author and a sub-analysis made by the second and third author. About 1084 emails were examined. The emails were obtained from the national eHealth organization Inera AB. Emails received between the years of 2016-2017 were anonymized and only the content of the emails were analysed further. Each email was read several times, and categorized into a type of email. Nine categories were created during the reading of the emails (Table 1). All emails containing positive and negative feedback, or descriptions of experiences of using the e-service were grouped into a general feedback category that was further divided into sub-categories (Figure 1). No other categories contained sub-categories.

The study was conducted in accordance with the Helsinki Convention and its ethical principles for medical research involving human subjects [20]. The voluntary user feedback contained patient opinions about the eHealth service, and are presented here as anonymous data.

Results

By analysing the emails, commonly used topics have been sorted into categories, which are presented in Table 1. We will here highlight some of the more interesting categories.

Region dependent comments

Each of the 20 regions in Sweden who provide the majority of healthcare, were given their own choice in deciding what information to make accessible to patients and when [16]. This has resulted in major differences in what information can actually be accessed depending on which region you have received care in [7].

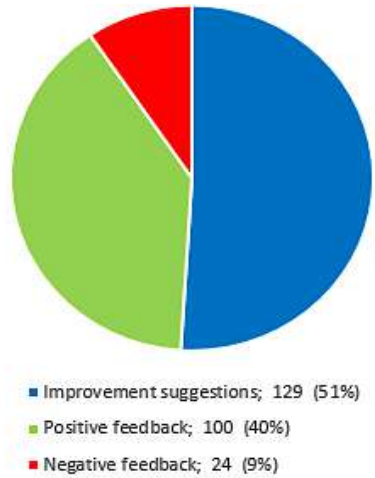


Figure 1. Presents how the feedback category is categorized, and the quantity of the three categories.

40% of the emails analysed in this study were related to users requesting access to information or features that their healthcare provider had decided to not make accessible. The majority of these emails were received during 2016 (308 emails compared to 125 in 2017), which may indicate that the situation was improved over time.

Wrong medical content

A total of 8% (N = 89) of the analysed emails concerned errors that the user had found in their record. We have not made an assessment of the severity of the identified errors, but the emails show that patients may act a quality control for the record content, but also that easier ways of reporting errors to the correct organizations are needed.

Table 1. Identified categories.

Category	Description	2016	2017	Total	
1. Region-dependent comments	The user is missing info or requesting features that do not exist due to decisions of the county council or healthcare region he/she belongs to.	308	125	433	40%
2. Feedback	Improvement suggestions are identified based on a problem the user has with the service, as well as positive and negative feedback comments.	108	145	253	23%
3. Wrong medical content	The user has detected wrong medical content, mostly in the record entries or in the medication list	36	53	89	8%
4. Seeking health care	The user wants to get in touch with his/her medical care provider/healthcare organization	16	25	41	4%
5. Requesting their health record	The user requests his/her health record in paper copies.	18	15	33	3%
6. Block or seal	The user wants to block or seal his/her health record, i.e. the entire service.	7	3	10	1%
7. Access their child	The user cannot see or access his/her child's record.	5	4	9	1%
8. Errors	Emails that are illegible and have appeared in error codes	35	22	57	5%
9. Other	Emails that are difficult to categorize because of its rare occurrence, and cannot be put into an own category	83	76	159	15%
	Total	***616***	***468***	***1084***	***100***

Feedback

This is the category, or topic, that emphasizes users' views about Journalen. The feedback category can in turn be broken down into improvement suggestions, only positive comments and only negative comments about the service, as presented in Figure 1. The feedback category is region independent unlike the first category in Table 1. Out of all feedback, 129 contains improvement suggestions, 100 only positive comments about the service, and 24 only negative comments. What is meant by 'only' positive and negative comments is when the researchers interpret that the sender has intended to only write something positive or only something negative about the service.

The users who expressed positive feedback (N = 100) (Table 2) believe that the service is a generally good service, which they embrace, and the majority of positive comments simply express that Journalen is a good service in general. About a quarter of the positive comments are praise for specific features, such as being able to access lab results, vaccinations, diagnosis, referrals, and so on. Many users want to share how good the service is, as it serves as a memory support after the appointments. Another recurring positive feedback describes how users appreciate being able to read Journalen in peace and quiet at home since it can be reached online. Journalen is also described as positive when it comes to clarifications, and 10% of the positive feedback emails describes how users check to see if any misunderstandings have occurred after an appointment. Without the e-service, they would never know this, if they would not request paper copies, which itself is a time-consuming process. Some users also expressed how good it is when all information is collected in one place.

Table 2. Categorized positive feedback about the service.

Category	Example of comments	Quantity	
A good service	*" I just want to say that the service of reading my online journal is amazingly good and useful"*	41	41%
Access to specific info	*"It is great and useful to be able to read the lab results"*	23	23%
Memory support	*"It is good to be able to read the health record, as many details are often forgotten after a doctor's appointment"*	12	12%
Peace and quiet	*"I find it calm to visit Journalen at home in peace and quiet, which is helpful because of my hearing impairment"*	10	10%
Clarifying	*"It is good to be able to read my health record online if there were any misunderstandings during the appointment"*	10	10%
One place	*"It is good that all information is collected in one place and easily accessible. Now I do not have to save and look for paper copies"*	4	4%

Emails containing only negative feedback was much rarer (N = 24). The majority of the negative comments concern Journalen as an eHealth service in general, the users express that the service is simply bad and difficult to navigate. Some comment that Journalen is not patient-safe, and that sensitive information should not be available on the internet. A few users believe that it should be up to each individual to be able to decide on their own medical information and be able to make changes in the information. Since this is not possible it is a bad system according to these users (Table 3).

Table 3. Categorized cons about the service.

Category	Example of comments	Quantity	
A bad service	*" I have been logged into it, and it is totally worthless. It is difficult to interpret and understand. It is a big joke!"*	17	71%
Not patient-safe	*"Such information as someone's medical information shout not be available online. It is not patient-safe"*	4	16%
Decide about the info	*"This is a really bad system. I want to be able to make changes in the content because it is about my identity! I want to be able to decide about the written info about me and how it is formulated"*	3	13%

The majority of the feedback category consisted of *improvement suggestions* (N = 129). Most users asked for older health records, and this can be explained by the fact that many regions decided to only give access to notes from the date Journalen was introduced (or shortly before this) which meant that much of what the users expected to see when the logged in was not actually there. This was followed by suggestions on how the interface could be improved to facilitate the navigation, such as making certain buttons more visible. Several users provided improvement suggestions for lab results, such as clarifying information of each lab result. Text reminders for new events in Journalen was also requested, and some wanted to see X-ray/MR results, which are currently not available at all in the service. Improved support for communication with the healthcare providers was also requested, for example having a direct phone number or other contact information visible. There were also some suggestions for clarifications of other specific information in Journalen, such as unclear instructions on how one's health record can be ordered in paper copies. Some of the users had opinions that they needed to see their children's information, and that some sections of Journalen required excessive number of mouse clicks to reach it. Some users considered that Journalen should be individualized and that users themselves should be able to determine what information should be accessible through the service.

A few had opinions about how the search feature in Journalen could be improved, and that only signed and validated information should be displayed to the patient, and that visible unsigned information can be considered a patient risk. Some healthcare providers choose to give immediate access to all notes, whether they are signed or not, whereas others keep unsigned notes hidden from the patient until they have either been signed or two weeks have passed, at which point the note is automatically locked for updates [21]. A few of the users considered that logs should be available to the patients, i.e. who have visited the patient's heath record. This feature is currently only activated by a few healthcare providers [22]. Another proposed improvement was to make a glossary of medical terms available in Journalen to all patients.

Discussion

The results presents users voluntarily given feedback about the service with identified pros and cons and experienced issues. There were few complaints and many positive comments. The improvement suggestions pointed to several issues and problems that the user experience, yet overall they mostly expressed positive feelings towards the service and wanted it to continue and be further developed to be even better. This is well in line with a recently published survey study among users of Journalen, in which 96% of the respondents believed that this was a good intervention. [22]. Compared to this figure, the emails analyzed in our study had a much higher expression of negative feedback, however this is likely related to the data source; when asked to voluntarily submit feedback a user who is very negative to an e-service may be more inclined to answer than someone who is positive.

The users' comments about their experienced benefits and problems can also be supported by an earlier interview study which aimed to provide in-depth understanding of cancer patients' attitudes and experiences of online health records [15], and other international studies within this field, for example the OpenNotes initiative in US which focus on patients ability to view clinical notes [3]. The study found that users of PAEHR services uses the service to see that they understood the information they received from the physician correctly. Another study [23] also found that one of the biggest reasons why patients read their health record entries are to be sure that they understood what the care provider said. This is also expressed by some of the users of Journalen in this study who believe that it is very good to be able to read their health records in retrospect to prevent misunderstandings (Table 2).

The study [15] emphasizes that most of the participants find that PAEHR services serve as a memory support, since it is easy to forget information from a physician's appointment. Similarly, several of the users of Journalen who emailed feedback expressed that it was a very good memory support to be able to read on their own what has been said and done during the appointments, as there is often a lot of information provided that is easy to forget about afterwards. Below is a quote from one of the users of Journalen (Table 2):

"It is good to be able to read the health record, as many details are often forgotten after a doctor's appointment"

In a study from the US [24], the purpose was to gain more knowledge about patient experiences where patients/users of a PAEHR service could provide feedback on their visit notes through a specific feedback tool within the service. The conclusion showed that patients can relate to personal, relational and safety benefits. Demand for similar features is also found in the results of this study. Below is a quote from one of the users of Journalen (Table 3):

"I want to be able to make changes in the content because it is about my identity! I want to be able to decide about the written info about me and how it is formulated"

Having other studies [3, 15, 23-24] with similar feedback reinforces the feedback given by the users of Journalen on how a PAEHR service can be improved and made as easy and clear as possible.

Methodological concerns

The data source needs to be considered when discussing the results of this study. We have analysed voluntary feedback that has been sent by the users of Journalen over two years. How representative this information is of the opinions of all users of Journalen is impossible to say, and therefore the numbers presented in Table 1 cannot be interpreted as valid for all users. As an example, the 8% of emails concerning errors in record does not mean that 8% of all records have errors in them nor that 8% of all users find errors. Yet, that so many individuals have made the effort to send an email about this issue could be considered an important indicator that this is an area to be further explored.

We also consider the fact that this information is voluntarily provided by users without being asked specific questions makes the data source especially interesting. This is feedback given without influence of how questions are formulated, fatigue from filling out a long survey or a willingness to please the researchers asking the questions.

Another limitation that could impact the outcome are the manually counting of the emails. However, the format of the email feedback required an aspect of human interpretation, sometimes the latent content of an email, the tone and sentiment of the message, was crucial for the result of the study. Using an automatic quantification by a software tool would have risked missing some of these subtleties of the material.

Voluntary feedback as formative evaluation

The feedback received through the email inbox have been continuously used by the developers of the e-service to improve it. One example is that since such a high rate of emails concerned missing information, it became clear that the users had trouble understanding *why* parts of their record where not shown. In order to make this more clear to the users, an interactive map of Sweden was implemented in the e-service showing which regions made what information accessible. When choosing a specific part of the record (e.g. diagnosis or lab results), the map will also automatically show which regions make this information available (figure 2). This easily accessible visualization of information access may be the reason for the decline in emails in this category during 2017 compared to 2016.

Figure 2 – map over which healthcare providers show lab results

This may be one way of using rapid, unstructured user feedback as part of formative evaluations to facilitate continuous improvement of an eHealth service.

Future research

The content of the voluntary feedback would be interesting to compare in more detail to results of a recently published survey among all users of Journalen [22]. How does the voluntary feedback correspond to the results obtained when users are asked direct questions about specific features? Repeated surveys should also be performed, as the eHealth service is continuously improved and a steady increase of users can be seen. We need more long-term follow-up of the users experiences, e.g. through re-occuring surveys among the users. As the improvement suggestions are handled in the continued

development of the service, we ought to be able to track and see improvements also in the users feedback. Exploring the usefulness of the voluntary feedback compare to other forms of feedback or evaluations for the developers of Journalen would be of great interest from a methodological perspective.

Conclusions

This research is the first study in Sweden that evaluates and analyses the users' voluntary submitted opinions about Journalen and uses these to emphasize improvement suggestions about the service. These data can be successfully used to continuously improve an already implemented eHealth service. The results shows that having healthcare providers that makes different information accessible through the same eHealth service, will cause confusion and questions from users, but most feedback received was positive, which is in line with other research into patients access to their EHR. The results also shows that the users are channeling their questions to Inera instead of to their healthcare provider, which means other communication paths are needed between patient and healthcare provider.

Acknowledgements

FORTE – the Swedish Research Council for Health, Working Life and Welfare funds the research project "PACESS" (2016-00623). This research was supported by Uppsala MedTech Science & Innovation (www.medtech.uu.se), a joint strategic initiative between Uppsala University and Uppsala University Hospital. We would also like to thank Inera AB (www.inera.se) for providing the data from the voluntary feedback.

References

1. European Commission. eHealth [internet]. Retrieved 2018 Mars 14 from: https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/2017/05/30/eHealth
2. Irizarry T, DeVito Dabbs A, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res. 2015; 17(6): e148. doi: 10.2196/jmir.4255
3. Payne TH, Beahan S, Fellner J, Martin D, Elmore JG. Health Records All Access Pass. Journaln of AHIMA 87, no.8 (August 2016): 36-39.
4. Inera AB. Journalen [internet]. Retrieved 2018 Mar 13 from: https://www.inera.se/tjanster/journalen
5. Inera AB. Journalen [internet]. Retrieved 2018 Oct 28 from: https://www.inera.se/tjanster/journalen/Journalen/#b252ed1d-a977-4dc5-ba63-a0e0ca9a028d
6. Kjellin Lagerqvist N. Inbox for User Feedback as a Method for Evaluating User Experience - A Method Evaluation in Scrum based Software Development [master's thesis]. Uppsala: Uppsala Universitet; 2015.
7. Hägglund M, Scandurra I. Patients' Online Access to Electronic Health Records: Current Status and Experiences from the Implementation in Sweden. Stud Health Technol Inform. 2017; 245:723-27.
8. Inera AB. Statistik Journalen [internet]. Retrieved 2018 Oct 27 from: https://www.inera.se/aktuellt/statistik/journalen_statistik/
9. Baller S, Dutta S, Lanvin B (eds.). The Global Information Technology Report 2016 – Innovating in the Digital Economy. Geneva: World Economic Forum; 2016.
10. Dutta S, Geiger T, Lanvin B (eds.). The Global Information Technology Report 2015 – ICTs for Inclusive Growth. Geneva: World Economic Forum; 2015.
11. Bilbao-Osario B, Dutta S, Lanvin B (eds.). The Global Information Technology Report 2014 – Rewards and Risks of Big Data. Geneva: World Economic Forum; 2014.
12. Ministry of Health and Social Affairs. Vision for eHealth 2025. Stockholm. Retrieved 2018 Jan 30 from: http://www.government.se/information-material/2016/08/vision-for-ehealth-2025/
13. Government Offices of Sweden. Nationell IT-strategi för vård och omsorg (Skr 2005/06:139). Stockholm. Retrieved 2018 Jan 30 from: http://www.regeringen.se/49b69c/contentassets/fefd0a1492f548c2ab1451e45b7f6463/nationell-it-strategi-for-vard-och-omsorg-skr.-200506139
14. Conway N, Webster C, Smith B, Wake D. eHealth and the use of individually tailored information: A systematic review. Health Informatics J. 2017; 23(3): 218-33.
15. Rexhepi H, Åhlfeldt RM, Cajander Å, Huvila I. Cancer patients' attitudes and experiences of online access to their electronic medical records: A qualitative study. Health Informatics J. 2018; 24(2):115-124. doi:10.1177/1460458216658778
16. Scandurra I, Lyttkens L, Eklund B. Implications of Swedish National Regulatory Framework of the Patient Accessible Electronic Health Record. Stud Health Technol Inform. 2016; 228:695-9.
17. Scandurra I, Pettersson M, Eklund B, Lyttkens L. Analysis of the Updated Swedish Regulatory Framework of the Patient Accessible Electronic Health Record in Relation to Usage Experience. Stud Health Technol Inform. 2017; 245:798-802.
18. Hägglund M, Scandurra I. A Socio-Technical Analysis of Patient Accessible Electronic Health Records. Stud Health Technol Inform. 2017; 244:3-7.
19. Hamad EO, Savundranayagam MY, Holmes JD, Kinsella EA, Johnson AM. Toward a Mixed-Method Research Approach to Content Analysis in The Digital Age: The Combined Content-Analysis Model and its Applications to Health Care Twitter Feeds. J Med Internet Res. 2016;18(3): e60. doi: 10.2196/jmir.5391
20. World Medical Association. WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects [internet]. Retrieved 2018 May 4 from: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/
21. Hägglund M, Moll J, Åhlfeldt R-M, Scandurra I. Timing It Right – Patients' Online Access to Their Record Notes in Sweden. Stud Health Technol Infor. 2018;247:336-340.
22. Moll J, Rexhepi H, Cajander Å, Grünloh C, Huvila I, Hägglund M, et al. Patients' Experiences of Accessing Their Electronic Health Records: National Patient Survey in Sweden. J Med Internet Res. 2018; 20(11):e278.
23. Bell SK, Mejilla R, Anselmo M, Darer JD, Elmore JG, Leveille S et al. When doctors share visit notes with patients: a study of patient and doctor perceptions of documentation errors, safety opportunities and the patient-doctor relationship. BMJ Quality & Safety. 2016;26: 257-58. doi: 10.1136/bmjqs-2016-005641
24. Gerard M, Fossa A, Folcarelli P, Walker J, Bell SK. What Patients Value About Reading Visit Notes: A Qualitative Inquiry of Patients Experiences With Their Health Information. J Med Internet Res. 2017;19(7): e237. doi: 10.2196/jmir.7212

Address for correspondence

Maria Hägglund, Department of Women's and Children's Health, Uppsala University, Dag Hammarskjölds väg 14B, 752 37 Uppsala, Sweden. Email: maria.hagglund@kbh.uu.se

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190402

Representation of the Transgender Population in Electronic Health Records: Implementation Strategy in the Public Health Care System of Buenos Aires City

Denise Levi[a], María Victoria Vazquez[a], María Victoria Giussi[a], S. Esteban[a], Analía Baum[a]

[a] *Health Information Systems Office, Ministry of Health, Buenos Aires City, Argentina*

Abstract

Electronic Health Records (EHR) face the challenge of collecting data about patient's gender identity in order to provide standardized and quality data to manage public policies in order to improve health disparities of the Trans population. Innovation in Health Information Technologies (HIT) develops in the midst of a cultural change process related to social representations of gender in favor of a diversity perspective. Understanding the health field as a complex adaptive system, the changes in the systems must consider multiple factors in every stage of the process. It is relevant to consider the people involved in it and the culture in which they are embedded. This article describes the implementation strategies of functionality that represents the transgender population in the EHR of the Public Health Care System of the Buenos Aires City.

Keywords:

Electronic Health Records; Transgender Persons; Cultural Evolution

Introduction

Transgender people are individuals whose gender identity differs from their sex assigned at birth, and may involve the modification of their appearance or body function through pharmacological, surgical or other means. In recent years, the Electronic Health Record (EHR) faced the challenge of representing the transgender population. Correctly identifying this population allows longitudinal health records that facilitate continuity of health care. On the other hand, it is indispensable to have standardized data in order to obtain quality metrics and create research opportunities that contribute to the management of public health policies for this vulnerable population.

On the legal level, in different countries transgender people can change their name and gender in their identity documents. However, the laws vary in different countries and localities. Since the process of opening a patient's chart in the EHR depends on the patient's identification, changing the name or gender of any individual represents a challenge when it comes to ensuring the longitudinality of the clinical record.

Articles from different countries mention how the systematic legal, economic, and social exclusion of trans people negatively affects every area of their lives: family, health, education, work, housing, and security. Their living conditions often face poverty, precarious housing or homelessness, and inequality in education, work and health.

Experiences of stigma and discrimination have been described in various areas, highlighting health in particular, protagonized by both administrative staff and health care professionals. The most feared situation that trans people describe is to be called by the name that appears in the person's legal documentation instead of the chosen one. This produces discomfort and ridicule that can lead to absenteeism of future healthcare appointments [1-3]. These experiences have a negative impact on patient satisfaction, the patient-professional relationship, and can also lead to a risk of verbal and physical assault by other patients in the waiting rooms.

In Argentina in 2012, National Law 26.743 "Gender Identity" was passed. This law granted people the right to be identified according to their self-perceived gender in the instruments that accredit their identity, among other rights [4].

In the United States in 2016, The Centers for Medicare & Medicaid Services (CMS) and The Office of the National Coordinator for Health Information Technology (ONC) announced that every entity certified under the Meaningful Use incentive program, will be requested to create structured fields allowing users to register, change, and access data on their gender identity [5]. Since then, several recommendations and guidelines have been developed in order to ensure efficient documentation of these data. It has been detailed how to request, store and display the data in the user's interface [6-8]. Furthermore, the incorporation of structured fields for the recollection of the chosen name, gender identity, sex assigned at birth and pronoun preference as demographic variables have been described [9-11]. Other research inquired about the perspectives of Trans people and healthcare professionals in relation to providing or having gender identity data [12-15]. Even controlled vocabularies for statistical and epidemiological analyzes such as the International Classification of Diseases 11th Edition (ICD-11) of the World Health Organization (WHO), proposed a new chapter related to sexual health conditions. The proposal was to include "Gender Inconsistency" as diagnostic category, instead of as a mental disorder as described in ICD-10 [16, 17]. The same happened with the Diagnostic and Statistical Manual of Mental Disorders (DSM) on its 5th edition [18]. Since 2009, people and organizations from different countries have set up an International Network for the Depathologization of Trans Identities. The goal has been to raise awareness and expose the psychiatrization of gender diversity and request the withdrawal of the aforementioned diagnostic categories from the ICD and DSM, previously classified as mental disorders [19].

A process of cultural change is currently gestating in favor of the recognition of the diversity of gender. It implies a modification of structures and social representations. It is important to take into consideration the cultural aspect, since it constitutes a system of shared meanings from which people interpret reality, define their way of acting and how to proceed. Otherwise, the mere IT tool that allows the recording of data in question would be insufficient. The goal of this paper is to describe the strategy that was used in the Public Health System of the Autonomous City of Buenos Aires for the implementation of functionality that represents the transgender population in the EHR.

Methods

Setting

The City of Buenos Aires has an area of 204 km² inhabited by 3,059,122 people. The health care system has 3 main sectors: The public sector, social security, and the private sector. These 3 can often coexist. The public health care system strives to offer medical access to all its inhabitants.

According to 2016 statistical data, 20% of Buenos Aires' population can only access healthcare through the public system. This percentage rises to 50% within the most unprivileged who live in the south of the city [20]. And it scales up to 78% among the transgender population [1].

Software

Since June 2016 an EHR has been gradually implemented within the outpatient setting [21]. In 2017, the "Self-perceived Name" functionality was incorporated, which allowed the user to to register a patient's preferred name in relation to their gender identity, when this names differs from their identity document.

Subsequently, in all the places where the person's name displayed in the EHR, the self-perceived name is displayed, with the initial letter of the identity document in parentheses next to the self-perceived name. All data related to the patient can be found within the personal data sheet. From this sheet, the name registered in the person's documentation can be extracted in order to be used in billing processes, prescription of medication or other legal operations.

Study Design

This is a descriptive study about the implementation strategies of a functionality that represents the transgender population in the EHR of the Public Health System of Buenos Aires City.

Results

The implementation strategy of this functionality included the following aspects:

1. Intersectoral approach

We convened the General Directorate of Coexistence in Diversity of the Undersecretariat of Human Rights and Cultural Pluralism, with dependence on the Vice-Government Office. Its responsibilities are to work on public policies and spaces for promotion, prevention and protection to reduce acts of discrimination and violence towards the people of the various groups that cohabit in the Autonomous City of Buenos Aires.

2. Profile stratification

The diversity of actors was stratified in different profiles to personalize the content. The stratification was decided according to the fulfilled function;

- Implementers: train end-users of the healthcare information system.
- Help Desk Operators: provide technical and functional support to the system users' queries.
- Administrative Staff: the patient's first point of contact when interacting with the healthcare system and are in charge of collecting the identification data.
- Healthcare professionals: professionals from different disciplines that provide healthcare.

3. Preparation of materials

Specific materials were developed for each profile, including: instructions explaining how to operate the new functionality, specific conceptual content of this topic (Table 1) and its legal framework. This was also made available on the Primary Health Care Web Portal, an intranet service that provides a communication space among healthcare workers.

4. Institutional stratification

The institutions that had the largest Trans population were identified in order to intensify the training.

5. Training

Two types of training were planned for all profiles:

- Face-to-face group training: a 3-hour workshop format was implemented. The aim was to raise awareness about gender issues. To accomplish this, different triggers were used to get in touch with the actors' perceptions which allowed to explain the concepts of gender binarism and heteronormativity. Then the concept of gender identity was introduced and it was differentiated from the concept of sexual orientation. Living conditions and health indicators for the Trans population were shared and the legal framework was deepened.
- Individual virtual training: several courses taken by administrative personnel and healthcare professionals were modified. The courses that were modified related to the use of applications for identifying people, granting and managing shifts, and the EHR application.

6. Communication

The communication strategy had 2 main recipients:

- Users of the information system: a pop-up was introduced in the login interface to the system notifying about the new functionality with a link to materials and instructions. In addition, a message to all users was posted on the Web Portal, emphasizing not only the opertation part, but also conceptual issues and regulatory framework.
- Users of the healthcare system: the Trans population was informed that they can register with their self-perceived name in the public healthcare system.

As of November 2018, there are a total of 170 Self-perceived Name records (Table 2).

Table 1 - Illustrative content about gender issues

Concept	Definition
Gender identity	The internal and individual experience and feeling of each person's gender which may or may not correspond to the sex assigned at the time of birth.
Biological Sex	The set of chromosomal, anatomical and physiological information that differentiates men and women. It is used to assign gender at birth.
Intersex	People whose anatomy does not fit the standards of female or male biological sex. It makes reference to the corporal diversity.
Gender expression	It involves everything that is communicated from our gender to others: clothing, hairstyles, body language, our interactions and social roles.
Gender binarism	It is a social construct that classifies people, limiting themselves to fulfilling feminine or masculine roles based on their anatomy.
Cisgender	People whose gender identity matches the sex assigned at birth.

Transgender	People whose gender identity differs from the sex assigned at birth.
Sexual Orientation	The ability of a person to feel emotionally, sexually and affectively attracted by a person of the same gender (homosexual), of the opposite gender (heterosexual), of more than one gender (bisexual).
Queer	Sociocultural movement that considers gender and sexual orientations as a social construct and seeks the disappearance of this categorization.
Heteronorm-ativity	The social rule that assumes the alignment of biological sex, gender identity and heterosexuality as desirable and normal.

Table 2– Self-perceived Name records in EHR

Year	Month	Records
2017	July-August	8
	September-October	13
2018	November-December	15
	January-February	12
	March-April	21
	May-June	36
	July-August	36
	September-October	29
	Total	**170**

Discussion

As seen from the diagnostic categories of ICD-10 and DSM-5, historically the transgender population has been considered in terms of abnormality, anomaly or disorder [19]. In recent years, questioning of gender binarism and heteronormativity have given rise to a broader perspective of diversity, thus beginning a process of cultural change. Even so, these changes are incipient. The lack of legal frameworks in many countries and the persistence of stigma and discrimination affect the quality of life of this population in various areas of their lives.

This paper describes the implementation strategies for functionality that allows the identification of transgender people according to their self-perceived gender identity. Multiple strategies were contemplated with an emphasis on raising awareness on this matter, going beyond the use of the system. The intersectoral approach made it possible to have the resources and means necessary for communication and training. It was also key to have specialists involved in the subject matter for the preparation of contents and materials in order to provide training that encouraged active participation. This helped to address people's concerns, questions, and understand their emotions given the complexity of the subject.

Initially, when face-to-face group trainings were conducted, the perceptions and previous knowledge about being a woman or a man were explored. The goal was to visualize the representations of social models of the binary gender system and understand them as a social construct were people are classified into two unique categories – feminine and masculine – based on biological differences, and also linked to social roles, behavioral expectations, sexual orientation, etc, which helps to maintain control and establish social order. After this collaborative work, the gender diversity perspective was introduced. This way of presenting the content was planned as a strategy to minimize resistance, using a deconstructive current. This current established that only by revealing the apparently hidden, and previous pragmatic elucidation of the success of an ideological system, there is a chance to actually glimpse the questioning or disarticulation of a socio-cultural structure and dynamics in view of its transformation or construction of a new meaning [22].

On the other hand, the strategy of stratifying profiles had its foundation in understanding health organizations as complex adaptive systems, shaped by dynamic networks of actors interconnected with different roles, functions, conflicts and interests [23]. Therefore, the contents were customized with a focus on the fact that implementers' and help desk operators' profiles should have tools to train all user profiles. The administrative profile is the patient's first contact with the healthcare system and they are in charge of collecting identification data. It was also taken into consideration that within the medical profile, the biologicist conception usually predominates. At the same time, these institutions are characterized by the concentration of power at their bases and not only in the hierarchical areas that are laid out on the formal organization chart. The governance in relation to both administrative and professional workers has a particular complexity. A great deal of emphasis was placed on the legal framework – the National Gender Identity Law No. 26,743 and the National Law No. 23592 Antidiscrimination – and how the system allows respecting the provisions of the law [9].

The stratification of institutions with the greatest Trans population was carried out with the aim of reinforcing training. Surprisingly, we found that they were the users with the highest awareness and had the best manners when it came to Trans patients. However, since they were familiar with the patients, many users indicated that they did not register Trans patients' data in the system since they were well known by every worker in the healthcare institution. In these cases, the training was directed to the importance of the registry of data, explaining that other institutions of the public healthcare system network can also access to this information.

Conclusions

Incorporating structured fields to the EHR to represent Trans people with standardized data that allow a correct identification of this population is essential to ensure the continuity of care. However, even though systems accompany the cultural changes, the computer tool alone do not transform people's practices. Therefore, it is important to consider the human factor and the culture in which people are embedded as a part of the strategy to implement functionality that meets the neds of the Trans population.

Limitations and Future Direction

Future research is needed to evaluate the implementation and perceptions of patients in relation to the quality of care in healthcare institutions, as well as the development of indicators to monitor the quality of the registry. In addition, other lines of work should aim to relieve specific information needs that are required at the EHR level for this population's healthcare.

Acknowledgments

We are grateful to the General Directorate of Coexistence in Diversity of the Undersecretariat of Human Rights and Cultural Pluralism, under the Deputy Chief of Government of the City of Buenos Aires, Argentina. Particularly to David Cohen and Cecilia Rodríguez for their review of this article that documents

the joint work done in the implementation of this new functionality.

References

[1] Fundación Huésped, ATTTA. "Ley de identidad de género y acceso al cuidado de la salud de las personas trans en Argentina". 2012

[2] Bradford, Judith et al. "Experiences of transgender-related discrimination and implications for health: results from the Virginia Transgender Health Initiative Study" American journal of public health vol. 103,10 (2013): 1820-9.

[3] Grant JM, Mottet LA, Tanis J. Injustice at Every Turn: A Report of the National Transgender Discrimination Survey. Washington, DC, National Center for Transgender Equality and National Gay and Lesbian Task Force, 2016

[4] Ley Nacional Argentina de IDENTIDAD DE GÉNERO Nro. 26.743. In 2012. p. 9–12.

[5] 2015 Edition Health Information Technology (Health IT) Certification Criteria, 2015 Edition Base Electronic Health Record (EHR) Definition, and ONC Health IT Certification Program Modifications. Final rule. Office of the National Coordinator for Health Information Technology (ONC), Department of Health and Human Services (HHS) Fed Regist. 2015 Oct 16; 80(200): 62601–62759.

[6] Deutsch, Madeline B et al. "Electronic medical records and the transgender patient: recommendations from the World Professional Association for Transgender Health EMR Working Group" *Journal of the American Medical Informatics Association : JAMIA* vol. 20,4 (2013): 700-3.

[7] Deutsch MB, Buchholz D. Electronic health records and transgender patients-practical recommendations for the collection of gender identity data. J Gen Intern Med. 2015;30:843–847

[8] Frangella J, Otero C, Luna D. Strategies for Effectively Documenting Sexual Orientation and Gender Identity in Electronic Health Record. Stud Health Technol Inform [Internet]. 2018 [cited 2019 Mar 28];247:66–70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29677924

[9] Cahill SR, Baker K, Deutsch MB, et al. Inclusion of sexual orientation and gender identity in stage 3 meaningful use guidelines: a huge step forward for LGBT health. LGBT Health. 2016;3:100–102

[10] Callahan EJ, Sitkin N, Ton H, Eidson-Ton WS, Weckstein J, Latimore D, et al. Introducing sexual orientation and gender identity into the electronic health record: One academic health center's experience. Acad Med. 2015;90:154–60

[11] Imborek, Katherine L et al. "Preferred Names, Preferred Pronouns, and Gender Identity in the Electronic Medical Record and Laboratory Information System: Is Pathology Ready?" *Journal of pathology informatics* vol. 8 42. 3 Oct. 2017, doi:10.4103/jpi.jpi_52_17

[12] Dunne, M J et al. "Interviews with Patients and Providers on Transgender and Gender Nonconforming Health Data Collection in the Electronic Health Record" *Transgender health* vol. 2,1 1-7. 1 Feb. 2017, doi:10.1089/trgh.2016.0041

[13] Patient Perspectives on Answering Questions about Sexual Orientation and Gender Identity: An integrative review. R. I. Bjarnadottir, W. Bockting, D. Dowding J Clin Nurs. 2016 Oct 5 Published online 2016 Oct 5. doi: 10.1111/jocn.13612

[14] Thompson "Patient Perspectives on Gender Identity Data Collection in Electronic Health Records: An Analysis of Disclosure, Privacy, and Access to Care" *Transgender health* vol. 1,1 205-215. 1 Oct. 2016, doi:10.1089/trgh.2016.0007

[15] Is It Okay To Ask: Transgender Patient Perspectives on Sexual Orientation and Gender Identity Collection in Healthcare. Allysha C. Maragh-Bass, Maya Torain, Rachel Adler, Anju Ranjit, Eric Schneider, Ryan Shields, Lisa M. Kodadek, Claire Snyder, Danielle German, Susan Peterson, *et al.* Acad Emerg Med. 2017 Feb 24 Published online 2017 Feb 24. doi: 10.1111/acem.13182

[16] World Health Organization, International Classification of Diseases 10th Edition, 1992.

[17] WHO "WHO releases new International Classification of Diseases (ICD 11)". Available in: https://www.who.int/news-room/detail/18-06-2018-who-releases-new-international-classification-of-diseases-(icd-11) [Accessed on March, 20, 2019]

[18] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders 5th Edition, 2013.

[19] Campaña Internacional "Stop Trans Pathologization". Available in: http://stp2012.info/old/es [Accessed on March 20, 2019]

[20] ASIS 2016, Gobierno de la Ciudad de Buenos Aires. Available in: http://www.buenosaires.gob.ar/sites/gcaba/files/asis_caba_2016_dic17_vf_1.pdf

[21] Giussi MV, Baum A, Plazzotta F, Muguerza P, González Bernaldo de Quirós F. Change management strategies: transforming a difficult implementation into a successful one. Stud Health Technol Inform. 2017;245:813-817. PMID: 29295211. DOI: 10.3233/978-1-61499-830-3-813. Available in: http://ebooks.iospress.nl/publication/48264

[22] AYALA ARAGON, Oscar Arnulfo. La deconstrucción como movimiento de transformación. *Ciencia, docencia y tecnología.* [online]. 2013. Available in: http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1851-17162013000200003&lng=es&nrm=iso&tlng=es [Accessed on March 20, 2019]

[23] Sittig, Dean F and Hardeep Singh. "A new sociotechnical model for studying health information technology in complex adaptive healthcare systems" Quality & safety in health care vol. 19 Suppl 3,Suppl 3 (2010): i68-74.

Address for correspondence

Levi Denise Paula

Ministry of Health, Autonomic City of Buenos Aires, Argentina.

E-mail: dlevi@buenosaires.gob.ar

Phone Number: +54 9 11 54987439

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190403

How Does GDPR Support Healthcare Transformation to 5P Medicine?

Bernd Blobel[a,b,c], Pekka Ruotsalainen[d]

[a] *Medical Faculty, University of Regensburg, Regensburg, Germany*
[b] *eHealth Competence Center Bavaria, Deggendorf Institute of Technology, Deggendorf, Germany*
[c] *First Medical Faculty, Charles University Prague, Prague, Czech Republic*
[d] *Tampere University, Tampere, Finland*

Abstract

Health systems advance towards personalized, preventive, predictive, participative precision (5P) medicine, considering the individual's health status, contexts and conditions. This results in fully distributed, highly dynamic, highly complex business systems and processes with multiple, comprehensively cooperating actors from different specialty and policy domains, using their specific methodologies, terminologies, ontologies, knowledge and skills. Rules and regulations governing the business process as well as the organizational, legal and individual conditions, thereby controlling the behavior of the system, are called policies. Trust and confidence needed for running such system are strongly impacted by security and privacy concerns controlled by corresponding policies. The most comprehensive policy dealing with security and privacy requirements and principles in any business collecting, processing and sharing personal identifiable information (PII) is the recently implemented European General Data Protection Regulation (GDPR). This paper investigates how GDPR supports healthcare transformation and how this can be implemented based on international standards and specifications.

Keywords:
European data protection, governing, privacy

Introduction

In the course of methodological paradigm changes, healthcare systems advanced from empirically describing health problems with one solution fits all to dedicated care, stratifying population for specific, clinically relevant conditions resulting in evidence-based medicine. Stratifying population by risk profiles, the current phase of healthcare transformation towards personalized, preventive, predictive and participative precision (5P) medicine considers the individual's health state, conditions and contexts, thereby integrating research and practice. Conditions and contexts include legal, social, environmental, occupational, or any other context. Disciplines or domains engaged in 5P medicine cover medicine, natural sciences, social sciences, engineering, etc., considering the individual from elementary particle to society. The required knowledge sharing and cooperation of multiple stakeholders from different domains using their methodologies, terminologies and ontologies establishes a legal, cultural and language challenge. As individual health state, conditions and contexts are highly dynamic, it is impossible to predefine the business systems, its processes and policies for meeting the business system objectives comprehensively, uniform and legally binding in a static way. Thus, 5P medicine requires the automated management of multiple dynamic domains including multiple dynamic policy domains. [1]

The paper investigates a) how the European General Data Protection Regulation (GDPR) [2] reflects architecture and policies of transformational healthcare systems and b) how GDPR must be implemented to support healthcare transformation. For that purpose, GDPR is structurally and functionally analyzed. For meeting principles and services required by GDPR, standards, specifications and products are recommended and explained in some details.

Methods

For analyzing such complex settings like 5P medicine under dynamically changing perspectives and contexts, a system-theoretical approach is used. According to IEEE 1471-2000, a system is a collection of components organized to accomplish a specific function or a set of functions to realize the business objectives of that business system intended by the involved stakeholders [3]. Systems interact with their environment and can be decomposed to subsystems or composed to super-systems in a recursive way. A system is defined by the system's architecture, i.e., its components, their functions and relations on the one hand, and the system's behavior represented by the system's policy on the other hand. The term policy implies any set of rules for selecting components and functions as well as constraints of the relations according to a business case, thereby controlling the behavior of that system. It represents the perspectives of all domains involved, i.e., process policies, legal constraints, individual preferences, resource management, etc. This behavioral description applies to any of the aforementioned subsystem or super-system.

A security and privacy policy according to ISO 22600 Health informatics – Privilege management and access control [4] is a complex of legal, organizational, functional, social, ethical and technical aspects to be considered in the context of privacy and security. It defines a framework, privileges and obligations, but also consequences and penalties when the regulations are ignored.

Ecosystems are structured systems and communities of living and non-living components, which follow specific rules (policies) and interact as unit among themselves or with their physical environment.

The approach is based on the Generic Component Model (GCM) introduced by the first author in the mid-nineties [5, 6].

How Does GDPR Reflect 5P Medicine?

While the EU Data Protection Directive (Directive 95/46/EC) [7] defined just direct or indirect identifiers as personal data, GDPR extends in Art. 4(1) that definition including "one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity" of a natural person [2]. That way, GDPR defines data characterizing an individual's health status, context and conditions in the sense of 5P medicine as personal data. The extended definition of processing in GDPR Art. 4(2) meets contrary to the one in Directive 95/46/EC, the process of 5P medicine as well. GDPR Art. 6(3) requires explicit policies, which have to be represented formally and machine-processable and bound to the information objects or process steps (Art. 4(20) and Art. 26(1)). Those policies must represent the current context, and – because that context is changing – must be managed dynamically (Art. 26(1)). For guaranteeing comprehensive interoperability, policies must be represented using terminologies and ontologies of the addressed audience (Art. 7(2) and Art. 12(1)), that way meeting the requirements of 5P medicine ecosystems. Design and management of that highly complex and highly dynamic ecosystem have to address the security and privacy perspectives according to the ISO 23903 Interoperability Reference Architecture [8], ISO 22600 [4] mentioned before, or ISO 21298 Health informatics – Structural and functional roles [9]. Properly managing the dynamic system in that respect proactively, a permanent risk analysis must be performed, turning the Data Protection Officer from a checkbox marker to a risk manager, directly intervening in the business system and processes throughout the complete system lifecycle presented in the next section.

How to Define a GDPR-Compliant 5P Medicine Ecosystem?

A GDPR-compliant 5P medicine ecosystem must be designed and implemented in an automated process as an architecture-centric, ontology-based, policy-driven multi-domain business system. This has to be done in a formalized and standardized way. Starting point of an appropriate solution is the ISO Interoperability Reference Architecture model and framework developed by the first author. For representing a systems model and framework manageable by engineers, the formal representation of an n-dimensional concept space deploying universal type theory, thereafter refined to a parametrized Barendregt Cube [10, 11], has been transformed in a 3-dimensional system engineering model (Figure 1a). One dimension covers the generic granularity levels or composition/decomposition of any system. The second dimension addresses the system development process following standardized approaches such as ISO 10746 Information technology – Open distributed processing – Reference model [12], SOA (Service-Oriented Architecture) methodology, or the Unified Process (formerly Rational Unified Process). The third dimension concerns the different perspectives on the multi-disciplinary business system in question represented by the variety of domain experts involved in the business process. The concepts have to be represented for each component using the related domain ontologies. The different domains can be properly refined into subsystems with different (sub)ontologies. In the context of GDPR-compliant ecosystems, the policy domain combines different perspectives on ruling the system such as the medical process policy domain, the contextual policy domain, and the administrative/ organizational policy domain managing resources, etc (Figure 1b). The contextual policy domain can be furthermore refined into the legal and regulatory domain, the personal policy domain, and the conditional/contextual domain representing environmental, social, occupational or other contexts.

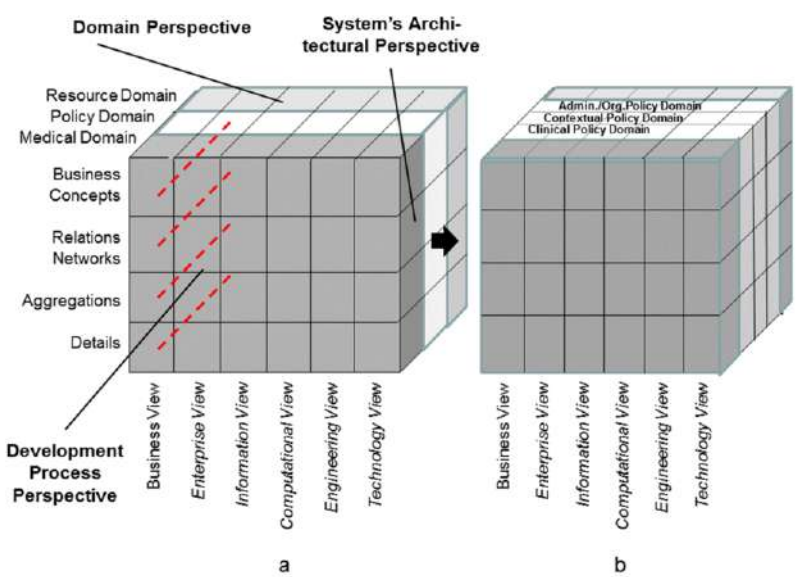


Figure 1. ISO Interoperability Reference Architecture [9]

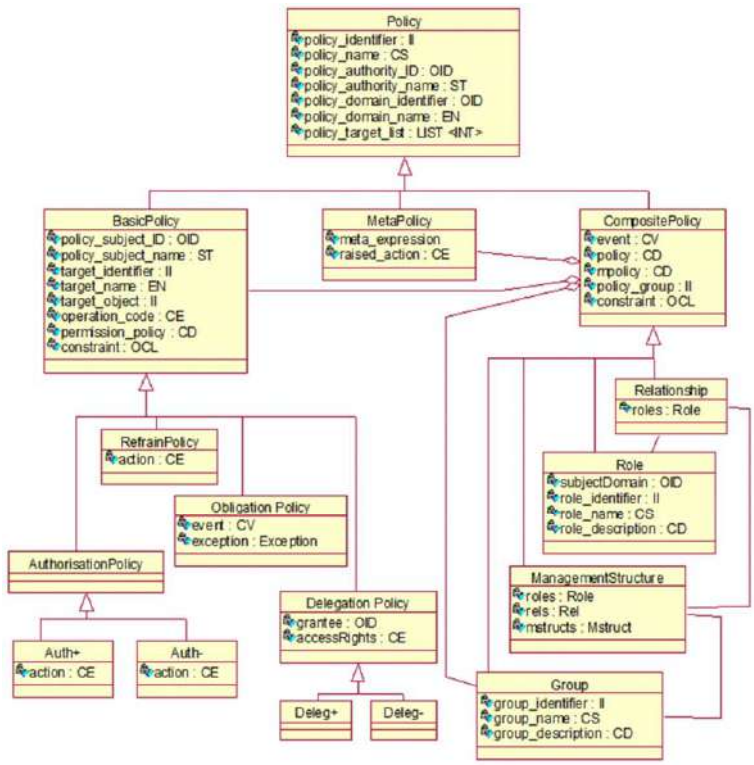


Figure 2. Policy Ontology according to ISO 22600 [4]

For representing policies, the first author has developed a policy ontology based on Damianou's Ponder Policy Specification Language [13]. That policy ontology has been standardized in ISO 22600 Health informatics – Privilege management and access control [4] (Figure 2). The policy ontology defines the entities to be instantiated (left hand side) and the policy management processes (right hand side). Using the presented approach, legally correct policies can be created using specific tooling, e.g., to support the subject of care – usually being a layman in the healthcare domain – to express her will and to understand the policies of other actors using the domain-specific jargon.

The other requirement of GDPR is the binding of policies to information objects and/or activities in the business process to dedicatedly managing them. This mechanism allows for definition and enforcement of data governance, i.e., of constraints on the WHO, HOW, WHEN, WHERE, WHAT FOR and WHY – or in other words, on the actor (which is not necessarily a person) and the context – for accessing and using PII. Such services also address, e.g., the data subject's right of being informed about the processing of personal data as well as rectifying those data if needed. That way, the limitations of that data subject's right offered ín GDPR could be minimized or fully overcome if it proves to be impossible or would involve a disproportionate effort. Alternative to policy binding, binding of labels referring to policies stored in policy repositories can be used. Both ways of binding policies or labels have been standardized in HL7's Healthcare Privacy and Security Classification System (HCS) [14]. As security and privacy labels have been defined: Confidentiality, Sensitivity, Integrity, Compartment and Handling Caveats. The first four are bound to information objects, and the last is bound to activities. The logical architecture to implement the described solution is shown in Figure 3. For enabling an assessment of trustworthiness of the offered service by the service user, a monitoring as well as a trust calculation service have been added to the architecture. Policy binding makes only sense when information is properly structured to enable the assignment of different policies to single, or groups of, information objects on the one hand, or different policies to single actions or related groups of actions on the other hand. Such data segmentation for privacy has been standardized at HL7 [15]. The entire system was demonstrated at HIMSS 2012 [16] and – with further extensions and improvements – also in the following years.

Discussion

Contrary to the Directive 95/46/EC, offering a limited scope and static controls for protection of personal health information, GDPR sets requirements, principles and methodologies to be applied to protect personal identifiable information of EU citizens and to document and demonstrate the compliance with the GDPR independent of the location and the technologies data collection, processing, communication and deployment of that information happen.

Acknowledging the nowadays organizational, methodological and especially also technological paradigm changes leading to highly distributed, complex, highly dynamic settings, turning the data subject to a consumer and changing her role and responsibility, GDPR excellently accommodates healthcare transformation. For that purpose, it has to establish the same

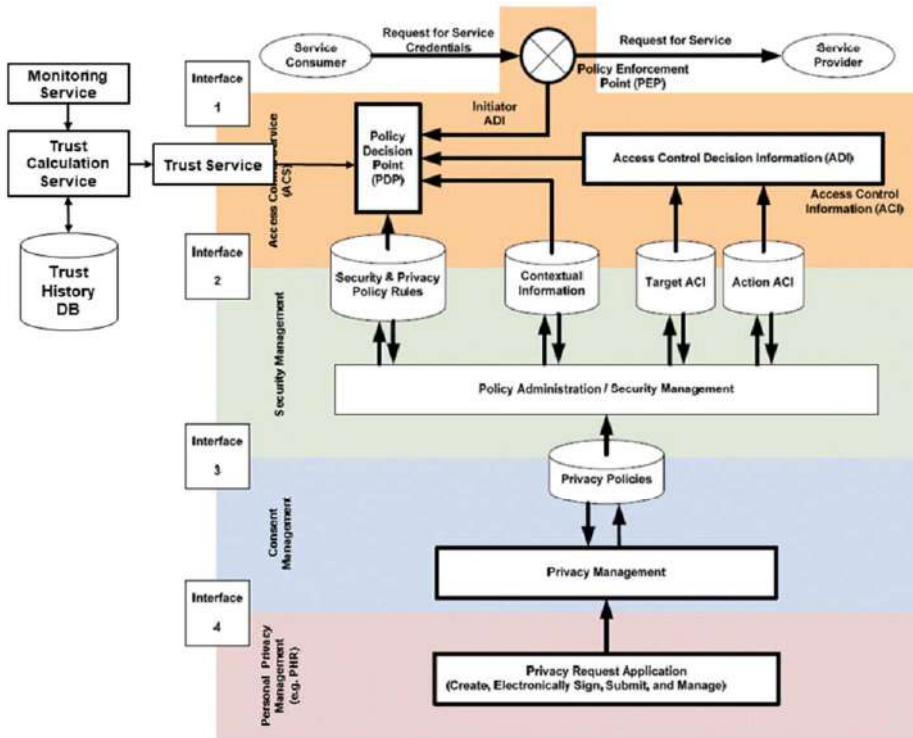


Figure 3. Authorization Reference Model [14]

principles of flexibility, dynamics, complexity, transparency, adaptation, intelligence, automation, etc. Defining principles and methodologies of data governance, GDPR and ISO/IEC 27001 Information technology – Security techniques – Information security management systems – Requirements [17] as well as ISO/IEC 27002 Information technology -- Security techniques -- Code of practice for information security controls [18] are closely interrelated. While GDPR focuses on the rights of affected persons, ISO addresses the related compliance issues.

Meanwhile, a bunch of standards, specifications and applications for adequately managing GDPR are available, like HL7 Artifacts to support managing GDPR. Here, the HL7 Security Labeling Service has to be mentioned, which contains a Policy Adjudication Engine for policy harmonization, an Ontology Reasoner to check the applicability of security and privacy policies against clinical policies, and the finally the Policy Inference Logic. Furthermore, specific trust services are provided. Most of those services are specified as implementable FHIR resources [19]. Examples are: FHIR consent resource [20]; FHIR AuditEvent or Provenance for storing information about data sources; FHIR AuditEvent enables tracking data communication including the deletion of data; FHIR Security Labels allow tagging data when the purpose of use has changed. For coding the purpose of use, specific vocabulary can be deployed.

Conclusions

Paradigm changes in health and social care lead to highly distributed and dynamic ecosystems, integrating multiple jurisdictional and policy domains, technologies, methodologies, knowledge and concept representation style, languages, cultural background and expectations, education and skills, etc., requiring advanced interoperability solutions. The interoperability challenge is not limited to ICT environment, but includes the entire ecosystem. Appropriate security and privacy solutions provide trust and therefore acceptance of health solutions and their IT support, as shown in Table 1.

Table 1. Relations of GDPR and 5P Medicine

GDPR	Details	Impacts to system architecture	Impacts to 5P medicine
New definitio n to PII	PHI characterizes health status and individual context/ conditions	Multi-domain business system architecture incl. development process	Individually tailored diagnosis and therapy including prediction and prevention
Explicit policies	Machine-processable, dynamically managed, ontology-based know-ledge-driven interopera-bility	Representation and transformation of a real-world business system to implementable artefacts, managing all concept levels from knowledge space to data models	Systems medicine incl. all perspectives from elementary particle to society with active participation of data subject
Business system and business process aware privacy	Permanent risk analysis, business system and process mngt, data governance management	Enterprise model and RM-ODP views for managing the system and selecting and correctly interrelating/ constraining existing resources	Democratizing health and social care, respecting personal wishes and expectations, combined with legal, ethical and fair principles

Thereby, security and privacy are not disabling but enabling new technologies. Security and privacy management will be increasingly model-driven, ontology-based and automated, using system intelligence such as AI and machine learning. Definition, harmonization and enforcement of policies must be automated as well. For that reason, policies must be represented comprehensively and formally. The presented system-theoretical, architecture centered modeling approach does not re-write problematic legislation [21], but supports use case specific analysis, management and design of needed comoonents and relations for 5P systems. That way, it enables intelligent, adaptive systems for advanced 5P medicine.

Acknowledgements

The authors are indebted to thank their colleagues from EFMI WGs "EHR" and "Security, Safety and Ethics" as well as from the IMIA WG "Security in Health Information Systems" for their valuable input. The work would not have been possible without the cooperation with ISO/TC215, CEN/TC251, HL7 International, Inc., but also without the support of the German Data Protection and Data Security Association (GDD).

References

[1] B. Blobel, P. Ruotsalainen, D.M. Lopez, F. Oemig, Requirements and Solutions for Personalized Health Systems. *Stud Health Technol Inform* **237** (2017), 3-21.

[2] European Parliament and Council, Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). EC, Brussels 2016. Available at http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

[3] IEEE Standards Association. ANSI/IEEE 1471-2000 - IEEE Recommended Practice for Architectural Description for Software-Intensive Systems. IEEE Computer Society, New York, NY, USA, 2000

[4] International Organization for Standardization. ISO 22600:2014 Health informatics – Privilege management and access control. ISO, Geneva, 2014.

[5] B. Blobel, Application of the Component Paradigm for Analysis and Design of Advanced Health System Architectures. *Int J Med Inform* **60,3** (2000), 281-301.

[6] B. Blobel, Architectural approach to eHealth for enabling paradigm changes in health. *Methods Inf Med* **49,2** (2010), 123-134.

[7] European Parliament and Council, Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. EC, Brussels 2016. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31995L0046

[8] International Organization for Standardization. ISO 23903 Health informatics - Interoperability Reference Architecture. In preparation.

[9] International Organization for Standardization. ISO 21298:2017 Health informatics – Structural and functional roles. ISO, Geneva, 2014.

[10] F. Kamareddine, T. Laan, R.A. Nederpelt, Modern Perspective on Type Theory. Kluwer Academic Publishers, New York, 2004.

[11] R. Bloe, F. Kamareddine, R. Nederpelt, The Barendregt Cube with Definitions and Generalized Reduction. *Inf Comput* **126,2** (1996), 123-143.

[12] International Organization for Standardization. ISO/IEC 10746-3:2009 Information technology-Open distributed processing-Reference model: Architecture. ISO, Geneva, 2009.

[13] N. Damianou, N. Dulay, E. Lupu, M. Sloman, The Ponder Policy Specification Language. In: M. Sloman, J. Lobo, and E. Lupu (Eds.): *POLICY 2001, LNCS 1995*, 18-38. Springer, Berlin Heidelberg, 2001.

[14] Health Level 7 International, Inc., HL7 Healthcare Privacy and Security Classification System (HCS) – Release 3. HL7 International, Ann Arbor, May 2013.

[15] Health Level 7 International, Inc., HL7 Implementation Guide: Data Segmentation for Privacy (DS4P), Release 1. HL7 International, Ann Arbor, September 2013.

[16] Data Segmentation for Privacy HIMSS 2012, Final V2. Available at https://vdocuments.mx/data-segmentation-for-privacy-himss-2012-fina-lv2.html

[17] International Organization for Standardization. ISO/IEC 27001:2017 Information technology – Security techniques – Information security management systems – Requirements. ISO, Geneva, 2017.

[18] International Organization for Standardization. ISO/IEC 27002:2013 Information technology -- Security techniques – Code of practice for information security controls. ISO, Geneva, 2013.

[19] Health Level 7 International, Inc., FHIR Resources. Available at https://www.hl7.org/fhir/

[20] Health Level 7 International, Inc., FHIR consent resource. Available at http://hl7.org/implement/standards/ fhir/consent-examples.html

[21] J.A. Bovenberg, M. Almeida. Patients v. Myriad or the GDPR Access Right v. the EU Database Right. *European J Human Genetics* **27** (2018), 211-215.

Address for correspondence

Bernd Blobel, PhD, FACMI, FACHI, FHL7, FEFMI, FIAHSI, Professor, Medical Faculty, University of Regensburg, Regensburg, Germany; Email: bernd.blobel@klinik.uni-regensburg.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190404

An Extensible De-Identification Framework for Privacy Protection of Unstructured Health Information: Creating Sustainable Privacy Infrastructures

Stefano Braghin[a], Joao H Bettencourt-Silva[b], Killian Levacher[a], Spiros Antonatos[a]

[a] Data Privacy Group, IBM Research, Dublin, Ireland
[b] Health and Person-Centred AI Group, IBM Research, Dublin, Ireland

Abstract

The volume of unstructured health records has increased exponentially across healthcare settings. Similarly, the number of healthcare providers that wish to exchange records has also increased and, as a result, de-identification and the preservation of privacy features have become increasingly important and necessary. Governance guidelines now require sensitive information to be masked or removed yet this remains a difficult and often ad-hoc task, particularly when dealing with unstructured text. Annotators are typically used to identify such sensitive information but they may only be effective in certain text fragments. There is at present no hybrid, sustainable framework that aggregates different annotators together. This paper proposes a novel framework that leverages a combination of state-of-the-art annotators in order to maximize the effectiveness of the de-identification of health information.

Keywords:

Privacy, De-Identification, Natural Language Processing

Introduction

Information exchange is a routine activity among practitioners in healthcare organizations. Much of the information that physicians require to deliver care is recorded and exchanged internally in patient records and, externally, via formal communication letters between referring and consulting physicians across organizations. Other forms of exchange include, for example, legal reasons or claims between providers and payers. As electronic medical records evolved and larger amounts of data were recorded, the value of the secondary uses of healthcare data became apparent [1; 2]. Both the new uses as well as improvements in exchange systems mean that information can more easily be misplaced or accessed by third parties without justifiable reasons.

Early efforts took place in the UK in the 1990s [4]. In the early 2000s, the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule was introduced in the US in order to regulate and safeguard the use and disclosure of protected health information (PHI), which is sensitive and identifiable [5]. The same rule sets out clear definitions of the types of information that are considered to be PHI, such as a patient's name or social security number, as well as guidelines on how to remove such information, such as the Safe Harbor method [5]. De-identification is the process in which PHI in a given patient record is either removed or masked such that it is not possible to link that particular record back to the identity of the original patient. Approaches for estimating the likelihood that de-identified information can be re-identified have been proposed [6] and blanket protection policies and methods such as the Safe Harbor, although effective, may still leave organizations susceptible to re-identification.

De-identification approaches may focus on structured data, for example, estimating the population size cut-offs for geographical areas so that no data suppression or further aggregation is necessary [8]. Other approaches, such as the anonymisation of clinical profiles in the form of International Classification of Diseases (ICD) codes, have also shown that it is possible to share information while safeguarding the privacy of the underlying individuals [11]. However, the rising need for sharing free-text records has boosted the need for Natural Language Processing (NLP) methods and techniques [9]. A systematic review of de-identification techniques revealed that the majority of systems focused on structured data and only a few addressed automatic text de-identification [10]. Furthermore, most systems performing annotation relied on pattern matching, rules, or dictionaries and the latter has shown better performance overall but lacks generalisation [10].

Free-text de-identification tools such as MIST [15] and the various systems in [10] have been more recently evaluated in [14]. The results of this evaluation showed the most successful system (MIST) had a precision of 97.8% and a recall of 95.1%. These evaluation metrics are defined in Figure 1. A different study [13] compared the performances of human annotators against a system using pattern matching based on a dictionary, regular expressions and heuristics. The results showed low precision (75%) but high recall (96%), meaning that the output is mostly de-identified, but also that there is a large amount of over de-identification. Similarly, in [12], the authors tested the performance of human annotators against MIST and an in-house identifier based on MALLET [18] that replicates MIST's design. The results were reassuring, reaching 95% precision in some entity types after adding pre and post processing to the workflow, in addition to model training. Similarly, [16] proposes a system called Medical De-identification System, designed to process HL7 messages in order to remove HIPAA and non-HIPAA specified identifiers. This system relies on scrubbing instead of providing a configurable masking system, yet the authors report results in line with the state-of-the-art systems.

$$P = \frac{detected\ relevant\ entities}{detected\ entities},\quad R = \frac{detected\ relevant\ entities}{relevant\ entities},\quad F_1 = 2\frac{P*R}{P+R}$$

Figure 1 Evaluation metrics: Precision (P), Recall (R) and the harmonic mean of P and R (F_1).

De-identification remains a task that cannot be perfectly handled by automated systems [14] yet, in practical scenarios, such systems have the potential to assist humans in performing de-identification if all the limitations are known and

acknowledged by the users. Contrary to previous work, the approach presented in this paper leverages an ensemble of annotators, taking advantage of the strength of each specific annotator. Moreover, this approach exposes a more flexible configuration to the user, thus allowing easier transition between domains (such as between clinical and financial/billing).

This paper presents the first hybrid framework for automatic text de-identification that combines different annotators together in order to maximise de-identification performance and increase generalisation. The methods section describes the framework in detail, how it was built and their individual components. The results section reveals the performance of each of the annotators, both individually and in combination, based on a representative corpus of clinical notes.

Methods

We have designed and implemented a de-identification pipeline that consists of five main steps, shown in Figure 2. The first step includes input handling and content extraction. The second step performs the pre-processing of the extracted text followed by an annotation step, where the individual annotators perform entity detection. The fourth step includes the post-processing of the results, false positives (i.e. incorrectly detected entity types) elimination and prioritisation. Finally, the de-identification step ensures the protection of the identified entities.

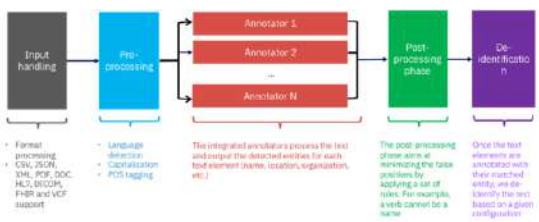


Figure 2 The high-level architecture of our de-identification pipeline

Input Handling

As we need to deal with a diverse set of input formats, our pipeline supports the following formats: CSV, JSON, XML, XLS/XLSX, PDF, DOC/DOCX, HL7, DICOM, VCF as well as plain text documents with no specific format. The user can specify which input elements contain free text information, for example, specific DICOM tags or HL7 segments. The input handling step is responsible for content extraction from the various input formats and preservation of the semantics that are necessary during the reconstruction phase.

Pre-Processing

The pre-processing phase includes three key steps: language detection, correct input capitalization and part-of-speech tagging.

Step 1 – Language Detection: the first step of the pipeline is to perform language detection. Based on the detected language, the system decides to proceed with the remaining steps depending on which annotators are available. For the vast majority of our use cases, we have to deal with input based on the English language but since we are integrating with international platforms, this step is essential. We rely on Apache OpenNLP[1] language detection components,[2] that support 103 languages.

Step 2 – Capitalization for Caseless Input: The input to the pipeline may be caseless; it can be either all lowercase or all uppercase. This has a significant impact on most of the annotators, hence this step is necessary for normalizing the input. We rely on Stanford CoreNLP's truecase annotator to handle such cases.[3]

Step 3 – Part-Of-Speech Tagging: Part-of-speech (POS) tagging is the last part of the pre-processing phase. Each token of the input is tagged with its part-of-speech information. This information is later used in the post-processing phase to alleviate false positives. Several annotators come with POS tagging functionality, such as Apache OpenNLP or Stanford CoreNLP. The POS tags used in this phase must be part of the Penn Treebank POS tagset.[4] For the POS taggers that do not follow this tagset, we manually created the mapping. POS tagging functionality can be extended in our pipeline.

Annotation Phase

During this phase, each annotator is called separately for the same input and then the individual results are gathered and merged. There is no inter-dependency among the annotators so this phase can be fully parallelized for performance.

Post-processing Phase

The post-processing phase includes several steps intended to increase the accuracy and quality of the end results.

Eliminating False Positives Based on POS Information

Since the entities to be de-identified are comprising of either nouns or adjectives, we can eliminate false positives based on the part-of-speech information. For example, if our input is "Will increase Lisinopril to 40 mg daily." and one of the annotators detected that "Will" is a name (this can happen due to blind dictionary matching) then this result can be filtered out since "Will" is neither a noun nor an adjective in the context of this sentence.

POS information need to be ignored for certain entities otherwise true positives would also be removed. Numeric entities, like phone numbers or credit card numbers, e-mail addresses, IP addresses, URLs and datetime entities can be classified as neither nouns nor adjectives, but they must not be discarded. Therefore, we provide configuration options to make sure that the pipeline handles these special cases accordingly.

Priority System

Since we invoke multiple annotators, there may be conflicting results for the same input token(s). In order to handle the potential conflicts, we introduced a priority system per annotator and per entity type. For each annotator, we assign a weight (a value in range [0, 100], with 1 as default) to each of the entity types supported. The initial weight assignment was performed based on the evaluation results of each individual annotator. The priority system also allows us to disable entities for a specific annotator. A weight value of 0 for an entity will force the pipeline to ignore the results for that entity for that annotator. The priority system allows us to disambiguate entries like “Alzheimer” which can be identified both as a name and a medical term, given that medical term classifier has higher priority over the name classifier for the conflicting terms.

[1] https://opennlp.apache.org/

[2] https://opennlp.apache.org/news/model-langdetect-183.html

[3] https://stanfordnlp.github.io/CoreNLP/truecase.html

[4] https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

Merging of Connected Entities

There are cases where an entity is split into separate results by an annotator. For example, let us consider that the input is "Health Center of Washington". An annotator can detect two entities: "Health Center" as an organisation and "Washington" as a location. However, we want to detect only one entity and that is the entire sequence as an organization. The pipeline enables the merging of these two results into a single entity. Users can specify, through a configuration, the left and right-hand side types of the connected entities along with the connecting particles and/or pronouns, such as “of”, or “on”.

Blacklisting

The final step of the post-processing phase requires checking the results, per entity type, against a configurable blacklist. These contain tokens that are not to be considered matches. For example, for date and time entity type, we assigned that the token “currently” is a blacklist term. It is also possible to blacklist entity types from the final report. As an example, UMLS[5] dictionary entries if they are not going to be de-identified. This feature is enabled in our system in order to reduce over-classification, hence reducing the loss in utility.

Identifying Missing Repetitions of Identified Entities

There are cases where an instance of an entity is recognized once but then subsequent instances are not recognized. The pipeline includes a mechanism to detect unrecognized occurrences for entities previously identified by the annotators. The mechanism assigns types to the missed entities based on previous occurrences. Newly annotated entity types will be processed using the previously described steps (earlier in Post-processing) in order to remove false positives.

De-identification Core

The final step of the pipeline is the de-identification of the detected entities. This step ensures that any PI/PHI is sufficiently protected and that the usability of the end result is at an acceptable level. This can be performed, for example, by an expert accessor as defined in the HIPAA Expert Determination Method.

Data masking is applied on the detected entities and it replaces the original values with fictionalized ones. Our pipeline offers masking capabilities for all supported data types, as listed in Table 1, as well as some generic masking providers, such as redaction, nullification, hashing, randomization, truncation and numeric value shifting. Replacing values with their entity types (for example, e-mail address "test@domain.com" can be masked to "EMAIL-0") is also an option. Two key properties are supported**. Extensive format and semantic preserving masking**: format and semantic preservation is a key aspect of masking, since in many cases the data recipient needs to operate on the masked data. Approaches such as hashing, encryption or redaction, would limit (or even eliminate) the value of the masked data and allow users only to perform basic operations (e.g. counting). Our approach provides utility-preserving masking for each identified type, such as replacing Social Security Numbers with fictional numbers following the same standard, or replacing gender-specific names with other names of the same gender. **Consistent masking**: this feature ensures that the original values are always masked to the same fictionalized value. Consistency is an important requirement for data analytic purposes in healthcare, since it is the only accurate way to measure uniqueness and to do identity matching.

Annotators

Our de-identification pipeline includes a list of existing annotators. This section describes the available annotators. However, as mentioned before, the real power of the toolkit is in its configurability and extensibility. The annotators listed below were selected based on availability and quality criteria.

PRIMA Annotator - The PRIMA annotator is a set of 38 classifiers that cover basic PI/PHI entities based on our work for structured data [17]. PRIMA embeds three types of classifiers: regular expression-based, dictionary-based and custom-logic-based. Entity types like e-mails, addresses and URLs can be detected via regular expressions. Other entity types, like names, cities, countries and ICD codes, are detected using dictionaries. For the rest of the entity types, we use custom code for type detection. For example, credit cards are firstly detected by a specified prefix but the correctness of the check digit is also verified. A list of the currently supported entity types is shown at Table 1. PRIMA exposes the necessary interfaces to easily extend and localize the available functionality, as described in [17]. PRIMA is able to parse HL7 and FHIR messages as it has been designed to be compatible with healthcare systems.

Table 1: Supported entity types for PRIMA annotator.

Demographics	Location	Healthcare	Technology	Banking	Telco
Date and time	City	ICD v9	E-mail	SWIFT	IMSI
Name	Country	ICD v10	URL	IBAN	IMEI
surname	Continent	NDC	MAC address	CC number	Phone number
Religion	Address	ATC	IP Address	CC type	
Marital status	Latitude/Longitude	CPT			
Ethnicity	US states	Hospital name			
National ID	US counties	DICOM metadata			
SSN US	US/UK ZIP Codes	Patient ID			
NINO		MR number			
Occupation					

The application of regex-based or dictionary-based classifiers over free-text can miss entities spanning across multiple tokens. For example, consider the input “Health Center of NY is currently expanding its ICU”, along with the objective to de-identify the organization term “Health Center of NY”. Applying classifiers only for every distinct token will not return any results, even if our dictionaries include terms like “Health Center of NY”. To tackle this issue, we implemented a shingle filter mechanism. For each token, we examine the token itself and the next 1 to N tokens. For example, with N=4 the filter will produce the shingles “Health”, “Health Center”, “Health Center of”, “Health Center of NY”, ..., "currently expanding its ICU". For each shingle, all available PRIMA classifiers are invoked. This mechanism enables detecting entities that span across multiple tokens. Shingles, however, may still return overlapping results yet the subsequent steps in this pipeline are used to merge them with results from the other annotators.

Apache OpenNLP - Apache OpenNLP supports the most common NLP tasks, such as tokenization, sentence segmentation, part-of-speech tagging, named entity recognition, chunking, parsing, language detection and coreference resolution. Apache OpenNLP comes as an offline Java library. Trained models are also available for several languages in addition to English.

Stanford CoreNLP - Stanford CoreNLP [19] provides a set of human language technology tools. It can give the base forms of

[5] https://www.nlm.nih.gov/research/umls/

words, parts-of-speech tagging, named entity recognition, normalize dates, times, and numeric quantities, mark up the structure of sentences in terms of phrases and syntactic dependencies, indicate which noun phrases refer to the same entities, sentiment analysis, extract particular or open-class relations between entity mentions, and more. Stanford CoreNLP comes with models for the most commonly spoken languages.

SystemT - SystemT [20] is a declarative information extraction system that has been designed and developed to perform the task of extracting structured information from unstructured or semi-structured data. It is based on the basic principle underlying relational database technology: complete separation of specification from execution. SystemT uses a declarative rule-based language and an optimizer that generates high-performance algebraic execution plans.

Advanced Care Insights - The Advanced Care Insights[6] (ACI) Service uses healthcare annotators that accelerate natural language processing capabilities to identify medical and social information in physician notes, discharge summaries, and pathology reports. ACI identifies, normalizes, and codes medical and social facts, including symptom, disease, procedures, allergies, medications, smoking status, lab results, ejection fraction, and various daily living assistance terms. Note that, like the other annotators, the ACI service can be used in a HIPAA compliant manner.

Extensibility and Integration -There are cases where we need to detect entities based on either dictionaries, regular expressions or custom logic. In order to address such cases, the pipeline supports the loading of dictionary-based and regex-based identifiers via configuration and external files. There is also provisioning for the user to load their own annotators and POS taggers as well as localize the resources used for the detection and de-identification components. This can be done by implementing a Java interface and instructing the pipeline on how to use and configure the annotators via a configuration file.

The pipeline also includes NLP Tools from the Cognitive Computation Group of the University of Illinois[7], Watson Natural Language Understanding[8] and out-of-vocabulary classifier. For brevity, we will not elaborate on each of these annotators.

Results

Entities and Coverage

The pipeline supports the identification and masking of 38 entities. However, this number can be increased by plugging in custom functionality. The out-of-the-box supported entity list is based on the HIPAA Safe Harbor compliance standard but was augmented based on use cases we encountered during development and deployment.

We have classified each entity either as “direct”, which means that it directly identifies an individual, or “indirect”. Some indirect identifiers can be treated as direct in specific contexts, for example, if occupation is “chief of medicine” then it is a direct identifier. However, we do not cover such cases as disambiguating these cases requires additional, contextual, information that might not be available.

Figure 3 describes the supported entities along with their classification and supported annotators. With respect to Stanford, SystemT and Apache OpenNLP annotators, the non-supported entities can be covered after creating custom models to the respective annotation core. The same applies for PRIMA, with the difference that PRIMA also accepts simpler regex-based and dictionary-based identifiers.

As shown in Figure 3, no single annotator covers all entities effectively, at least not without considerable customisation and custom model training. However, our pipeline allows the concurrent execution of all available annotators, so all entity types can be covered. Note that certain entities are marked as "*" in Figure 3. The reason behind this is that such entity types do not have a universally accepted format, contrary to other entity types such as IP or e-mail addresses. These entity types, e.g. medical device identifiers and health beneficiary numbers, can vary between manufacturers and countries and it is therefore non-trivial to provide general models.

Entity	Classification	Stanford	ACI	SystemT	OpenNLP	PRIMA	Overall
Name	Direct	S	S	S	S	S	S
Location	Direct	S	NS	S	S	S	S
Dates	Indirect	S	NS	S	S	S	S
Phone numbers	Direct	NS	S	S	NS	S	S
Email	Direct	NS	S	S	NS	S	S
Social Security Numbers	Direct	NS	NS	NS	NS	S	S
Medical Record Numbers	Direct	NS	NS	C*	NS	C*	C*
Health Beneficiary Numbers	Direct	NS	NS	C*	NS	C*	C*
Organizations	Indirect	S	NS	S	S	S	S
Account numbers	Direct	NS	NS	C*	NS	C*	C*
Certificate / License numbers	Direct	NS	NS	C*	NS	C*	C*
Medical Device Identifiers	Direct	NS	NS	C*	NS	C*	C*
URLs	Indirect	NS	NS	S	NS	S	S
IP Addresses	Direct	NS	NS	C	NS	S	S
Occupation	Indirect	NS	NS	C	NS	S	S
Symptoms (optional)	Indirect	NS	S	C	NS	NS	S
Procedures (optional)	Indirect	NS	S	C	NS	NS	S

Figure 3. List entities supported by each of the annotators. S – Supported; NS – Not Supported; C – Customisable, not prebuilt.

Evaluation and Outcomes

We evaluated the performance of our pipeline in terms of precision and recall based on a random sample of 400 manually annotated medical notes, each referring to an individual patient. The sample was drawn from a large dataset from a US-based healthcare provider and ethical approval was obtained. The full dataset includes medical notes for more than 100,000 patients each of which has at least a medical note. The sampling processes removed machine generated notes and notes not containing any PHI. The notes contained entities for age, date, location, medical record numbers (MRN), names, phones and URLs. We compared our pipeline against NLM Scrubber.[9]

NLM Scrubber is a freely available, HIPAA compliant, clinical text de-identification tool designed and developed by the National Library of Medicine. It is the de-facto standard for de-identification of clinical notes in practical use cases and this allowed us to directly compare similar entity types. For the purpose of evaluation.

The results are summarized in Table 2. Using our pipeline, precision, recall and F_1 are consistently better across all entities. NLM Scrubber does not include models for ages and URLs. The recall is lower than NLM Scrubber only in the case of location and name entities but it is important to note that NLM Scrubber has very low precision on these entities, which means that NLM Scrubber will incorrectly identify as names or location “words” that are not actual names or location, respectively.

[6] http://www-03.ibm.com/software/products/en/advanced-care-insights

[7] Cognitive Computation Group @ Illinois, http://bilbo.cs.illinois.edu/

[8] https://www.ibm.com/watson/services/natural-language-understanding/

[9] https://scrubber.nlm.nih.gov/

Table 2 Performance of pipeline versus NLM Scrubber for all entities that need to be de-identified. Highlighted cells show best values.

	NLM Scrubber			Our Pipeline		
Entity Type	**Precision (%)**	**Recall (%)**	**F_1 (%)**	**Precision (%)**	**Recall (%)**	**F_1 (%)**
Age	0	0	0	93.18	78.85	85.42
Date	97.98	96.52	97.24	99.87	98	98.93
Location	31.25	100	47.62	88.24	88.24	88.24
MRN	53.85	95.45	68.85	100	100	100
Name	68.45	97.92	80.57	94.77	93.96	94.36
Phone	100	93.88	96.84	100	97.50	98.73
URL	0	0	0	100	100	100

Discussion

A multitude of highly specialized annotators are available and significant research has been undertaken to improve the accuracy of specific entities, such as names or medical terms. However, there is at present no existing framework to combine annotators so as to construct a robust pipeline that is able to satisfy multiple governance, regulatory and business purposes. This approach aims to create a flexible system that adapts to a plethora of scenarios. Similarly, a system able to support real-world heterogeneous data poses several challenges not captured by standard benchmarks. Clinical notes typically include typos, mixed cases (different across notes), and partially structured machine generated data (e.g., lab notes with alerts).

The approach proposed in this paper focuses on the identification of entity types, since regulatory requirements are not global or constant and on the premise that ensemble approaches outperform individual annotators. The pipeline proposed in this paper outperformed NLM Scrubber (see Table 2) and provided similar or better results than other approaches [12]. The latter proposes a framework and model that can be refined and where identifiers are customised. The generated models in [12] are trained and updated, however, this would typically mean a loss in precision when new entities are added to the classification model. The main difference between our approach and the one presented in [12] is that our models can return multiple types with weights in addition to the combination of information during post-processing with any other information available (e.g. using POS information to reduce false positives).

Combining positive with negative models helped increasing accuracy, for example, disambiguating between names (sensitive) and medical conditions (not sensitive according to HIPAA). Future development would include automating the priority system, for example, using semi-supervised learning techniques.

Conclusions

Automatic text de-identification is becoming increasingly important as larger volumes of narrative data containing PHI are being used. This paper presents a new hybrid system that combines different annotators and demonstrates how this new approach has a substantial effect on performance and generalisability. This is a useful tool for any data privacy expert following the HIPAA Expert Determination Method. To the best of our knowledge, this is the first framework that combines annotators together to provide precision and recall values better than any of the single annotators alone. Our results are also in line with, when not better than, other free text de-identification systems described in the literature. Further work is underway to evaluate this approach on larger corpora and to cover new data types.

References

[1] D. Blumenthal, Stimulating the adoption of health information technology, *N Engl J Med* **360** (2009), 1477-1479.

[2] C. Safran, M. Bloomrosen, W.E. Hammond, S. Labkoff, S. Markel-Fox, P.C. Tang, D.E. Detmer, and P. Expert, Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper, *J Am Med Inform Assoc* **14** (2007), 1-9.

[3] A. Wright, C. Soran, C.A. Jenter, L.A. Volk, D.W. Bates, and S.R. Simon, Physician attitudes toward health information exchange: results of a statewide survey, *J Am Med Inform Assoc* **17** (2010), 66-70.

[4] T.H. Payne, D.E. Detmer, J.C. Wyatt, and I.E. Buchan, National-scale clinical information exchange in the United Kingdom: lessons for the United States, *J Am Med Inform Assoc* **18** (2011), 91-98.

[5] HHS, Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, in, 2012.

[6] K. Benitez and B. Malin, Evaluating re-identification risks with respect to the HIPAA privacy rule, *J Am Med Inform Assoc* **17** (2010), 169-177.

[7] M. Kayaalp, Modes of De-identification, *AMIA Annu Symp Proc* **2017** (2017), 1044-1050.

[8] K. El Emam, A. Brown, and P. AbdelMalik, Evaluating predictors of geographic area population size cut-offs to manage re-identification risk, *J Am Med Inform Assoc* **16** (2009), 256-266.

[9] P. Zweigenbaum, D. Demner-Fushman, H. Yu, and K.B. Cohen, Frontiers of biomedical text mining: current progress, *Brief Bioinform* **8** (2007), 358-375.

[10] S.M. Meystre, F.J. Friedlin, B.R. South, S. Shen, and M.H. Samore, Automatic de-identification of textual documents in the electronic health record: a review of recent research, *BMC Med Res Methodol* **10** (2010), 70.

[11] R. Heatherly, L.V. Rasmussen, P.L. Peissig, J.A. Pacheco, P. Harris, J.C. Denny, and B.A. Malin, A multi-institution evaluation of clinical profile anonymization, *J Am Med Inform Assoc* **23** (2016), e131-137.

[12] L. Deleger, K. Molnar, G. Savova, F. Xia, T. Lingren, Q. Li, K. Marsolo, A. Jegga, M. Kaiser, L. Stoutenborough, and I. Solti, Large-scale evaluation of automated clinical note de-identification and its impact on information extraction, *J Am Med Inform Assoc* **20** (2013), 84-94.

[13] I. Neamatullah, M.M. Douglass, L.W. Lehman, A. Reisner, M. Villarroel, W.J. Long, P. Szolovits, G.B. Moody, R.G. Mark, and G.D. Clifford, Automated de-identification of free-text medical records, *BMC Med Inform Decis Mak* **8** (2008), 32.

[14] A. Stubbs, C. Kotfila, and O. Uzuner, Automated systems for the de-identification of longitudinal clinical narratives: Overview of 2014 i2b2/UTHealth shared task Track 1, *J Biomed Inform* **58 Suppl** (2015), S11-19.

[15] J. Aberdeen, S. Bayer, R. Yeniterzi, B. Wellner, C. Clark, D. Hanauer, B. Malin, and L. Hirschman, The MITRE Identification Scrubber Toolkit: design, training, and assessment, *Int J Med Inform* **79** (2010), 849-859.

[16] F.J. Friedlin and C.J. McDonald, A software tool for removing patient identifying information from clinical documents, *J Am Med Inform Assoc* **15** (2008), 601-610.

[17] S. Antonatos, S. Braghin, N. Holohan, Y. Gkoufas, and P. Aonghusa, *PRIMA: An End-to-End Framework for Privacy at Scale*, 2018.

[18] A.K. McCallum, MALLET: A Machine Learning for Language Toolkit, in, 2002.

[19] C.D. Manning, M. Surdeanu, J. Bauer, J.R. Finkel, S. Bethard, and D. McClosky, The Stanford CoreNLP Natural Language Processing Toolkit, in: *ACL*, ACL, 2014.

[20] L. Chiticariu, M. Danilevsky, Y. Li, F. Reiss, and H. Zhu, SystemT: Declarative Text Understanding for Enterprise, in: *Proceedings of the 2018 Conference of the North American Chapter,* ACL, New Orleans, Louisiana, 2018, pp. 76-83.

Address for Correspondence

Joao H Bettencourt-Silva. e-mail at jbettencourt@ie.ibm.com.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190405

A Glimpse at the Australian Health Information Workforce: Findings from the First Australian Census

Kerryn Butler-Henderson [a] and Kathleen Gray [b]

[a] University of Tasmania, Launceston, Tasmania, Australia,
[b] Health and Biomedical Informatics Centre, The University of Melbourne, Melbourne, Victoria, Australia

Abstract

The Australian Health Information Workforce is a critical discipline in the health sector as the investment in digital technologies increases. Yet historically there was no standardized reporting about the workforce and its six professional areas: clinical coding, costing analysts, data analysts, health informaticians, health information managers and health librarians. This paper presents the findings from the inaugural Australian Health Information Workforce Census. Analysis of 1,596 responses indicates this is an aging (56.1% ≥45 years) workforce with a large (78.1%) female population. Working in permanent (82%), public hospital (72%) roles, in professional or managerial roles (84%). The majority (93.2%) of respondents hold a tertiary qualification in health information, one-quarter of these at masters or doctoral level. Fewer than 30% of respondents hold a health information credential from a professional or industry association. The data from the ongoing national census will inform workforce planning and enable forecasting of the future workforce needs.

Keywords:

Censuses, Informatics, Health Manpower.

Introduction

The health information workforce is a specialized discipline responsible for the development, maintenance and governance of the systems for the management of health data, health information, and health knowledge [1]. A 2013 examination of the workforce established it is poorly defined with a lack of accurate data about the professions that constitute the workforce [2]. The Health Workforce Australia report [2] made six recommendations: 1). delineate the workforce, 2). improve data collection, 3). support strategic partnerships between stakeholders, 4). consider the future configuration of workforce, 5). address known health information workforce shortfalls, and 6). promote health information training and careers. Subsequently the 2016 National Health Information Workforce Summit [3] agreed to the following alliance strategy "Strategy A: We need a census of the health information workforce and regular collection of data". The Health Information Workforce Census Project ("Census Project") was developed to address these recommendations and strategy, with the project aim to provide timely evidence about the nature and scope of the Australian health information workforce through a national census.

The Australia Health Information Workforce Alliance [4] states the workforce consists of the following professional areas: (1) Clinical coding; (2) Clinical costing analysts; (3) Health data analysts; (4) Health informaticians; (5) Health information managers; and (6) Health librarians. Yet, there is no empirical research to support this delineation. Existing empirical research has only focused on part of the workforce. Half of the workforce do not have an Australian and New Zealand Standard Classification of Occupations (ANZSCO) classification, so they are not captured by the national household census [5]. And where information is captured, past research suggests it may not be an accurate representation [4]. There have been a small number of Australian profession specific surveys [4, 6-8] that only survey professional members of peak bodies and/or one or two professional areas of the workforce. Never before has the entire workforce been examined with a scientifically developed and validated instrument.

There is also a lack of workforce data internationally. The most scientifically rigorous health information workforce research to date was undertaken in Canada [9], which included the health informatics, health information management, and clinical coding professional areas. This study estimated the hiring requirements for these professionals in 2008 and 2013 and identified roles for which there is a risk of skills shortages in Canada. Yet, the study used data from existing data collections systems and not from the individual, and did not examine the educational background of the population, the day to day operations of the role, the professional development requirements of individuals, peak body membership, credentials, or the future intentions of the workforce.

The increase in both digitization and adoption of technological approaches has created a growing demand for workforce specialists.[10], and the converging interests of the global health information workforce is now being recognized [11]. But without a clear picture of who the workforce is, the health system is unable to forecast future workforce needs.

A 2015 focus group suggested that Health Workforce Australia recommendation 1 and 2 could only be addressed by an integrated, scientific data collection system, such as a routine census [12]. Such a system would enable the collection of data to inform the other Health Workforce Australia recommendations. This paper describes the partnership between the authors and Australian peak bodies, the Australasian College of Health Informatics (ACHI), Australian Library and Information Association (ALIA) Health Libraries Australia (HLA), Health Informatics Society of Australia (HISA), Health information Management Association of Australia (HIMAA), and key government organizations, the Australian Digital Health Agency (ADHA) and the Victorian Department of Health and Human Services Workforce Branch, to develop a national census. The aim of the census itself is to quantify and qualify the Australian Health Information Workforce, specifically to delineate and count the workforce,

consider the future configuration of the workforce, identify health information workforce shortfalls, and identify current health information training and career pathways to meet future workforce demands. The results from the first Australian census shall be presented below.

Methods

The census instrument was developed through a large national study involving Australian and New Zealand stakeholders [13]. A Delphi approach was adopted to enable the staged development of the instrument through seven rounds: data elements (rounds 1-3), question and response format (rounds 4-6), and lastly the pilot testing of the instrument (round 7). Representatives from the six partner organizations, and the Clinical Coding Society of Australia and Health Informatics New Zealand were invited to join an Expert Panel, who were closely consulted through both face to face focus groups and online surveys at each stage during the project. A larger Consultation Group was formed with more representatives from these organizations, and other identified stakeholder organizations in Australia. This group was engaged through online surveys at all stages throughout the project once the Expert Panel had finalized their input. The sixth round of the project included a consultation with the Executive from each of the partner organizations to obtain their feedback on the instrument, and the seventh round was the overall pilot testing of the instrument by anyone who had been involved in the previous rounds. This testing included both testing of the questions within the instrument and the user experience of the online census.

The census is housed within the REDCap (Research Electronic Data Capture) system hosted at the University of Tasmania. This is a secure online research data capture system, that includes a survey capability. This system was selected as it will allow the ongoing capture of repeated census data collection. REDCap configures surveys to be displayed on any type of device and is compatible with all browser types. The census was also available in paper format on request.

Throughout the development of the census instrument, and in the lead up to the census month, there was extensive communication with stakeholders and individuals who registered through the census website. Recipients were encouraged to distribute the information throughout their networks in an attempt to reach as much of the workforce as possible. Figure 1 outlines the eligibility criteria for participation.

> To complete the Australian Health Information Workforce Census, a participant must:
>
> 1. Work (including volunteer or actively seeking) in a role where the primary function is related to developing, maintaining, or governing the systems for the management of health data, health information, or health knowledge. AND
>
> 2. Work (including volunteer or actively seeking) for/with an organisation that operates in Australia, where the role relates to the Australian operations, and relates to the health sector.

Figure 1– Participation eligibility statement

The census was deployed in May 2018, with ongoing communication throughout the month to the stakeholder list. There was large national engagement and international interest, with the census website receiving nearly 5,000 visits during May. Once the census closed on the 31 May 2018, the data was cleaned and analyzed.

Results

There were 1,597 participant responses included in the analysis. This response rate is estimated to be approximately 20% of the actual workforce, based on comparisons to previous reports about subsections of this workforce [4, 6-8].

There was representation across the six professional areas identified by the Health information Workforce Alliance [4]. A summary of the distribution is provided in Figure 2.

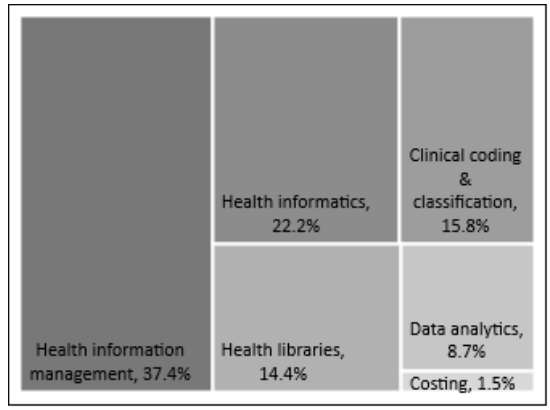


Figure 2–Distribution of professional areas

The census results indicate three quarters of the Australian workforce were born in Australia, with 98.3% either an Australian citizen or permanent resident. The majority of respondents live in Victoria (38.1) and New South Wales (24.1%). 0.8% of respondents indicated they reside outside of Australia.

The gender distribution amongst respondents was strongly skewed towards one gender, with 78.1% of respondents identifying as female, 21.6% as male and 0.3% as other. 56.1% of respondents are aged 45 years or older (Figure 3). Only 1.9% of respondents identified they are an Aboriginal Australian and/or Torres Strait Islander person. 3.4% of respondents specified they had a disability of health condition that limits their participation in activities.

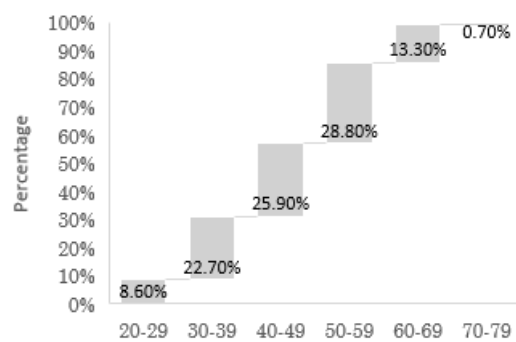


Figure 3 – Age distribution of participants

The majority of respondents work in permanent roles (82%) in the public hospital sector (72%). 15.1% of respondents are actively seeking employment in a health information role, half of which already work in a heath information role. Only 1.4% of respondents identified as self-employed. Respondents indicated they have been in their current role for 7.2 years, with nearly half (47.9%) of respondents reporting they have held their current position for more than 5 years. The majority (84.8%) of respondents indicated they are working in a managerial or professional role, with 13.5% identifying they work in a clerical role. The average weekly remuneration (before taxation) was in the $1,500-$1,999 bracket, with 51.1% of respondents reporting earning between $1,000-$1,999 per week (before taxation). The average number of hours worked per week was 32.6 hours. Respondents indicated they plan to remain working in the Australian health information workforce for an average 14.6 years.

The majority (93.2%) of respondents hold a tertiary qualification in health information. Table 1 summarizes the distribution of highest health information qualifications. Over a quarter (26.0%) of the workforce hold a Masters or Doctoral qualification in health information.

Table 1 – Distribution of educational level

Educational level	Percentage
Certificate I-IV	8.3
Diploma	5.5
Associate degree	1.4
Bachelor degree	40.3
Bachelor honours	3.2
Graduate certificate or diploma	15.3
Master degree	21.7
Doctorate	4.3

11.5% of respondents did not undertake any professional development in the last 12 months. Work base learning (40.3%) and professional activities (24.1%) were the most popular professional development activities (Figure 4). Nearly three quarters (71.8%) of respondents do not hold a health information credential (Figure 5), but 12.3% maintain their health practitioner registration under the National Registration and Accreditation Scheme. Over half (55.5%) of respondent maintain a professional or industry association membership (Figure 6).

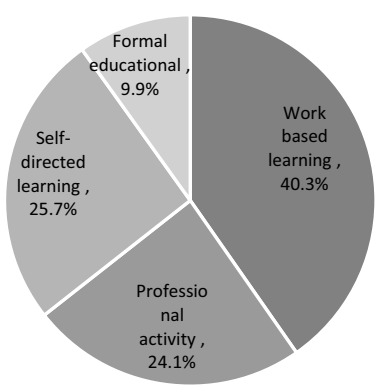


Figure 4 –Distribution of professional development

Discussion

It must first be acknowledged that the sample size for this first census cannot be determined because we simple do not know the size of the workforce in Australia. As the census shall be repeated in 2020 and every three years thereafter, it is anticipated participation will increase. However, an estimate indicates the 2018 result represent approximately 20% of the workforce and the results did provide many insights into the Australian workforce.

This is the first study to establish there is an international health information workforce servicing Australia's health sector needs, Whilst only 1% of the workforce is located outside of Australia, this number is likely to be significantly larger. It will be difficult to reach that workforce to complete the census, but it is important they are included to measure their impact on the Australian workforce.

The results indicate this is an aging workforce. The Australian Intergenerational Report [14] suggests by 2055 there will be a decline in the traditional workforce population (15-65 years old), with an increase in average life expectancy for those aged over 65 years. With an already aging workforce (≥45 years [15]), the Australia health information workforce is at risk of not meeting the productivity demands of the health sector. Investment in a younger workforce is required now to meet these future needs. This is compounded by a largely permanent workforce who remain in a position for several years. Without constant movement across the workforce, opportunities for advancement is reduced, resulting in further loss from the workforce as younger professionals seek opportunities elsewhere.

Diversity is lacking in this workforce across several areas. The census reported a very large female population. Further analysis of the gender data is required to identify if there is a gender pay disparity, and the percentage of women in senior health information positions in health. Further investment is also required to increase participation in this workforce by Aboriginal Australians and Torres Strait Islander people. The results from this census indicate this population is currently under-represented in the workforce compared to their representation in the Australian population.

In terms of how the workforce views itself as a professional workforce, the results showed there is still a critical technical and clerical element. Yet, the majority of respondents have a tertiary qualification. The ongoing census will be necessary to monitor if the discipline evolves from operational roles to information strategist professional roles as digital transformations and automation changes the landscape [16]. Further research is required to examine the level of education and career opportunities.

Opportunities exist for peak bodies and educational providers to provide professional development opportunities. This may increase individual's enthusiasm to obtain professional credentials and join their professional or industry membership organizations.

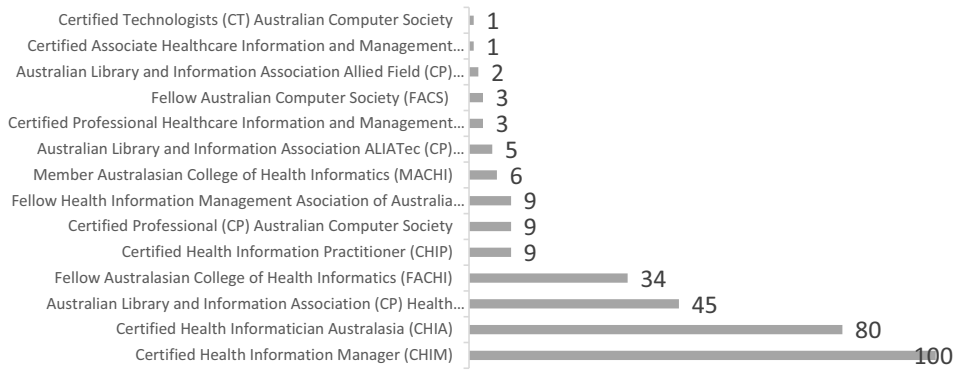


Figure 5 –Distribution of professional credentials

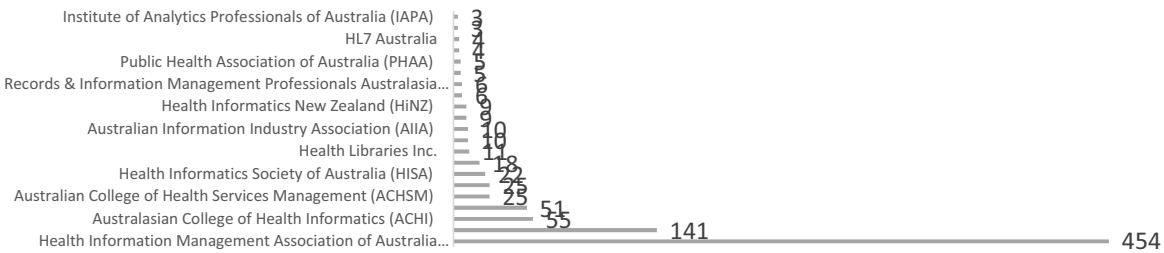


Figure 6 –Distribution of professional and industry membership

Conclusions

In a data driven, technological world, the health information workforce has emerged to be a critical area in the health sector. Prior to the Australian Health information Workforce Census there was no standardized and valid means of measuring the workforce. This results from the Australian Health information Workforce Census will enable workforce planning and enable forecasting of the future workforce needs.

Acknowledgements

Funding for this project was provided by the University of Tasmania.

The authors wish to acknowledge the Australian Health Information Workforce Census Management Group members (Vicki Bennett, Julie Brophy, Meredith Makeham, Christopher Pearce, Ann Ritchie, Louise Schaper, and Monica Trujillo) and associated partner organizations.

The Census Project has been approved by the University of Tasmania Social Science Human Research Ethics Committee.

Further information about the Australian Health Information Workforce Census can be found at: http://www.utas.edu.au/business-and-economics/hiwcensus

References

[1] K. Butler-Henderson and K Gray, *Health information Workforce Census*, (2018). http://www.utas.edu.au/business-and-economics/hiwcensus (accessed November 17, 2018).

[2] Health Workforce Australia, *Health information workforce report*, 2013.

[3] Health Information Management Association of Australia, *Health information workforce summit 2016 summary and outline of agreed actions*, 2016.

[4] Health Information Management Association of Australia, *Health information workforce summit 2015 report*, (2016) http://www.himaa2.org.au/index.php?q=node/2898, (accessed November 17, 2018).

[5] F Martín-Sánchez & K Gray , Recognition of health informatics in Australian standard classifications for research, occupation and education, (2014), *Studies in Health Technology and Informatics*, 204: 92-97.

[6] M Kammermann, *Census of Australian health libraries and self-described health librarians working in settings other than libraries final report*, (2016), https://www.alia.org.au/sites/default/files/CENSUS%20of%20Aus%20Hlth%20Libs%202012-14_Final%20Report_2016.pdf, (accessed November 17, 2018).

[7] M Legg & B Lovelock, *A review of the Australian health informatics workforce*, Health Informatics Society of Australia, Melbourne, 2009.

[8] Australian Institute of Health and Welfare, *The coding workforce shortfall*, cat. no. HWL 46, AIHW, Canberra, 2010.

[9] Canada Health Infoway, Canadian Health Information Management Association, Canadian Institute for Health Information, and COACH, *Health Informatics and Health Information Management: human resources outlook, 2014 – 2019*, (2014), https://www.echima.ca/uploaded/pdf/reports/HI-HIM-HR-Outlook-Report-Final-w-design.pdf, (accessed November 17, 2018).
[10] M Whittaker, N Hodge, RE Mares & A Rodney, Preparing for the data revolution: identifying minimum health information competencies among the health workforce, (2015), *Human Resources for Health* 13: 17.
[11] CJ Gibson, BE Dixon, & K Abrams, Convergent evolution of health information management and health informatics – a perspective on the future of information professionals in health care, (2015), *Applied Clinical Informatics,* 6: 163-184.
[12] K Butler-Henderson, R Lawrance, S Low, N Donnolley, & J Lee, Researching the health information workforce, (2016), *Health Information Management – Interchange*, vol. 6 (2), pp. 32-35.
[13] K Butler-Henderson, K Gray, D Greenfield, S Low, C Gilbert, A Ritchie, M Trujillo, V Bennett, & LK Schaper, The development of a national census of the health information workforce: expert panel recommendations, in *Integrating and Connecting Care*, R. A, S. LK, and W. S, Editors. (2017), Studies in Health Technology and Informatics: The Netherlands. p. 8-13.
[14] Commonwealth of Australia, 2015 Intergenerational report. Australia in 2055, (2015), Commonwealth of Australia: Parks, Australia.
[15] L Brooke, *Human resource costs and benefits of maintaining a mature age workforce, (*2003), International Journal of Manpower, 24(3): p. 260-283.
[16] K Butler-Henderson, *Health information management 2025. What is required to create a sustainable profession in the face of digital transformation?*, (2017), University of Tasmania: Launceston, Australia., himaa2.org.au/sites/default/files/Butler-Henderson_HIM_In_2025_Jan2017.pdf, (accessed November 17, 2018).

Address for correspondence

Dr Kerryn Butler-Henderson
Locked Bag 1317
University of Tasmania
Launceston 7250
Tasmania, Australia
Kerryn.Butlerhenderson@utas.edu.au
+61363243329

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190406

NewCope: A Theory-Linked Mobile Application for Stress Education and Management

Laura Carter[a], Deevakar Rogith[a], Amy Franklin[a], Sahiti Myneni[a]

[a] *National Center for Cognitive Informatics and Decision Making in Healthcare*
University of Texas School of Biomedical Informatics at Houston, TX, USA

Abstract

The negative effects of long-term stress on health outcomes are well-documented. Emerging technologies that harness mobile technologies have been linked to positive effects on stress management. However, the ways in which existing inter- and intrapersonal theories of behavior change are integrated into development processes of these mHealth technologies for stress coping are limited. In this paper, we present a novel theory-driven approach to develop and implement a sustainable mobile application for stress education and management. Specifically, we integrate the taxonomy of Behavior Change Techniques and user engagement framework to model and adapt theory-driven techniques in the context of mobile technologies. A total of 12 behavior change techniques were incorporated into our mobile application. Initial user evaluation and usability testing was conducted. Results indicate heuristic modifications could improve overall delivery of content, and potential user satisfaction is likely. We conclude that this novel approach may have implications well beyond stress management.

Keywords:
Psychological stress, behavior therapy, telemedicine.

Introduction

Psychological stress (in lay terms, stress) occurs when the demands on an individual exceed the ability to cope [1]. The negative effects of long-term stress on human health has been well-documented [2]. Stress includes mental health symptoms such as anxiety and depression [3]. The global cost of mental illness is estimated at one trillion dollars and affects 300 million people [4]. Americans' average stress levels are 4.8 out of 10, 45% experience sleeplessness due to stress, and 75 % experienced 1 or more symptoms of stress in the previous month, including nervousness or anxiousness [5]. Stress costs the U.S. enterprise economy up to 2.6% of the GDP annually and in Europe the cost can be as much as 3.3% [6]. Stress can be managed using self-care behaviors, such as avoiding alcohol, finding social support, eating and sleeping well, and staying active [7]. Literature shows that evidence-based practice can reduce perceived stress [8]. Recently, mHealth applications promoting behavioral change were found to have positive effects on workers' perceived stress [9].

Emerging trends in mobile smartphone apps can provide new frontiers to management of chronic conditions such as stress [10]. 77% of Americans own mobile smartphones, and 90% own a mobile phone of some kind [11]. With mobile network capabilities and affordable smartphones reaching well beyond other infrastructure in countries with low- and mid-range economies, mHealth and health promotion efforts in these areas are able to be offered in new ways via mobile phone technology, such as personalized care or public health [12]. Literature suggests that cost effectiveness of mHealth solutions is likely [13]. Platforms for mHealth or eHealth with designs that allow for innovation in aspects such as use, utility, financial impact, sustainability, scalability, and impact will be needed for advancement [14,15]. Mobile health technology can improve user adherence in a variety of health behaviors [16,17]. It is known that mHealth also can deliver education and patient self-care, among other health benefits [18]. Evidence-based smartphone apps for mental health have been shown to be effective and to have the potential to reach beyond current treatment limitations [19]. Large-scale clinical trials establishing effectiveness and efficacy of using app based modalities are sparse and often indecisive. Despite the overall lack of evidence-based smartphone apps in existence, there is potential for successful delivery of evidence-based stress management strategies [20]. Efforts have been made to understand the utility and manifestation of behavior change techniques, however, specific software development approaches to generate synergistic collaborations between technology developers and health research are limited [21]. In this paper, we present an integrative approach to infuse theory-driven techniques and engagement features to develop novel, evidence-based, and engaging mHealth applications. Specifically, we utilized the taxonomy of Behavior Change Techniques (BCT taxonomy) [22] and the Patient engagement framework [23] to consolidate uniquely identified behavior change concepts and user engagement features in a mobile format. In the next sections of the paper, we describe in detail our approach for the development of a mobile application, NewCope, and initial results of the system evaluation through user testing and usability analysis.

Related Works

Existing work focusses on reviews of smartphone apps that deliver stress management behavior therapy. For example, the "Tension Tamer" app, which uses systolic blood pressure as a measure of stress, focused on breathing awareness meditation [24]. Similarly, heartrate has been used as a measure of stress in the "It's Time to Relax" app, which focuses on mindfulness [25]. Similarly, the "Virtual Hope Box" app was designed based on Cognitive Behavioral Therapy (CBT) [26]. "Headspace," which utilizes mindfulness and meditation, has been used for stress management delivery in multiple mHealth stress management studies [27, 28].

Additionally, other mHealth solutions have been developed to address the needs facing delivery of stress management therapy. Method of Levels transdiagnostic cognitive therapy

was used in the development of the "MindSurf" app, which showed positive results in preliminary function and usability testing [29]. And importantly, the Virgin Pulse Global Initiative (VPGI) was an mHealth solution that overcame existing paradigms to show a positive outcome for workers completing the mHealth behavioral program [9].

Methods

Theoretical Rationale

Taxonomy of Behavior Change Techniques

Michie et al. [22] defined a set of theory-linked techniques that can be used to integrate, implement, and harness multiple theoretical constructs to unify inter- and intra-personal processes of human behavior into technology features that can assist and support sustainable positive health behaviors. Their taxonomy of 93 theory-linked techniques is the first step towards creating a model that provides a snapshot of intervention content in the context of theory-driven behavior change constructs [22]. The taxonomy provides a common vocabulary to understand the ways that sociobehavioral and cognitive constructs of the existing behavior change theories can be operationalized in a specific intervention. The complete list of techniques of the taxonomy with definitions and detailed examples can be found here [22].

Patient Engagement Framework (PEF)

The PEF has been developed by the Healthcare Information and Management Systems Society (HIMSS) through a cumulative layering of five phases: "inform me," "engage me," "empower me," "partner with me," and "support my e-community." A total of nine features have been specified at the highest engagement level, including 'information and way-finding', 'e-tools', 'forms', 'patient-specific education', 'patient access and use', 'patient generated data', 'interoperable records', 'collaborative care and community support'[23]. This framework was used in our study to sketch the functionality of the mHealth infrastructure which potentially facilitates the adoption of the tool, self-management, goal setting and reinforcement, peer support, and patient-provider communication.

Technology-Theory-Engagement Mapping

A mapping process was conducted to identify BCT techniques in conjunction with user engagement features that could be used to operationalize the delivery of stress management-related content to users. The BCT techniques allowed us to understand the characteristics of the information that should be delivered to the user. We also specified the level of intended user engagement to characterize the granularity and complexity of the system features. Tables 1 and 2 provide a summary of the BCTs and engagement functions that were selected for implementation at feature level and content level.

Development

NewCope was built on the opensource Ionic Framework (ICL version 4.1.1) [30]. Native Ionic features such as modals, cards, slides and buttons were used in the app design, in addition to custom coded features such as the progress bar and imbedded "fun fact" questions. The NewCope design does not necessitate the collection of any Personal Health Information (PHI) including email address or other personally identifying user information, so Google's Firebase database was used for the sake of simplicity and ease of set-up. User authentication was handled via a user-generated username and password.

NewCope has five main component pages, including a homepage with primarily informational resources, a dashboard page that provides feedback to the user, a self assessment page, a journal page, and a page that offers information about the app, help to the user, and a place to contact the researchers.

Table 1– Theory Mapping for App Content and Functions

BCT name	Definition	Example
Instruction on how to perform the behavior	Advise or agree on how to perform the behavior	Education
Social comparison	Draw attention to others' performance to allow comparison with the person's own performance	Education
Self-talk	Prompt positive self- talk before and during the behavior	Daily task
Valued self-identity	Advise the person to write or complete rating scales about cherished value or personal strength as a means of affirming the person's identity as part of behavior change strategy	Journal
Information about health consequences	Provide information about health consequences of performing the behavior	Education
Incentive	Inform that a reward will be delivered if and only if there has been effort and/or progress in achieving the behavioral outcome	Daily task
Self-monitoring of behavior	Establish a method for the person to monitor and record their behavior(s) as part of a behavior change strategy	Progress
Behavior practice/rehearsal	Prompt practice or rehearsal of the performance of the behavior one or more times in a context or at a time when the performance may not be necessary, in order to increase habit and skill	Daily task
Focus on past success	Advise to think about or list previous successes in performing the behavior	Daily task
Social support (emotional)	Advise on, arrange, or provide emotional social support for performance of the behavior	Daily task
Credible source	Present verbal or visual communication from a credible source in favor of or against the behavior	Education
Restructuring the social environment	Change, or advise to change the social environment in order to facilitate performance of the wanted behavior	Daily task

Table 2- Engagement Features and Ontology Relations

Engagement phase	Description	Technology feature
Engage (Level 2)	e-Tools	Perceived stress scale integration
Self-monitoring		Daily task
Goal setting		Stress feedback
Create synergy and extend reach (Level V)	Education	Journal
Self-management	User-specific feedback	Daily task
Daily reminders (e.g. Lifestyle tip)		Progress summary

The health education materials provided in the application relate to stress prevalence [31, 32], guides meditation, stress management techniques associated with relatable people as

#hashtag signs. The dashboard includes a checklist of three daily tasks designed to help the user adopt and sustain positive health changes that can aid them in stress coping with user-specific feedback, thus promoting self-monitoring habits. As can be seen in Figure 1, examples of tasks on the daily checklist are "listen to music," "go through old photo albums," "call a loved one", and "write your personality strengths." These tasks are related to the specific BCT techniques, "Restructuring the social environment", "Valued self-identity", and "Social support (emotional)" respectively.

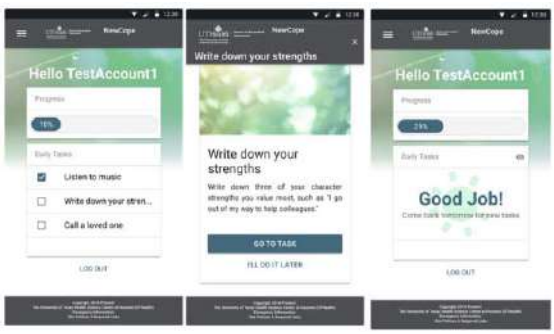


Figure 1– "NewCope" Dashboard

The self-assessment page utilizes the Perceived Stress Scale to evaluate the user's perceived stress level as low, medium or high [33]. The monitoring instrument addresses specific BCT techniques, "Feedback and monitoring" and "Goals and planning." The journal page is designed for use both in daily tasks that the user is asked to complete, such as "write down your strengths," which embodies "Self-belief", and for general use by the user, which further supports "Identity." Journal entries are designed to appear on "sticky note" graphic elements that can be edited, deleted, and rearranged by the user.

Evaluation

The prototyped app was reviewed for usability and accessibility and for use by human participants. One usability expert evaluated the user interface and generated a list of heuristic violations according to the Nielsen–Shneiderman Heuristics [34]. For each violation a severity rating (scale 1 – 4 where 1-Minor, 2-Moderate, 3-Major, 4-Catstrophic) was assigned. The design was refined to meet shortfalls that were most relevant to user testing. A secondary usability evaluation was conducted with five participants, selected based on recruitment by the researchers. The participants were recruited to test the app via a protocol of six timed tasks and the System Usability Scale (SUS) [35]. The UTHealth Internal Review Board approved the study for human participants.

Results

Initial Evaluation

The results from the heuristic evaluation revealed a total of 47 violations, eight of which were minor, 14 moderate, 19 major, and six catastrophic. Most violations represented issues with consistency (see Table 3).

Table 3– Heuristic Evaluation Summary

Violations	Minor	Moderate	Major	Catastrophic	Total
Consistency	5	3	7	1	16
Control		1	1	1	3
Flexibility		1	1		2
Help		2	1	1	4
Language	2	1	4	1	8
Match		4	1	1	6
Memory			1		1
Minimalism	1				1
Visibility		2	3	1	6
Total	8	14	19	6	47

The results from the user evaluation were as follows: All users were successful in completing all tasks, though some required assistance. The percentage of independent user completion was between 40 and 80% for all tasks, with user satisfaction maintaining a score between 1.2 and 1.8 for all tasks. The average completion time for tasks was between 8.0 and 126.8 seconds (see Table 4).

Table 4– Task Completion Summary

	Task Area					
	Create Ac-count	Self Assess-ment	Checklist	Video	Journal Entry	Article
User completed independently	80%	40%	60%	80%	80%	80%
Average completion time (sec)	50.4	126.8	8.0	88.4	38.4	59.0
Average user satisfaction 1-high 5-low	1.6	1.8	1.8	1.5	1.2	1.2

As seen in Table 5, the SUS scale results for positively correlated items showed an average of 4.04 out of 5, while negatively correlated items showed an average of 1.28 out of 5. Average item scores for positively correlated items were between 3.6 and 5.0, while average negatively correlated item scores were between 1.0 and 1.4.

Table 5– SUS Summary

Positive Items	**Frequency**	**Easy**	**Integration**	**Learnable**	**Confident**
Mean	3.6	3.8	3.8	5.0	4.0
Range	1-5	2-5	2-5	5	2-5
Negative Items	**Complex**	**Technical**	**Inconsistent**	**Cumbersome**	**Acquisition**
Mean	1.4	1.2	1.4	1.4	1.0
Range	1-2	1-2	1-3	1-2	1-1

Discussion

While the sample size limits the statistical validity of the quantitative data, we do observe the clustering of responses on the SUS items, particularly the negatively correlated items. This potentially indicates that participants in the trial were more consistently not put off by the system than that there was a consistent desire among the participants to use the system. As participants were not screened for stress level or desire for stress management, this result may not be unexpected. In fact, our findings from the evaluation determined that the average user stress level was 21.2 out of 40, which is classified as "moderate stress." Qualitative user data has indicated that users were often uncertain if task completion for the purposes of the trial included clicking buttons, with 5 of 11 trial administrator interventions based on helping the user to decide on button pressing behavior. This was primarily seen in the Self-assessment task (n = 3). Additionally, qualitative data shows that users had trouble locating the task checklist on the

dashboard screen when described as a "checklist." This accounted for the 40% administrator assistance rate on the View Checklist task, and 20% of the administrator assistance rate on the Self-Assessment task. User suggestions included using larger font sizes for the buttons, moving the "quit assessment" button further away from the "next question" button, and having Google prompt the user to automatically remember the password. User results do not, however, indicate a barrier to delivering BCTs via a smartphone app. All heuristic and user-identified barriers to the usage of the app were based on the user experience and user interface critiques. Design features can be further refined to create a smoother user experience while maintaining the necessary BCT components.

Methodical integration of known behavior change techniques has resulted in the convergence of constructs from multiple theories of behavior change including the Transtheoretical Model of Change [36], Social Cognitive Theory [37], and Health Belief Model [38], thereby ensuring our application harnesses inter- and intra-personal processes of behavior change [39]. Further, predetermined engagement levels allowed rapid prototyping with a generalized understanding of software features and malleable modes of theory-driven content delivery. This, while leaving room for clinical partner and wellness service integration, which is vital for large-scale dissemination of health promotion technologies.

Limitations

There are several limitations to our study. We have used a specific theoretical taxonomy to guide the development process. While the taxonomy is integrative, participatory and crowdsourced design activities that employ social listening can result in a holistic development approach. Additional features ,including online communities and smart technology synching, are important to improve the user experience and sustained use of NewCope. Our evaluation results are limited by the size of the participant pool and did not account for long term user engagement measures. Future studies should consider large-scale evaluations for better understanding of the acceptance of technologies such as NewCope.

Conclusions

Mental well-being, chronic disease management, and general health are affected by an individual's ability to manage stress, playing a role in one's happiness and productivity. Our paper describes a novel approach for the integration of evidence-based, theory-driven techniques in the digital era. The proposed approach is generalizable to other health conditions and can have implications well beyond stress management. Integration of theoretical models of human behavior into emerging technological domains such as mobile applications is important to facilitate effective dissemination of evidence-based strategies at scalable and sustainable levels. Attrition with the use of technology-based health promotion can be addressed with the proper implementation of user engagement principles.

Acknowledgements

Research reported in this publication was partially supported by the National Cancer Institute of the National Institutes of Health under Award Number R21CA220670 and funds by the Patient Safety Quality Care project. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] R.S. Lazarus, and S. Folkman. Stress, Appraisal and Coping, Springer, New York, 1984.
[2] National Institute of Mental Health. 5 Things You Should Know About Stress, (n.d). https://www.nimh.nih.gov/health/publications/stress/index.shtml (accessed October 25, 2018).
[3] American Heart Association Lower Stress: How does stress affect the body? (n.d.). http://www.heart.org/en/healthy-living/healthy-lifestyle/stress-management/lower-stress-how-does-stress-affect-the-body (accessed October 25, 2018).
[4] WHO | Mental health in the workplace, *WHO*. (n.d.). http://www.who.int/mental_health /in_the_workplace/en/ (accessed October 25, 2018).
[5] American Psychological Association. Stress in America, (2017). https://www.apa.org/news/press/release /stress/2017/state-nation.pdf (accessed October 11, 2018).
[6] J.P. Brun. Work-related stress: scientific evidence-base of risk factors, prevention and costs, (n.d.). https://www.who.int/occupational_health/topics/brunpres0307.pdf?ua=1(accessed October 22, 2018).
[7] Centers for Disease Control and Prevention. Coping with Stress, (2018). https://www.cdc.gov/features/copingwithstress/index.html (accessed October 21, 2018).
[8] S.B. Stillwell, A.L. Vermeesch, and J.G. Scott, Interventions to Reduce Perceived Stress Among Graduate Students: A Systematic Review With Implications for Evidence-Based Practice, *Worldviews on Evidence-Based Nursing*. **14** (2017) 507–513. doi:10.1111/wvn.12250.
[9] D. Meyer, M.W. Jayawardana, S.D. Muir, D.Y.-T. Ho, and O. Sackett, Promoting Psychological Well-Being at Work by Reducing Stress and Improving Sleep: Mixed-Methods Analysis, *J Med Internet Res*. **20** (2018). doi:10.2196/jmir.9058.
[10] S.M. Coulon, C.M. Monroe, and D.S. West, A Systematic, Multi-domain Review of Mobile Smartphone Apps for Evidence-Based Stress Management, *Am J of Preventive Medicine*. **51** (2016) 95–105. doi:10.1016/j.amepre.2016.01.026.
[11] Pew Research Center. Mobile Fact Sheet, (n.d.). http://www.pewinternet.org/fact-sheet/mobile/ (accessed November 2, 2018).
[12] M. Kay, J. Santos, M. Takane. MHealth: new horizons for health through mobile technologies, WHO Global Observatory for eHealth and the World Health Organization, Geneva, 2011. http://www.who.int /goe/publications/goe_mhealth_web (accessed November 2, 2018).
[13] S.J. Iribarren, K. Cato, L. Falzon, and P.W. Stone, What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions, *PloS One*. **12** (2017) e0170581. doi:10.1371 /journal.pone.0170581.
[14] J. Depasse, A. Caldwell, D. Santorino, E. Bailey, S. Gudapakkam, D. Bangsberg, and K. Olson, Affordable medical technologies: bringing Value-Based Design into global health, *BMJ Innov*.**2** (2016) 4–7. doi:10.1136/bmjinnov-2015-000069.
[15] World Health Organization. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment, (2016). http://www.who.int /reproductivehealth/publications/mhealth/digital-health-interventions/en/ (accessed October 25, 2018).
[16] Y. Gandapur, S. Kianoush, H.M. Kelli, S. Misra, B. Urrea, M.J. Blaha, G. Graham, F.A. Marvel, and S.S. Martin, The role of mHealth for improving medication

adherence in patients with cardiovascular disease: a systematic review, *Eur Heart J Qual Care Clin Outcomes*. **2** (2016) 237–244. doi:10.1093/ehjqcco/qcw018.
[17] N.L. Ifejika, E.A. Noser, J.C. Grotta, and S.I. Savitz, Swipe out Stroke: Feasibility and efficacy of using a smart-phone based mobile application to improve compliance with weight loss in obese minority stroke patients and their carers, *International Journal of Stroke*. **11** (2016) 593–603. doi:10.1177/1747493016631557.
[18] M. Alghamdi, H. Gashgari, and M. Househ. A systematic review of mobile health technology use in developing countries, *Stud Health Technol Inform.* **213** (2015) 223-226. doi:10.3233/978-1-61499-538-8-223
[19] P. Chandrashekar, Do mental health mobile apps work: evidence and recommendations for designing high-efficacy mental health mobile apps, *Mhealth*. **4** (2018). doi:10.21037/mhealth.2018.03.02.
[20] T. Donker, K. Petrie, J. Proudfoot, J. Clarke, M.-R. Birch, and H. Christensen, Smartphones for Smarter Delivery of Mental Health Programs: A Systematic Review, *J Med Internet Res*. **15** (2013). doi:10.2196/jmir.2791.
[21] E. de Korte, N. Wiezer, M. Bakhuys Roozeboom, P. Vink, and W. Kraaij, Behavior Change Techniques in mHealth Apps for the Mental and Physical Health of Employees: Systematic Assessment, *JMIR Mhealth Uhealth*. **6** (2018). doi:10.2196/mhealth.6363.
[22] S. Michie, M. Richardson, M. Johnston, C. Abraham, J. Francis, W. Hardeman, M.P. Eccles, J. Cane, and C.E. Wood, The Behavior Change Technique Taxonomy (v1) of 93 Hierarchically Clustered Techniques: Building an International Consensus for the Reporting of Behavior Change Interventions, *Ann Behav Med*. **46** (2013) 81–95. doi:10.1007/s12160-013-9486-6.
[23] HIMSS Patient Engagement Framework. https://www.himss.org/himss-patient-engagement-framework (accessed November 24, 2018).
[24] Z.W. Adams, J.C. Sieverdes, B. Brunner-Jackson, M. Mueller, J. Chandler, V. Diaz, S. Patel, L.R. Sox, S. Wilder, and F.A. Treiber, Meditation smartphone application effects on prehypertensive adults' blood pressure: Dose-response feasibility trial., *Health Psychology*. **37** (2018) 850–860. doi:10.1037/hea0000584.
[25] C. Carissoli, D. Villani, and G. Riva, Does a Meditation Protocol Supported by a Mobile Application Help People Reduce Stress? Suggestions from a Controlled Pragmatic Trial, *Cyberpsychology, Behavior, and Social Networking*. **18** (2015) 46–53. doi:10.1089/cyber.2014.0062.
[26] N.E. Bush, D.J. Smolenski, L.M. Denneson, H.B. Williams, E.K. Thomas, and S.K. Dobscha, A Virtual Hope Box: Randomized Controlled Trial of a Smartphone App for Emotional Regulation and Coping With Distress, *PS*. **68** (2016) 330–336. doi:10.1176/appi.ps.201600283.
[27] J. Laurie, and A. Blandford, Making time for mindfulness, *Int J Med Inform*. **96** (2016) 38–50. doi:10.1016/j.ijmedinf.2016.02.010.
[28] L. Wen, T.E. Sweeney, L. Welton, M. Trockel, and L. Katznelson, Encouraging Mindfulness in Medical House Staff via Smartphone App: A Pilot Study, *Acad Psychiatry*. **41** (2017) 646–650. doi:10.1007/s40596-017-0768-3.
[29] T.A. Carey, J. Haviland, S.J. Tai, T. Vanags, and W. Mansell, MindSurf: a pilot study to assess the usability and acceptability of a smartphone app designed to promote contentment, wellbeing, and goal achievement, *BMC Psychiatry*. **16** (2016) 442. doi:10.1186/s12888-016-1168-z.
[30] Ionic Framework. https://ionicframework.com/ (accessed October 25, 2018).
[31] Need stress relief? Try the 4 A's, *Mayo Clinic*. (n.d.). https://www.mayoclinic.org/healthy- lifestyle/stress-management/in-depth/stress-relief/art-20044476 (accessed October 11, 2019).
[32] The Busy Lifestyle. 6 New Stress Statistics from around the World (It's Time to Take It Easy), *The Busy Lifestyle*. (2018). https://thebusylifestyle.com/stress-statistics-worldwide/ (accessed October 11, 2019).
[33] S. Cohen, T. Kamarck, and R. Mermelstein, A Global Measure of Perceived Stress, *Journal of Health and Social Behavior*. **24** (1983) 385–396. doi:10.2307/2136404.
[34] J. Zhang, T.R. Johnson, V.L. Patel, D.L. Paige, and T. Kubose, Using usability heuristics to evaluate patient safety of medical devices, *Journal of Biomedical Informatics*. **36** (2003) 23-30. doi:10.1016/S1532-0464(03)00060-1
[35] System Usability Scale (SUS). https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html (accessed November 15, 2018).
[36] J.O. Prochaska, and W.F. Velicer. The transtheoretical model of health behavior change, *Am J Health Promot.* **12** (1997) 38-48. doi:10.4278/0890-1171-12.1.38
[37] R.A. Schwarze, and A. Luszczynska, Social cognitive theory, in: M. Conner, and P. Norman (Eds.), Predicting Health Behaviour, 2nd edition, Open University Press, Maidenhead, 2005: pp. 127–169.
[38] G. Hochbaum, I. Rosenstock, and S. Kegels. Health Belief Model, United States Public Health Service, Washington D.C., 1952.
[39] K. Glanz, B.K. Rimer, and K. Viswanath, eds., Health Behavior and Health Education: Theory, Research, and Practice, 4th edition, Jossey-Bass, San Francisco, 2008.

Address for Correspondence

Laura Carter, The University of Texas School of Biomedical Informatics at Houston, 7000 Fannin, Houston, TX 77030, laura.e.carter@uth.tmc.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190407

User-Centered Development of a Behavioral Economics Inspired Electronic Health Record Clinical Decision Support Module

Sara Kuppin Chokshi[a], Andrea Troxel[a], Hayley Belli[a], Jessica Schwartz[b], Saul Blecker[a], Caroline Blaum[a], Adam Szerencsy[b], Paul Testa[b], Devin Mann[a,b]

[a] *Department of Population Health, New York University School of Medicine, New York, NY, USA,*
[b] *Medical Center Information Technology, NYU Langone Health, New York, NY, USA*

Abstract

Changing physician behaviors is difficult. Electronic health record (EHR) clinical decision support (CDS) offers an opportunity to promote guideline adherence. Behavioral economics (BE) has shown success as an approach to supporting evidence-based decision-making with little additional cognitive burden. We applied a user-centered approach to incorporate BE "nudges" into a CDS module in two "vanguard" sites utilizing: (1) semi-structured interviews with key informants (n=8); (2) a design thinking workshop; and (3) semi-structured group interviews with clinicians. In the 133 day development phase at two clinics, the navigator section fired 299 times for 27 unique clinicians. The inbasket refill alert fired 124 times for 22 clinicians. Fifteen prescriptions for metformin were written by 11 clinicians. Our user-centered approach yielded a BE- driven CDS module with relatively high utilization by clinicians. Next steps include the addition of two modules and continued tracking of utilization, and assessment of clinical impact of the module.

Keywords:

user computer interface, clinical decision support, electronic health record, behavioral economics

Introduction

Changing ingrained physician behaviors is difficult [1]. Integrating behavioral economic (BE) strategies into electronic health records (EHR) using various clinical decision support (CDS) tools is a novel approach to improving guideline adherence that also seeks to minimize negative impacts on clinical workflow and cognitive load. This study's hypothesis is that employing a user-centered approach to design a CDS module that incorporates BE strategies will result in a low burden tool to support provider adherence to guideline-based recommendations.

Maximizing the potential of CDS through behavioral economics and user-centered-design

Large, systematic reviews of EHR-based CDS have demonstrated a moderate ability to reduce morbidity, utilization, and costs [2, 3]. These modest improvements, however, are undermined by the well-documented problems of alert fatigue and poor workflow integration, which together blunt the potential of the EHR and CDS to improve healthcare outcomes [4]. New approaches are needed to complement the traditional alerts, reminders, and other CDS tools that disrupt clinical workflow, increase cognitive load, and stress the limited capacity of clinicians to rationally process and evaluate the diverse and competing demands on clinician attention. Behavioral economics has shown success as an approach to support evidence-based decision-making with little additional cognitive burden to clinicians; however BE strategies are rarely leveraged in current CDS tools.

A user-centered design approach employing design thinking strategies is common in digital development projects outside of healthcare. Prior research, including that of the research team, show the value of taking a user-centered approach to development of CDS [5-8]. A user-centered approach employs design thinking exercises designed to identify opportunities to incorporate BE in CDS related to the target user – in this case physicians – to stimulate idea generation. Design thinking strategies are well-suited to exploring physician motivations and workflow opportunities which require understanding for successfully incorporating BE to create feasible, usable decision support solutions likely to achieve adoption.

Choosing Wisely guideline for older adults with diabetes is an opportunity to explore this potential

The objective of this study was to develop a scalable, EHR module that incorporates BE strategies to promote appropriate diabetes care in older adults based on the American Geriatric Society's Choosing Wisely (CW) guideline, which aims to reduce over-treatment to benefit older adults with diabetes.

The CW guideline for older adults with diabetes recommends clinicians "avoid using medications other than metformin to achieve hemoglobin HbA1c<7.5% in most older adults; moderate control is generally better" [9]. This offers an opportunity to explore the potential of BE in the EHR to support guideline adherence for providers. In spite of this guideline, which recommends less aggressive target A1c levels based on older age and lower life expectancy, a substantial number of older adults with diabetes continue to be prescribed more aggressive therapies that may not only be unnecessary but also harmful [10-12].

This paper 1) outlines a user-centered approach to the development of a BE inspired EHR-CDS module for promoting guideline-based treatment of older adults with diabetes, and 2) reports on module activity metrics from the 4.5 month development phase of the project (June-October 2018).

Methods

Overview

This study employed a pragmatic (emphasis on real-world clinical workflows), user-centered approach to develop a new BE inspired CDS module (BE-EHR) to improve provider adherence to the CW guideline targeting over-treatment among older adults with diabetes.

To accomplish this, we conducted:

1. semi-structured interviews with key informants (n=8);
2. a two-hour design thinking workshop with a multidisciplinary group of clinicians, informaticists, EHR analysts, product designers and others to derive and refine initial module ideas; and
3. semi-structured group interviews with clinic leaders and clinicians at each of two "vanguard" sites to elicit feedback on three draft module components (inbasket refill, medication preference list, navigator tab).

Setting and population

The intervention was developed by a multi-disciplinary research team at a large academic medical center and deployed in two "vangaurd" ambulatory primary care practices serving patients with a diverse range of socio-demographic characteristics. The BE-EHR module was deployed for internal medicine and endocrinology physicians and nurse practioners.

Two "vanguard" practices were selected purposefully based on key characteristics: a primary care focus, a relevant patient population (i.e. adults over age 76 with diabetes), and a willingness of practice leadership to serve as a test site for implementation of module prototypes and provide periodic feedback.

Life expectancy algorithm

In order to build a user-centered CDS tool that triggers appropriately for the target patient population, algorithms were built into the BE-EHR module to drive the timing and content of module alerts that incorporate both patient life expectancy (high, medium, low) and target glycemic index per the CW guideline. These categories were defined in the algorithm as follows:

1. healthy older adults with an HbA1c target range of 7-7.5% and long life expectancy (defined here as 10+ years);
2. those with moderate comorbidity and a life expectancy of 3+ to 10 years, with a target range of 7.5-8%; and
3. those with multiple comorbidities and life expectancy of less than or equal to 3 years, with a target range of 8-9%. [13]

Analysis

Key informant, group interview data, and insights generated at the design thinking workshop were recorded by research staff in the form of field notes and summarized by usability theme and, when appropriate, by module component for rapid iteration of the prototype. Once findings were incorporated into the module prototype, the module was deployed in the two vanguard sites. EHR-based reports were built to track frequencies of module component firings and action taken (navigator tab component only). These metrics were calculated per number of unique patients and number of unique providers at the vanguard sites throughout the development phase (first 133 days of implementation). EHR reports also served to confirm the module and related outcomes reporting tools were firing as expected.

Results

Module development

Key informant and vanguard clinic group interviews identified the refill protocol and medication preference list as promising candidate CDS tools for the BE-EHR module based on their compatibility with provider workflows. These components were refined based group interview findings combined with current best practices in CDS development (e.g. avoid interruptive alerts) [14].

For example, the medication preference list component leveraged the behavioral economics principle of availability bias (tendency to rely on easy-to-access examples) by placing the CW recommended medication (metformin) at the top of the medication preference list (Figure 1).

Findings from the design thinking workshop included refinements to both the refill protocol and the medication preference list components. Two new ideas were generated that incorporated a combination of behavioral economic principles: one new EHR CDS component (a navigator tab for older adults with diabetes illustrated in Figure 2), and a non-EHR based supportive "campaign," as an adjunct to the BE EHR-CDS module to raise awareness of the CW guideline. The campaign, requiring significant additional design work, was not deployed in the development phase.

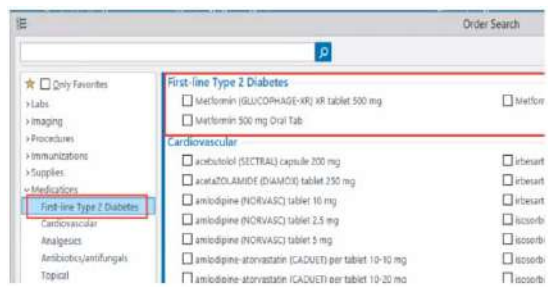


Figure 1- Medication preference list module component

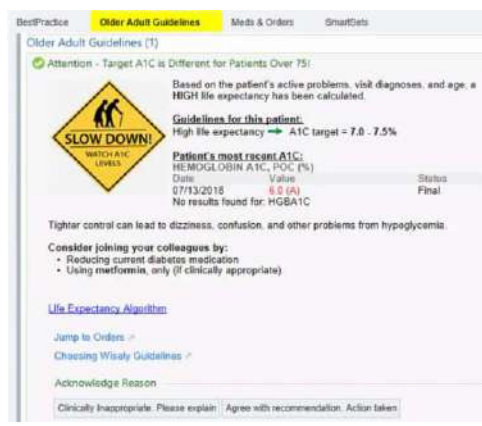


Figure 2- Navigator tab module component

Once prototyped, user feedback from site visits confirmed the compatibility of the navigator tab with clinical workflows and contributed to refinement of design and content.

Module activity

Table 1 shows the components of the nudges implemented during the development phase along with frequency of module firings. In the first 133 days at the two vanguard clinics, the navigator tab fired 299 times for 193 unique patients and 27 unique clinicians across sites. The inbasket refill alert fired 124 times for 84 unique patients and 22 unique clinicians (Table 1). Fifteen prescriptions for Metformin (guideline recommended medication if appropriate) were written to 14 unique patients by 11 unique clinicians – 31% of all clinicians (n=35) – who

prescribed metformin to the target population from the preference list at the two vanguard sites during this time period.

Table 1- BE-EHR module components activity

Module component	# fire	# patients	# clinicians
Inbasket refill	124	84	22
Navigator tab	299	193	27
	# Rx	**# patients**	**# clinicians**
Preference list Rx	15	14	11(31%)
Metformin Rx total	123	98	35

Discussion

Design thinking, employed as an approach to facilitate successful implementation and sustainability of the BE-EHR-CDS intervention, is a useful tool in building an intervention with institutional and clinician buy-in. Our user-centered approach to design yielded a behavioral economics driven CDS module with relatively high engagement by clinicians. The incorporation of behavioral economic principles into EHR CDS tools shows promise as a strategy to improve guideline adherence by addressing stubborn barriers, such as alert fatigue and poor compatibility with workflow, that can prevent CDS from having the desired impact on clinician behavior.

A pilot study incorporating finalized versions of the three module components developed and implemented in this vanguard phase as described here, along with additional components, is recently underway in 4 additional clinics. New module components include, for example, an email message to be sent to clinicians by clinic medical directors with content leveraging the behavioral economic principle of social comparisons (comparing their proportion of patients within CW A1c target range with that of peers at their clinic and institution wide) has been developed and will be sent out monthly throughout the 6 month pilot period.

The proposed BE-EHR module serves as a highly scalable platform for embedding a BE-based CDS into any EHR system. Importantly, this module can be easily applied to many other conditions in older adults and other populations where combining BE with EHR-based clinical decision support will be useful for improving guideline adherence such as CW recommendations related to preventative screening procedures (e.g. colonoscopies), or increasing compliance to tobacco cessation.

Limitations

While the design of the BE-EHR module reflects input from a large number and wide variety of key informants and end users, utilization numbers reflect only the two vanguard practices. This is appropriate, however, for a development phase of such a project and in line with the user-centered design approach which values early, iterative feedback prior to widespread tool deployment. Additionally, key informant and group interviews were recorded with field notes taken by research staff rather than audio-recorded and transcribed. This was a conscious decision by the research team to serve the quick CDS development timeline and to allow for rapid iteration of module components. Finally, measurement of user engagment with or adoption of the module components is difficult given they are built, by design, not to be interruptive (requiring user interaction to resolve); process metrics collected therefore focus on opportunities for users to view module components.

Conclusions

The resulting BE-EHR module establishes a platform for exploring the ability of BE concepts embedded within the EHR to affect guideline adherence for other CW target areas. Moreover, it represents an exciting new channel for influencing provider behavior through less cognitively burdensome methods. Evidence and lessons learned from this study can potentially inform the design, testing, and implementation of similar interventions for other CW target conditions and beyond.

Acknowledgements

Research reported in this publication was supported by the National Institute of Aging under Award Number R21AG057382. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] D. Saddawi-Konefka, D.J. Schumacher, K.H. Baker, J.E. Charnin, and P.M. Gollwitzer, Changing Physician Behavior With Implementation Intentions: Closing the Gap Between Intentions and Actions, *Acad Med* **91** (2016), 1211-1216.

[2] T.J. Bright, A. Wong, R. Dhurjati, E. Bristow, L. Bastian, R.R. Coeytaux, G. Samsa, V. Hasselblad, J.W. Williams, M.D. Musty, L. Wing, A.S. Kendrick, G.D. Sanders, and D. Lobach, Effect of clinical decision-support systems: A systematic review, *Ann Intern Med* (2012).

[3] L. Moja, K.H. Kwag, T. Lytras, L. Bertizzolo, L. Brandt, V. Pecoraro, G. Rigon, A. Vaona, F. Ruggiero, M. Mangia, A. Iorio, I. Kunnamo, and S. Bonovas, Effectiveness of computerized decision support systems linked to electronic health records: a systematic review and meta-analysis, *Am J Public Health* **104** (2014), e12-22.

[4] M. Khalifa and I. Zabani, Improving utilization of clinical decision support systems by reducing alert fatigue: Strategies and recommendations, *Stud Health Technol Inform* **226** (2016), 51-54.

[5] D.A. Feldstein, R. Hess, T. McGinn, R.G. Mishuris, L. McCullagh, P.D. Smith, M. Flynn, J. Palmisano, G. Doros, and D. Mann, Design and implementation of electronic health record integrated clinical prediction rules (iCPR): A randomized trial in diverse primary care settings, *Implement Sci* **12** (2017), 37.

[6] L.J. McCullagh, A. Sofianou, J. Kannry, D.M. Mann, and T.G. McGinn, User centered clinical decision support tools: adoption across clinician training level, *Appl Clin Inform* **5** (2015), 1015-1025.

[7] D.M. Mann, J.L. Kannry, D. Edonyabo, A.C. Li, J. Arciniega, J. Stulman, L. Romero, J. Wisnivesky, R. Adler, and T.G. McGinn, Rationale, design, and implementation protocol of an electronic health record integrated clinical prediction rule (iCPR) randomized trial in primary care, *Implement Sci* **6** (2011), 109.

[8] Chokshi, S., & Mann, D. Innovating from within : A process model for user-centered digital development in academic medical centers, *JMIR Hum Factors* (in press).

[9] AGS, Choosing Wisely: An initiative of the ABIM Foundation, *American Geriatric Society*, 2017.

[10] K.J. Lipska, J.S. Ross, Y. Miao, N.D. Shah, S.J. Lee, and M.A. Steinman, Potential overtreatment of diabetes mellitus in older adults with tight glycemic control, *JAMA Intern Med* **175** (2015), 356-362.

[11] F.A. McAlister, E. Youngson, and D.T. Eurich, Treated glycosylated hemoglobin levels in individuals with diabetes mellitus vary little by health status: A retrospective cohort study, *Medicine* **95** (2016), e3894.
[12] R.G. McCoy, K.J. Lipska, X. Yao, J.S. Ross, V.M. Montori, and N.D. Shah, Intensive treatment and severe hypoglycemia among adults with type 2 diabetes, *JAMA Intern Med* **176** (2016), 969-978.
[13] A.G.S.C.W. Workgroup, American Geriatrics Society identifies five things that healthcare providers and patients should question, *J Am Geriatr Soc* **61** (2013), 622-631.
[14] J.A. Osheroff, JM Teich, B. Middleton, E.B. Steen, A. Wright, and D.E. Detmer, A roadmap for national action on clinical decision support, *JAMIA* **14** (2007), 141-145.

Address for correspondence

Sara Chokshi, DrPH
227 E. 30th St
New York, NY 10016
Sara.Chokshi@nyulangone.org

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190408

The Role of Personal Health Information Management in Promoting Patient Safety in the Home: A Qualitative Analysis

George Demiris[a], Shih-Yin Lin[b], Anne M. Turner[b,c]

[a] Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA, USA,
[b] Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, USA,
[c] Department of Health Services, University of Washington, Seattle, WA, USA

Abstract

Patient safety is a critical component of health care services; however, it has beent mostly conceptualized for the hospital sector. As home health care expands, it is important to examine the concept of patient safety in the home and identify opportunity for personal health information management (PHIM) tools to support and maximize patient safety. The goal of this study is to explore how PHIM can be a facilitator for patient safety in the home. We explore a comprehensive framework of patient safety in the home and identify the role of PHIM in this context. We analyzed the coded transcripts of in-depth interviews with 88 older adults (60 year and older), 56 family members or informal caregivers and 27 clinicians. Findings demonstrate the physical, emotional, social and functional dimensions of patient safety in the home and concrete ways for informatics tools to maximize safety aspects.

Keywords:

Patient Safety; Home Care Services; Health Records, Personal

Introduction

Home health care is the fastest growing segment of the US health care system with a rapidly increasing number of patients receiving care services in their home and a well-documented shift from acute to home care for patient populations across the lifespan. As patients with complex needs and their families are faced with challenges of care coordination in the residential setting, patient safety in the home needs to be examined to identify barriers and facilitators in this unique context. A significant body of literature has examined patient safety in clinical settings, mostly in the hospital setting, and various frameworks and instruments have been developed to define, measure and track patient safety in in-patient settings. Less emphasis has been placed on a formal examination of patient safety in the home setting, although recently more efforts are underway to address this area.

As Lang et. al. [1] point out, applying the notion of patient safety as conceptualized for the hospital sector (with a sole focus on error detection and adverse events) is too narrow and therefore, insufficient for home care. Safety in the home setting requires a broader definition to acknowledge the complexity of health conditions, residential settings and family dynamics and the unregulated and at times unpredictable process of health care delivery within the home. There are structural and procedural differences between the hospital and home setting that affect patient safety considerations. One example is the fact that home health staff often work alone in the field during a home visit with support resources available remotely (in contrast to clinical teams within the hospital that operate as teams of care physically co-located). Often, physicians rely to home care nurses to make assessments and communicate findings while they may have little or no direct contact with patients. Home health care nurses spend more time on documentation and addressing reimbursement issues than hospital nurses [2]. Additionally, patient autonomy is respected in both settings but overseeing the implementation of a behavioral or pharmaceutical or other intervention varies based on the opportunities for observation and reinforcement. Medication adherence is one such area where hospital-based care may be more effective in eliminating barriers. Patient variables such as reading skill, cognitive ability and financial resources have been found to affect patients' ability to safely manage medication regimens in the home [3] whereas these parameters are less instrumental in effective medication administration in the hospital setting. Finally, family members often play an essential role in the health care of a patient in the home and need to play an essential role in the discourse on patient safety [4]. As significant elements of the health care services delivered to patients become responsibilities of family or informal caregivers, challenges arise related to training, supervision, competencies and information flow during transitions of care that require coordination across multiple stakeholders and settings of care.

Efforts to address patient safety in the home include initiatives that target specific aspects of home care such as fall prevention [5], medication management [6], unplanned hospital admissions [7], functional outcomes [8], wound [9] and pressure ulcer management [10]. However, a more comprehensive conceptualization of patient safety in home care is needed in order to identify opportunities for tailored interventions. Lang has developed a model to outline four dimensions of safety in the home [11]: The physical dimension involves the physical attributes of the home care setting, including environmental hazards such as home layout and infrastructure, clutter, and unsanitary conditions. Key processes of care that affect home safety — medication management, infection control, nutrition, fall prevention, complex clinical care, and care coordination — are also included in this category. The emotional dimension of home care safety involves stress, trauma, and discomfort related to receiving and providing care. Finally, the social and functional dimensions of home care safety involve the community and the network of support, and the effects of health conditions on activities of daily living [12].

As with any health care process, delivery of care in the home setting involves information gathering and exchange among various stakeholders and between different systems of care. Home care patients and their families are called to manage personal health information generated from multiple sources and often between systems and across sites of care facing challenges of interoperability and threats to continuity of care. Personal health information management is critical to ensuring care coordination and high quality of care and affects all four dimensions of patient safety.

Personal health information management (PHIM), defined as the entire set of activities that "support consumers' access, integration, organization, and use of their personal health information" [13] is at the core of patient safety considerations in home care where care transitions, personal relationships, unsupervised caregiving work, and environmental attributes introduce unique challenges. The goal of this paper is to explore how PHIM can be a facilitator for patient safety in the home. We aim to discuss a comprehensive framework of patient safety in the home and identify the role of PHIM in facilitating patient safety based on a secondary data analysis of interview data with community dwelling older adults to highlight these considerations.

Methods

We conducted a qualitative analysis of interview data using the theoretical framework by Lang et. al. [11] to highlight the underlying elements of patient safety in the home. The qualitative data were generated as part of the SOARING study (soaringstudy.org), a five-year investigation of the personal health information needs and practices of adults 60 years and older living in a metropolitan area in the Pacific Northwest in the United States. The SOARING project focuses on information needs and personal health information management of older adults; patient safety in the home is not examined in that parent project but its large data set is used for a secondary analysis for the patient safety focus of the study presented here. Secondary data included interviews conducted over a three-year period with older adults and their friends and family members that were involved in their health and health information management [14]. Interviews focused on health, health information seeking, technology use and health information management. Purposive sampling ensured a diverse representation of gender, living situation, race, and income. Living situations included independent living, retirement communities, assisted living and homelessness. Inclusion criteria for participation included age 60 or over, ability to communicate in English, and lack of cognitive impairment. Additionally, interviews were conducted with health care providers who provide care to older adults. The interview protocol covered a broad range of topics including needs and preferences pertaining to personal health information management, and the personal stories of aiming to maintain independent aging. We analyzed the coded transcripts which were collected from in-depth in person interviews with 88 older adults (60 year and older) and phone interviews with 56 identified by older adults as friends and family who were involved in their health and health information management, and 27 clinicians (physicians, nurses, social workers). Older adult participants had an average age of 76 years, were predominantly female (69%), and white (72%). In terms of living arrangements, they lived in various settings: in independent private residences (25%), independent-shared dwellings (27%), retirement communities (27%), assisted living facilities (19%), and homeless (2%). The content analysis was informed by the theoretical framework by Lang et. al. [11] in order to highlight the underlying elements of patient safety in the home. We reviewed transcripts for their pertinence to patient safety issues in particular along the dimensions of care (physical, emotional and social and functional). For each of these themes we described the role of information technology as a patient safety facilitator.

Results

Table 1 provides a summary of the dimensions of patient safety in the home with example quotes from our corpus of qualitative data generated by interviews with older adults and their families and providers demonstrating instances where a specific dimension may become a challenge for patient safety.

The use of the framework by Lang et. al. [11] enabled us to confirm its validity by identifying the underlying dimensions in our own data and also inform how PHIM may become a facilitator in that context. The physical dimension of patient safety in the home context includes multiple sub-themes. There are numerous key processes of care that may include safety risks when information flow or oversight are deficient. These processes include medication management, infection control, fall prevention, nutrition, complex clinical care coordination and transitions of care. Challenges associated with medication management and adherence have been well documented and are expected to intensify with aging as patients may experience cognitive or functional limitations [15]. The physical dimension also includes environmental hazards that pertain to aspects such as the home infrastructure and physical layout of the residential space as well as potential clutter.

The emotional dimension of patient safety includes concepts such as burden, stigma, self-view, confidence and resilience, all psychological constructs that affect self-efficacy, communication and coping, ultimately impact patient autonomy and safety.

The social dimension includes elements such as family and community support and social isolation and loneliness. As Lang notes, patients may face increased risk to their safety simply because of the desire to remain at home "at any cost." Illnesses bring about inherent limitations in daily functioning, which may multiply as an illness progresses. However, the extent to which safety risks increase, and for whom, is a function of the patient's social matrix.

Finally, the functional dimension of patient safety includes numerous potential limitations such as visual, hearing, mobility, and cognitive limitations that may affect one's ability to safely manage their health condition in the home and the degree to which assistance or additional tools may be necessary to safely execute health related procedures and tasks.

Discussion

Our findings validate the dimenions of patient safety in the home setting as outlined by Lang's framework [11] and demonstrate how information management may act as mediator for these patient safety attributes. Informatics can play a role in facilitating information flow and exchage among

stakeholders to address and maximize patient safety. Various informatics tools such as passive monitoring sensors, personal health records, wearables, symptom management tracking and reporting apps can support the physical, emotional, social and functional dimensions of patient safety. Patient and family characteristics, available resources and the participating entities in transitions of care and coordination of services have to be considered when designing strategies to safeguard patient safety in home health care. Additionally, our findings demonstrate that home care patients may often struggle with information management challenges while trying to cope with complex care needs. Educational initiatives to increase awareness of patient safety among clinicians, family members and patients themselves can be tailored to address the unique challenges and opportunities as outlined in this framework. System designers can integrate informatics elements into the system overall functionalities to promote safety and well-being of home care patients. Personal health records and other PHIM tools may benefit from incorporating a patient safety perspective in their design in order to improve processes of health care and meaningfully engage or even empower home care patients and their families.

Conclusions

The findings of this qualitative analysis demonstrate patient safety concerns and challenges as experienced by older adults, family members and health care providers. Patient safety in the home needs to be examined beyond previous work that has defined and measured safety solely in the hospital setting. This is to our knowledge the first comprehensive inquiry of patient safety as a broad concept in the home setting engaging multiple stakeholder groups and examining the potential of informatics tools to serve as safety facilitators. Personal health information management and informatics tools can be implemented to promote patient safety in an inclusive approach that covers the physical, functional, social, emotional and social dimensions of patient safety. Our findings can inform the design of tailored patient safety interventions in home health care.

Acknowledgements

This research was supported through funding from the Agency for Healthcare Research and Quality R01HS022106. The findings and conclusions expressed here are those of the authors and do not necessarily represent the views of AHRQ.

References

[1] A. Lang, N. Edwards, and A. Fleiszer , Safety in home care: a broadened perspective of patient safety, *Int J Qual Health Care* **20** (2008), 130-135.

[2] A. Anthony, and P. Milone-Nuzzo, Factors attracting and keeping nurses in home health care, *Home Healthc Nurse* **23** (2005), 372-377.

[3] C.H. Ellenbecker , S.C. Frazier, and S. Verney, Nurses' observations and experiences of problems and adverse effects of medication management in home health care, *Geriatr Nurs* **25** (2004), 164-170.

[4] A. Harrison, and M. Verhoef, Understanding coordination of care from the consumer's perspective in a regional health system, *Health Services Research*, **37** (2002), 1031-1054.

[5] S. Sperling, K. Neal, and K. Hales. A quality improvement project to reduce falls and improve medication management, *Home Health Care Serv Q* **24** (2005), 13-28.

[6] S. Meredith, P. Feldman, and D. Frey, Improving medication use in newly admitted home health care patients: a randomized controlled trial, *J Am Geriatr Soc* **50** (2002), 1584-1491.

[7] R.H. Fortinsky, E.A. Madigan, and T.J. Sheehan, Risk factors for hospitalization among Medicare home care patients, *West J Nurs Res* **28** (2006), 902-917

[8] L.D. Scott, K. Setter-Kline, and A.S. Britton, The effects of nursing interventions to enhance mental health and quality of life among individuals with heart failure, *Applied Nurs Res* **17** (2004), 248-256.

[9] J. Fellows, and L. Crestodina, Home-prepared saline: a safe, cost-effective alternative for wound cleansing in home care, *J Wound Ostomy Continence Nurs* **33** (2006), 606-609.

[10] M.J. Strauss, J. Gong, and B.D. Gary, The cost of home air-fluidized therapy for pressure sores. A randomized controlled trial, *J Fam Pract* **33** (1991), 52-59.

[11] Lang A, Macdonald MT, Storch J, et al. Researching triads in home care: Perceptions of safety from home care clients, their caregivers, and providers. *Home Health Care Management & Practice.* 2014;26(2):59–71.

[12] D. Carpenter, T. Famolaro , S. Hassell, B. Kaeberle, S. Reefer, C. Robins, and S. Siegel, *Patient Safety in the Home: Assessment of Issues, Challenges, and Opportunities*. Cambridge, Massachusetts: Institute for Healthcare Improvement, 2017.

[13] A. Civan, M.M. Skeels, A. Stolyar, and W. Pratt, Personal health information management: consumers' perspectives. *AMIA Annu Symp Proc* (2006),156-60.

[14] A.M. Turner, K. Osterhage, A. Hartzler, J. Joe, L. Lin, N. Kanagat, G. Demiris. Use of Patient Portals for Personal Health Information Management: the Older Adult Perspective. *AMIA Annu Symp Proc* (2015), 1234-1241.

[15] E.P. Peron, S.L. Gray, J.T. Hanlon. Medication use and functional status decline in older adults: a narrative review. *Am J Geriatr Pharmacother* **9** (2011), 378-391.

Address for correspondence

Corresponding author:
George Demiris
University of Pennsylvania
Claire Fagin Hall, Room 324
418 Curie Blvd, Philadelphia 19104 PA, USA;
gdemiris@upenn.edu

Table 1– Dimensions of Patient Safety in the Home in the Context of Personal Health Information

Safety Dimension		**Definition** *Selected Participant Quotes* *OA-Older Adult Participant* *P-Provider* *FC-Family Member)*	**Informatics tools as facilitators**
Physical	Medication Management and Adherence	Challenges include risks for wrong dosage, missed dosage, wrong medication, side effects, administration of medication *So knowing what their* [patients'] *medicines are, is actually a huge barrier. I literally have to track down their mental health provider. Page, call, recall, receive faxes, take phone calls, just so that I can get a med list.* (P)	IT tools to provide: • Reminders • Automatic dispensing • Tracking of medication administration • Communication of adherence challenges • Notifications • Reporting of Side Effects
	Infection Control	Challenges to infection surveillance, prevention and control in the home setting can jeopardize patient safety and well-being. Tools are needed to measure and study the risks for home-care acquired infections (including valid definitions, measures and tools for surveillance). *Another thing I'm starting to do is infection control. I need to start tracking who has infections that mostly is communicable, like flu or pneumonia, or something like that.* (P)	IT tools that track: • Physiological variables associated with infection (e.g. temperature) • Photographic or video-assessment of wounds or other infection sites • Access to information • Symptom tracking and reporting
	Fall Prevention	Falls are a major risk for home care patients, especially frail elderly patients and those with rehabilitation needs after a major health event. *a fear of falling because I know that – I've – it's been several years since I fell, but I've fallen twice and the older I get, the less inclined I – I mean the more inclined I am to not want to fall anymore because – yeah. I mean your bones get more brittle and I don't want to break anything, let alone a hip. So yeah.* (OA)	IT tools to facilitate • Fall Prevention (e.g., motion sensors to detect mobility patterns, gait monitors to assess gait characteristics over time) • Fall Detection (e.g. sensors that detect weight and pressure, acoustic sensors that register fall, wearables that track falls)
	Nutrition	Nutritional needs may have a great level of complexity. Considerations include nutritional guidelines informed by the diagnosis or symptom management needs of the patient as well as potential challenges in accessing and preparing meals that adhere to specific dietary restrictions (including potential practical barriers to grocery shopping, transporting and storing food items, meal preparation and feeding). *If we plan to do something, we need to know when he needs to eat. Or we're always arranging any activities around the schedule of his insulin and his dietary needs also. We're aware of that, so if we're cooking or – we're all watching him because we know the signs now of when something's not right.* (FC)	IT tools that facilitate • Tracking of dietary choices • Access to information about nutritional values • Reminders

Table 1 (cont.)

	Care Coordination and Transition	Patients with complex care needs experience transitions from various care settings (e.g., discharge from hospital to a rehabilitation site to home health care) and challenges include barriers to information flow across settings. Missing or incomplete information and lack of coordination may introduce significant patient safety risks. *But it has happened several times that I'd come home and find him unconscious on the porch, … he's had an incident with his insulin. And so, I've asked them to please give me the information or tell me what they want me to do or what we're supposed to do.* (FC)	IT tools that provide a platform for access, sharing and transferring of data sets across stakeholder groups and systems of care
	Environmental Hazards	Environmental hazards include elements of the home infrastructure and physical layout, challenges of clutter in the home and unsanitary conditions. The actual residential infrastructure may introduce risks to patient safety. Infrastructure elements may include attributes such as temperature, light and level of accessibility that may hinder or pose a risk as residents carry out activities of daily living. *I wish I had a bath tub instead of a shower so it's a sit-down shower and that's just hard for me.* (OA) *I mean, her apartment is such a mess. So, that's another thing that can interfere with people's ability to keep track. They absolutely don't know where anything is.* (P)	IT tools that facilitate environmental scanning and assessment to identify accessibility issues as well as ongoing monitoring of hazards or triggers for adverse events.
Emotional	Emotional Challenges	Emotional challenges such as lack of confidence and resilience, issues of self-view and perceived stigma of frailty or vulnerability may affect the ability of a patient to engage in effective disease management to maximize patient safety. *But I had this terrible night where I really wanted to die –my daughter is still kind of out there somewhere. She texts me, but I don't see her anymore and I really wanted to die.* (OA)	IT tools to promote assessment of mental health issues and psycho-social aspects of mental health and well-being
Social	Social Support	Social support includes family and friends and other community members who interact with a home care patient and can provide practical or emotional support. Lack of social support may lead to isolation and loneliness, both linked to disease trajectory and clinical outcomes and as such may affect overall patient safety. *It occurs to me that that's probably something if you can help the patient to understand how necessary it is that the people around them have a list of what to watch out for and what to do about it.* (FC)	IT tools that facilitate: • virtual connections and social interactions with friends and family • sharing of information, messaging, communication • assessment of ongoing social interactions, visits, calls and detection of trends of social isolation • tools to promote social inclusion in virtual and physical communities
Functional	Functional Limitations	Functional limitations (including visual, hearing, mobility and cognitive limitations) may affect patient safety as they can interfere with the ability to independently carry out activities of daily living and also engage in disease management and symptom control activities. *The fire alarm going off and me not being able to go up and down the stairs that well because you can't use the elevator.* (OA) *And I have come in here and the stove has been left on. That's very scary to me.* (FC)	Informatics solutions that provide audio-assistance such as voice-interfaces to process visual information, that facilitate hearing aids or provide alternative (such as visual) displays of information, and cognitive orthotics (reminders, voice-interfaces to provide contextual information for activities)

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190409

Towards Emotion-Sensitive Conversational User Interfaces in Healthcare Applications

Kerstin Denecke[a], Richard May[a], Yihan Deng[a]

[a] Bern University of Applied Sciences, Institute for Medical Informatics, Bern, Switzerland

Abstract

Perception of emotions and adequate responses are key factors of a successful conversational agent. However, determining emotions in a healthcare setting depends on multiple factors such as context and medical condition. Given the increase of interest in conversational agents integrated in mobile health applications, our objective in this work is to introduce a concept for analyzing emotions and sentiments expressed by a person in a mobile health application with a conversational user interface. The approach bases upon bot technology (Synthetic intelligence markup language) and deep learning for emotion analysis. More specifically, expressions referring to sentiments or emotions are classified along seven categories and three stages of strengths using treebank annotation and recursive neural networks. The classification result is used by the chatbot for selecting an appropriate response. In this way, the concerns of a user can be better addressed. We describe three use cases where the approach could be integrated to make the chatbot emotion-sensitive.

Keywords:

Conversational user interface, sentiment analysis, deep learning, natural language processing

Introduction

A fundamental shift in healthcare takes place, driven by an aging population and the increasing incidence of chronic conditions that are induced by behavior. Changing individual behavior is increasingly at the center of healthcare. The reactive system where a patient's acute illnesses are treated is evolving to a treatment more centered on patients, prevention, and the ongoing management of chronic conditions. Beyond, mobile health applications are increasingly used by patients offering the opportunity to collect health data, to continuously monitor the personal health and to be accompanied by a personal health coach all over the day.

To realize such applications, conversational user interfaces (CUI) gained in interest for mobile health applications in the last years [1]. Different terms have been used for a conversational user interface or agent such as: chatbot, machine conversation system, virtual agent, dialogue system, and chatterbot. A CUI-based system is a computer program that interacts with users using natural language (written or spoken). The purpose of such system is to simulate a human conversation. There are two types of chatbots: Unintelligent chatbots interact using a predefined conversation flow. Intelligent chatbots use machine learning to automatically generate responses on the fly. The chatbot architecture integrates a language model and computational algorithms to generate an informal chat communication between human and computer using natural language. Interacting with intelligent agents is not a new topic, but reliable linguistic functionality, availability as services and inclusion of intelligence through machine learning and deep learning has increased its popularity.

CUI have been used in health related applications for example to achieve a health behavior change [2] or to support disease self-management. Lokman and Zain [3] introduced a chatbot that serves as a virtual dietitian for diabetic patients. The chatbot asks questions and gives at the end a diet advice suitable for the current diabetic situation. The conversation is going along a path that is remembered by the system to consider all answers in the decision making.

Similar to this virtual dietitian, the majority of existing medical chatbots only allows for constrained user input (e.g. multiple choice of several options instead of natural language input) to avoid misunderstandings that can occur within natural language interpretation. However, a more realistic, natural interaction requires a natural language user interface without predefined answers, but with integrated information extraction and natural language processing capabilities to recognize and interpret the content correctly. In this work, we are focusing on such chatbots with unconstrained natural language input capabilities. In these applications, users communicate with the system in their own words. In this way, it is possible to express personal emotions and sentiments in certain situations, for example when a person is not feeling well. However, this requires CUI-based applications that can analyze and interpret emotions from user input which are so far rarely available.

The objective of this work is to introduce a concept for analyzing emotions and sentiments resulting from interactions with a CUI. More specifically, our ultimate goal is to equip a health chatbot with capabilities to determine emotions and sentiments expressed in a chat with a CUI-based application in order to better focus suggestions and to increase the impact of recommendations of the application on the behavior of a user (e.g. increase compliance, react on negative emotions).

The main questions addressed in this paper are:

- How emotion and sentiment analysis could be integrated in mobile health applications with CUI?
- Which use cases exist for emotion-sensitive CUI in healthcare?

Conversational agents and sentiment analysis

There are several mobile health applications using CUI available. Some of them have been studied with respect to efficacy in clinical trials; most of them not [1]. Amoto et al. [4] introduced a chatbot-based recommender systems HOLMeS (Health On-Line Medical Suggestions). The system is designed to autonomously interact with a user by understanding natural language in a chat and acting as a human physician. It provides general information on itself or the affiliated medical center, collects patient information and enables a patient to book an

appointment in the affiliated medical center. It is implemented using the IBM Watson Conversation Service and trained via the Bluemix platform. VPBot, a SQL-based chatbot that simulates a patient that medical students can interview. VPBot was successfully exploited in Harvard Medical School's virtual patient program [5]. In such use cases for education or information provision, emotions and sentiment are not extremely relevant or do not even occur. Emotion and sentiment in this context comprises subjective, emotional statements e.g. description of the health status ("I am feeling well"), outcome of a treatment or experiences with it ("The therapy session was helpful"), emotions ("I feel enirely loss").

However, there are use cases where it becomes essential to identify and analyse expressions bearing emotions. Woebot [6] is a chatbot designed for supporting cognitive behavioral therapy. The chatbot allows to enter emotions by selecting terms from a list of suggestions. This limits the user to comprehensively express his or her actual emotions and feelings. In that case, natural language processing and sentiment analysis could be very useful to assess and address the user's mood and situation carefully. When it comes to mobile applications that aim at encouraging behavior changes, behavior models and motivation strategies have to be considered.

Aberg [7] analyzes different motivation strategies realized as chatbots as a mean to motivate people to live more sustainable lives. The effect of motivational factors from behavioral psychology were tested, and as well as the impact on people's food consumption habits. The findings of this paper were based on three chatbot prototypes; one that is built on the motivational factor of information; a second one that is implemented on the motivational factor of goal-setting, and a third one that follows the motivational factor of comparison. The result from the user interviews indicates that chatbots can affect and motivate people to consume food in a more sustainable way.

Analyzing emotions and sentiments resulting from interactions with CUI has so far only rarely been addressed. There are multiple ways to enable a chatbot to choose an emotion category for a response. On the one hand, the chatbot can be equipped with a personality and background knowledge. On the other hand, training data can be used to find the most frequent response emotion category for an emotion in a given response and use this as the response emotion.

Previous research by Skowron proposed affect listeners, i.e. conversational systems that can respond to user's utterances on a content-, but also on an affect-level [8]. Zhou et al. [9] describe an emotional chatting machine that can generate appropriate responses fitting in content and emotion to a user's response. The architecture consists of a recurrent neural network enabled with GRU cells with attention mechanism. It contains three different mechanisms for generating responses with a specific emotion: External knowledge serves to model emotions explicitly using an external emotion vocabulary. Internal memory captures emotion dynamics and finally, different emotion categories are represented as embedded vector. Socher et al. introduced a sentiment treebank that includes fine grained sentiment labels for parsing trees of sentences. On this treebank, they applied recursive deep models to predict sentence level sentiment. With a relatively complicated treebank annotation, the proposed method has better recognized the negated sentiment and achieved more than 80% overall accuracy [10].

Sentiment and emotion analysis in a medical context has been mainly addressed for web content. Denecke and Deng reviewed the state of the art and studied the challenges of sentiment analysis in medical settings [11]. They found out that given the varying usage and meanings of terms, sentiment analysis from medical documents requires a domain-specific sentiment source and complementary context-dependent features to be able to correctly interpret the implicit sentiment. The challenges of sentiment and emotion analysis in health chatbots have not yet been considered so far. Further, health applications equipped with emotion and sentiment analysis are still missing. In contrast to existing work, our aim is not to create a chatbot that formulates its responses with certain emotion terms, but to develop a method to analyse a user statement to select an appropriate, motivating or encouraging response given a specific user emotion.

Material and methods

Synthetic intelligence markup language (SIML)

SIML, pronounced "si mal", is used to build the chatbot's brain in our application (https://simlbot.com). It is a derivative of Extensible Markup Language (XML) and is able to react to user input, collect and manage data, learn from it and generate new content for conversations. The official parser was developed for C# and can be used on all Windows platforms supporting .NET Framework 4.5 or higher as well as under Linux and Mac (Mono). SIML contains two specifications: SIML Classic and SIML Modern/OSCOVA. All specifications are interpreted differently and are aimed at different target groups. SIML Classic is used in our application because we resist on using internal natural language processing as SIML Modern offers.

In SIML, all data is arranged in the form of a large decision tree and can be addressed by pattern matching. SIML files consist of a large number of tags. The basic structure is created by concepts and models, which are enclosed by the SIML tag. Concepts are the basic unit of knowledge storage in SIML. Each concept is a rule for matching an input and converting to an output, and consists of a pattern, which matches against the user input, and a response, which is used in generating the chatbot answer. In SIML, it is possible to use regular expressions and loops within a model, to interpret JavaScript and to create own tags to call specific methods of the program code. Another function of SIML is the usage of mappings which we will use to select an appropriate response for a user statement containing emotions and sentiments.

Sentiment and emotion analysis

Traditional sentiment analysis approaches classify texts or sentences according to their overall sentiment which can be positive, negative, neutral or even have more fine-grained sentiment categories. They often base upon a bag of word representation. Additionally, the polarity shifting and syntactic structures are transformed into rules to regularize the composition of sentiment at sentence level. However, the performance of bag of words based methods has not reached an accuracy of 80%.

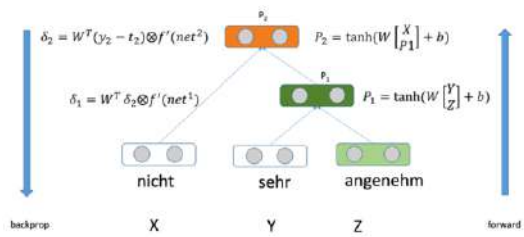


Figure 1. Compositional function of recursive neural networks and backpropagation of errors. Light green color indicates the positive polarity of joy, dark green represents intensified polarity and joy while orange shows negated polarity of joy.

A tree-based composition of a sentiment treebank has been proposed by Socher et al. [10] in order to improve the accuracy. The recursive neural network that is used in their approach requires a well-formed tree annotation corpus as training data. A recursive neural network is a type of deep neural network created by applying the same set of weights recursively over a structured input, in this case a linguistic parse tree of a sentence. The annotation of a sentiment treebank is quite laborious and sophisticated, since the polarity of the sentence (polarity at root) must be determined by a composition of all subtrees. The annotator needs to judge the polarities at different levels and provides also an overall value to the tree root. Our approach for an emotion analyzer is inspired by the work of Socher et al. [10], see Figure 1 and 3.

The classification of user statements in the context of a chatbot is an important step for automatic, appropriate answering of user statements by a chatbot. The concrete task is to categorize a user statement into predefined emotional categories. We currently employ seven axes of emotions (disgust, joy, surprise, anger, fear, sadness, contempt), see Figure 2. The polarity is transformed to a respective emotional tree. For each emotion class, a treebank classifier is trained. The emotion of one sentence is represented as the following polarity vector:

$P_i = (x_{anger}, x_{disgust}, x_{joy}, x_{surprise}, x_{fear}, x_{contempt}, x_{sadness})$

where P_i represents the i-th user statement of the chatbot conversation. The value x represents the normalized strength of the emotion where we distinguish 3 classes (low, medium, high). To obtain a vector of polarity for a user statement, seven instances of the classifier based on recursive neural networks pass through the seven polarity treebanks for initial pre-training.

Before the classification starts, the user statement is linguistically parsed into a syntactic tree. Each node of the tree is represented by word vectors learned from a corpus of German Wikipedia and medical forum entries. During training, one feedforward step is conducted to learn the structure (bottom-up) whereas a back-propagation step is performed to adjust the errors (top-down).

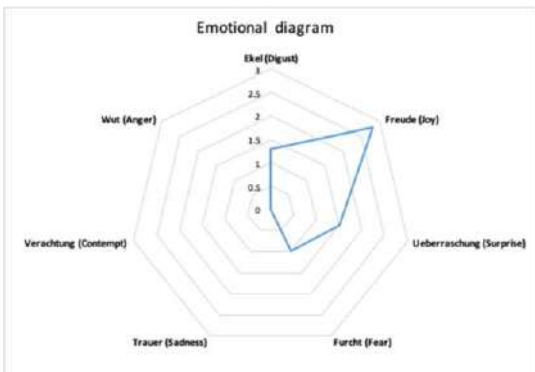


Figure 2: Emotional diagram with seven axes: disgust, joy, surprise, fear, sadness, contempt, anger

In Figure 2, we can see one example of recursive neural networks that we will apply for the emotion detection. The leaf nodes represent the emotion-related terms in the parsed sentence. In this example, the polarity of "angenehm" (pleasant) has been intensified by "sehr" (very) and finally negated by "nicht" (not). Apparently, the syntactic tree has better represented the semantic scope with negation. For the negation parsing, we have chosen FastContext[1] [12] with German negation corpus.

The objective of the optimization is maximizing the probabilities of the correct prediction while minimizing the cross-entropy error between the predicted distribution y_i and target distribution t_i. For the classification into one of the three polarity strength classes, the posterior probability over labels given the word vector via $Y^x = \text{softmax}(W \cdot X)$

where $W \in \mathbb{R}^{3Xd}$ is the emotion classification matrix.

Results

Concept for an emotion-sensitive chatbot

Our concept for integrating emotion analysis into a SIML-based chatbot is shown in Figure 4. A user input resulting from an interaction with the chatbot application is analyzed by the above mentioned emotion analysis algorithm. The resulting polarity vector with strength per emotion class is used to select the corresponding answers based on predefined selecting metrics. SIML mappings assign an emotion label after receiving the emotion vector. Depending on the emotion, we use loops within the models to find matching answers to the user input in the chatbot's knowledge base. These answers can be randomly selected from a pool of matching answers to make the conversation more natural. The answer is displayed on the user interface of the application. In this way, the conversation can be continued coherently considering the emotional fluctuation of users. At the current step, we focus on recognizing the primary emotional polarity of a user input. For treebank construction, we will select 100 sentences for each of the seven classes of emotions (see Figure 3) and parse them into a tree structure. These tree structures will afterwards be annotated with polarity values ranging from 1 to 3 by human annotators.

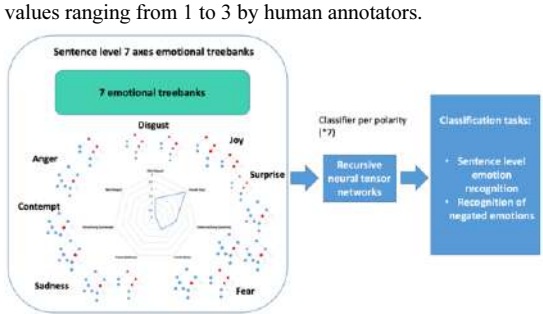


Figure 3. Training of the emotion analyzer based on annotated treebanks

Use case scenarios

Traditional models of care delivery basically base upon face-to-face interactions between clinicians and patients. However, new technologies are augmenting this interaction model and fundamentally transforming the ways in which clinicians deliver care to individuals. Mobile apps, for example, can facilitate tracking and monitoring. These remote and self-care-oriented technologies may help creating a truly interactive healthcare ecosystem for patients. The following use cases base on mainly two paradigms: Engaging individuals by interacting with a CUI and utilizing remote and self-care-oriented technologies to support and empower individuals. Emotion and sentiment analysis will help to realize these two paradigms.

[1] https://github.com/jianlins/FastContext

Addressing concerns during medication management

eMMA, the electronic medication management assistant [14], is a mobile application to support in the medication management. In its current version, it is designed to collect compliance information from a patient, to provide information on interactions between food and medications and to answer his or her questions on the medication [14]. Equipping eMMA with capabilities for automatically recognizing and classifying concerns of a user mentioned in the application could help to address appropriately the concerns of the user. In this way, the compliance could be improved. The chatbot will not convince the user, but recognizes concerns which are stored for discussion with the physician and provides information for engaging the user to take the medication as recommended by the physician. In case the system detects serious concerns and symptoms, it could suggest to contact the physician.

Addressing problems and concerns during self-anamnesis

Ana is a system for collecting the medical history within the context of music therapy using a CUI [15]. Equipping Ana with emotion and sentiment analysis capabilities would enable the system to identify situations when the user runs into problems and concerns because of the questions he or she has to answer. In a usability study with that system [15], the users confirmed, that they liked the possibility of entering free text to communicate with the chatbot. However, they asked for better interactions with the chatbot, in particular when she states that she is not feeling well. Such expressions could be detected and analyzed by our method for providing an appropriate reaction.

Determine cognitions and suggest behavior changes

Another use case is to recognize the tone of a client's response to suggestions made by a chatbot. The suggestions might be for example therapeutic suggestions within a mobile cognitive behavior therapy (CBT) or motivational suggestions within a mobile application targeting at achieving a personal health goal. Within the HABIT project, a mobile application for CBT has been developed that uses a chatbot interface for delivery of CBT [16]. CBT shares the idea that behavior change may be affected via cognitive change [17]. A chatbot system for CBT requires facilities to assess a client's mood overall, or the tone of the response to a particular suggestion in the therapeutic chat. For example a negative answer, high emotional, neglecting a suggestion of the therapist might be interpreted as low acceptance of the immediately prior recommendation. Our emotion analyzer could support in automatically analyze the responses of a user, supporting the therapist in analyzing the chatlog afterwards or enabling chatbot responses that address these moods and tones of user statements.

Discussion

In the general domain, the objective of enriching a chabot with emotions is to achieve a more human-like interaction style. In contrast, in our scenarios the objective is to determine the emotion of a patient for future interpretation by the treating physician or by addressing concerns directly by the chatbot.

Sentiment analysis in the medical domain differs from sentiment analysis in other domains [11]. We address this fact, by training the treebank classifier on a data set of health-related content. Additionally, health-related lexicons could be integrated in the parsing process. Asgar et al. [18] suggest a bootstrapping model and a dataset of health reviews to learn a health-related sentiment lexicon. Rane et al. introduce their concept for using sentiment analysis to improve emotional health of a user [19]. More specifically, they analyze the sentiment or emotion of a user and display specific media to counter the emotion. They used a Naïve Bayes classifier based on supervised learning that was trained on a twitter data set. Besides the different technology, our approach aims at considering the peculiarities of sentiment in the health care domain, which is lost when training is realized with a non-domain-specific data set.

The use of ontologies to support interpretation of several types of data (audio, video, and text) from a chatbot's environment has been analyzed in previous work. Formal concept-based rules are suggested to express affective behavior aiming at improving the empathy of bots [12]. The proposed technique relies on semantic technologies such as OWL and SWRL languages. Affective states are exploited to improve the bot's empathy in the interaction based on an emotion ontology. In a chatbot conversation, user statements might be short. This makes it even more challenging to find appropriate methods. It is still unknown to what extent the previous statements in a conversation and the user context can contribute to the interpretation of expressed emotions.

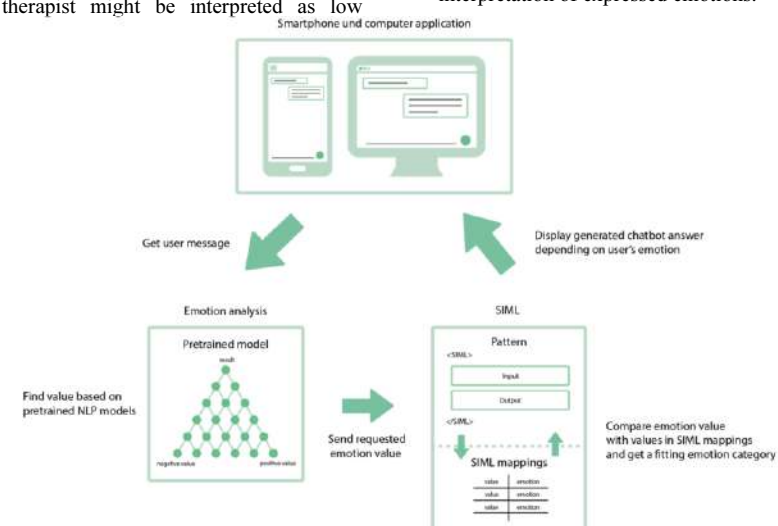


Figure 4: System architecture

Conclusions

In this paper, we introduced a concept for an emotion analyzer for natural language statements as they result from a chatbot conversation. This approach is currently implemented. In future work, we will evaluate the above mentioned method. For this purpose, we will compare the recursive neural networks with the traditional kernel based method in terms of sentence level polarity recognition and the detection of negation. We will run the sentiment analysis using SVM or convolutional neural network on the sentence level annotation to serve as benchmark [20]. Further, we will extend one of the introduced mobile applications with the emotion analyzer.

References

[1] L. Laranjo, A.G. Dunn, H. Ly Tong et al. Conversational agents in healthcare: a systematic review. Journal of the American Medical Informatics Association, 25(9), 2018, 1248–1258

[2] A. Fadhil and S. Gabrielli. Addressing challenges in healthy lifestyles: The Al-chatbot approach. In Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare, PervasiveHealth '17, New York, USA, 2017, pages 261-265

[3] A.S. Lokman and J.M. Zain. An architectural design of virtual dietitian (vidi) for diabetic patients. Proceedings of the International Conference on Computer Science and Information Technology, 2009, pages 408-411

[4] F. Amato, S. Marrone, V. Moscato, G. Piantadosi, A. Picariello, C. Sansone: Chatbots Meet eHealth: Automatizing Healthcare. In: Impedeovo D, Pirlo G (eds.) Proc. of the Workshop on Artificial Intelligence with Applications in Health, CEUR Workshop Proceeedings, volume 1982, 2017, pages 40-49, http://ceur-ws.org/Vol-1982/

[5] B.P. Kerfoot, H. Baker, T.L. Jackson, W.C. Hulbert, D.D. Federman, R.D. Oates, and W.C. DeWolf. A multiinstitutional randomized controlled trial of adjuvant web-based teaching to medical students. Academic Medicine: Journal of the Association of American Medical Colleges, 81:224-30, 2006

[6] K.K. Fitzpatrick, A. Darcy, M. Vierhile. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health. 2017 Jun 6;4(2):e19. doi: 10.2196/mental.7785

[7] J. Aberg: Chatbots as a mean to motivate behaviour change. Master's thesis in Interaction Technology and Design. Umea University, Sweden, 2017

[8] M. Skowron: Affect listeners: acquisition of affective states by means of conversational systems. COST'09 Proceedings of the Second international conference on Development of Multimodal Interfaces: active Listening and Synchrony, 2009, pages 169-181

[9] H. Zhou, H. Huang, T. Zhang, X. Zhu and B. Liu. Emotional Chatting Machine: Emotional Conversation Generation with Internal and External Memory. AAAI Conference on Artificial Intelligence, 2018

[10] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. Manning, A. Ng and C. Potts. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. Conference on Empirical Methods in Natural Language Processing (EMNLP 2013), 2013

[11] K. Denecke and Y. Deng. Sentiment analysis in medical settings: New opportunities and challenges. Artificial Intelligence in Medicine, 2015, 64(1):17-27 10.1016/j.artmed.2015.03.006

[12] A. Justo, F.R. Jensen, R. Bonacin, I. Calado, J.C. Dos Reis. Exploring Ontologies to Improve the Empathy of Interactive Bots. In Proceedings of the 27th IEEE International Conference on Enabling Technologies (WETICE'18). Paris, France, 2018, pp. 261-266

[13] W.W. Chapman, D. Hillert, S. Velupillai S, et al. Extending the NegEx lexicon for multiple languages. Stud Health Technol Inform. 2013;192:677-81.

[14] M. Tschanz, T.L. Dorner, J. Holm, K. Denecke: Using eMMA to manage medication. IEEE Computer, vol. 51, issue 08, pages 18-25, 2018

[15] K. Denecke, S. Lutz Hochreutener, A. Pöpel, and R. May: Self anamnesis with a Conversational User Interface: Concept and Usability Study. Methods of Information in Medicine, 2018 (to appear)

[16] J. Warren, E. Tempero, I. Warren, A. Sathianathan et al.: Experience Building IT Infrastructure for Research with Online Youth Mental Health Tools, Proceedings of the ASWEC conference 2018

[17] K.S. Dobson, D.J.A. Dozois. Historical and Philosophical Bases of the Cognitive-Behavioural Therapies. In Dobson KS (Ed.) Handbook of Cognitive-Behavioural Therapies, The Guilford Press, 2001

[18] M.Z. Asghar, S. Ahmad, M. Qasim, S.R. Zahra, F.M. Kundi. SentiHealth: creating health-related sentiment lexicon using hybrid approach. Springerplus. 2016 Jul 20;5(1):1139. doi: 10.1186/s40064-016-2809-x. eCollection 2016

[19] P. Rane, K. Bhansali and S. Nair: Sentiment Analysis to Improve Emotional Health of User. International Journal of Computer Applications 120(1):21-24, June 2015

[20] J. Shi, J.F. Hurdle. Trie-based rule processing for clinical NLP: A use-case study of n-trie, making the ConText algorithm more efficient and scalable. J Biomed Inform. 2018 Sep;85:106-113. doi: 10.1016/j.jbi.2018.08.002

Address for correspondence

Kerstin Denecke

Bern University of Applied Sciences, Institute for Medical Informatics, Quellgasse 21, 2501 Biel, Switzerland,

E-Mail: kerstin.denecke@bfh.ch

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190410

Blended Learning for French Health Students: Does Acceptance of a Learning Management System Influence Students' Self-Efficacy?

Lionel Di Marco[ab], Hassina El Kechaï[c], Donald K Martin[d], Pierre Gillois[b]

[a] *Midwifery Department, Faculty of Medicine, Grenoble-Alpes University, France,*
[b] *CNRS / TIMC-IMAG Lab., UMR 5525 (Themas), Grenoble-Alpes University, France*
[c] *TECHNE Lab., EA 6316, Poitiers University, France*
[d] *CNRS / TIMC-IMAG Lab., UMR 5525 (SyNaBi), Grenoble-Alpes University, France*

Abstract

A learning management system (LMS) used for the initial training of health professionals has been rejected by students. Our study aimed to explain the reason of this rejection. We performed this evaluation on a sample of health students in 2012 and 2017 (n=144). We used scales from the literature (Technology Acceptance Model, General Self-Efficacy Scale, LMS-Self-Efficacy Scale), and studied the social representation of the LMS. The system seemed accessible and useful, but unfortunately with similarities to the system used in a traumatic student environment. Health students using the system did not have a lower self-efficacy. Although the LMS seemed relevant to students, its initial rejection might have been due to a confounding context that created confusion in the acceptability of the tool. To conclude, there is a need to create new dematerialized course formats but with strong tutorship to improve the usage of the technologies by students.

Keywords:

computer-assisted instruction, educational models, medical education.

Introduction

Context

In France, the first year of health studies contest (PACES in French) is stressful and typically is aimed to be a highly selective filter, which is not the optimal means to encourage learning [1]. This first year ends with a ranking examination that is used to select around 10% of students to continue their studies in the health disciplines of either medicine, midwifery, pharmacy, or dentistry. In 2006, the members of the advisory board of Grenoble Medical School decided to change the pedagogical paradigm to assure more fairness and improved learning for health students during PACES.

Since 2006, the paradigm is a blended-learning system that includes flipped classrooms [2]. This is a stepwise program on a 4-week cycle (Fig. 1). For example, on week 4 of the university calendar, students learn the knowledge capsule of the 4th course on their Learning Management System (LMS), they ask online questions to the lecturer of the 3rd course of the year, they participate in face-to-face explanations with the lecturer of the 2nd course of the year, and they manipulate the concepts of the first course of the year that have been learned during the first three weeks.

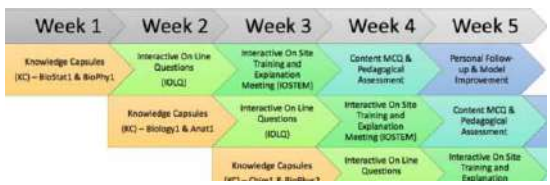


Figure 1– Stepwise learning method in sequences of 5 activities

Between 2011 and 2017, all the curricula of health studies in Grenoble were reformed in the same way (medicine, pharmacy, midwifery, and paramedical studies). However, the initial reactions of staff differed to those of the students. For example, students found this LMS was not appropriate; particularly the students of year 2 who stated that using this LMS created social isolation [3]. Students also declared that they felt themselves inefficient while learning Knowledge Capsules. But the teaching staff found the Knowledge Capsules relevant during face-to-face sessions. These two different feedbacks prompted us to consider the link between the acceptance of the LMS and students' self-efficacy. Our main objectives were to understand the initial rejection of the LMS and to improve this pedagogical method used by more than 5,000 health students.

Concept: Acceptance

The "Technology Acceptance Model" (TAM) described by Davis [4] proposed a model of the way that users accepted and used an information system in order to diagnose design issues. This model has been validated in distance learning contexts [5]. Davis defined the perceived usefulness as the degree to which a person believes that using a particular system will enhance job performance. He defined the perceived ease of use as the degree to which a person believes that using a particular system will be free from effort. These two concepts impact on the user's intention to use the system.

According to Venkatesh and Davis [6], the use of information and communication systems by professionals depends on a subjective standard, which is the perception by the user that the group to which he belongs considers that he should, or not, use the system. They suggested TAM2, a model derived from TAM, as adapted to mandatory use contexts. It is therefore particularly suitable for health students in Grenoble, since it is mandatory for them to use their LMS as part of their studies.

Concept: Self-Efficacy

General Self-Efficacy covers all the tasks that a health student is required to do during his/her training and is defined by Luszczynska [7] as "the belief in one's competence to tackle novel tasks and to cope with adversity in a broad range of stressful or challenging encounters as opposed to specific self-efficacy, which is limited to the particular task at hand". There is a benefit in correlating the general self-efficacy to the self-efficacy focused on technological tasks [8]. According to Martin et al. [9], Learning Management System Self-Efficacy is defined as self-assessment of one's skill in using the LMS. It is based on a model developed by Compeau and Higgins [10] for studying the relationship between specific self-efficacy and students' performance during distance or blended learning. Different dimensions were analyzed, including the skills in accessing courses, the skills in performing tests, the skills in synchronous and asynchronous communication and the skills in using advanced tools (blogs, collaborative work, etc.).

Hypotheses

We decided to analyze the link between the acceptability of the LMS and students' self-efficacy, its alteration between 2012 and 2017, and students' social representation of the LMS. Our assumption was that if students do not perceive the LMS as useful, usable, and acceptable, they will not perceive it as a helping and motivating tool, which could reduce their self-efficacy. We hypothesized that:

- General Self-Efficacy is poorer in students with poorer technology acceptance.
- If 2017 students have a better acceptance of the LMS than 2012 students, their General Self-Efficacy will be better.
- The social representation of the LMS is generally negative, although some aspects such as practicality may be positive.

Methods

Overview

During the PACES, students can choose to pursue studies in one to four of the health curricula proposed (medicine, pharmacy, midwifery, or dentistry). To access each chosen curriculum, the second semester of the PACES includes some specific courses and a specific ranking examination. Classically, students choose 2 or 3 curricula. When they fail to access the curriculum they prefer, their perception of the LMS could be linked to the exam failure and to variations of self-beliefs such as self-efficacy.

Participants

We decided to analyze the population with the longest experience using the LMS and a high rate of multiple choices of curricula. Students who met these criteria were year 2 midwifery students of the 2011-2012 class (those that are the closest to PACES), and all the 2016-2017 midwifery students (to study the alteration of the acceptance). Students of the years 3, 4, and 5 of the 2011-2012 class were excluded because of they did not use the LMS.

Materials

Many scales to evaluate the Technology Acceptance Models are described in the literature. Most of them were created for the evaluation of Information and Communication Technologies (ICT) tools used in business contexts. However, some [11][12] have been validated in learning contexts. These scales analyze Perceived Ease Of Use (PEOU), Perceived Usefulness (PU), and Perceived Acceptability (PA).

We also used the validated General Self-Efficacy Scale (GSES) [13] to assess general self-efficacy. We partially used the LMS-self-efficacy scale (LMSSES)[9], by retaining 12 items (of the original 24) and removing those items that corresponded to functions not offered by the studied LMS.

Finally, we used prototypical analysis to study social representations of the LMS. According to the extended TAM and the concepts that could influence acceptability of a technology, we decided to evaluate the way students think about the "job relevance" of the LMS (i.e. the way they think the LMS is useful in their student activities), and also the "results demonstrability" (i.e. the way the LMS shows them "learning results").

We used the prototypical analysis [14] to identify the semantic field linked to these concepts. From these semantic fields, students gave a value to each induced word according to the significance it had for them. They had to prioritize words in order to create a structure of the social representation with a core area, surrounding items and contrasted elements. Prototypical analysis was performed by calculating the citation frequencies and the rank assigned to each word or expression by each respondent. This analysis was represented graphically by rank and frequency.

Procedure

This study was a user-centered evaluation of the LMS, based on a self-administered questionnaire by students performed after a complete semester of blended learning. Thus, the questionnaire was presented in the week after the last exams of the first semester. This was a paper-based questionnaire for the 2012 population, and an online questionnaire for the 2017 population. Three reminders to complete the questionnaire were sent weekly and the study ended just after the beginning of the second semester.

Analysis

To validate our assumption, we had to study the population characteristics that could influence the analyzed concepts. We used the Spearman correlation coefficient, to demonstrate:

1. that perceived usefulness and perceived ease of use are correlated to the perceived acceptability, so as to validate the TAM2 model in this population;
2. that the TAM2 score is correlated to the general self-efficacy;
3. that the TAM2 score is correlated to the LMS self-efficacy; and
4. that the upper-TAM2-score students and the lower-TAM2-score students (subdivided by the mean TAM2-score of the population) present a different correlation between TAM2 and self-efficacy.

Results

The characteristics of our population are shown in Table 1. These two populations are neither comparable in terms of age nor in terms of the different studied scores, except for the GSES. Both populations have a comparable declared mastery of ICT tools.

Table 1 – Comparison between 2012 and 2017 midwife students (characteristics, self-efficacy, and acceptance scores)

	2012 Students (n=26)	**2017 Students (n=118)**	**p-value**
ICT control *median*	Good	Good	NS
Age *median (IQR)*	20(1)	22(1)	<0.001
GSES *mean (s.d.)*	2.97 (0.35)	2.97 (0.37)	NS
LMSSES *mean (s.d.)*	3.42 (0.36)	3.56 (0.43)	0.013
PEOU *mean score (s.d.)*	4.91 (1.06)	5.67 (1.06)	<0.001
PU *mean score (s.d.)*	4.19 (1.32)	5.18 (1.10)	<0.001
PA *mean score (s.d.)*	4.32 (1.17)	5.18 (1.16)	<0.001

Each tool had a satisfactory Cronbach's alpha (2012-GSES α=0.803; 2012-LMSSES α=0.744; 2012-TAM2 α=0.901; 2017-GSES α=0.764; 2017-LMSSES α=0.820; 2017-TAM2 α=0.906).

Concerning the theoretical model, correlations between acceptance and self-beliefs are shown in Figure 2. All components of the TAM2 model were correlated (0.4241<r<0.7293, p<0.01).

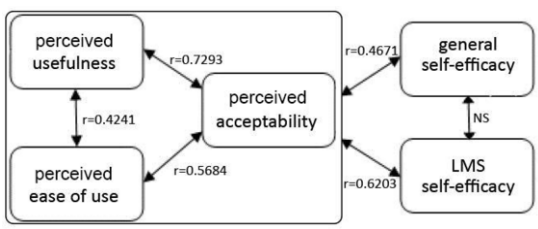


Figure 2– Theoretical model: correlation between acceptance and self-efficacy in 2012 students (p<0.01)

Upper-TAM2-score students presented better affirmed computer skills than lower-TAM2-score students (Good vs Deficient, p=0.005) and higher LMS-self-efficacy mean score (3.57 vs 3.2, p=0.003). These groups were similar in terms of mean age, sex, affirmed regular use of internet, and the mean general-self-efficacy score. Finally, there was no correlation between the TAM2-score of each group and their self-efficacy (r=-0.0405, p=0.8543).

With regard to the social representation of the LMS's job relevance (Fig. 3), the most significant core items were "organization" and "accessible". The most significant potential surrounding item with a negative value was "as during PACES".

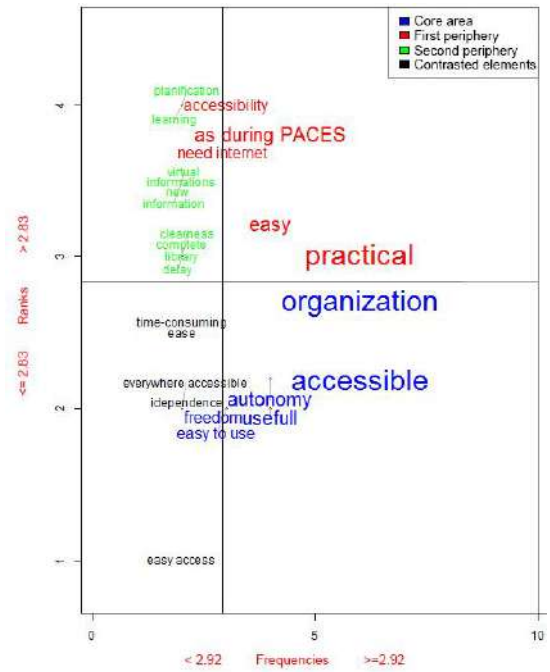


Figure 3– Prototypical analysis: social representation of LMS's job relevance

With regard to the social representation of the LMS's "results demonstrability" (Fig. 4), positive core items were "lengthy" and "organization". A potential surrounding item had a very negative value: "work alone".

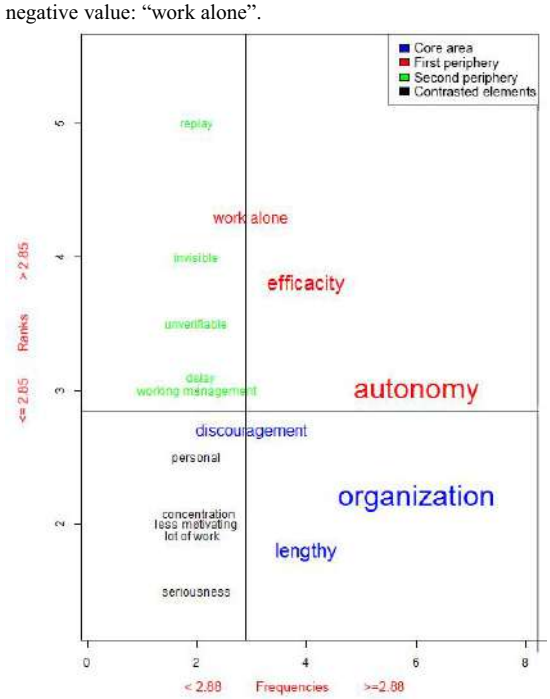


Figure 4– Prototypical analysis: social representation of LMS's "results demonstrability"

Discussion

Main results

Only one hypothese was deemed valid.

If the theoretical model is consistent, each sample does not have a link between acceptability and self-belief. It seems there is no link between students' acceptance of the LMS and their self-efficacy. Indeed, the general self-efficacy is not significantly different between students who identify the LMS as being less useful, less easy to use and less acceptable compared to students who identify the LMS as being more useful, more easy to use, and more acceptable. This is confirmed by the analysis of the 2017 population, who declared having a better technology acceptance, in all its dimensions, but had a similar general self-efficacy to the 2012 population.

The way to learn is negatively linked to the first year of health studies (PACES) because of the similarities between the LMS used during both years. It explains the fear of isolation, according to the surrounding item "work alone". But the social representation of the LMS proves that the system has many benefits, which are linked to the organization and the usefulness. Indeed, the LMS was created in that way to propose many features that facilitated the organization of study, including a learning-activities agenda, an online survey to ask questions to the lecturer while working on dematerialized lectures, and discussion forums. Moreover, these features were reinforced through the discussions with teaching staff and peer-experienced students, who explained each feature of the system and how students can organize their work.

Limitations of this study

This study has the following limitations. Only Year 2 students used the LMS, hence it was not possible to have a larger group in 2012 and therefore the group sizes are different. During the study period, user groups and attitudes probably changed, but it was necessary to analyze the GSES for the new generation. Scales used in our study are validated scales and reflect the measures in each group.

The choice of using TAM2 is also debatable. Some other technology acceptance models have been described (TAM, TAM3, UTAUT), but TAM2 has the advantages of being neither simplistic nor too complex and of having been validated within the framework of mandatory uses. Moreover, it presents precursors of perceived utility that were fundamentally interesting for our study ("job relevance" and "results demonstrability").

Internal validity

From 2012 to 2017, acceptance of the technology has increased significantly. This is probably due to enhancements of the LMS and changes in tutorship. Moreover, the 2017 population includes students older than that of the 2012 population. This could impact on declared perceived usefulness, ease of use, and acceptability as a confounding factor. Indeed, users' experience is linked to perceived usefulness[6]. This bias probably has a minor impact on our results because students' mean age also impacts on general self-efficacy; older students are skilled students, with some level of responsibilities at the hospital. Finally, cultural differences between student generations regarding acceptance of this particular technology could also be a confounding factor. The additional 5 years of experience with the LMS system probably increased students' perceived acceptability. However, the use of validated and contextual scales (TAM2, GSES, LMSSES) limits all of those biases.

External validity

In its entirety, the study linked the various concepts even if most of the correlations are moderate. The pedagogical method (not only the LMS as a tool) has already been evaluated [15] concerning its intention for use and the impact on learning strategies. The method is well accepted, but students have an intention to be too neutral in using the system, probably due to distractibility during lectures. So, we suggest improvements to make the LMS more acceptable for students, by (1) creating adapted types of interactive multimedia lectures (SCORM format activities for example) to be different from PACES and to have more rapidly perceived learning results [16][17]; (2) making changes in the presentation session of the LMS to improve the perceived ease of use of the tool.

Implication for health professional education

Our study showed that it is important to train students to use their learning environment. The intention to use a LMS and its acceptability are correlated to learners' satisfaction, which itself depends on quality of the transmitted information and ease of use [18][19]. Blended Learning is an interesting way of teaching healthcare [20]. It doesn't interfere with deep learning strategies[15]. These strategies are needed to help students make links between knowledge, skills, and habits. Our study showed that the LMS used in the method proposed by the Grenoble-Alpes University has a negative social perception for health students. To change this, the faculty of medicine must create an adapted context of use for this LMS, which would include new formats for courses (different than those used in PACES), and improved tutorship for using and understanding the method and the tool. Especially since students show an improvement in their computer skills after using an LMS [21].

Conclusions

Our study shows that the LMS is well considered by students concerning work organization, autonomy, and accessibility. Moreover, it seems that this rejected tool has no impact on students' self-efficacy. But since students are forced to use this LMS for their training, it is necessary to improve its conditions of use. This must involve creating learning objects different from those currently existing.

Acknowledgements

The authors are thankful to midwife students of Grenoble-Alpes University for their participation in this study. The authors are also thankful to Prof. Florence Martin, Prof, Ralf Schwarzer, Prof. Venkatesh, and Prof. Davis for their authorization to use their rating scales.

References

[1] M.P. Gustin, C. Vinciguerra, S. Isaac, C. Burillon, and J. Etienne. Approches d'apprentissage et réussite en première année commune des études de santé (PACES) en France. *Pedag Med* **17** (2016), 23-43.

[2] M.J. Lage, G.J. Platt, and M. Treglia. Inverting the Classroom: A Gateway to Creating an Inclusive Learning Environment. *J Econ Educ* **31** (2000), 30-43.

[3] L. Di Marco, N. Bosson, and P. Chaffanjon. Influence sur les croyances sur soi d'une réforme pédagogique utilisant les TICE. *Pedag Med* **10** (2009), S77.

[4] F. Davis. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quart* **13** (1989), 19-340.
[5] C.S. Ong, J.Y. Lai, and Y.S. Wang. Factors affecting engineers' acceptance of asynchronous e-learning systems in high-tech companies. *Inform Manage* **41** (2004), 795-804.
[6] V. Venkatesh, F. Davis. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Manage Sci* **46** (2000), 186-204.
[7] A. Luszczynska, B. Gutierrez-Dona, and R. Schwarzer. General self-efficacy in various domains of human functioning: Evidence from five countries. *Int J Psychol* **40** (2005), 80-89.
[8] E. Alqurasi. Self-Efficacy In Online Learning Environments: A Literature Review. *Contemp Issues Educ Res* **9** (2016), 45–52.
[9] F. Martin, J. Tutty, and Y. Su. Influence of Learning Management System Self-efficacy on e-Learning Performance. *JSCH* 5(2009), 26-35.
[10] D. Compeau, and C. Higgins. Computer self-efficacy: development of a measure ad initial test. *MIS Quart* **19** (1995), 189-211.
[11] R.G. Saadé, F. Nebebe, and W. Tan. Viability of the "Technology Acceptance Model" in Multimedia Learning Environments: A Comparative Study. *IJELLO* **3** (2007), 175-84.
[12] S.Y. Park. An Analysis of the Technology Acceptance Model in Understanding University Students' Behavioral Intention to Use e-Learning. *Educ Technol Soc* **12** (2009), 150-62.
[13] R. Schwarzer, and M. Jerusalem. *Generalized Self-Efficacy scale*. In: Measures in health psychology Causal and control beliefs, England, NFER-NELSON, 1995.
[14] J.C. Abric. *La recherche du noyau central et de la zone muette des représentations sociales*. In Méthodes d'études des représentations sociales, Ramonvile Saint Agne, Editions érès, 2003.
[15] L. Di Marco, A. Venot, and P. Gillois. Does the acceptance of hybrid e-learning affect learning approaches in France? *JEEHP* **14** (2017), 24.
[16] D.A. Cook, M.H. Gelula, D.M. Dupras, and A. Schwartz. Instructional methods and cognitive and learning styles in web-based learning: report of two randomised trials. *Med Educ* **41** (2007), 897-905.
[17] E. Jamet, and A. Arguel. La compréhension d'un document technique multimédia peut-elle être améliorée par une présentation séquentielle de son contenu ? *Trav Hum* **71** (2008), 253-70.
[18] S Eom. Understanding e-Learners' Satisfaction with Learning Management Systems. *Bull IEEE Tech Comm Learn Technol* 16(2014), 10-13.
[19] N.F. Abdel-Maksoud. The Relationship between Students' Satisfaction in the LMS "Acadox" and Their Perceptions of Its Usefulness, and Ease of Use. *J Educ Learn* **7** (2018),184-90.
[20] M.R. Sajid, A.F. Laheji, F. Abothenain, Y. Salam, D. AlJayar, and A. Obeidat. Can blended learning and the flipped classroom improve student learning and satisfaction in Saudi Arabia? *Int J Med Educ* **7** (2016), 281-5.
[21] H.R. Farhan. *LIS Students' Perceptions of the Use of LMS: An Evaluation Based on TAM (Kuwait Case Study)*. In: Alexandrov D., Boukhanovsky A., Chugunov A., Kabanov Y., Koltsova O. (eds) Digital Transformation and Global Society. DTGS 2018. *CCIS*, vol 859. Springer, Cham

Address for correspondence

Lionel DI MARCO, Midwifery Department, Faculty of Medicine, Grenoble-Alpes University, CS 90338, 38434 Echirolles cedex, France, +33476765257, **lionel.di-marco@univ-grenoble-alpes.fr**

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190411

Development of a Digital Tool to Assist the Training of Health Professionals in the Determination of Brain Death

Carlos Eduardo Rochamedica Correia[a], Clécio de Oliveira Godeiro Junior[a]

[a] Department of Integrated Medicine, Rio Grande do Norte Federal University, Natal, RN, Brazil

Abstract

Due to technological advancement of medicine, patients have been maintained through mechanical ventilation and vasoactive drugs despite complete and irreversible brain injuries. Accurate diagnosis of brain death (BD) reduces costs, shortens family's suffering, and increases availability of intensive care beds and organs for transplantation. Guidelines were created to standardize BD diagnostic parameters, but knowledge of medical students and medical professionals has been demonstrated to be insufficient. To assist health professionals' in BD determination, a digital training tool that contained images, videos and interactive content was developed for desktops and mobile devices. Software to create and animate 3D models (MakeHuman™ and Blender™) and a game development platform (Unity) were used. Versions for all the major operating systems (iOS™, Android™, macOS™, Windows™ and Linux™) are being made available through online repositories and mobile application stores.

Keywords:

brain death; education, medical; mobile applications.

Introduction

Technological advances in the field of intensive care medicine, with artificial ventilatory support and use of vasoactive drugs, has contributed to the maintenance of hemodynamic and respiratory activity in patients with severe and irreversible central nervous system injury. Recognizing and diagnosing this situation is critical, since such therapeutic efforts can result in increased hospital costs, contributing to the shortage of intensive therapy beds, as well as prolonging the suffering of patients' families [1].

Additionally, patients with irreversible brain damage are recognized as potential organ and tissue donors. With the advent of transplantation in the 1960s, it is necessary to redefine the concept of death, avoiding controversies in obtaining organs [1].

Mollaret and Goulon, in their 1959 classic publication, were the first to define irreversible coma ("Le coma dépassé") [2]. In 1968, a committee created at Harvard Medical School established criteria for determining irreversible coma, with the definition of brain death (BD) [1]. The American Guidelines for Determination of Death, published in 1981, were based on Harvard Criteria [3]. The American Academy of Neurology published a review of the guidelines in 1995, and the more detailed descriptions of the steps of the examination specified parameters for the apnea test positivity, and six hour intervals between evaluations [4]. These latter guidelines are still valid.

Although there is some divergence, the American criteria are often referenced as guidelines in other countries. In 2002, Wijdicks reviewed the process for BD determination in 80 countries: 69% reported presence of legal standards on organ transplantation and 88% had practice guidelines for BD [5]. Some aspects varied among countries, such as the number of physicians needed to declare BD (more than 1 in 50%), apnea test (not performed in 24%), need for complementary diagnostic tests, observation time between examinations and the mandatory qualifications of the physicians. A 2015 publication found that among 91 countries, 70% had specific legislation, with institutional protocols established at 77%. Deviations in the guidelines from the American Academy of Neurology were observed in 53% of the countries [6].

In Brazil, Federal Council of Medicine published three resolutions with criteria to BD determination that were based on the American guidelines, most recently in December 2017. According to Brazilian law, the two clinical exams that are required, must be conducted with a minimum one hour interval hour, by different physicians, and consist of a confirmatory test proving absence of electrical or metabolic brain activity, or cerebral circulatory collapse. Physicians who perform the tests must have at least one year of experience in the care of patients in coma and follow or perform at least ten determinations of BD or participated in a training course [7].

Studies that evaluated the knowledge of physicians and medical students about BD diagnosis further demonstrate the need for training. Among the undergraduate students in Brazil, there is a lack of knowledge about BC, and few feel apt to perform the examination [8-11]. Other countries, such as South Africa [12] and the United States report similar challenges [13]. The knowledge of medical professionals also has been demonstrated to be insufficient, even those working within intensive or emergency units [14-17]. This low performance may be related to the limited number of opportunities to perform the exam during training, observed even among residents in neurology or neurosurgery [18].

This lack of practical experience can be reduced with the use of realistic simulation mannequins. The use of this didactic resource has been demonstrated to be effective in BD determination training, and has been positively rated by students [19]. The BD scenario is one of the most accessible in neurology for this modality of teaching, since BD requires testing of absent reflexes. The simulation of situations in which the findings would not be compatible requires the use of more advanced simulators that incorporate capacity for movement and pupillary reaction. The main barriers to the more frequent use of realistic simulation are cost, lack of structure (employees, physical space, supplies and technical support), longer time required for application, and lack of experience with the method. Cost remains the main barrier, as

only the high-fidelity, and more expensive, mannequins can simulate respiratory movements during the apnea test [20].

The use of computers to create virtual environments and virtual patients has emerged as a powerful educational alternative for health professionals. Training that is not dependent on face-to-face methodologies can reach a larger number of students, that may be in remote locations, without increasing costs [21].Additionally, there is increased opportunity for students to manage their own learning [22] with the growing availability of mobile devices, such as smartphones and tablets. Most students and physicians have a smartphone, and most of them use specific medical applications [23, 24]. Mobile devices can be used to quickly access scientific literature, self-assessment applications, calculators, and multimedia educational content, proving to be valuable tools in health education [25, 26].

Recognizing the educational gap in medical education about the diagnosis of BD, the use of a digital teaching tool that can be delivered through mobile devices, would be a valuable contribution to the training of health professionals. Thus, in this paper we review how we developed an application for computers and mobile devices that can assist the training of students and health professionals in BD determination.

Methods

Creating a Virtual Scenario

We used Blender ™ [27], a free and open source software, to create a virtual simulation experience consisting of three-dimensional (3D) models. Complex 3D figures can be created from basic geometric shapes (planes, cubes, spheres, cylinders, etc.), modified by the addition, subtraction or repositioning of vertices, edges and faces, and by combination with other shapes. Materials or textures can be assigned to the surfaces of the models providing color and brightness to the objects.

Using visual references as images searched on the Internet, 3D representations of object found in an intensive care unit were created. Examples included vital signs monitor, mechanical ventilator, hospital bed, infusion pumps, flow meters, cannula, as well as other equipment used in this hospital environment.

We found creating 3D human models to be a much more challenging task. We used MakeHuman™, also free and open source software, [28] to generate customized characters, Parameters that could be selected included gender, height, age, weight, skin and hair color. A male model was chosen as a patient and a female as a medical examiner. The created templates were exported to Blender™.

To demonstrate the clinical examination of BD a scene was created of an intensive care unit (see Figure 1). The male patient was placed on a hospital bed and wore a disposable diaper. The female medical examiner wore blue clothing and disposable gloves and was positioned on the left side of the bed. Medical equipment that was being used to monitor the patient included electrodes, oximeter, probe to measure esophageal temperature and was connected to a monitor, triple-lumen subclavian catheter, bladder catheter with urinary drainage bag and orotracheal tube connected to a mechanical ventilator. The final elements created for the scenario included walls and a floor, infusion pumps with intravenous administration sets and solutions bags, medical gas system (valves, flow meters, hoses and reservoirs), side table with material used in the examination (syringes, bowls, catheters, "cold" solution bottle, otoscope, flashlight and cotton swabs), negatoscope with tomography film, bed control, cover sheet and blanket, and clipboard inside a rack fixed to the bed. The final version of the file with both models and all of the objects totaled approximately 1.6 million vertices.

Figure 1 – Creation of Virtual Scenario of an Intensive Care Unit Using Blender ™

After adding light sources to the scene, Blender™ can be used to create a rendered image by calculating the color, shading and lighting of objects' surfaces, in a chosen angle of view of a virtual camera. Depending on the fidelity parameters, it can take several minutes on a typical home computer to create a single image. Thus, the time to process and obtain a single sample of the approximately 3,500 images used for the project would exceed 200 hours. The time render to images was reduced by 25 times due to the use of an useful Blender™ feature, that "bakes" an object's texture and color shading information, while maintaining quality of the images.

Creating Animations with 3D Models

Within Blender™, the process known of "rigging" allows for animations to be created through manipulations of 3D models., The changes to size, shape, or position of the 3D models in successive frames creates a perception of object movement.

To control the rigging process , the 3D models are given an "armature", which is a skeleton that can have its shape, size or location modified in relation to other objects. Human models are provided with more complex armatures, composed of several bones and joints, capable of reproducing a wide variety of poses.

More discrete movements such as breathing, blinking, and heartbeat are difficult to control through armatures. For these, we used the shape keys feature to generate various versions of the same object that can re-position in the different vertices.

In Blender's™ timeline windows, we could select different frames, numbered sequentially. By default, every 24 frames in the timeline equaled one second of animation. When modifying the armatures poses or the shape keys in different frames, Blender™ was set-up to automatically calculate the expected shape of the objects in the intermediate frames. The rendered images of each frame was combined in sequence to create a video.

Seven virtual cameras were used, with different viewing angles, to better visualize the steps of the examination (See Figure 2).

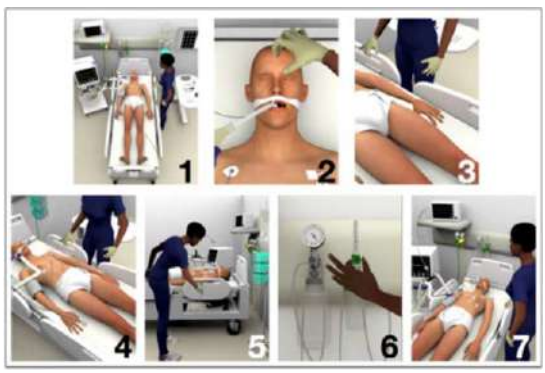


Figure 2 – Different Viewing Angles of Cameras Used in Animation Sequences

Creating and Editing Videos

Blender™ also provided a useful tool for creating and editing videos. The rendered images of all frames were combined and exported as video files (using H.264 MPEG-4 format), one video was created for each step of the clinical examination. The video editor allowed the addition of layers and transitions effects. The application of this feature was demonstrated in the development of the video for the apnea test, where an accelerated clock sequence was superimposed on the patient's image to represent the 10-minute observation time.

Editing Images

We used the GIMP™ (GNU Image Manipulation Program), a free and open source software [29], to edit and retouch the images used as textures in Blender™ and create the application (opening screen, icons, backgrounds, etc.).

Sample images of complementary exams were downloaded from Radiopaedia.org [30], a free collaborative radiology educational web resource. The images did not contain patients' data, and credit for reproductions followed Radiopaedia's user-contribution agreement.

Using a Game Engine

Developed by Unity Technologies, we used Unity™ (also known as Unity3D™ or Unity Engine™), as our cross-platform game engine [31]. Unity has a user-friendly interface and within a single project can generate application versions for all the major operating systems (Windows™, MacOS™, Linux™, iOS™ and Android™). Although it has paid subscriptions, it offers a free license for beginners whose project does not result in annual revenue over US $ 100,000 [32], fitting in with the proposal of our work.

For our project, Unity™ was used to create a game or application containing images, sounds, videos and 3D models (collectively called Assets), combined with user interface elements such as windows, text and buttons. These elements were divided into several scenes, which could be alternated according to the player's interaction with the elements on the screen.

A total of 20 scenes were created, with interactive elements that could be clicked (or touched) to display videos, text windows with relevant information or change to other scenes. Screens representing the Term of Declaration of Brain Death, similar to what was available within the Brazilian guidelines, played the role of "command center" in the application structure, from which all stages of the examination could be accessed through numbered icons. Users were allowed to revisit them as often as necessary to ensure understanding.

Some steps were chosen to be demonstrated in a more interactive way, requiring the user to perform gestures of clicking (or touching the screen of mobile devices) and dragging to interact with the content. Examples include:

- In the pupillary light reflex test, the user moved the examiner's hand with a flashlight to illuminate the patient's eyes, with fixed pupils. An icon with a question mark showed the expected response in normal situations, with reduction of pupil size to light exposure.
- In examining the corneal-eyelid reflex, the user moved the examiner's hand with a cotton swab to touch the patient's cornea, without reactivity to the stimulus. It was also possible to observe the normal find, with blinking as the response.
- In the oculocephalic reflex test, the user was instructed to horizontally drag the mouse pointer (or finger on the screen) to move the patient's head to the left or right. The patient's eyes followed the movement of the head. In the demonstration of the normal reflex, the eyes moved in the opposite direction of the head.

For more complex responses such as those shown in the steps described above, Unity requires the creation of scripts, text files containing commands written in programming language and instructions detailed to accomplish the task.

At the end of the work, executable files were created for the main operating systems. Only small differences could be seen between versions of application, such as screen resolution or some icons images (e.g. mouse pointer for computers and a drawing hand for mobile devices).

Results

The "Brain Death Determination: A Digital Guide" application beta version was created for Windows™, MacOS™, Linux™, iOS™ and Android™ operating systems.

The user is guided through the Brazilian Term of Declaration of Encephalic Death, from identification of the patient, definition of cause of coma, exclusion of confounding factors, measures for clinical stabilization of the patient, and all steps of the neurological examination with tutorials videos and interactive content. During the process, the user is alerted to important aspects and possible pitfalls during examination. In some stages it is possible to display the expected findings in normal patients. A brief explanation and sample image are given on the complementary tests used in the BD determination. At the end, the user is congratulated and invited to review the steps as many times as necessary. See Figure 3 for screenshots of examples of the application.

No images and videos of actual patients were used. This study was approved by the university ethics board (3.079.856).

The application is currently in the test period to evaluate bugs and area of improvements. Once testing is complete the application will be available for download in Internet repositories and mobile app stores (App Store™ and Google Play™).

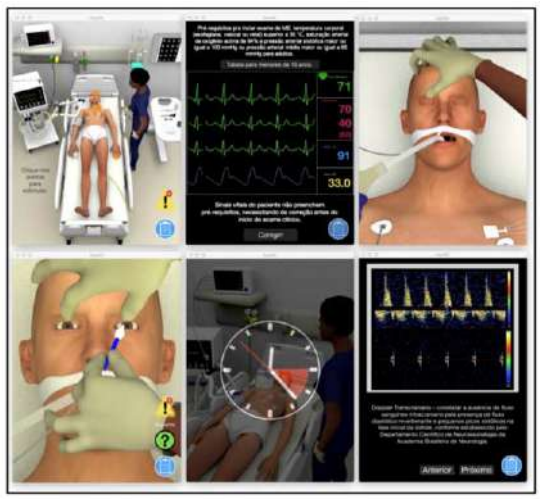


Figure 3 - Screenshots of the Application "Brain Death Determination: A Digital Guide"

Discussion

Knowledge concerning BD has proved insufficient among medical students and health professionals, highlighting the need for more effective teaching practices. A digital tool with multimedia tutorials, available in several platforms, could be a valuable tool to assist the training in the determination of BD.

With the lack of opportunities for students and residents to follow the BD examinations in practice, we choose a game-like approach to integrate content. We selected this approach for students based on its appeal, strength in experiencing of realistic simulation with mannequins, and frequent use as a methodology for training courses. The use of computers and mobile devices has the additional benefits of reaching a greater number of students with reduced cost.

The program could be easily adapted for use in other countries, with language modification and adjustments to the local guidelines. There is also the opportunity to provide updated versions to meet specific needs (e.g. providing a module for self-evaluation).

In recent years, similar applications have been used to assist in the training of different medical skills, proving to be valuable educational tools. A mobile application simulator has been demonstrated to be an effective education tool for of medical students, improving the performance in several operations and practical procedures, such as male urinary catheterization and chest tube insertion [33, 34]. For laparoscopic cholecystectomy, the same application was useful for learning cognitive aspects of procedure when used in combination with virtual reality in a multimodal training approach for general surgeons and medical students [35]. In a blinded randomized controlled study, the use of a web-based otoscopy simulator increased the diagnoses of otologic diseases by medical students, with a 24% higher score than the control group using standard otology lectures [36].

Limitations

The true potential of the application in practice still needs to be evaluated through a randomized blinded study that measures the impact of the acquisition of knowledge that involves a comparison of the application with other traditional teaching methodologies.

Conclusions

Using open-source software to create and animate 3D models (MakeHuman™ and Blender™), and a game development platform that offers free license for beginners (Unity), we were able to create a low-cost application to assist health professionals' training in BD determination. Available for all major operating systems (iOS™, Android™, macOS™, Windows™ and Linux™), this teaching tool can be used for students through their desktops or mobile devices.

Acknowledgements

The authors would like to express their thanks to Prof. José Diniz Júnior and Prof. Mário Emílio Teixeira Dourado Júnior (Rio Grande do Norte Federal University, Natal, RN, Brazil) for manuscript revision.

References

[1] Committee of the Harvard Medical School to Examine the Definition of Brain Death, A definition of irreversible coma, *JAMA* **205** (1968) 337–40.

[2] P. Mollaret and M. Goulon, The depassed coma (preliminary memoir), *Rev Neurol (Paris)* **101** (1959) 3–15.

[3] No author, Guidelines for the determination of death: Report of the medical consultants on the diagnosis of death to the President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research, *JAMA* **246** (1981) 2184–6.

[4] The Quality Standards Subcommittee of the American Academy of Neurology, Practice parameters for determining brain death in adults (summary statement), *Neurology* **45** (1995) 1012–4.

[5] E.F.M. Wijdicks, Brain death worldwide: Accepted fact but no global consensus in diagnostic criteria, *Neurology* **58** (2002) 20–25.

[6] S. Wahlster, E.F.M. Wijdicks, P. V. Patel, D.M. Greer, J.C. Hemphill, M. Carone, and F.J. Mateen, Brain death declaration: Practices and perceptions worldwide, *Neurology* **84** (2015) 1870–1879.

[7] Conselho Federal de Medicina. Resolução no. 2.173, de 23 de novembro de 2017, Brasil: Diário Ofical da União 15 dez 2017, *Seção* 1 (2017) 274-276.

[8] F.H.A. Dibo, A.F. Gravena, R.A. de Freitas, C.M. Dell'Agnolo, E. de Almeida Benguella, S.M. Pelloso, and M.D. de Barros Carvalho, Brain death: Knowledge of future Brazilian physicians, *Transplant Proc* **49** (2017) 750–755.

[9] F.P. dos Reis, B.H.P. Gomes, L.L. Pimenta, and A. Etzel, Brain death and tissue and organ transplantation: The understanding of medical students, *Rev Bras Ter Intensiva* **25** (2013) 279–83.

[10] A.G.V. Bitencourt, F.B.C.S. Neves, L. Durães, D.T. Nascimento, N.M.B.C. Neves, L. de A. Torreão, and S. Agareno, Avaliação do conhecimento de estudantes de medicina sobre morte encefálica, *Rev Bras Ter Intensiva* **19** (2007) 144–150.

[11] R.C. Afonso, D.A.B. Buttros, D. Sakabe, G.C. Paranhos, L.M.C. Garcia, M.B. Resende, and B.H. Ferraz-Neto, Future doctors and brain death: What is the prognosis?, *Transplant Proc* **36** (2004) 816–817.

[12] S. Sobnach, K.R. Wiese, M.E. Tselanyane, M. Borkum, D. Kahn, and R. Segobin, Medical student's knowledge about brain death: A South African contribution, *Transplant Proc* **48** (2016) 1904–1906.

[13] I. Tawil, S.M. Gonzales, J. Marinaro, T.C. Timm, S. Kalishman, and C.S. Crandall, Do medical students understand brain death? A survey study, *J Surg Educ* **69** (2012) 320–325.

[14] J.V. Magalhães, K.N. Veras, and C.M. de M. Mendes, Avaliação do conhecimento de médicos intensivistas de Teresina sobre morte encefálica, *Rev Bioética* **24** (2016) 156–164.

[15] M. Sheerani, M.Z.S. Urfy, B. Khealani, J. Patel, Qamarunnisa, and S. Rath, Brain death: Concepts and knowledge amongst health professionals in province of Sindh, Pakistan, *J Pak Med Assoc* **58** (2008) 352–356.

[16] A.E. Schein, P.R. Carvalho, T.S. Rocha, R.R. Guedes, L. Moschetti, J.C. La Salvia, and P.C. La Salvia, Evaluation of intensivists' knowledge on brain death, *Rev Bras Ter Intensiva* **20** (2008) 144–148.

[17] C.H. Marck, T.J. Weiland, S.L. Neate, B.B. Hickey, and G.A. Jelinek, Australian emergency doctors' and nurses' acceptance and knowledge regarding brain death: A national survey, *Clin. Transplant* **26** (2012) 254–260.

[18] A. Kashkoush, A. Weisgerber, K. Dharaneeswaran, N. Agarwal, and L. Shutter, Medical training and the brain death exam: A single institution's experience, *World Neurosurg* **108** (2017) 374–378.

[19] B.J. MacDougall, J.D. Robinson, L. Kappus, S.N. Sudikoff, and D.M. Greer, Simulation-based training in brain death determination, *Neurocrit Care* **21** (2014) 383–391.

[20] S. Hocker, and E.F.M. Wijdicks, Simulation training in brain death determination, *Semin Neurol* **35** (2015) 180–187.

[21] K. Masters, and R. Ellaway, e-Learning in medical education Guide 32 Part 2: Technology, management and design, *Med Teach* **30** (2008) 474–489.

[22] R. Ellaway, and K. Masters, AMEE Guide 32: E-Learning in medical education Part 1: Learning, teaching and assessment, *Med Teach* **30** (2008) 455–473.

[23] P. O'Connor, D. Byrne, M. Butt, G. Offiah, S. Lydon, K. McInerney, B. Stewart, and M.J. Kerin, Interns and their smartphones: Use for clinical practice, *Postgrad Med J* **90** (2014) 75–79.

[24] K.F.B. Payne, H. Wharrad, and K. Watts, Smartphone and medical related app use among medical students and junior doctors in the United Kingdom (UK): A regional survey, *BMC Med Inform Decis Mak* **12** (2012) 1.

[25] L. Briz-Ponce, J.A. Juanes-Méndez, F.J. García-Peñalvo, and A. Pereira, Effects of mobile learning in medical education: A counterfactual evaluation, *J Med Syst* **40** (2016).

[26] M. Mi, W. Wu, M. Qiu, Y. Zhang, L. Wu, and J. Li, Use of mobile devices to access resources among health professions students: A systematic review, *Med Ref Serv Q* **35** (2016) 64–82.

[27] Blender Foundation, Blender [computer program]. Version 2.79a. [2018].

[28] The MakeHuman Team, MakeHuman [computer program]. Version 1.1.1. [2018].

[29] The GIMP Development Team, GIMP (GNU Image Manipulation Program) [computer program]. Version 2.10.6. [2018].

[30] Radiopaedia.org, the wiki-based collaborative Radiology resource, (2018). https://radiopaedia.org/ (accessed November 12, 2018).

[31] Unity Technologies, Unity [computer program]. Version 2017.3.1f1 Personal. [2017].

[32] Unity Terms of Service - Unity, (2018). https://unity3d.com/pt/legal/terms-of-service (accessed November 25, 2018).

[33] R.D. Bartlett, D. Radenkovic, S. Mitrasinovic, A. Cole, I. Pavkovic, P.C.P. Denn, M. Hussain, M. Kogler, N. Koutsopodioti, W. Uddin, I. Beckley, H. Abubakar, D. Gill, and D. Smith, A pilot study to assess the utility of a freely downloadable mobile application simulator for undergraduate clinical skills training: A single-blinded, randomised controlled trial, *BMC Med Educ* **17** (2017) 247.

[34] P. Haubruck, F. Nickel, J. Ober, T. Walker, C. Bergdolt, M. Friedrich, B.P. Müller-Stich, F. Forchheim, C. Fischer, G. Schmidmaier, and M.C. Tanner, Evaluation of app-based serious gaming as a training method in teaching chest tube insertion to medical students: Randomized controlled trial, *J Med Internet Res* **20** (2018) e195.

[35] K.-F. Kowalewski, J.D. Hendrie, M.W. Schmidt, T. Proctor, S. Paul, C.R. Garrow, H.G. Kenngott, B.P. Müller-Stich, and F. Nickel, Validation of the mobile serious game application Touch Surgery™ for cognitive training and assessment of laparoscopic cholecystectomy, *Surg Endosc* **31** (2017) 4058–4066.

[36] C. Stepniak, B. Wickens, M. Husein, J. Paradis, H.M. Ladak, K. Fung, and S.K. Agrawal, Blinded randomized controlled study of a web-based otoscopy simulator in undergraduate medical education, *Laryngoscope* **127** (2017) 1306–1311.

Address for correspondence

Carlos Eduardo Rocha Correia, email: cercorreia@yahoo.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190412

Patterns of Interaction Between General Practitioners and Their Patients by Means of a Messaging System Within the Electronic Health Record Regarding Messages Asking for a Referral to a Specialist: A Descriptive Study

María Emilia Espósito[a], Santiago Esteban[a], Sergio Terrasa[a], Gabriel Villalón[a]

[a]Department of Family and Community Medicine, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina.

Abstract

The continuity of care is the relationship between a physician and a patient that extends throughout different episodes of illness. This care includes that in some opportunities the patient must consult a medical specialist. In this qualitative study, we set out to explore and describe the different requests for referral authorization to a specialist by patients affiliated with private health insurance, in order to understand the different patterns of interaction between general practitioners (GP) and their patients. The conversation analysis let us identify categories (structural and functional) after a critical reading of messages between GPs and their patients through an Electronic Health Records (EHR) messaging system. This work allowed us to know more in depth the process of continuity of care. In most cases, the doctor answered affirmatively to the request for a new referral to a specialist

Keywords:

Referral and consultation, electronic health records, continuity of patient care.

Introduction

In 2011, the World Organization of National Colleges, Academies and Academic Associations of General Practitioners / Family Physicians (WONCA) reviewed the definition and desirable attributes of GPs or family doctors. Some of these characteristics are a physician-patient relationship established over time through effective communication and the responsibility to provide longitudinal and continuous care, determined by the needs of the patient [1].

This last characteristic, the continuity of care, is defined as the relationship between a physician and patient that extends throughout different episodes of illness. It has a positive effect on the satisfaction of patients and doctors, for example, it is associated with greater adherence in people with chronic diseases, fewer consultations to emergency services and fewer hospitalizations [2-4].

Hospital Italiano de Buenos Aires, general practitioners and management of referrals to specialists

The Hospital Italiano de Buenos Aires (HIBA) is a private hospital that has a private health insurance (HI-HIBA) that is based on the concept of GPs as gatekeepers: if a patient needs or wants to have an appointment with a specialist, his GP has to allow that referral.

Regardless of the mechanism of access to the specialist, with few exceptions, it is desirable that the GP makes a formal assessment of the problem before the referral. Inappropriate referrals not only generate overload to the system, but they can condition unnecessary diagnostic cascades and harmful overdiagnoses for the patient, unnecessary expense of resources and delays in the care of patients that clearly would benefit from getting a quick consultation [5-7].

The decision to make a referral to a specialist is based on multiple factors: the disease to be treated, the recommendations of clinical practice guidelines or experts, the availability of specialists, the ability and predisposition of the GP to solve the problem, his satisfaction with his job and remuneration, etc [8].

In 2006, the HIBA began the Personal Health Portal project (PHP), designed to provide new functionalities to the people who are treated at the Hospital. The PHP is a Personal Health Record linked to the EHR hospital, which provides services and access to unified data in multiple devices, allowing patients to interact or consult their medical or administrative information. Within the tools provided by this system, there is a messenger that allows patients to write to their GP. This is one of the most frequent ways of requesting referral authorization. Patients also can send messages by calling a call-center.

Health Care Quality Improvement Program and its indicators in the Department of Family and Community Medicine of the Hospital Italiano de Buenos Aires

The Department of Family and Community Medicine of the HIBA (DFCM) has a Quality Improvement Program (QIP) that use certain indicators and economic incentives as main strategies. Since 2017, we began measuring two continuity of patient care indexes for each GP: the Usual Care Provider index (UPC) and an index developed by Bice & Boxerman (COC) [9]. Other indicators used by the QIP are related to the accessibility to scheduled appointments and the response to messages sent by patients through the PHP or call-center. This last indicator is operationalized as the rate at which a GP consult the EHR of patients who have sent messages asking for a referral during the previous seven days.

We consider that given that this indicator is measured purely quantitatively, without evaluating the quality of the responses provided by the GPs, it does not allow to distinguish the answers framed in the concept of "continuity of care" and those that are not.

With this study we set out to explore and describe the different requests for referral authorization to specialists from patients affiliated with the HI-HIBA, with the aim of understanding the different patterns of interaction between physicians and patients.

Methods

We made a qualitative and quantitative description of different requests for referral authorization made by patients and the answers provided by GPs working in the DFCM-HIBA.

Population, source and data collection

We defined a population of "requests for referral authorization". For the purposes of this investigation, the latter was operatively defined as the requests for referral authorization to specialists made by patients through the messenger service of the PHP in order to avoid extra payment.

First stage: identification of words most used by patients in the referral request messages

The research began by requesting a random sample of 150 patients who had sent a message to their GP between 01/01/2016 and 12/31/2016. This data was requested from the Information Management Area of the Research Department of the HIBA. The answers given by their GP during the first seven days after the first message was sent, were obtained from the EHR.

From the sample of 150 patients who had sent at least one message to their GP in the selected period, a total of 1019 exchange messages were obtained. Of these, 372 messages (36.5%) made reference to the request for referral authorizations. Of these 372, a total of 288 had been sent by the patients.

Of the total of 288 messages sent by patients for the request for referral authorization, 93 (32.2%) were generated using a structured option in the PHP and not as free text addressed to the GP. However, in the free texts of the messages, the following keywords were identified: *derivación* (referral), *derivaciones* (referrals), *especialidad* (speciality), *especialista* (specialist), *interconsulta* (interconsultation), *autorización* (authorization), *autorizar* (authorize), *hacer una orden* (make an order).

Second stage: searching messages specifically related to the request of referral to specialists through the roots of words identified in the previous step

Once the most used words were identified in referral requests, we made a second request to the Information Management Area with the roots of words found in the previous step, independently of the number of repetitions. We also requested the registration of the authorizations made by the GP in the EHR.

Table 1 - Roots of the words found during the first stage that were used during the second stage to identify the free text messages linked to the request for a referral

Root	Words
Deriv	*Derivación / Derivaciones / Derivar / Derivarme*
Interc	*Interconsulta / Interconsultas*
Autoriza	*Autorización / Autorizar*
Especiali	*Especialista / Especialidad*
Pase	*Pase*
Orden	*Orden*

In this second stage, we obtained a total of 2600 messages between patients and doctors, which allowed us to carry out a qualitative analysis of 214 conversations in which the patient had requested his or her GP to authorize a referral to a specialist. For this analysis, the main researcher reviewed the conversations between patients and physicians, as well as the clinical notes recorded in the EHR during the episodes of care. This analysis allowed us identifying patterns of interaction between GPs and their patients, making it possible to identify categories that were built from the critical reading and the discourse analysis.

Results

The qualitative analysis of 214 conversations related to the request for referral authorization to specialists allowed the elaboration of two types of categories: structural and functional.

Structural categories

Three categories were identified that refer to the structure of the analyzed conversations: a) the content of the message sent by the patient requesting the referral authorization; b) the answer made by the doctor; c) the referral authorization made by the doctor through the EHR.

a) The content of the message sent by the patient requesting the referral authorization to a specialist

The content of the message was defined as all the information included in the message sent by the patient to his/her GP requesting a referral authorization. There were identified 136 messages in which the only content detected was the request for referral authorization. Of these 136, 77 messages had been sent from the PHP messenger in a free text format, such as:

> Female, 61 years old: *"...hola Dr. le pido derivación para endocrinología y coloproctología..."* ("... hello Dr., I'm asking you for referral to endocrinology and coloproctology ...").

The other 59 messages were created from a "ask for referral authorization" structured option in the PHP, which sent an automatic message with the following text:

> *"El paciente (...) solicita renovar su derivación con el servicio de endocrinología ambulatoria".* ("The patient (...) requests to renew his referral to ambulatory endocrinology service".)

Other 78 messages with more content were identified, in which the reason for requesting the referral authorization was expressed. These messages were analyzed to identify the different reasons for referral request (see functional categories below).

b) The answer made by the doctor in response to the request for referral authorization

This category was defined as all the information contained in the message sent by the GP in response to the request for referral authorization.

It was observed that of the 214 requests, only 57 were answered by the GP through a message in the PHP and, in all these cases, the content of this messages was only the notification of the given authorization or a personal greeting to the patient, with the eventual addition of recommendation of a professional:

> 49 years old female GP to 20 years old female patient: *"...te dejo la derivación con Nutrición..."* ("... I leave you the referral authorization to Nutrition ...").
>
> 47 years old male GP to 72 years old male patient: *"...va a ser cuestión de tiempo la recuperación del olfato. Cualquier cosa, la Dra. XXX es de mi confianza..."* ("... it's going to be a matter of time the olfactory recovery. In any case, Dr. XXX is of my trust ...").

c) Referral authorization made by the doctor through the EHR

When analyzing the referral authorization made by the doctor through the EHR, we identified different possible scenarios and two possible responses: loading or not the authorization in the EHR.

In most cases, the GP performed the authorization. The only cases in which we observed that the authorization had not been carried out were in the context of requests for renewal of

previous referrals, when the authorization were already loaded and still valid in the EHR, in the presence of some acute symptom or when the GP decided to make an appointment with the patient.

Functional categories

We identified four functional categories that refer to the purpose of conversations between patients and physicians: a) the reason for requesting referral authorization expressed in the patient's message; b) the time of evolution of the health problem related to the request; c) the level of care where the health problem can be solved (primary care vs. not primary care); d) the trigger for the request for referral authorization.

a) The reason for requesting referral authorization expressed in the patient's message

From the total analyzed sample (214 conversations), 78 were identified where the reasons for requesting a referral could be detected.

One of the options in this category is when the reason given by the patient for requesting for a referral to the specialist is a new diagnosis made in the context of an unscheduled consultation in the emergency department:

> Male, 61 years old: *"...estoy con una tendinitis en el brazo derecho. Me estoy atendiendo con la guardia de traumatología pero ahora me derivaron a un especialista de mano. Necesitaría una derivación para esa especialidad..."* ("... I have a tendinitis in the right arm. I was dealing with it in the trauma emergency department but now I was referred to a hand specialist. I would need a referral authorization to that specialty ... ").

We also detected referral requests for health problems previously known to both the patient and his doctor and which had previously motivated consultations with a specialist:

> Female, 62 years old: *"...me enviarías una derivación para la Dra. de otorrino, es por mis acúfenos, tengo turno el 2 de junio..."* ("...would you send me a referral authorization to the otolaryngologist, it is because of my tinnitus, I have an appointment on June 2...").

Another option that we observed were cases in which the persistence of symptoms, having previously consulted the GP about this problem, make patients to request referral authorization to a specialist:

> Male, 72 years old: *"...todavía no recupere el sentido del olfato y el gusto, derivame por favor a un otorrinolaringólogo de tu confianza para evaluarme..."* ("... I still do not recover the sense of smell and taste, please refer me to an otolaryngologist from your trust to evaluate me ...").

Similarly, we observed that in some cases, a new or recurrent appearance of an acute symptom triggers a referral request:

> Male, 62 years old: *"...me resucitó el dolor de espalda. Hace días que no se va. No es demasiado intenso pero es persistente... Me podés derivar con alguien para verlo?"* ("... my back pain is back since days ago. It is not too intense but it is persistent... Can you refer me to someone to see it?").

Patients also usually express that they request for referral authorization to perform control of a specific health problem related to a speciality. This situation could be verified by reviewing the EHR:

> Male, 62 years old: *"... por favor solicito derivación a cardiología para poder asistir a control..."* ("... please, I request referral to cardiology to be able to attend to control...").

On other occasions, it was observed that the patient has been suggested by another person than the GP to consult with a specialist to evaluate the result of a complementary study (see further category "referral request trigger"):

> Male, 62 years old, referring to an electrocardiogram: *"...me piden que me haga la derivación para cardiología, por el informe del ECG tiene que verme un Cardiólogo..."* ("...they told me to ask you for the referral authorization to cardiology, for the EKG report a cardiologist has to see me...").

At last, as we expressed previously, in some cases the reason for requesting a referral was not expressed. This type of requests was frequent for dermatology, and somewhat less frequent for otorhinolaryngology, general traumatology and nutrition:

> Male, 52 years old: *"Hola, necesito una derivación para dermatología, ya que el dia 18/03 tengo un turno con el Dr. XX"* ("Hello, I need a referral to dermatology, because I have an appointment with Dr. XXX on the 18th of March").

b) The time of evolution of the health problem related to the request

In relation to the time of evolution of the health problem that motivated the referral request, we were able to differentiate acute evolution problems analyzing the messages and the EHR:

> Male, 43 years old: *"...me están doliendo muchos los tobillos y también las piernas, pienso que por el efecto del tobillo, qué especialista en pie me recomendás?"* ("... my ankles are hurting a lot and also my legs, I think that because of the ankle effect. Which foot specialist do you recommend me?").

We also recognized requests related to chronic problems:

> Female, 61 years old: *"Sería tan amable de hacerme una derivación a hipertensión. Es solo por control" ("Would you authorize me a referral to hypertensology? Is it just for a control").*

c) The level of care where the health problem can be solved (primary care vs. not primary care)

We documented referral requests to specialists for problems that could be handled in primary care by the GP. These situations could make us infer that the patient would not recognize the GP as someone with the capacity to solve that problem:

> Male, 52 years old, with a recent diagnosis of type II diabetes without pharmacological treatment: *"...me podrías dejar una derivación para consultar un especialista en diabetes"* ("...could you leave me a referral to consult a diabetes specialist?").

On the other hand, we have also identified health problems that exceed primary care capacities, such as some surgical problems:

Female, 62 years old: *"...estoy pensando en operarme la hernia umbilical. ¿A quién me recomendás? ¿Podrás hacerme una derivación?"* ("I'm thinking on solving my umbilical hernia. Who do you recommend? Can you authorize the referral?").

d) The trigger for the request for referral authorization

When analyzing the different conversations, who triggered the request for referral authorization was detected as a category. In most cases was the patient himself:

> Male, 29 years old: *"...este mes tengo turno para el dermatólogo por unos granitos en la cara que no se me van y con el otorrinolaringólogo ya que estoy constantemente, mes tras mes, con la garganta irritada. Me parece que necesito derivaciones para ambos..."* ("... this month I have appointments with a dermatologist because of some spots on my face that do not go away

and with the otorhinolaryngologist since I am constantly, month after month, with sore throat. It seems to me that I need referrals for both... ").

In other cases the patient requests a referral after a suggestion made in the past by the GP:

Male, 21 years old: *"Doctor, siguiendo su indicación, necesito que me haga una derivación a una nutricionista"* ("Doctor, following your instructions, I need you to refer me to a nutritionist").

Another possibility is that the referral has been suggested by another professional of some ambulatory specialty:

Male, 54 years old: *"Necesitaría una orden de derivación para cirugía plástica a quien me derivó el dermatólogo…"* ("I would need a referral authorization to plastic surgery to whom the dermatologist referred me…").

Also, the referral could be suggested by a physician in the emergency department:

Female, 51 years old: *"… tengo turno con un especialista en mano mañana por la tarde por una molestia, me lo indicaron en demanda espontánea. Me dicen que necesito que me ingreses una derivación…"* ("... I have an appointment with a hand specialist tomorrow afternoon, I have been referred in the emergency department. They told me that I need you to authorize the referral...").

Other health professionals previously involved in the patient's care, for example, a psychologist from another institution, could have made the referral:

Male, 26 years old: *"…me trato con un psicólogo particular. El mismo me sugirió comenzar tratamiento por depresión en el hospital. Ya tengo turno para la admisión. Necesito derivación a psiquiatría…" ("... I´m seeing a private psychologist. He suggested me to start treatment for depression in the hospital. I already have an appointment. I need a referral to psychiatry...").*

Also, a primary care physician could have referred a patient to a specialist during an emergency department consultation:

Male, 52 years old: *"…necesito la derivación para dermatología. Estuve el sábado en la guardia y me recomendaron ver al especialista..."* ("... I need the referral authorization to dermatology. I was on the emergency department on Saturday and they recommended me to see the specialist...").

Finally, it was identified that in some cases who triggers the referral request is a non-medical personnel, for example, an electrocardiograph technician:

Male, 62 years old, after an electrocardiogram requested by his GP: *"…me piden que me haga la derivación para Cardiología, por el informe que tiene que verme un Cardiólogo…"* ("... they asked me to make the referral for Cardiology, for the report a cardiologist has to see me ...").

Discussion

This study allowed us to deepen how the interaction between physicians and patients occurs regarding referral requests to specialists. From the qualitative analysis of the conversations, we were able to develop categories that allowed us to characterize conversations and to understand the process of continuity of patient's care.

It is striking that, contrary to what we thought, in most cases, the GP responds in an affirmative way to the request for a new referral to a specialist regardless of the speciality, the reason expressed in the message and the time of evolution of the health problem. This generates uncertainty regarding the role of the GP as "gatekeeper".

This controversy is also observed in countries with more rigid health systems in relation to the patient's contact with the second level of care [10]. Among the arguments in favour of the role of primary care physician as gatekeeper are: that promotes less use of the health system, a shorter delay in appointments with specialists, a greater system efficiency and less overdiagnosis and overtreatment. On the other hand, some of the arguments against are: the increase in costs due to delayed diagnosis, the lack of respect for the patient's choice, the increase in workload for primary care physicians and conflicts in the physician-patient relationship [11].

It is very likely that this controversy exists internally in GPs, being one of the factors that leads to respond affirmatively in almost all referral requests, jeopardizing the continuity of care. Another factor could be that GPs working in HIBA are used to referral many health problems that could be handled in primary care, such as dermatology or otorhinolaryngology issues.

On the other hand, some affirmative answers to the referral requests were adequate and framed in the continuity of care. For example, those requested for surgical specialities and specialities that perform procedures that are not usually carried out by GPs of the DFCM of the HIBA.

We consider that some requests for referral authorization are difficult to answer in a negative way. Such is the case of the request for referral to a specialist to follow up a diagnosis made in the emergency department, where the patient is evaluated by another doctor. The same applies to those referrals suggested by a specialist, by another health professional (for example a psychologist) and when the GP is licensed, and the replacement physician must answer the request.

Our study has certain limitations and weaknesses. We have not been able to adequately analyze the conversations in which the GP decides to communicate with the patient by telephone to assess the referral request. Most of these interactions are not recorded in the EHR.

We did not include in our protocol messages sent through the call-center service. This implies having left out of the analysis a certain population, mostly elderly, that does not use the messenger system of the PHP.

Despite the previous limitations, we have not been able to identify similar studies carried out in our country or in our region and we believe that this work constitutes a fundamental contribution to understand the process of referral requesting to specialists.

In the future, it would be interesting to know the opinion of patients about the role of the GP as gatekeepers, to evaluate their satisfaction with this system and to examine whether this function improves the quality of care of them.

Conclusions

This work allowed us to understand the process of referral requesting and the arrival of the patient to the appointment with the specialist. It lays the basis for new research on the continuity of patient care and the role of the general practitioner as a gatekeeper.

Acknowledgements

This study was funded by the Department of Family and Community Medicine of the Hospital Italiano de Buenos Aires.

References

[1] Allen, J., Gay, B., Crebolder, H., Heyrman, J., Svab, I. and Ram Maastricht, P. (2011). The European Definition of General Practice / Family Medicine. 3rd ed. [ebook] WONCA EUROPE. Available at: http://www.woncaeurope.org/sites/default/files/documents/Definition%203rd%20ed%202011%202011%20with%20revised%20wonca%20tree.pdf

[2] Nutting, P. A. Continuity of Primary Care: To Whom Does It Matter and When?, The Annals of Family Medicine 1.3 (2003): 149-55

[3] Dreiher J, y col. The association between continuity of care in the community and healt outcomes: a population-based study. Isr J Health Policy Res 2012;1(1):21.

[4] Pourat N. y col. In California, Primary Care Continuity Was Associated With Reduced Emergency Department Use And Fewer Hospitalizations. Health Affairs 2015;34(7):1113-20.

[5] Brekke, Kurt R., Robert Nuscheler, and Odd Rune Straume. Gatekeeping in Health Care. Journal of Health Economics 26.1 (2007): 149-70.

[6] Ciapponi A. Goodbye to prostate cancer screening with the current technologies. Evid Act Pract Ambul. 2012;15(2):Abril-Junio 2012.

[7] Velazquez A. The overdiagnosis and the problems of preventive medicine. Evid Act Pract Ambul. 2014;17(3):Julio-Septiembre 2014.

[8] Godager, Geir, Tor Iversen, and Ching-To Albert Ma. Competition, Gatekeeping, and Health Care Access. Journal of Health Economics 39 (2015): 159-70.

[9] Reid R, Haggerty J, McKendry R. Defusing the confusion: Concepts and measures of continuity of healthcare [Canada]. Ottawa, Canada: Canadian Foundation for Healthcare Improvement/Fondationcanadienne pour l'amélioration des services de santé; 2002.

[10] Forrest CB. Primary care in the United States: primary care gatekeeping and referrals: effective filter or failed experiment?. BMJ 2003;326:692-5.

[11] Greenfield G, Foley K, Majeed A. Rethinking primary care's gatekeeper role. BMJ. 2016;354(September):1–6.

Address for correspondence

Maria Emilia Esposito. Correspondence to:
mariaemilia.esposito@hospitalitaliano.org.ar

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190413

Training Leaders in Health Informatics

Susan H. Fenton, Angela Ross, Debora Simmons

The University of Texas School of Biomedical Informatics, Houston Texas, USA

Abstract

Despite the widespread adoption of electronic health records (EHRs) in the U.S. over the past decade, significant improvements, especially in patient safety, have yet to be realized. This finding, along with health informatics workforce data and an identified gap in the offerings of an educational program, led to a proposed professional doctorate in health informatics. Developed via stakeholder focus groups, the program was approved by the public university system, the state-level educational authority, and the regional accreditation body, with final approval in July 2018. Unique features of the program include a prolonged practice project demonstrating a return on investment, as well as online and face-to-face delivery components. This program aims to develop evidence-based professionals who improve the health of people and populations through the application of health informatics. Applications and interest in the first class are high.

Keywords:
Education, Health Informatics

Introduction

In 2004, President George W. Bush outlined a plan to ensure that most Americans would have electronic health records within the next 10 years. The President believed that better health information technology was essential to his vision of a health care system that puts the needs and the values of the patient first and gives patients the information they need to make clinical and economic decisions – in consultation with dedicated health care professionals [1].

President Barack Obama continued to support President Bush's mandate during his administration, including incentives for implementing electronic health records (EHRs) in the American Recovery and Reinvestment Act [2]. As a result of this, the health informatics industry has flourished with the need for an educated informatics workforce[3]. This workforce is needed to assess, design, implement and evaluate the electronic health record (EHR) and patient care technologies, as well as effectively utilize the data generated; all with the goal of preventing medical errors and providing the information needed to improve human health.

As of 2017, 96% of all U.S. Non-federal Acute Care Hospitals have adopted a certified EHR [4]. Documented benefits of EHRs include enhanced patient care; real-time, multi-user access to the patient record; alerts for medication errors and critical lab values, among others[5]. However, in spite of these benefits of the introduction and widespread use of EHRs in hospitals, medical errors are perceived to have increased and are estimated to be the nation's third leading cause of death[6]. These medical errors also include unanticipated and undesirable consequences from implementing an EHR. In 2016, Johns Hopkins researchers examined four separate studies that analyzed medical death rate data from 2000 to 2008. Then, using hospital admission rates from 2013, they extrapolated that based on a total of 35,416,020 hospitalizations, 251,454 deaths stemmed from a medical error, which the researchers say now translates to 9.5 percent of all deaths each year in the U.S[6]. In March 2019, Fortune magazine published "*Death by 1,000 Clicks"* describing multiple failures of EHRs, with even those who led the national U.S. effort stating EHRs "have not fulfilled their potential."[7]

The health care industry is in need of transformational change in the education required for executive-level health informaticists who practice at the most advanced levels. The 2019 Health Information Management Systems Society (HIMSS) Leadership and Workforce Survey found that 84% of all respondents reported to be in a management role, with 47% associating themselves with an "Executive Management" position[8]. The health informatics executives are an integral part of the health care team, providing the foundational technology infrastructure to support efficient, effective care. The recommendation that health informatics executives practicing at the highest levels should receive doctoral level preparation emerges from multiple factors evident by the expansion of scientific knowledge in health informatics and growing concerns regarding patient safety, the continued need for improved patient outcomes, and quality care delivery.

The 2019 Health Information Management Systems Society (HIMSS) U.S. Leadership and Workforce Survey ranked the following priority needs concerning quality care delivery and health informatics/health IT in health care organizations:

- Cybersecurity, Privacy, and Security
- Improving Quality Outcomes Through Health IT
- Clinical Informatics and Clinician Engagement
- Culture of Care and Care Coordination
- Process Improvement, Workflow, Change Management
- User Experience, Usability and User-Centered Design
- Data Science/Analytics/Clinical and Business Intelligence
- Leadership, Governance, Strategic Planning
- Safe Info and Tech Practices for Patient Care
- HIE, Interoperability, Data Integration and Standards[8]

It is worthwhile noting that leadership and user experience are new to the top 10 in 2019. The health informatics workforce data indicates there is an increased demand for health informatics in several health care delivery models. In 2017, HIMSS noted that efforts to advance clinical IT in non-Hospital Provider environments (ambulatory and long term Post-Acute care facilities) can be challenging when there is no educated, trained, and designated health IT workforce [9]. Both the 2018 and 2019 reports concluded that the influence of executive management health IT professionals continues to expand in organizations[8,10].

The United States has implemented the position of the Chief Medical Informatics Officer (CMIO), as well as the Chief Nursing Informatics Officer (CNIO), in its largest healthcare organizations. For the CMIO, the American Board of Medical Specialties administers the clinical informatics subspecialty exam for physicians[11]. Likewise, the American Nurses Credentialing Center administers an informatics nursing certification, the RN-BC[12]. In the UK, it appears that the position of Chief Clinical Information Officer (CCIO) is emerging[13]. The nursing certification content outline does not include management or leadership skills. The US subspecialty exam includes one item related to leadership needed for fostering change and managing large-scale information projects. The UK CCIO role is proposed as a leadership role with a clinical focus. While there is no doubt of the need for these clinical leadership roles in health informatics, they often do not include the techinical or management background needed for the chief information officer (CIO) or chief technology officer (CTO). In addition, while these clinical roles are essential in care delivery organizations, they may not be found in pharmaceutical or insurance or other health information technology organizations. The advent of digital health in the US and around the world will only expand the types and numbers of organizations engaging in the health informatics field. Effective informatics leaders are needed in all of them.

Therefore, **it is critical that the executive-level health informatics professional is educated and prepared to manage complex organizations across the continuum of care and throughout the healthcare industry**.

In addition to the available workforce data, the University of Texas School of Biomedical Informatics (SBMI) identified a gap in their current educational offerings after multiple applications for the doctor of philosophy degree from industry professionals who could not articulate a research area for their studies. These applicants felt the need to continue their education to a terminal degree, as is the case for the MD, PharmD, DNP, and DPT. However, historically, the only terminal degree option in the informatics field has been the PhD. With all of these issues in mind, a group of applied informatics faculty proposed the development of a practice doctorate, or professional degree, in health informatics (DHI).

Methods

The first step in the development of the DHI was approval of the concept at a faculty retreat. Following this, the task of developing the proposal was assigned to the Associate Dean for Academic Affairs, with assistance of the applied faculty and staff members. The process for public institutions in Texas necessitates several steps with the format of the proposal proscribed in regulation.

Development of the proposal began with submission of a letter of intent to develop a new degree program to the university and The University of Texas System administration. This was followed by the convening of focus groups of Texas Medical Center and other SBMI stakeholders. These stakeholders are persons who mentor SBMI master's degree students for capstone projects or have employed SBMI graduates or are SBMI alumni. Focus groups were held in early 2017 to ascertain whether the proposed program was needed, as well as determine topics that were deemed most important for health informatics executives and managers. Eleven (11) people either participated in the focus groups or were interviewed individually if the focus group meetings were not convenient. The titles of those participating are as follows:

- One Executive Officer with a nursing background from the U.S. Defense Health Agency
- Three Directors of Information Technology, Information Technology Development, and Information Systems Operations from large systems in the Texas Medical Center, respectively.
- One Vice President of Clinical Informatics for a large integrated health care system.
- One Division Director of Clinical Informatics for a multi-state health care delivery system
- One Vice President of Applied Clinical Informatics and Chief Nursing Informatics Officer from a tertiary care system.
- One Executive Director of Clinical Quality Informatics for a multi-hospital system
- One Director of Nursing Operations
- One Informatics Manager with a public health system.
- One Associate Chief Medical Informatics Officer

Once the focus group results were analyzed, development of the proposal began. The format required includes evidence related to Job Market Need; Existing Programs; Student Demand; Student Recruitment; Enrollment Projections; Accreditation; Admission Standards; Program Degree Requirements; Curriculum; Candidacy/Dissertation Requirements; Use of Distance Technologies; Program Evaluation Plans; Faculty Availability and Teaching Load; Student Financial Aid; Library and Other Teaching Resources; Support Staff; potential Proposal Evaluators; and a 5-year Financial Plan.

Results

Focus Group Results

The focus groups and persons interviewed individually were very supportive of the need for and subsequent development of a doctorate in health informatics (DHI). Selected comments include:

"There needs to be an understanding of analytics and linking that back to scientific inquiry to better collaborate with research Ph.D. students & make things workable for the end user. It is important to work in an evidence based space so we can continue to inform the community and fill the gap in this doctorate; fill the gap so people are trained appropriate in practical application."

"Terminal degree completion for specific specialties is important. Practice prepared topics can make the research information usable in the field and can provide real-time solutions."

"Practice based trends/evidence are important. Students should be driving things back to the researchers – feeding things back to Ph.D. researchers to identify trends and help validate their work."

"We are looking for someone with the ability to think strategically and be forward thinking to look at the impact of regulatory requirements. How can we leverage or utilize technology that addresses new needs that are coming in the future?"

"The number one skill needed is change management experience as we implement new systems. You need to be able to change a culture in an organization."

"While there is a discussion of workflow, formality and discipline, the structure is not there. How do you structure a proper workflow that is then communicated across the organization – it is just not there. Discipline is lacking. When you look at staff and come to see how they have become an "Informatician", background is lacking."

"Many informaticians are heavily focused on implementation and do not look back at the practice of it to do an evaluation or look for evidence-based practice."

The final proposal document with all required attachments totaled 62 pages.

The proposal was approved by The UT Board of Regents in August 2017. From there, it was submitted to the Texas Higher Education Coordinating Board (THECB). The THECB required a desk review of the proposal and later conducted a site visit with three external reviewers. Final approval was obtained in July 2018.

Program Development

The professional doctorate in health informatics is a degree sorely needed for the current and future success of health care across the globe. In the U.S., a master's degree extends what is learned at the bachelor's degree, but does not provide the depth of a doctoral degree. Many informatics professionals, currently holding master's in informatics degrees and working in health care organizations wish to continue their training. However, they do not wish to conduct the research necessary for a PhD. The only knowns professional doctorate in health informatics and the requirements are described herein.

The objectives of the program are to:

- Implement evidence-based practice to improve human health.
- Assume leadership positions throughout the healthcare industry having integrated health informatics with organizational leadership and ethics.
- Design, implement and evaluate health information technology quality improvement projects in health care systems.
- Synthesize health informatics and patient care technologies to effect improvements in health care delivery and patient safety.
- Employ effective communication and collaboration skills to identify and implement best practices in health care delivery.

Admission Standards

The admission standards include:

- Bachelor's degree with documented executive/management health care experience
- Master's degree preferred
- Personal statement that includes short and long-term goals after completion of the practice doctorate program
- Proposed area of interest for the translational practice project
- *Letter of Support from the health care organization willing to facilitate the translational practice project*
- Resume or curriculum vitae
- Three letters of recommendation from supervisors (two letters should be from supervisors), colleague etc.
- Interview with the SBMI Application, Progression, and Graduation Committee (APG) (by invitation upon review of application)
- Cumulative GPA of 3.0 or higher

The maximum number of students admitted to the first cohort will be thirty (30), with no set minimum. That is, if only 15 applicants are qualified for the program, only 15 students will be accepted into the first cohort.

Curriculum

The student learning objectives for the practice doctorate in health informatics are as follows:

1. The student will design and implement a project plan to address an operational health informatics problem.
2. The student will apply the appropriate scholarly foundations and evidence-based practices when designing and implementing the project.
3. The student demonstrates broad knowledge in the field of health informatics and advanced knowledge in a sub-discipline, such as privacy and security, data analytics, revenue cycle management, or other specified, in conjunction with the advisory committee.
4. The student will present and disseminate his/her project and the evaluation using strong verbal and written communication skills.

The curriculum is found in Table 1.

Table 1 - Curriculum

Course #	Course Title	Semester Credit Hours
Year 1 – Fall		
BMI 5300	Introduction to Biomedical Informatics	3
BMI 6324	Health Information Technology Policy	3
Year 1 – Spring		
BMI 6328	Healthcare Delivery in an EHR-enabled Environment	3
BMI 6311	Advanced Decision Analysis	3
Year 1 - Summer		
BMI 7350	Scholarly Foundations of Advanced Health Informatics Practice	3
Year 2 – Fall		
BMI 6305	Social Dynamics and Health Information	3

Course #	Course Title	Semester Credit Hours
BMI 6316	Change Management in Health Informatics	3
BMI 7170	Project Advisement	1
Year 2 - Spring		
BMI 7360	Advanced Project Management	3
BMI 7351	Evidence-Based Health Informatics Practice	3
BMI 7170	Project Advisement	1
Year 2 - Summer		
BMI 7361	Vendor Relations and Contract Negotiation	3
BMI 7170	Project Advisement	1
QUALIFYING EXAMINATION		
Year 3 – Until Graduation		
BMI 7070	Fellowship in Health Informatics	6
BMI 9950	Project Evaluation and Writing	6

The courses beginning with 6000 are electives in the master's and PhD degree programs. Those beginning with 7000 are unique to the DHI program. BMI 5300 is a leveling course required for all students new to SBMI. The selection of the additional courses was influenced and guided by the focus group feedback. As can be seen from the course titles, the program curriculum is focused on developing the skills needed for informatics managers and executives to implement evidence-based health informatics practices and technologies to improve care delivery and outcomes.

In addition, during their course of study, students will identify and seek a certification relevant to their translational practice project.

Translational Practice Project

The proposed practice doctorate will conclude with a project evaluation report to be written upon completion of a translational practice project. The proposed translational practice project requirements will consist of:

- Background and Review of Relevant Literature/Evidence
- Project Overview
- Theoretical Framework/Logic Model
- Purpose Statement/Significance of Project
- Evaluation Design, including Return on Investment
- Implementation/Gather Evidence
- Recommendations
- Future Implications

Program Delivery

The program was designed to be delivered executive-style. Thus, all of the didactic course content will be available online. However, students will be required to travel to Houston, Texas for onsite meetings. These will be held over long weekends, beginning on a Thursday and ending on a Saturday. This will enable cohort building, meetings with advisors, and value-add activities such as guest lectures from leaders in the field of health informatics. First year students will meet 5 times throughout the year, while second through fourth year students will meet on campus once per semester.

Program Evaluation

This program is the first of its kind in the country. As such, there is no accreditor for a practice doctorate in this discipline at this time.

During the first and second cohorts, formative evaluations in the form of student feedback sessions will be conducted. Course evaluations will also be reviewed. Longitudinally, the program will be evaluated as follows:

1. Acceptance rates each year and longitudinally – More applications for the program than available slots will be an indication of the demand for this type of program. Decreasing rates of acceptance as the number of slots available each year increases over the first five years will indicate strong need for the program.
2. Qualitative Assessment at 2 and 4 years – Graduates will be interviewed and surveyed by program faculty at 2 and 4 years after graduation to determine which skills and competencies learned in the program have proven useful and which have not.
3. Aggregate Return on Investment of practicum projects – Demonstrating the cost-effectiveness of the practicum projects demonstrates the benefit the program can provide to the health care industry.

In addition, the state of Texas requires an annual report from all doctoral programs and an external review every 7 years from external reviewers.

Discussion

The DHI has not yet been implemented. However, given the evaluation proposed and the inclusion of a Return on Investment as an essential component of the translational practice project, value is expected to be documented as soon as the first students complete their studies. Whether these savings are expressed in actual dollars saved or adverse patient events prevented or efficiencies in the application of health informatin technology remains to be seen.

Upon approval in the early fall of 2018, the school issued a press release and opened an application on the website. (https://sbmi.uth.edu/news/story.htm?id=61f010ca-127e-4778-ab02-d7925d562f7e) An informational webinar held in early October 2018 had 27 participants from across the U.S. and even overseas.

One of the ways to judge the perceived need for a program such as this is by the expressed interest. By the March 1, 2019 deadline, twenty-nine (29) persons had started applications and thirteen (13) had submitted completed applications for review. As of this submission, applicant interviews were underway to choose the initial cohort.

Conclusions

This is a new terminal degree program aiming to educate health informatics and information technology executives as they seek to become evidence-based professionals who improve the health of people and populations through the application of health informatics. To the knowledge of the authors and sponsoring school, this is the only degree of this type in the world. As such, the program seeks to advance standards of care and models of delivery and health care policy via the implementation of the translational health informatics practice project. This proposed program offer health informatics

executives a unique educational opportunity that is not currently offered anywhere else in the U.S. or the world. The practice doctorate program will serve diverse communities, as executives within the program will represent a variety of health care facilities, companies, and other organizations (such as insurance and pharmaceutical companies) throughout the Texas Medical Center and beyond. The in-depth translational projects will seek to advance quality of care in those facilities and improve health care delivery, while reducing costs, for patients, their families, clinicians and the sponsoring organizations. The level of interest to date suggests that the health care industry is ready for a practice doctorate of health informatics or DHI.

Acknowledgements

The authors have no conflicts. The outline of the program contained in this document is subject to change. The authors which to thank their faculty colleagues at SBMI, as well as the SBMI Office of Academic Affairs staff for their support in bringing this dream to reality.

References

[1] George W. Bush, State of the Union, (2004). https://georgewbush-whitehouse.archives.gov/infocus/technology/economic_policy200404/chap3.html (accessed November 18, 2018).
[2] United States Government, American Recovery and Reinvestment Act of 2009, 2009. http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=111_cong_bills&docid=f:h1enr.pdf (accessed March 2, 2009).
[3] W. Hersh, The health information technology workforce: estimations of demands and a framework for requirements, *Appl Clin Inform*. **1** (2010) 197–212. doi:10.4338/ACI-2009-11-R-0011.
[4] Office of the National Coordinator, Non-federal Acute Care Hospital Electronic Health Record Adoption, (n.d.). /quickstats/pages/FIG-Hospital-EHR-Adoption.php (accessed October 24, 2018).
[5] J. King, V. Patel, E.W. Jamoom, and M.F. Furukawa, Clinical Benefits of Electronic Health Record Use: National Findings, *Health Services Research*. **49** (2014) 392–404. doi:10.1111/1475-6773.12135.
[6] M.A. Makary, and M. Daniel, Medical error—the third leading cause of death in the US, *BMJ*. **353** (2016) i2139. doi:10.1136/bmj.i2139.
[7] E. Fry, and F. Schulte, Death by a Thousand Clicks: Where Electronic Health Records Went Wrong, *Fortune*. (2019). http://fortune.com/longform/medical-records/ (accessed March 26, 2019).
[8] HIMSS, 2019 HIMSS U.S. Leadership and Workforce Survey, HIMSS, Chicago, IL, 2019. https://www.himss.org/sites/himssorg/files/u132196/2019_HIMSS_US_LEADERSHIP_WORKFORCE_SURVEY_Final_Report.pdf (accessed March 25, 2019).
[9] HIMSS, 2017 HIMSS U.S. Leadership and Workforce Survey, HIMSS, Chicago, IL, 2017. https://www.himss.org/library/2017-himss-leadership-and-workforce-survey (accessed November 18, 2018).
[10] HIMSS, 2018 HIMSS U.S. Leadership and Workforce Survey, HIMSS, Chicago, IL, 2018. https://www.himss.org/2018-himss-leadership-and-workforce-survey (accessed November 6, 2018).
[11] Clinical Informatics Becomes a Board-certified Medical Subspecialty Following ABMS Vote | AMIA, (n.d.). https://www.amia.org/news-and-publications/press-release/ci-is-subspecialty (accessed March 26, 2019).
[12] Informatics Nursing Certification (RN-BC), *ANA*. (n.d.). https://www.nursingworld.org/our-certifications/informatics-nurse/ (accessed March 26, 2019).
[13] H. Sood, K. NcNeil, and B. Keogh, Chief clinical information officers: clinical leadership for a digital age, *BMJ*. **358** (2017) j3295. doi:10.1136/bmj.j3295.

Address for correspondence

Susan H. Fenton, PhD,
UTHealth School of Biomedical Informatics
7000 Fannin, Suite 690
Houston, Texas 77030
USA
713 500 3591 tel
susan.h.fenton@uth.tmc.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190414

Layered Privacy Language Pseudonymization Extension for Health Care

Armin Gerl[a], Felix Bölz[a]

[a] *Chair of Distributed Information Systems, University of Passau, Passau, Bavaria, Germany*

Abstract

Enforcement of General Data Protection Regulation strengthens privacy in Europe and especially emphasizes protection of special categories of data as required in health care. Layered Privacy Language intends to model privacy policies to enforce them. Hereby, a special focus lays on the Policy-based De-identification process, which is based on anonymization and privacy models. Motivated by a health care scenario, this work shows pseudonymization capabilities are essential for health care. An overview of pseudonymization methods is given, showing a great variety of methods for different use cases. Therefore, a pseudonymization extension for Layered Privacy Language is introduced to define several pseudonymization methods. Furthermore, pseudonymization is added to Policy-based De-identification process of the overarching privacy framework of Layered Privacy Language. An example policy configuration is given demonstrating the introduced pseudonymization extension on the given health care example. The results are discussed, concluded, and future work is introduced.

Keywords:

Anonyms and pseudonyms, confidentiality, privacy

Introduction

The enforcement of the General Data Protection Regulation (GDPR) in May 25th 2018 strengthens privacy in Europe. The legal principles of Privacy by Design and Privacy by Default postulate that privacy has to be taken into consideration during design of all technical systems and default settings (e.g. privacy policy in an electronic system) have to be privacy friendly by default [Art. 25, 1]. Special categories of personal data, e.g. data concerning health or genetic data, are hereby especially protected and must not be processed unless special conditions are met like a legal basis or explicit consent [Art. 9, 1]. The Layered Privacy Language (LPL), including its overarching framework, intends to model and enforce privacy policies 'from consent to processing' [2]. Hereby, LPL models privacy policies, which are presented and possibly personalized by the individual Data Subject. If agreement and/or consent is given on the privacy policy, it will be pre-processed and stored. On a request of personal data for a specific purpose the Policy-based De-identification process is conducted. This determines if the requesting entity is authorized to request the data and, if necessary, anonymizes the data. Thus, anonymization techniques like generalization, suppression, and deletion are used to enforce privacy models like k-anonymity [3], l-diversity [4], or diffential privacy [5].

In this work, we introduce a scenario that demonstrates that privacy models solely are not suitable for health care. In health care, capabilities for pseudonymization have to be introduced in LPL, section *'Data Analysis in Health Care'*. Therefore, we give an overview of pseudonymization techniques in section *'Pseudonymization'*. In section *'Integration in LPL'* we describe extension of LPL, as well as integration of the pseudonymization techniques within existing Policy-based De-identification process. Section *'Health Care Scenario Results'* discusses extension of LPL according to the introduced scenario. In *'Discussion'* possibilities and limitations of the approach are shown. Lastly, the work is summarized and future work given in *'Conclusion'*.

Data Analysis in Health Care

In health care, patients' personal data is stored and processed for different purposes like billing, analysis, or research. Especially, information on patients' condition, e.g. blood sugar, symptoms, diseases and treatments, which fall under the special categories of personal data [Art. 9, 1], are stored alongside regular personal data like name, age or address, of patients. In the following, we assume a scenario in which a dataset is analysed to identify a possible epidemic. Here, influenza, commonly known as 'the flu', is divided into different types. *Type A* is generally responsible for large flu epidemics, *Type B* is less harmful, and *Type C* does not cause epidemics. There are additional sub-types, which are irrelevant for this example.

Table 1– Example health care dataset describing influenza types for patients. Attributes are assigned to privacy categories.

Name (*EI*)	**Age (*QI*)**	**Zip Code (*QI*)**	**Virus (*SD*)**
John	28	94032	Flu Type C
Max	25	94032	Flu Type C
Mary	22	94034	Flu Type A
Harry	20	94032	Flu Type B
Theresa	24	94034	Flu Type C

We assume a health care dataset like shown in Table 1, consisting of attributes of different privacy categories. In general, data is classified in four categories. *Explicit Identifiers (EI)*, attributes which identify a person uniquely. *Quasi Identifiers (QI)* which in combination identify a person. *Sensitive Data (SD)*, consisting of sensitive but not identifiable information about the record owner. *Non-Sensitive Data (NSD)*, data not in any of the categories above [6]. Based on this classification, which is also present in LPL, privacy models anonymize a dataset. We assume in this case that we want to protect the identity and sensitive information, which could also be maliciously utilized for de-identification [4]. Therefore, a

privacy model like l-diversity could be choosen and applied on the dataset.

The *EI* in l-diversity is anonymized by deletion (denoted by '*'). The *QI* and *SD* are anonymized, such that each group (age, zip code, and symptom) contains at least three identical entries. The values for attribute 'age' are generalized by decades, which leads to generalization of each value (e.g. '20' to '20 – 29'). For zip code, suppression is applied beginning with the last character, which is suitable for German zip codes. Lastly, attribute 'virus' is generalized from specific to more general description, e.g. 'Flu Type A' to 'Flu'.

The resulting anonymized dataset (see Table 2) may be interpreted as an epidemic in the general area of 9403* despite original data shows only one case of influenza *Type A* in area 94034, and one case of influenza *Type B* in 94032. Therefore, original data does not indicate epidemic outbreak but anonymized dataset might indicate outbreak of influenza (which type is unspecified). Although not shown explicitly in this example, *EI* is especially important in time series data in which same patient has to be identified over several records. This would be completely destroyed or falsified by anonymization techniques [6].

Table 2– De-identified data set with applied 3-diversity.

Name (*EI*)	Age (*QI*)	Zip Code (*QI*)	Virus (*SD*)
*	20 - 29	9403*	Flu
*	20 - 29	9403*	Flu
*	20 - 29	9403*	Flu
*	20 - 29	9403*	Flu
*	20 - 29	9403*	Flu

This leads to the conclusion that anonymization techniques solely are not sufficient. Pseudonymization techniques, on which we give a broad overview in the next section, are required and have to be explored for LPL.

Pseudonymization

In general, original dataset *D* is pseudonymized to *D'*. It consists of previously mentioned categories of data.

$$D = (EI, QI, SD, NSD)$$

In *D'* attributes are replaced with uniquely identifiable pseudonyms. Depending on use case, only *EI* have to be pseudonymized. Application of the different categories of data depends on use case. Typically, *EI* and *QI* are pseudonymized. However, *SD* and *NSD* may be pseudonymized.

$$D' = (EI', QI' SD', NSD')$$

In the following, overview of pseudonymization methods is presented.

Tokenization

Tokenization swaps distinct values with a token. Generation of the token can vary and may include a pseudo-random seed [7] or facilitate keys [8]. A distinction is made between dependent and independent tokenization. Dependent tokenization retains a relationship with original data in contrast to independent tokenization.

$$Independent: TokenGenerator \rightarrow Token$$

$$Dependent: TokenGenerator(Value) \rightarrow Token$$

As a result, pseudonyms based on independent tokens are more secure, because re-identification with given pseudonyms only is not possible (e.g. injection attacks) [9]. In general, generation of token is based on random seeds, cryptographic methods, or hashing [6]. An overview is given in the following.

Random Seeds

A pseudonym can be generated based on (pseudo-) random seeds. Here, it is essential to prevent token collisions. To gain more privacy, the random seed can be combined with secret keys. An example is the patient ID generator by K. Pommerening [7].

Cryptographic Methods

Encryption approaches are either symmetric or asymmetric. Symetric approaches utilize same key for encryption and decryption. For example, Heurix-Neubauer et al. [8] and Noumeir et al. [10] utilize DES and AES. Asymmetric approaches encrypt with public key and decrypt with private key. Rottondi et al. [11] presented pseudonymization based on the RSA. The keys must be stored securely to sustain feasibility of this approach [8]. To definitively prevent de-pseudonymization the private key can be dropped. In general, the disadvantage of crypthographic methods is increased computational cost. Additionally, generated token may not be of fixed length and might get quite excessive.

Hashing

A hash function usually maps an input set to a smaller target set. Thus, most hash functions are not injective. Contrary to cryptographic methods, generated tokens have same size regardless of entrys' length. However, due to the hash function's nature, different inputs may become the same pseudonym. Therefore, collision-resistant hash algorithms are preferred for pseudonymization. Furthermore, hashing methods can be classified into keyed and non-keyed hashing. Noumeir et al. [10] and Brekne et al. [9] utilize non-keyed MDX and SHA-X algorithms. The disadvantage of non-keyed hashing is that generated tokens can be linked between different datasets if same hashing function is used. Keyed methods extend cryptographic based hashes by a key, e.g. hash message authentication code (HMAC) [12]. Assuming the key remains secret, dictionary attacks or similar attacks are prevented.

Value-Preserving Techniques

There are some further methods for tokenization, which preserve particular operations on de-identified datasets increasing utility. These algorithms are developed for specific use cases. Examples are prefix-preserving IP address anonymization [13] and distance-preserving pseudonymization [14].

Implementation Patterns

To gain additional features, like increased degree of privacy or possibility for de-pseudonymization, tokenization can be combined with additional methods. Note that the following methods comparatively perform very weak if they are used solely in a security perspective, especially for small datasets. Thus, they should always be combined with other approaches [9,15].

A *bijective mapping* stores generated token and the original value.

$$Value \leftrightarrow Token$$

Thus, it allows authorized de-pseudonymization of the token. Therefore, storage has to be encrypted to avoid unauthorized de-pseudonymization [7]. When new data entries are added to

D, which have to be pseudonymized, then each value will be looked up in the encrypted mapping store and replaced by corresponding token. If no corresponding token is found new mapping is created and appended. Therefore, *D* with less distinct values is processed faster than *D* with various distinct value. Processing performance during de-pseudonymization depends on used data structure for storage.

Limited token generation mimics original values by exclusively calculating pseudonyms with same character set and same structure. As a result, privacy increases. For instance, an input date will also result structured as date after pseudonymization [6]. *Data permutation* swaps entries based on permutation function. For each new entry, changes have to be tracked by simple mapping or by updating the function. The pseudonymization can be reverted if permutation is known [9]. *Noise addition* is used to gain more privacy by adding a pseudo-random noise to input entries [15]. However, anomaly detection or similar techniques may counter noise. *Salting* adds entropy (or salt) to token generation process to increase computing costs of insider dictionary attacks and therefore reduces risk of de-pseudonymization [12]. The salt value can either be generated deterministically from a given entry as in PBKDF2 (hashing) [12], or as a random value as in RSA-OAEP modification (encryption) [11]. For equal values salt is reused.

Reasoning for Various Pseudonymization Methods

Pseudonymization method has to be selected based on intended usage. Here, several aspects have to be considered. Beginning with properties of the data value, required utility of the token, and usage of hashing and encryption methods. The requirement for later de-pseudonymization using bijective mappings or usage one-way pseudonymization for increased security should also be taken into consideration. Moreover, multiple techniques can be added to reach the desired level of privacy/security. Due to various possibilities for pseudonymization we concluded the requirement for a general specification of pseudonymization methods within LPL and its overarching privacy framework, which we discuss in the following.

Integration in LPL

Layered Privacy Language (LPL) is intended to model privacy policies combining both legal and computer science views on privacy. The original LPL [2] has been further extended by a *UI Extension* to support privacy icons for personal privacy policy user interface [16] and *Art. 12-14 GDPR Extension* [17]. Furthermore, LPL is accommodated by an overarching privacy framework that enables *Policy-based De-identification* utilitzing anonymization methods and privacy models. To facilitate various pseudoynmization methods within LPL, both policy and *Policy-based De-identification* process have to be extended. This is shown in the following.

Pseudonymization Extension for LPL

In the following, we describe *Pseudonymization Extension* for LPL based on original formalization by Gerl et al. [2]. Note that *UI Extension* [16] and *Art. 12-14 GDPR Extension* [17] are not considered as they do not interfere with integration of pseudonymization in LPL. The root element of LPL is the *LayeredPrivacyPolicy*-element *lpp*, which remains unchanged.

$$lpp = (version, name, lang, ppURI, upp, ds, P)$$

It consists of LPL version number, privacy policies name, language defined to display descriptions, link to legal privacy policy, *UnderlyingPrivacyPolicy*-element, *DataSource*-element and a set of *Purpose*-elements.

The *Purpose*-element *p* is extended allowing definition of pseudonymization methods on specific sets of data.

$$p = (name, optOut, required, descr, DR, r, pm, D, PSM)$$

Therefore, the purpose consists of its name, status flag specifying if user can opt-out or opt-in on the purpose, a flag whether the purpose is required to be accepted, description in defined language *lang*, set of *DataRecipient*-elements *dr*, *Retention*-element *r*, *PrivacyModel*-element *pm*, and a set of *Data*-elements *d*. The set of *PseudonymizationMethod psm* is appended, which will be further detailed. Note that it is also possible to assign no *psm*, if no pseudonymization is required for the purpose.

A *PseudonymizationMethod psm* represents one pseudonymization configuration, which will be applied on the dataset.

$$psm = (name, attrName, NOD, descr, header, PSMA)$$

It is a tuple with the following attributes:

- *name*: Defines pseudonymization approach. Consists of pre-defined set of available methods.
- *attrName*: Textual representation of name for newly created attribute, which holds pseudonyms.
- *NOD*: Set of *Name*Of*Data*-elements *nod*, which represent attributes to be pseudonymized. It refers the *name* attribute of the *Data*-element. It must at least consist of one valid *name*.
- *descr*: Human-readable description of pseudonymization in defined language *lang*.
- *header*: Human-readable header of pseudonymization in defined language *lang*.
- *PSMA*: Set of *PseudonymizationMethodAttribute*-elements *psma* describing further configurations of pseudonymization approach.

Each *PseudonymizationMethodAttribute*-element *psma* configures the defined approach:

$$psma = (key; value)$$

This key-value tuple defines attributes necessary for all possible methods.

LPL has been extended by adding *PseudonymizationMethod*-element *psm*, and *PseudonymizationMethodAttribute*-element *psma*, which allows definition of various pseudonymization methods (see Figure 1). In the current state, the extension supports various hashing approaches (SHA-X, MDX), keyed hashing (HMAC SHA-X), keyed hashing with entropies and mapping (PBKDF2 HMAC SHA-X), symmetric cryptography (AES, DES, 3DES, RC4, Blowfish), and random seeds with mapping. Further approaches can easily be added by defining them as *psm* and *psma* elements.

Secrets, which can lead to re-identification, are not stored in the LPL model. According to GDPR, these information must be stored separately [Recital 29, 1]. The concrete key and secret management belongs to the controller, entity which administrates stored data.

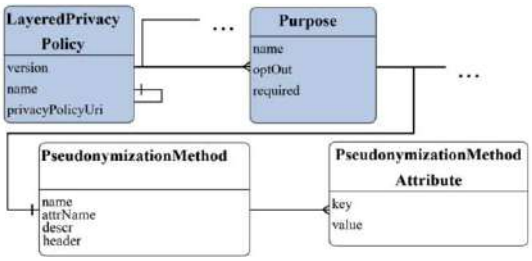


Figure 1– LPL structure extended by the pseudonymization elements. Further elements and attributes are omitted for the scope of this paper.

Policy-based De-identification

Policy-based De-identification based on LPL requires requesting entity to provide the purpose, a set of data , and set of data sources for the request. The requesting entity will then be authenticated. The purpose, entity and set of data will be authorized against properties of respective privacy policies *lpp* of the data sources. During *Minimum Anonymization* personalized privacy settings of each data source will be applied. Lastly, a common privacy model is derived from all related *lpp* and applied on the dataset.

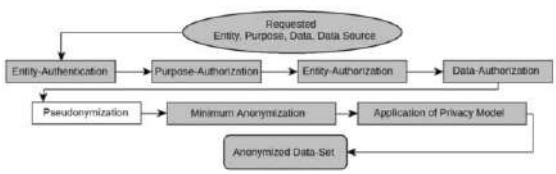


Figure 2– Modified Policy-based De-identification processes of the LPL framework when data is requested [2].

This process chain will be extended by adding the *Pseudonymization* process before *Minimum Anonymization* (see Figure 2). An inverted order, anonymization before pseudonymization, would result in less utility preservation, due to possible anonymization of values that should be preserved by tokenization with bijective mapping. Each pseudonymized *Data*-element will be altered after pseudonymization. For example, classification will be changed to *NDS* to avoid any further anonymization by later processes.

Health Care Scenario Results

With pseudonymization extension we can now de-identify the influenza dataset (see Table 1) and simultaneously add possibility of re-identification using bijective mappings. Patient names and virus type in the influenza example will both be stored as pseudonyms or IDs. For instance, if new insights regarding patients' health arise, IDs can be de-pseudonymized to deliver these new information to affected patients by gaining their identity and virus information.

The following configuration, using HMAC-SHA-1 algorithm and patient ID generator [7] combined with mappings, would achive such pseudonymization:

psm_0 = ("HMAC-SHA-1", "Name -ID", {"Name"}, "Description_0", "HMAC-SHA-1 pseudonymization", {})

psm_1 = ("PID", "Virus-ID", {"Virus"}, "Description_1", "PID pseudonymization", {})

It is noteworthy that the key, necessary for HMAC, and mapping are not defined within the LPL privacy policy to comply with GDPR [Recital 29, 1].

The results of the pseudonymization, applied on the raw data (see Table 1), defined by *psm_0* and *psm_1* are given in Table 3, which is accessible by the data analyist. Note that in Table 3 anonymization or privacy models are not applied. The mappings, necessary for re-identification, can be examined in Table 4, which should be secured and only be accessible by authorized personal. The extension also supports an AES encryption for mappings to increase security properties.

Table 3– De-identified dataset with pseudonymization. Pseudonyms shortened for better readability.

Name-ID (*NSD*)	Age (*QI*)	Zip Code (*QI*)	Virus-ID (*NSD*)
EA4B255	28	94032	KM93N2O
ADE0D85	25	94032	KM93N2O
2412F8F	22	94034	93I8M72
FF85768	20	94032	0O9INMW
71624DB	24	94034	KM93N2O

Table 4– Mappings which can restore content replaced by pseudonyms.

Name-ID	Name
EA4B255	John
ADE0D85	Max
2412F8F	Mary
FF85768	Harry
71624DB	Theresa

Virus-ID	Virus
93I8M72	Flu Type A
0O9INMW	Flu Type B
KM93N2O	Flu Type C

If we now compare anonymized data (see Table 2) and pseudonymized data (see Table 3), faulty conclusion cannot be made anymore. It can be observed that different types of viruses are present, which can uniquely be assigned to individuals. Still it can be seen that there is clustering. But with bijective mappings it can be shown that often appearing virus *KM93N2O* is the relatively harmless influenza *Type C*. Therefore, conclusion that there is no influenza epidemic, can be drawn. Furthermore, pseudonymization of SD can mitigate *attribute linkage* [4], identification by sensitive attribute. Due to pseudonymization, the attacker cannot infer privacy by identifying individuals by unique sensitive attributes or utilize external knowledge because provided information is not readable by the attacker.

Discussion

Based on verification of pseudonymization extension of LPL against introduced health care scenario, we conclude that anonymization methods and privacy models are insufficient for the given health care scenario. This is also valid for other anonymization techniques and privacy models. The application of pseudonymization methods tackles presented flaws by preserving essential data utility which would be lost otherwise. This has been exemplarily validated on the given scenario. To the extent of our knowledge, LPL is the first privacy language that models and enforces pseudonymization methods. Implementation is currently limited and in the state of work-in-progress. However, potentially various pseudonymization methods (which follow described configuration structure) may be added. For instance, asymmetric cryptographic functions and non-keyed hashing approaches with use of mappings for re-identification are appropriate additions. LPL certainly enables arbitrary combinations of pseudonymization and

anonymization. It is however important to choose suitable methods (pseudonymization, anonymization, privacy models) based on intended purpose, otherwise privacy may be broken. The selection of appropriate de-identification should be executed by an expert, e.g. trained data protection officer.

Conclusions

Privacy is an emerging topic which has to be considered within various fields. In health care it is especially necessary that data privacy is considered because processed data falls under the special categories of data. LPL intends to model privacy policies to inform *Data Subjects* and enforce privacy by design. In LPL, solely anonymization and privacy models were defined and used for de-identification. However, it was shown that pseudonymization is necessary for health care scenarios. *Pseudonymization Extension* of LPL offers possibility to define pseudonymization methods that have to be applied on the dataset. For example bijective mappings would allow an authorized de-identification, which is not possible for anonymization only techniques. Additionally, tokenization allows clear distinction of original values. *Policy-based De-identification* process joins pseudonymization, anonymization, and privacy models as a holistic approach. LPL and its overarching framework are continiously developed and researched. Evaluation of *Policy-based De-identification* process based on real health care data is sort after. Also, evaluation of LPL in other domains like *IoT*, *Cloud,* and *Mobility* are aimed for. The framework itself should be extended for fullfilment of *Data Subject Rights* to support *Data Protection Officers*. Suitable user interfaces for different domains and user groups (e.g. elderly or children) are developed to allow GDPR-compliant consent.

References

[1] Regulation (EU) 2016/679 of the European parliament and of the council (General Data Protection Regulation), European parliament and council, April 27, 2016.

[2] A. Gerl, N. Bennani, H. Kosch, and L. Brunie, LPL, towards a GDPR-compliant privacy language: formal definition and usage, In: *Transactions on Large-Scale Data-and Knowledge-Centered Systems XXXVII*, Springer, 2018, 41-80.

[3] P. Samarati and L. Sweeney, Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression, In: *Technical Report*, SRI International, 1998, 101-132.

[4] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, L-diversity: privacy beyond k-anonymity, *ACM Trans Knowl Discov Data* **1** (2007), 3.

[5] C. Dwork. Differential privacy: a survey of results, In: M. Agrawal, D. Du, Z. Duan, and A. Li (Eds), *Theory and Applications of Models of Computation*, Springer-Verlag, Berlin, Heidelberg, 2008, 1-19.

[6] N. Venkataramanan and A. Shriram, *Data Privacy: Principles and Practice*, Chapman and Hall/CRC, 2016.

[7] M. Lablans, A. Borg, and F. Ückert, A RESTful interface to pseudonymization services in modern web applications, *BMC Med Inform Decis Mak* **15** (2015), 2.

[8] J. Heurix and T. Neubauer, Privacy-preserving storage and access of medical data through pseudonymization and encryption, In: *International Conference on Trust, Privacy and Security in Digital Business*, Springer, Berlin, Heidelberg, 2011, 186-197.

[9] T. Brekne, A. Årnes, and A. Øslebø, Anonymization of ip traffic monitoring data: attacks on two prefix-preserving anonymization schemes and some proposed remedies, In: *International Workshop on Privacy Enhancing Technologies*, Springer, Berlin, Heidelberg, 2005, 179-196.

[10] R. Noumeir, A. Lemay, and J.M. Lina, Pseudonymization of radiology data for research purposes, *J Digit Imaging* **20** (2007), 284-295.

[11] C. Rottondi, G. Mauri, and G. Verticale, A data pseudonymization protocol for smart grids, In: *Online Conference on Green Communications (GreenCom), 2012*, IEEE, 2012, 68-73.

[12] H. Aamot, C.D. Kohl, D. Richter, and P. Knaup-Gregori, Pseudonymization of patient identifiers for translational research, *BMC Med Inform Decis Mak* **13** (2013), 75.

[13] J. Fan, J. Xu, M.H. Ammar, and S.B. Moon, Prefix-preserving IP address anonymization: measurement-based security evaluation and a new cryptography-based scheme, *Computer Networks* **46** (2004), 253-272.

[14] F. Kerschbaum, Distance-preserving pseudonymization for timestamps and spatial data, In: *Proceedings of the 2007 ACM workshop on Privacy in electronic society*, ACM, 2007, 68-71.

[15] Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Technique (WP216), Data Protection Working Party, April 10, 2014.

[16] A. Gerl, Extending layered privacy language to support privacy icons for a personal privacy policy user interface, In: *Proceedings of the 32nd International BCS Human Computer Interaction Conference*., Belfast, Ireland, 2018, 1-5.

[17] A. Gerl and D. Pohl, Critical analysis of LPL according to articles 12-14 of the GDPR, In: *Proceedings of the 13th International Conference on Availability, Reliability and Security*, ACM, Hamburg, Germany, 2018, 1-9.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190415

The Burden and Burnout in Documenting Patient Care: An Integrative Literature Review

Emily Gesner[a], Priscilla Gazarian[a], Patricia Dykes[b]

[a] *Department of Nursing, University of Massachusetts Boston, Boston, MA,*
[b] *Department of Medicine, Brigham and Women's Hospital, Boston, MA*

Abstract

The implementation of the electronic health record (EHR) across the globe has increased significantly in the last decade. Motivations for this trend include patient safety, regulatory requirements, and healthcare cost containment. However, the impact of regulatory requirements and new EHRs on clinicians has increased the incidence of documentation burden and may lead to burnout syndrome. It is important to understand the extent of documentation burden and potential solutions such as EHR user-interface redesign and the use of scribes to assist healthcare providers across the world.

Keywords:

Occupational Burnout, Electronic Health Records, Healthcare Providers

Introduction

The adoption of the electronic health record (EHR) in developed countries across the world has increased to almost 95% [1-3]. The motivation for health policies to implement the EHR is to improve the quality of care and reduce healthcare costs [4]. An example of a health policy that encourages EHR implementation is the Health Information Technology for Economic and Clinical Health (HITECH) Act in the United States (U.S.) [5]. One of the more significant programs within HITECH is the Promoting Interoperability (PI) Program which was formerly known as the Meaningful Use Program. The HITECH financial incentives and punitive measures were motivators for healthcare organizations to adopt EHR [5]. In addition, the World Health Organization (WHO) has encouraged the implementation of EHR in developing countries because of studies that show the improvement of patient outcomes [6].

The incidence of burnout syndrome in nurses in the U.S. ranges from 10% to 70% [7]. Also, more than half of physicians in the U.S. are reporting symptoms of burnout [8]. Studies have shown that the incidence of burnout syndrome is inversely proportional to patient outcomes [9]. Nursing and physicians are particularly vulnerable to burnout syndrome due to decreased job satisfaction, emotional, and physical exhaustion [10-12]. Clinicians report that they spend more than half of their day documenting patient care [13]. Studies also have found that when clinicians spend more time on clerical tasks, they have a decreased rate of job satisfaction which correlates to increased symptoms of burnout [11]. The purpose of this integrative review is to synthesize research evidence related to the extent of the documentation burden of healthcare providers and the potential solutions.

Methods

Using the strategies outlined by Whittemore & Knafl [14], four databases, (CINAHL, PubMed, ScienceDirect, and Web of Science) were searched as well as bibliographies of relevant articles and a broad internet search of the topic. The keywords used alone or in combination were "EHR documentation", "electronic documentation", "burden", "nursing", "nurse burnout", "physician burnout" and "clinician burnout", subject headings in CINAHL such as "burnout" were also used to broaden the search. All duplicate articles in the search were removed. Inclusion criteria included the following: Nursing and physician providers including Physician Assistants and Nurse Practitioners, EHR/electronic documentation of patient care, all healthcare settings. Due to the rapid change of technology and health policies throughout the world, we included research articles published between the years of 2013 and 2018. The types of studies included were experimental, quasi-experimental, descriptive, mixed-methods and qualitative. The exclusion criteria included any articles that focused on the process of transition from paper documentation to an EHR or from a legacy EHR to a new EHR system. Also excluded were non-physician/nursing providers (e.g. Chiropractors), news articles, articles not published in English and non-peer reviewed manuscripts.

Results

The initial search yielded 106 articles. After the inclusion/exclusion criteria were applied and duplicates removed, 26 articles remained. After the full-text articles were reviewed, a total of 10 articles [6; 11; 12; 15-21] were included in the final review (Figure 1). Common themes were identified from each article and tabulated (Table 1). A common theme identified in the articles is time pressure for documentation which relates to the increased stress clinicians experience when completing the time-consuming patient care documentation [11; 12; 15-20]. The main population found in the literature was physicians [6; 11; 12; 15-18; 20; 21]. Other populations identified in articles were advanced practice registered nurses (APRNs), ancillary care professionals in outpatient clinics and one article identified patients [6; 19; 21]. The solutions to documentation burnout identified were the use of clerical support or scribes to perform the documentation for clinicians [11; 12; 15-19], optimization of the EHR to enhance workflows [6; 12; 20] and additional training to improve documentation efficiency and quality [20].

Time Pressure for Documentation

All included articles [6; 11; 12; 15-21] discussed a possible link between the pressure or increased stress to document combined with the lack of time allocated for documentation and clinician burnout. Ehrenfeld & Wanderer [16] identified some evidence linking burnout to time pressure for documentation but not enough to characterize a strong link between the two. Most of the articles discussed the linkage between increased stress (e.g., documentation as one contributing factor to that stress) and the high rate of clinician burnout [11; 12; 15-19]. A common theme was that the requirements for data entry in the EHR took time away from patient care which led to poor job satisfaction [11; 12; 16-18]. Guo, Chen, & Mehta [20] discussed how decreasing "click-burden" or the amount of mouse clicks required in the EHR led to increased job satisfaction due to the decreased amount of time required to document.

Use of Clerical Support as a Solution

A proposed solution to increasing face-time and job satisfaction is the use of clerical support or "scribes" to perform the documentation for the clinician [11; 12; 15-19]. Four articles discussed how the increased amount of time required to document patient care could be relieved by simply assigning the task of documenting the care to another person who would observe the patient care along with the clinician [11; 15]. The use of scribes was identified only to assist physicians. Nurses were suggested as a possible staff member that could take the role of a scribe for the physicians [11; 16; 18]. There were no solutions offered in the articles related to assisting nurses with documentation burden.

Optimizing EHR Workflow through Education and Technical Enhancements

Jawhari et al. [6] discussed how lack of standardization increased documentation time which implied an increase in documentation burden. However, several articles [11; 12; 17; 22] suggested that education combined with technical enhancements in the EHR would improve the efficiency and quality of documentation. Robinson and Kersey [17] provided an educational intervention that improved the clinicians EHR workflow while also improving job satisfaction. Guo, Chen, & Mehta [20] provided a multi-faceted intervention which included technological options and EHR enhancements combined with a structured educational program. Studies that included interventions to improve the user-interface in combination with education showed significant improvement to an end user's documentation efficiency, quality of documentation and job satisfaction [17; 20].

Discussion

The main theme of the articles included in this review relates to how EHR documentation can contribute to burnout, and the potential solutions to mitigate this problem. Central to this theme was the discussion that increased documentation time takes away from providing direct patient care [6; 11; 12; 15-19; 22]. Solutions included the use of scribes, reconfiguration of the EHR, increased education about the EHR including education to improve clinician workflows, and utilizing technology in the EHR to assist with the completion of documentation.

Nurses at the Point of Patient Care

In terms of solutions for documentation burden, several articles discussed how nurses could be used as the scribes for the physicians [11; 16; 18] but this presents a greater problem [11]. Nurses must have the time to document their own care and would not have time to do documentation for the physicians. Nurses spend up to 60% of their shift documenting patient care and it is a considerable source of stress [10; 13; 23]. A recent study states that nurses have been shown to document up to 875 times per 12-hour shift which equates to documenting 1 data point per minute [24]. It is not feasible to add additional documentation responsibilities. Nurses are highly educated and over skilled for conducting only clerical work. Considering that the literature does suggest that higher amounts of required documentation increase the rate of burnout, it can be hypothesized that the rate of burnout syndrome in nursing would increase if their documentation requirements included documenting both nursing and medical care. The use of scribes may also affect the clinician-patient relationship by adding another person to listen and document during a patient's visit or in the patient's hospital room. Instead of adding more staff to fix the problem, the solutions may have to come from a modification of both regulatory requirements (e.g., lessen documentation requirements) and optimization of the EHR.

Emerging Technology to Decrease Burden

Another potential solution to the documentation burden presented in the literature was optimization of the EHR and education [20]. Emerging technology should be included in this category because of how innovations such as natural language processing (NLP) or artificial intelligence (AI) could address this growing problem (i.e., using NLP to integrate free text into clinical decision support). Though NLP has been used in various capacities in medicine, the application to clinical notes is still limited [24]. NLP could decrease documentation burden by allowing nurses or physicians to write a quick note which NLP would scan for information needed for clinical decision support (CDS) and convert unstructured data to structured data adequate for reporting to government agencies [25]. Artificial intelligence is currently being used to recognize progressions of disease in imaging but the possibilities to use AI in other capacities through machine learning to understand how the clinician documents and provide suggestions through algorithms or automate some documentation is conceivable [26; 27]. While AI or NLP may be possible interventions in the future, presently they may not be feasible for wide use due to software or financial constraints [28]. Based on this review, it is more feasible at this stage to optimize the EHR configuration to increase usability by decreasing the number of clicks and find low-technological ways to reduce documentation demand. Any reconfiguration of the EHR requires a well planned and structured education program for clinicians. The structured education program can provide communication regarding the changes to the EHR which may impact clinician workflows. The impacts may not be negative and provide a means to improve the quality and efficiency of documentation. The drive to implement the EHR and collect data has global implications which should be considered because it is these policies that have driven the EHR implementation for some of the major countries (e.g. China, France, Denmark, etc.) for the last decade [4].

International Health Policies surrounding EHR Documentation

The WHO identified that health outcomes and care efficiency are significantly impacted when an EHR is not implemented in a healthcare facility. They encouraged the implementation of the EHR in developed and developing countries. The motivation of the WHO to raise this topic on the global political agenda is to improve patient outcomes [22]. However, the downstream effect of this act may be that there is increased

pressure to perform more patient care documentation [11; 12; 15-20]. Health policies regarding the EHR and patient care documentation requirements vary. The articles reviewed did not discuss policies that may influence the type and amount of required patient documentation. The Commonwealth Fund profile on healthcare systems in nineteen countries discuss what policies influence the implementation of EHR documentation in each country but did not address the regulatory requirements surrounding documentation [4]. Studies evaluating international policies that encourage EHR implementation were not found during this literature review. However, that does not mean that they do not exist. It would be beneficial to analyze these policies for downstream effects especially as interoperability continues to expand into international space.

Conclusions

The studies included in this review have discussed the theme of documentation time pressure as a possible cause of clinician burnout syndrome. The possible solutions to relieve this pressure included the use of clerical support, EHR optimization and education. The focus of most of the studies to date involved physicians or APRNs. The evidence supports a link between documentation time pressure and burnout syndrome among all clinicians. The solution of using nurses as scribes may not be feasible due to their own documentation demands. It may be more reasonable to assess the EHR itself and possible technological innovations that can decrease documentation time burden. Despite any policies that may be in place across the globe, the problem of documentation burden is something that must be taken seriously as we continue to advance technology and monitor the incidence of clinician burnout syndrome.

Recommendations for Future Study

Future studies should include physicians and nurses in inpatient and outpatient settings. It should identify factors that can influence patient care documentation. It is important to discover if regulations and other health policies surrounding required documentation influence clinician burn out syndrome. Also, studies should include new technological interventions such as AI and NLP to understand if they make a positive or negative impact on the clinician's workflow and documentation time demands.

Tables

Table 1 – Articles reviewed

Author(s)/ year	Participants	Themes Identified	Proposed Solutions
Adnt, Beasley, Watkinson, & et al., (2017).	142 physicians in a single system in Wisconsin	Time required to use and document in the EMR; clerical burden	Use of Scribes Additional training for EHR documentation
Ehrenfeld & Wanderer (2018)	6 articles discussion the relationship between clinician burnout and the EHR	Clerical burden (e.g. burden of data entry)	Use of Scribes
Golob, Como, & Claridge (2015)	Trauma surgeons, orthopedic surgeons and neurosurgeons at a level 1 trauma center over the year of 2014	Time pressure for documentation	Clerical assistance for documentation (e.g. scribes), changes in workflow
Guo, Chen, & Mehta (2017)	Physicians at New York Presbyterian Hospital	EHR optimization	CDS (safety alerts) Use of EHR technology to optimize clinician workflow
Harris, Haskell, Cooper, Crouse, & Gardner (2018)	371 APRN's across Rhode Island	Time pressure for documentation	Use of clerical support staff (e.g. scribes).
Jawhari, J., Keenan, L., Zakus, D et al., (2016)	Outpatient clinic staff (physicians, nurses, etc.) in clinic in Kibera	Lack of EHR optimization	Optimize EHR; Providing training
Linzer, Poplau, Bobbit, et al., (2016)	Inpatient and outpatient physicians, nurse practitioners	Shorter visits, documentation time pressure, insufficient support staff	Use of Scribes IT Support and training
McMurray, Hicks, et al., (2013)	Outpatient clinic staff and 3 patients	Lack of EHR interoperability; Documentation burden	EHR optimization

Author(s)/ year	Participants	Themes Identified	Proposed Solutions
Robinson, Kersey (2018)	3500 physicians from 35 specialties	Lack of EHR optimization; education to improve documentation	Provide training for documentation best practice Optimize EHR workflow
Shanafelt, Dyrbye,Sinskey, et al, (2016)	Physicians in the U.S. across all specialties and settings	Time required to use and document in an EMR; clerical burden (e.g. data entry requirements)	Use Scribes Optimize EHR workflow Increase staff training

Figures and Graphs

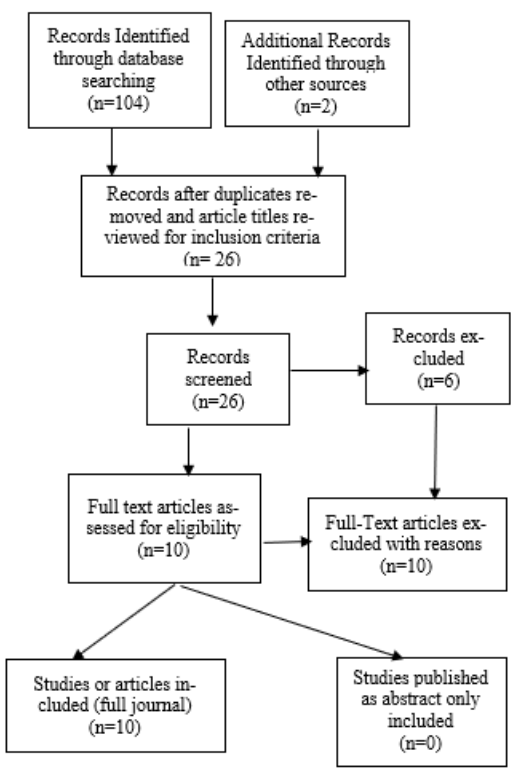


Figure 1- Documentation Search Strategy

References

[1] O.A.H.A. (AHA), AHA Annual Survey Information Technology Supplement, in, 2015.

[2] C. Schoen, R. Osborn, D. Squires, M. Doty, P. Rasmussen, R. Pierson, and S. Applebaum, A survey of primary care doctors in ten countries shows progress in use of health information technology, less in other areas, *Health Aff (Millwood)* **31** (2012), 2805-2816.

[3] H. Landi, Survey: Infrastructure, Interoperability Key Barriers to Global HIT Development, in: *Healthcare Innovation*, 2018.

[4] C. Fund, International Profiles of Healthcare systems, in, 2017.

[5] V. Palabindala, A. Pamarthy, and N.R. Jonnalagadda, Adoption of electronic health records and barriers, *J Community Hosp Intern Med Perspect* **6** (2016), 32643.

[6] B. Jawhari, L. Keenan, D. Zakus, D. Ludwick, A. Isaac, A. Saleh, and R. Hayward, Barriers and facilitators to Electronic Medical Record (EMR) use in an urban slum, *Int J Med Inform* **94** (2016), 246-254.

[7] A. Lyndon, Burnout Among Health Professionals and Its Effect on Patient Safety, *Perspectives on Safety* (2016).

[8] S. Kumar, Burnout and Doctors: Prevalence, Prevention and Intervention, *Healthcare (Basel)* **4** (2016).

[9] L. Poghosyan, S.P. Clarke, M. Finlayson, and L.H. Aiken, Nurse burnout and quality of care: cross-national investigation in six countries, *Res Nurs Health* **33** (2010), 288-298.

[10] N. Khamisa, K. Peltzer, and B. Oldenburg, Burnout in relation to specific contributing factors and health outcomes among nurses: a systematic review, *Int J Environ Res Public Health* **10** (2013), 2214-2240.

[11] T.D. Shanafelt, L.N. Dyrbye, C. Sinsky, O. Hasan, D. Satele, J. Sloan, and C.P. West, Relationship Between Clerical Burden and Characteristics of the Electronic Environment With Physician Burnout and Professional Satisfaction, *Mayo Clin Proc* **91** (2016), 836-848.

[12] B.G. Arndt, J.W. Beasley, M.D. Watkinson, J.L. Temte, W.J. Tuan, C.A. Sinsky, and V.J. Gilchrist, Tethered to the EHR: Primary Care Physician Workload Assessment Using EHR Event Log Data and Time-Motion Observations, *Ann Fam Med* **15** (2017), 419-426.

[13] W.J. Duffy, M.S. Kharasch, and H. Du, Point of care documentation impact on the nurse-patient interaction, *Nurs Adm Q* **34** (2010), E1-E10.

[14] R. Whittemore and K. Knafl, The integrative review: updated methodology, *J Adv Nurs* **52** (2005), 546-553.

[15] M. Linzer, S. Poplau, S. Babbott, T. Collins, L. Guzman-Corrales, J. Menk, M.L. Murphy, and K. Ovington, Worklife and Wellness in Academic General Internal Medicine: Results from a National Survey, *J Gen Intern Med* **31** (2016), 1004-1010.

[16] J.M. Ehrenfeld and J.P. Wanderer, Technology as friend or foe? Do electronic health records increase burnout?, *Curr Opin Anaesthesiol* **31** (2018), 357-360.

[17] K.E. Robinson and J.A. Kersey, Novel electronic health record (EHR) education intervention in large healthcare organization improves quality, efficiency, time, and impact on burnout, *Medicine (Baltimore)* **97** (2018), e12319.

[18] J.F. Golob, Jr., J.J. Como, and J.A. Claridge, The painful truth: The documentation burden of a trauma surgeon, *J Trauma Acute Care Surg* **80** (2016), 742-745; discussion 745-747.

[19] D.A. Harris, J. Haskell, E. Cooper, N. Crouse, and R. Gardner, Estimating the association between burnout and electronic health record-related stress among advanced

practice registered nurses, *Appl Nurs Res* **43** (2018), 36-41.
[20] U. Guo, L. Chen, and P.H. Mehta, Electronic health record innovations: Helping physicians - One less click at a time, *Health Inf Manag* **46** (2017), 140-144.
[21] J. McMurray, E. Hicks, H. Johnson, J. Elliott, K. Byrne, and P. Stolee, 'Trying to find information is like hating yourself every day': the collision of electronic information systems in transition with patients in transition, *Health Informatics J* **19** (2013), 218-232.
[22] R.C. Hermann, S. Mattke, D. Somekh, H. Silfverhielm, E. Goldner, G. Glover, J. Pirkis, J. Mainz, and J.A. Chan, Quality indicators for international benchmarking of mental health care, *Int J Qual Health Care* **18 Suppl 1** (2006), 31-38.
[23] C.P. Chang, T.T. Lee, C.H. Liu, and M.E. Mills, Nurses' Experiences of an Initial and Reimplemented Electronic Health Record Use, *Comput Inform Nurs* **34** (2016), 183-190.
[24] S. Collins, B. Couture, M.J. Kang, P. Dykes, K. Schnock, C. Knaplund, F. Chang, and K. Cato, Quantifying and Visualizing Nursing Flowsheet Documentation Burden in Acute and Critical Care, *AMIA Annu Symp Proc* **2018** (2018), 348-357.
[25] J.D. Osborne, M. Wyatt, A.O. Westfall, J. Willig, S. Bethard, and G. Gordon, Efficient identification of nationally mandated reportable cancer cases using natural language processing and machine learning, *J Am Med Inform Assoc* **23** (2016), 1077-1084.
[26] M.I. Fazal, M.E. Patel, J. Tye, and Y. Gupta, The past, present and future role of artificial intelligence in imaging, *Eur J Radiol* **105** (2018), 246-250.
[27] A. Atutxa, A. Perez, A. Casillas, A. Atutxa, A. Perez, and A. Casillas, Machine Learning Approaches on Diagnostic Term Encoding With the ICD for Clinical Documentation, *IEEE J Biomed Health Inform* **22** (2018), 1323-1329.
[28] F. FitzHenry, H.J. Murff, M.E. Matheny, D. Aronsky, N. Gentry, R. Reeves, E.M. Fielstein, V.P. Messina, and T. Speroff, A Natural Language Processing (NLP) Case Study: The Challenges and Opportunities with Large-Scale NLP of the Electronic Health Record, *Journal of Healthcare Information Management* **29** (2015), 28-35.

Address for Correspondence

Emily Gesner DNP, RN-BC
University of Massachusetts Boston
College of Nursing and Health Sciences
100 William T, Morrissey Blvd, Boston, MA 02125
Emily.Gesner001@umb.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190416

Experts Views on the Use of Mobile Devices to Support Patients with Mild Learning Disabilities During Clinical Consultations

Ryan Colin Gibson[a], Matt-Mouley Bouamrane[b], Mark Dunlop[a]

[a] *Department of Computer & Information Sciences, University of Strathclyde, Glasgow, United Kingdom*
[b] *Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Scotland*

Abstract

Due to several factors including time and budget constraints, General Practitioners (GPs) are often under-trained on the communication needs of patients with learning disabilities (LDs). As such, they may find it difficult to extract accurate information from these patients. Digital technologies have the potential to alleviate communication barriers, yet their use in this context remains vastly unexplored. Hence, we conducted 2 focus groups with 12 experts in LDs to investigate how tablet applications may be used to promote the information exchange process between GPs and patients with mild LDs. The experts identified an initial set of design criteria for the future implementation of these technologies and were enthusiastic about the potential impact they may have on primary care. In addition, they also discussed a potential model for extracting medical information from this population, which focused on breaking the overall consultation down into smaller, less cognitively challenging segments.

Keywords:

Mobile Applications; Intellectual Disability; Primary Health Care

Introduction

People with learning disabilities (LDs) are more susceptible to a range of conditions and comorbidities [1] and therefore have a higher demand for healthcare services compared to that of the general population. Despite this, the standard of care being provided is often inadequate [1] and this has a detrimental impact on both the length and quality of their lives.

To determine the overall scale of the problem, researchers at the University of Bristol conducted an inquiry into the premature deaths of people with LDs [1]. They examined the cause of death of 247 patients with LDs across 5 primary care trusts in the South West of England and found that approximately 50% were avoidable. Of these deaths, 27.5% were directly amenable to better care and this suggests that such patients are being subjected to serious health inequalities.

Previous studies [1–5] have investigated the various barriers to providing primary and secondary care for people with LDs, some of which may contribute to the findings made by Heslop et al. [1]. This literature covers a span of 2 decades, and with a number of barriers appearing consistently throughout, it is clear that effective support for this population has not been identified. Some of these obstacles include: difficulties identifying and accessing appropriate services; under-trained staff on the health and communication needs of patients with LDs; inflexible procedures; and insufficient collation and use of health care data.

Central to many of the identified barriers is communication. In primary care, this is extremely problematic since Howells suggests that "*the art of general practice lies in the ability to communicate with patients*" [6].

Nevertheless, patients with LDs have a number of impairments that affect their ability to convey medical information [2,3,7,8]. In addition, General Practitioners (GPs) often lack the skills required to adjust their consultation methods to limit the effect these impairments may have on the appointment [3,5].

Consequently, the overall goal of our research is to investigate the use of Alternative and Augmentative Communication (AAC) applications to promote the exchange of information between GPs and patients with mild LDs. AAC technologies are used to enhance an individual with disabilities capacity to communicate by offering those who cannot speak a platform to convey their needs (alternative), or by supplementing the vocabulary of those who can (augmentative). This contrasts with traditional information applications, which often treat accessibility as an afterthought.

Throughout the paper, we will present the results of an exploratory study in which 12 experts discussed how tablet AAC applications can improve consultations involving patients with mild LDs. The requirements listed will assist in the future development of medical AAC applications that target the needs of these stakeholders.

Background

In this section, we define the term "mild learning disability" and introduce some of the impairments common to this population that may have an adverse effect on the consultation process. We then discuss the available guidelines on how to communicate effectively with patients who have LDs, before giving an overview on the current use of digital technologies to promote the health of these patients.

Mild Learning Disability Characteristics

An individual may be diagnosed with a learning disability if they satisfy the following 3 criteria: their intellectual functioning is impaired; their social functioning is impaired; and the aforementioned conditions occur before adulthood [9]. LDs typically manifest across a scale ranging from mild to severe; however, those with mild LDs are generally able to communicate their everyday needs but may struggle with more complex concepts such as describing symptoms. A number of impairments tend to coexist with LDs that affect an individual's capacity to communicate their medical needs.

These include: cognitive impairments that affect vocabulary and sentence formulation skills, meaning the patient may not

possess the language required to accurately describe symptoms; reduced receptive skills that may affect their ability to understand the GP; limitations in their abstract thinking and long-term memory which may affect their ability to provide an accurate medical history; and a restricted knowledge of the human body, meaning they may not even recognise the presence of symptoms [2,3,7,8].

Guidelines in Consulting with Patients who have LDs

National and international guidelines e.g. [10] have been developed to assist medical professionals in conducting consultations with this population. Much of the advice regarding communication focuses on carrying out reasonable adjustments to cater to the individual needs of patients [10]. Some of the key recommendations include: extracting information directly from the patient; establishing the patient's preferred method of communication as early as possible e.g. by reviewing a clinical passport [11] if available; targeting a range of communication modalities based on the needs and preferences of the individual; and avoiding the use of medical jargon. GPs should also consider: utilizing gestures to emphasize communication; being vigilant for any additional information conveyed by the patient's body language; making sure the person has understood the information they have received; providing additional time for the patient to consider any information conveyed; and supplying information in advance of the consultation to help the patient prepare for the appointment.

Existing Health Applications for People with LDs

Researchers in the past have explored the use of digital technologies in a number of areas of health including: dentistry [12]; psychiatry [7]; and patient profiling [13]. Once again, this literature highlights the importance of exchanging information in a manner suited to the patient's individual needs. In particular, Menzies et al. recognized that the sole use of speech was not sufficient in conveying dental information to patients with cognitive disabilities [12]. Instead, they found that imagery/videos were particularly effective in describing the potential procedures to be carried out and the tools used within them. Furthermore, the professionals involved in this study requested features that assist in determining the patient's preferred method of communicating the terms "yes", "no" and "stop" – three aspects deemed crucial to their care. Prior et al's. study explored this functionality in further depth [13]. They developed a digital aid that extracts vital information from the patient (such as their communication needs, allergies etc.) prior to treatment. This information may then assist medical professionals in providing improved care, since they will be able to utilize the best practices when interacting with a patient. Bostrom & Eriksson investigated the possibility of providing healthcare data in advance of appointments [7]. They found that questionnaires could be successful in highlighting potential psychiatric conditions providing the information presented is accessible to stakeholders.

Methods

To determine the feasibility of embedding AAC applications in primary care, and to identify initial requirements that cater for the needs of patients with mild LDs, we conducted 2 focus groups with 12 experts in LDs (found in Table 1). We recruited experts in this study, as opposed to GPs, since they have extensive knowledge about the needs of people with LDs – a characteristic often not found in traditional medical professionals [5].

The LD nurses also understood the procedures involved in the consultation process, meaning the experts were better suited to identify how the proposed technology can support such patients. The set of features discussed will be expanded on during future studies that incorporate the views of both adults with mild LDs and GPs.

Table 1 – Expert Demographics

Expert IDs	Profession	Sex
1.1, 1.2, 1.3	Academics in the health and well-being of people with LDs.	F,F,F
1.4	Employee of an advocacy charity for people with LDs. Has mild LDs.	F
1.5	Employee of an advocacy charity for people with LDs.	F
1.6	Former LD nurse. Manager of a resource center for people with LDs.	F
1.7	Digital Inclusion Assistant.	M
2.1, 2.3, 2.5	Community LD nurses.	F,F,F
2.2	Employee of an advocacy charity.	F
2.4	Employment support officer	F

The focus groups were designed to achieve 2 goals: (1) improve the accessibility of co-design techniques that may be employed within future workshops; and (2) identify an initial set of features for the development of the application. This paper will primarily focus on the results pertaining to goal 2. All 12 participants were required to complete the 4 activities shown in Figure 1 - the details of which have been described in the "Data Collection" subsection. These activities were identified during a review of previous literature that aimed to explore the use of co-design processes with participants who have LDs. They were selected to address 3 specific aspects of the proposed application: appropriate imagery to capture medical symptoms; its overall functionality; and the design of the interface including the layout of each screen.

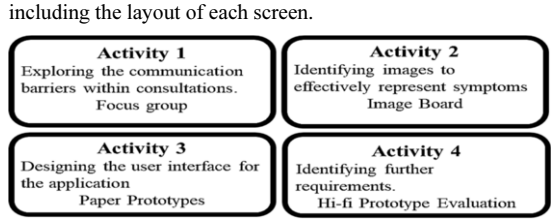


Figure 1 – The 4 Co-design Tasks Presented to the Experts

Invitations to participate in the study were issued (via email & telephone) to various charities, universities and hospitals throughout Scotland in May 2018. Seven experts from the city of Glasgow and 5 from Dundee consented to take part and formed focus groups 1 and 2 respectively. The focus groups were carried out in June 2018. Ethical approval to conduct this study was obtained from the Department of Computer and Information Sciences Ethics Committee at the University of Strathclyde.

Data Collection

The first task completed by the experts was a focus group that aimed to explore the primary barriers to effective health care for patients with LDs, as well as the potential use of digital technologies in mitigating these barriers.

The questions presented focused on the following 4 themes: (1) preparing for an appointment; (2) positive and negative encounters with GPs; (3) aptitude in using touch screen

technologies; and (4) how technology may be used to support the patient throughout the consultation.

Open-ended questions were primarily used to promote discussion and the session was conducted on a semi-structured basis to ensure the experts were able to raise, and expand upon, topics unforeseen by the authors [14].

The second task involved employing the image board methodology [15] to identify appropriate pictures to be included within the application. The experts were required to review images that depict common symptoms experienced by people with LDs and then separate these into one of two categories: those that accurately capture the condition; and those whose meaning is more obscure. A discussion then occurred as to why some images were more effective in capturing this information than others. Each symptom was portrayed using 3 separate styles of images - photorealistic, cartoon drawings, and simplistic black and white drawings to determine the style best suited to people with mild LDs. These styles were selected since they are often used in health-related resources for people with LDs.

The penultimate task consisted of a basic paper prototyping process. This involved placing/drawing elements onto a paper representation of a tablet based on the experts' views of the functionality and layout of each screen.

The fourth task involved the evaluation of a previously developed tablet application to try and discern the requirements that were not identified during task 3. This process was modelled around a "think-aloud" [16] session where the participants were required to complete 2 exercises and describe their reasons behind the actions being performed during real-time.

Data Analysis

The focus groups were recorded with participant consent and transcribed verbatim. The transcriptions were then subjected to a framework analysis to ensure a structured summary of the key features/requirements discussed was obtained. An initial thematic framework was developed by the first author based on the themes that emerged from a previous scoping review of the technologies used to support patients with mild LDs during clinical consultations.

The transcripts did not conform entirely to this framework and further codes were created to address this issue, at which point similar codes were grouped together to form overarching themes. The framework was reviewed by the second author and any discrepancies were resolved by the third author. The first author then tagged the transcriptions using the final framework and the relevant excerpts were transferred to their appropriate positions in the framework analysis table made available via the following doi: 10.15129/76f97730-a5fa-49da-973f-995373cee7ad. The requirements presented in the next section are based upon the main themes/sub-themes that emerged during this process.

Results

In this section, we present the key requirements identified by the experts. The quotes used to support these features are referred to using the participant ID listed in Table 1.

Simplifying the Consultation Process

The experts were of the opinion that the consultation process is often too complex for people with mild LDs.

Patients generally have to contemplate or provide information on aspects that are difficult to understand and must achieve this using methods that may be unsuited to their needs.

Consequently, the experts suggested developing technologies that help to break this process down into manageable sections, as discussed by participant 2.5:

"Could you not have something like that for the parts of the body - saying what part of the body the pain is in first of all? Once you've narrowed it down, have a different set of cards to say what type of pain is it? Is it hot pain? Does it [feel] cold? Is it sharp like a needle or something?

The participants in focus group 2 also recognised that the application should explore conditions that do not involve pain: participant 2.5:

"I suppose the problem is if [you] start with body parts and then go on to what's wrong with that body part, general symptoms of tiredness [for example] wouldn't be [picked up]. Do you know what I mean cause they might just feel totally drained all the time."

In summary, the experts suggested a potential model that may be utilized by GPs to explore the health of patients with LDs, as shown in Figure 2.

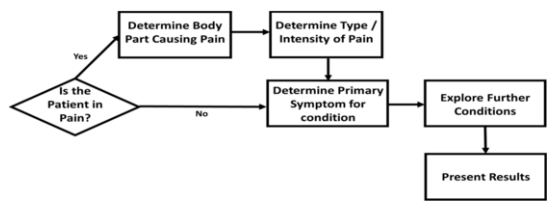


Figure 2 – Suggested Model for Diagnosing Patients with LDs

Utilizing Appropriate Modalities

The experts discussed two key strategies that may assist in promoting the accessibility of medical information, the first of which centres on the language used and the second focuses on the incorporation of images.

Comprehensible Language

Whilst describing appropriate language to be embedded within the application, the experts frequently cited common accessibility guidelines. This included: utilizing plain and simple language, along with short sentences that focus on solitary ideas; employing a minimum font size of 12; and offering the ability to playback textual information. In addition, the use of concrete examples was emphasized by participant 1.6 when describing particularly complex concepts. Finally, the experts in focus group 2 revealed that closed questions are most effective in extracting information from this population.

Identifiable Imagery

All experts throughout both focus groups emphasized the importance imagery may have in conveying medical information. Nevertheless, they were unable to agree upon the style of image that will be most effective in achieving this. For example, the participants in focus group 1 found that the more photorealistic images managed to capture the symptoms accurately, as described by participant 1.3:

"*I thought this tired one was quite good it was quite realistic - better than the sort of drawing of someone lying in their bed. I suppose that's a bit more cartoony, I think I prefer the actual person."*

In comparison, the experts in focus group 2 advocated for the use of the more simplistic black and white drawings, as discussed by participant 2.3:

"I prefer the egg head kind of ones 'cause they're not male or female. You know you might get a female with autism who's like that's not me 'cause [the picture is of a man]…And also,

less colour - just the black and white (colours) I think is more effective."

These excerpts suggest that a range of needs will have to be catered for by the images implemented within the application and this matches the views of participant 1.6:

"*It's quite difficult because when you think of people, some will really connect with some of them [the pictures] and some individuals will connect with others*."

Combining Modalities

In addition, the experts in focus group 1 revealed that the combination of text and images provided the most complete and accurate description of the symptoms presented as discussed by participant 1.4:

"You have headache at the bottom and I think if it didn't have headache at the bottom it would be quite confusing 'cause it could [mean something else]. So I think it's good with the headache heading."

Identifying Most Appropriate Communication Strategy

Participant 2.3 also discussed the benefits of using the application to identify the communication needs of the patient:

"My sister is a radiographer and sometimes there will be a little footnote somewhere [suggesting] some sort of learning disability and she's like "okay that's good to know but I want [more information]. You know, avoid saying this or use this approach".

This has the potential to increase communication significantly and matches the process described by both Menzies et al. [12] and Prior et al. [13].

Guiding the Patient

The experts in focus group 2 discussed two common scenarios that generate a heavy burden on healthcare services. Firstly, participant 2.4 suggested that some patients book medical appointments for the social experience as opposed to actually requiring treatment:

"*So [sometimes] they use health professionals inappropriately. You know, they make appointments with the doctor and they don't have any symptoms, they just want to talk to somebody. The doctor won't find the symptom cause there's not one there*."

The second involves patients prematurely booking appointments for conditions that have just occurred and will heal in due course, as discussed by participant 2.5:

"*For some of our clients, I don't see any point in [them] going to the GP. Sometimes it's something that's just happened and we expect it to be like that so [they shouldn't] go to the doctor."*

To overcome these issues, the experts discussed implementing a feature that makes use of the extracted information to suggest a course of action for the patient, as explained by participant 2.5:

"*Whether you can have solutions at the end to say well how long have you had a headache for? Right, try [taking] paracetamol or try drinking some water or a lie down or something. You know go and tell your care worker or your family first of all, so it could almost be like a filter*."

Consequently, the application could be used in the patient's home, before directing the individual to treatments out with primary care for minor ailments such as short-term headaches.

However, the app may also suggest that the individual contacts a medical professional, at which point the extracted information can be embedded in the consultation process.

Further Features

This subsection describes those features that were deemed to be important but do not fit into the previous 2 themes introduced.

In addition to presenting closed questions, participant 1.2 revealed that the amount of choice available should be limited, preferably to 2 options:

"I think as much as possible if you could have yes/no questions or like a tick and a cross to say is it painful [for example]. I think they might struggle if there's too many options."

The experts were also concerned about the user possessing the attentiveness required to complete the questionnaire, as discussed by participant 2.5:

"Even if they put down symptoms in different parts of the body and they gave up - if they take that to the GP, they could see some of things going on."

As such, they discussed the need to record the patient's progress to be completed at a later date or subsequently presented to the GP for review.

Discussion

Prior research has shown that digital technologies have the potential to increase the health of adults with LDs [7,12,13]. We add to this body of literature by highlighting the positive impact AAC applications may have within consultations involving this population. The experts were particularly enthusiastic about the technologies ability to support GPs in implementing many of the communication guidelines discussed in the "Background" section [10].

Previous research has explored extracting medical [7] or personal [11,13] information from the patient in advance of the appointment; however, the experts suggested that an application that combines both of these strategies should lead to optimal communication. Extracting medical information will enable practitioners to shape the questions to be presented, thus affording them more time to focus on aspects that may be crucial to a diagnosis. Furthermore, the patient may have more time to deliberate the questions being asked and subsequently construct an appropriate response. Obtaining personal information will enable the GP to utilize the strategies most suited to the patient's needs, which may ultimately increase their comprehension of the data being presented.

In accordance with the findings of previous literature [7,12,13,17,18], the experts highlighted the importance of combining images with accessible language to convey medical information. Nevertheless, they were unable to agree upon the style of image that captures this information best and instead revealed that a wide range of preferences must be catered for to meet the needs of people with mild LDs. Two strategies could be used to achieve this. First, several sets of images may be developed with the option to dynamically change between these sets e.g. when a user is unsure of the meaning conveyed by a particular image. However, this process may be cognitively challenging for people with LDs. As such, the second option involves the user completing an initial questionnaire that determines the most effective style of image to be embedded in the system, based on the individual's needs. Furthermore, the experts suggested that the application could assist in limiting the amount of unnecessary appointments attended by the patient.

This problem is also common throughout the general population, yet there is evidence to suggest that a higher percentage of people with LDs live with undiagnosed conditions e.g. [19]. Consequently, it is more important for these patients to seek medical care since more serious conditions may be the source of their current problem. The application can assist in this process by exploring all potential causes for the symptoms extracted, before suggesting a course of action.

A plethora of guidelines are available e.g. [10] to assist practitioners in conducting consultations with patients who have LDs, yet little research has been conducted into the specific questions to ask such patients. The experts discussed a potential model to achieve this by breaking down what is essentially a difficult process into more manageable parts. This process is shown in Figure 2 and consists of deducing whether the patient is pain; extracting the primary symptom causing their condition; and finally exploring any additional symptoms that may be present.

Conclusions

In this paper, we have presented one of the first studies to explore the potential use of tablet AAC applications to support patients with mild LDs during clinical consultations. Twelve experts in LDs participated in 2 focus groups throughout Scotland and subsequently identified a set of design criteria for the future development of such technologies. Developers will therefore be able to consider a variety of complex needs required by people with LDs and this criteria may be expanded on during future research with target stakeholders. In addition, this process has resulted in a potential model that may be utilized by GPs to extract symptoms from patients with mild LDs.

Acknowledgements

This research has been supported by the University of Strathclyde's Student Excellence Awards scheme.

References

[1] P. Heslop, P. Blair, P. Fleming, M. Hoghton, A. Marriott, and L. Russ, Confidential Inquiry into premature deaths of people with learning disabilities, (2012), http://www.webcitation.org/77EQhVVwP (accessed March 29, 2019).

[2] G.L. Krahn, L. Hammond, and A. Turner, A cascade of disparities: health and health care access for people with intellectual disabilities, *Ment Retard Dev Disabil Res Rev* **12** (2006) 70–82.

[3] N.J. Simpson, Bridging primary and secondary care for people with learning disabilities, *Advances in Psychiatric Treatment* **1** (1995) 207–213.

[4] A. Phillips, J. Morrison, and R.W. Davis, General practitioners' educational needs in intellectual disability health, *J Intellect Disabil Res* **48** (2004) 142–149.

[5] A. Ali, K. Scior, V. Ratti, A. Strydom, M. King, and A. Hassiotis, Discrimination and other barriers to accessing health care: perspectives of patients with mild and moderate intellectual disability and their carers, *PLoS ONE* **8** (2013) e70855.

[6] K. Dodd, and J. Brunker, 'Feeling Poorly': Report of a Pilot Study Aimed to Increase the Ability of People with Learning Disabilities to Understand and Communicate About Physical Illness, *British Journal of Learning Disabilities* **27** (1999) 10–15.

[7] P. Boström and E. Eriksson, Design for Self-reporting Psychological Health in Children with Intellectual Disabilities, in: Proceedings of the 14th International Conference on Interaction Design and Children, ACM, New York, NY, USA, 2015: pp. 279–282.

[8] J. Murphy, Perceptions of communication between people with communication disability and general practice staff, *Health Expect* **9** (2006) 49–59.

[9] WHO, Definition: intellectual disability, (2018), http://www.webcitation.org/77EQPqde1 (Accessed March 29, 2019).

[10] GAIN, Guidelines on Caring For People with a Learning Disability in General Hospital Settings, (2010), https://rqia.org.uk/RQIA/files/81/81662c46-b7bb-43a5-9496-a7f2d919c2a3.pdf (accessed November 8, 2018).

[11] R. Northway, S. Rees, M. Davies, and S. Williams, Hospital passports, patient safety and person-centred care: A review of documents currently used for people with intellectual disabilities in the UK, *J Clin Nurs* **26** (2017) 5160–5168.

[12] R. Menzies, D. Herron, L. Scott, R. Freeman, and A. Waller, Involving Clinical Staff in the Design of a Support Tool Improve Dental Communication for Patients with Intellectual Disabilities, in: Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility, ACM, New York, NY, USA, 2013: pp. 55:1–55:2.

[13] S. Prior, HCI Methods for Including Adults with Disabilities in the Design of CHAMPION, in: CHI '10 Extended Abstracts on Human Factors in Computing Systems, ACM, New York, NY, USA, 2010: pp. 2891–2894.

[14] B. Kaplan, and J.A. Maxwell, Qualitative Research Methods for Evaluating Computer Information Systems, in: J.G. Anderson, and C.E. Aydin (Eds.), Evaluating the Organizational Impact of Healthcare Information Systems, Springer New York, New York, NY, 2005: pp. 30–55.

[15] M. Zisook, and R. Patel, Exploring the Communication Needs of People with Disabilities Through Participatory Design, in: CHI'14, Assoc Comput Machinery, 2014: p. 4.

[16] J. Nielsen, T. Clemmensen, and C. Yssing, Getting Access to What Goes on in People's Heads?: Reflections on the Think-aloud Technique, in: Proceedings of the Second Nordic Conference on Human-Computer Interaction, ACM, New York, NY, USA, 2002: pp. 101–110.

[17] R.C. Gibson, M.-M. Bouamrane, and M. Dunlop, Design Requirements for a Digital Aid to Support Adults with Mild Learning Disabilities During Clinical Consultations: Qualitative Study with Experts, *JMIR Rehabilitation and Assistive Tech*. **6** (2019) e10449. doi:10.2196/10449.

[18] R.C. Gibson, M.-M. Bouamrane, and M. Dunlop, Mobile support for adults with mild learning disabilities during clinical consultations, in: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '18, ACM Press, Barcelona, Spain, 2018: pp. 1–8.

[19] NICE, Everyone with learning disabilities should have their mental health checked annually, NICE says, as fears thousands may be undiagnosed, (2019). http://www.webcitation.org/77EQ9e2Hv (accessed March 29, 2019).

Address for correspondence

Ryan Colin Gibson, Department of Computer & Information Sciences, University of Strathclyde, Glasgow, U.K.

ryan.gibson@strath.ac.uk

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190417

A Mobile Application for Patient Engagement to Support Interdisciplinary Care

Christian Haux, Julia Lutyj, Max Seitz, Petra Knaup

Institute for Medical Biometry and Informatics, University of Heidelberg, Heidelberg, Germany

Abstract

There is evidence on interrelationships between several dental and chronic diseases. However, dentists and general practitioners often lack information when treating such multimorbid patients. Engagement of the patient in the decision making process may help to fill this gap and improve intersectoral care. The Dent@Prevent project therefore aims to develop a mobile application that can be used by patients to report information about their health status for interdisciplinary care. In this paper, the user interface concept and evaluation of the prototype of this mobile application is presented.

Keywords:

Patient Participation, Mobile Applications, Intersectoral Collaboration

Introduction

Empirical evidence indicates interrelationships between dental diseases and chronic-systemic diseases, particularly between periodontitis and diabetes [1,2]. For example, it has been shown that the treatment of periodontitis can lead to a better glycemic control in diabetics [3].

This makes obvious that interprofessional collaboration between dentists and general practitioners (GPs) is important to provide good health care. However, interprofessional collaboration is often limited. In Germany, these professions are organizationally separated, and dentists and GPs often lack information about the general or oral health status of the patient from the other profession. For example, the oral health status is often not known to the GP [4]. Supporting interprofessional collaboration between dentists and GPs might enable an encompassing view of the patients' health status.

EHealth services and systems (e.g., referral portals and electronic health records) can support communication between health care providers. Nevertheless, there are no established or optimized solutions to support collaboration between dentists and GPs. This is necessary for conveying essential information as Electronic health records in general can lead to information overload for physicians [5].

Better involvement of the patient in the decision-making process may help to fill this gap and improve interprofessional care. Patient-reported outcome measures (PROMs) are gaining importance to measure outcomes of health care. Self-anamnesis and patient-centered parameters are used to get information about the general and oral health status of the patient. Patient-reported experience measures (PREMs) are used to measure patients' experiences with health care services [6].

The Dent@Prevent project aims to develop a mobile application for the collection of such patient-centered parameters [7]. The aim is to provide this data to dentists and GPs during the treatment process in order to improve interdisciplinary care. Further, the mobile application can be used to provide patients with additional individualized information about their general and oral health status and to make them aware of the interrelationships between diseases.

To support interdisciplinary care among dentists and GPs, PROMs and PREMs from both areas have to be requested from patients. This leads to a considerable amount of questions for the mobile application. It is crucial for the success of this project that the patient stays motivated during reporting parameters. Further, the application must be designed to meet the requirements of the intended user groups.

In this paper, the development and testing of a user interface concept and prototype of a mobile application for patients to report patient-centered parameters for interdisciplinary care of dentists and GPs is presented.

Methods

The application was developed in a five phases process.

First, the preliminary content of the application was determined in a scoping literature review [8,9]. The aim of the scoping review was to identify patient-centered parameters that are potentially relevant in the context of interrelationships between chronic-systemic and dental diseases.

The results of the scoping review were evaluated by dentists, GPs, scientists, and patients in an Online Delphi survey afterwards. The participants assessed the relevance of the preliminary patient-centered parameters, suggested changes, and determined the final parameters to be recorded by the patient with the mobile application.

A requirement analysis for the user interface concept was conducted subsequently. For this, personas were developed on the basis of a review of case reports in scientific literature that reflect the intended user groups. The case reports were found with a PubMed search using the search terms (((“case reports”[Publication Type]) AND “periodontitis”[Title/Abstract]) AND “diabetes”).

General information about the patients reported in the case reports was taken from the case reports to create the personas. Personas are an established method for representing a concrete model of a user [10]. The developed personas were used to deduct requirements for the user interface of the mobile application for the Dent@Prevent project. The requirements were differentiated into functional requirements and non functional requirements.

The user interface concept was designed using the prototyping tool Axure RP 8.1.0 [11]. The prototype that was developed with this tool was fully functional, so that it could be evaluated with users afterwards in a focus group discussion.

The objective of the focus group discussion was to evaluate the usability of the user interface concept. The participants were recruited in the department of endocrinology and metabolism at Heidelberg University Hospital and a pharmacy in Heidelberg. The user interface concept was installed on tablet PCs so that the participants of the focus group could test the application. Afterwards, the feedback was discussed with the participants. An interview guideline was created using questions from the System Usability Scale (SUS) [12]. Unlike the SUS questionnaires, the questions were asked by the moderator. Further, the experiences of the participants with the interdisciplinary cooperation between dentists and GPs were discussed during the focus group. The answers of the participants were recorded in writing.

The feedback from the participants of the focus group was incorporated in the user interface concept and a prototype of the mobile application was developed. The application was implemented for the operating systems Android and iOS. The framework, ResearchKit, was used to develop the application for iOS [13] and the framework, ResearchStack, was used to develop the application for Android [14]. ResearchKit is an open source framework for medical research that was released in 2015. ResearchStack is a similar framework for Android. The frameworks provide predefined modules and templates that can be adapted and customized for individual studies.

Results

The scoping literature review resulted in 23 publications that concern patient-centered parameters for the intersect of oral and chronic systemic diseases and interprofessional collaboration between dentists and GPs [8]. Nineteen patient-centered parameters were selected from the results for the Online Delphi, in which 22 GPs, dentists, scientists, and patients participated. The Online Delphi resulted in selecting 16 relevant patient-centered parameters.

The application contains 33 questions for anamnesis. They can be recorded with the mobile application to get information about the general and oral health status of the patient. For example, the patient is asked about existing diseases like diabetes and periodontitis. Further, the application contains validated risk questionnaires. To assess the risk for diabetes, questions from the FindRisk questionnaire [15] were included. To assess the risk for periodontitis, questions from the Periodontitis risk score [16] were included. Finally, the application contains parameters concerning patients perception on intersectoral care. Among others the patient is asked, which data he regards as important to be exchanged between dentists and GPs.

The PubMed search resulted in identification of 30 case reports. Using these, two personas were developed for the requirement analysis. The first persona covers younger patients, aged 20 to 40, that only suffer from either diabetes or periodontitis. The second persona covers multimorbid and older patients, aged greater than 40.

The functional and non-functional requirements that were deduced from these personas are depicted in table 1. One requirement for the application is that it should have a high usability for older users. Therefore, care was taken to ensure good readability by presenting the questions in a clear and uncluttered way.

Each question is displayed on a separate screen. Small font sizes were avoided as well as long texts, as they not only hamper the readability, but also cause a longer answer time. The most used answer formats are single-choice or multiple-choice answers and free text to supplement uncovered replies. Some questions used a Likert scale or an EQ-visual analogue scale (EQ-VAS). Answer options were vertically aligned to use the space on the smartphone display optimally. For navigation, a toolbar was integrated in the user interface. The users could switch backwards and forwards between questions. Further, they had the option to quit the survey and discard the results at any time.

The user interface concept was subsequently evaluated in a focus group discussion at the Heidelberg University Hospital in summer 2018. Six partly multimorbid participants joined the discussion. The feedback on the usability of the application was positive. The participants found the application intuitive and easy to use. Further, they found the time needed to answer the questions acceptable. However, the participants also provided some criticism. First, they stated that the preview of the next question at the bottom of the screen was distracting. Therefore, it was removed when the prototype of the application was developed (figure 1).

Table 1– Functional and non functional requirements for the mobile application

No.	Requirement	Description
Functional requirements		
1	Application for iOS and Android	The application should be available for iOS and Android
2	Start screen	The application should contain a start screen that contains a description of the study and the option to start the survey.
3	Informed consent	The application needs to support the collection of the informed consent of the patient.
4	Options when taking the survey	It should be possible to quit the survey at any time. All answers that were given so far must then be deleted. Questions should be skippable. Further, the user should have the option to switch between questions and correct answers.
Non functional requirements		
1	High usability	The application should be easy to use for different age groups. The font and font size should be easy to read.
2	Visualization	The questions should be presented in a clear and uncluttered way.
3	Self-descriptiveness	The application should be simple and easy to understand and allow intuitive operation.

Further, they found some questions difficult to answer. For example, the FindRisk questionnaire contains questions concerning waist circumference and body mass index (BMI). The participants assumed that some patients might not know their waist circumference or BMI. Therefore, the question concerning BMI was changed to allow users to enter their height and weight, and BMI is calculated from these parameters.

When asking the participants about their experiences with the interdisciplinary cooperation between dentists and GPs, they stated that dentists lack information about the patients' general diseases and vice versa. The participants assumed that better patient engagement in the process may help to improve the information exchange, and that they would use such a mobile application in their daily life. They appreciated if the mobile application could be used to provide information concerning the interrelationships between chronic and dental diseases to them.

The feedback from the focus groups was adopted in the prototype that was implemented using the ResearchKit framework for iOS and the ResearchStack framework for Android. The application consists of three modules. First, users give informed consent. Second, they fill in the questions for the patient-centered parameters. A third module that contains questions from the SUS was added to evaluate final application in dental practices. The defaults of the framework almost fit the specifications of the user interface concept. Some of the features of the user interface concept could not be implemented in the Android application using ResearchStack because the framework was lacking the corresponding features. For example, it was not possible to use a slider bar for the EQ-VAS answer formats. Instead, the input of a number was chosen (figure 2).

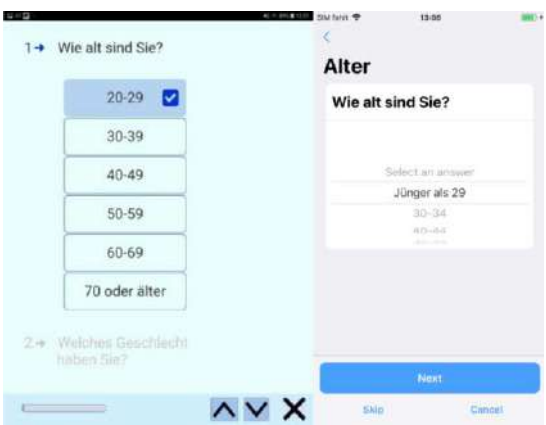


Figure 1– Example question of the user interface concept in Axure (left) and the prototype of the mobile application using ResearchKit (right). The preview of the next question at the bottom of the screen was removed in the prototype.

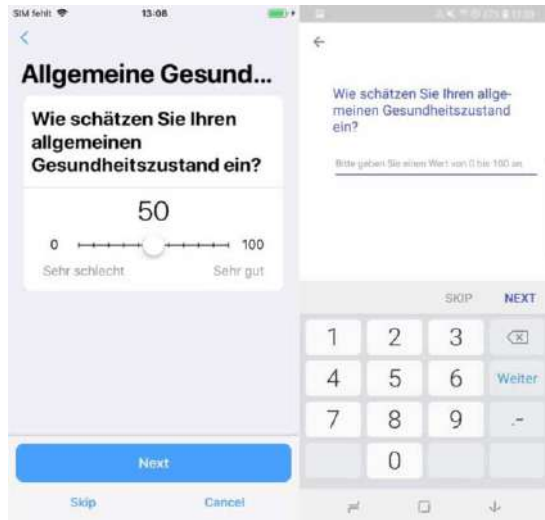


Figure 2– Difference between the ResearchKit framework (left) and the ResearchStack framework (right).

Discussion

The interprofessional collaboration between dentists and GPs is limited despite evidence on interrelationships between dental and chronic systemic diseases. Active patient involvement can play an important role in the implementation of an improved interdisciplinary treatment process.

This paper has described the development of a mobile application for patient engagement. By having a multi-step approach with designing first a user interface concept, having it be evaluated by users in the focus group, and implementing the prototype of the application, it was possible to integrate the feedback of users at an early stage of the development process to ensure that the application fits the requirements of its users.

Axure RP has proven a very comprehensive and complex prototyping tool for the user interface concept. The development of the user interface was rather time-consuming with the tool. The final development of the prototype using the ResearchKit and ResearchStack frameworks was done in a much shorter time. Therefore, it would have been possible to implement the user interface concept directly with the ResearchKit and ResearchStack frameworks for our application.

The choice to involve intended users in the focus group was advantageous – the users were integrated at an early stage, they could test the user interface concept on a device, and could give direct feedback. Further, problems or questions could be solved immediately when they raised during the discussion. However, it also had the disadvantage that the responsiveness of the users was different. As the moderator asked the questions to all participants in the group, some of the participants answered more than others. Therefore, the opinion of all users could not be collected completely, and users may have influenced each other [17]. Individual interviews with each participant may solve this issue.

By choosing the ResearchKit and ResearchStack frameworks for the development of the prototype, it was possible to save time during the development of the prototype because many of

the predefined modules and templates could be used directly. However, in some cases it was not possible to realize every aspect of the user interface concept with the defaults. Further, the functionality between the ResearchKit framework and the ResearchStack framework differ. For example, the ResearchStack framework lacks answer formats. Such limitations were solved by choosing a different answer type in the Android application. Another advantage of using these frameworks is that they are not dependent on any data storage or data exchange service.

The developed prototype will be tested in dental and general practices in the near future. Patients will be asked to fill in the questionnaire and rate the usability of the application afterwards. The answers of the participants will be stored locally on the mobile device in an aggregated way. Thus, data protection requirements can be adhered to.

Conclusions

The prototype of the mobile application for patient engagement to improve the interprofessional collaboration between dentists and GPs was successfully developed. The prototype will subsequently be tested by patients in dental and general practices.

Acknowledgements

Dent@Prevent receives funding from the Federal Joint Committee (G-BA) - Innovation Fund, Grant agreement number: 01VSF16052.

References

[1] J. Botero, C. Rodriguez, and A. Agudelo-Suarez, Periodontal treatment and glycaemic control in patients with diabetes and periodontitis: An umbrella review, *Aust Dent J*. **61** (2016) 134–148.

[2] J.M. Liljestrand, A.S. Havulinna, S. Paju, S. Mannisto, V. Salomaa, and P.J. Pussinen, Missing teeth predict incident cardiovascular events, diabetes, and death, *J Dent Res*. **94** (2015) 1055–1062.

[3] P. Preshaw, A. Alba, D. Herrera, S. Jepsen, A. Konstantinidis, K. Makrilakis, and R. Taylor, Periodontitis and diabetes: A two-way relationship, *Diabetologia*. **55** (2012) 21–31.

[4] K. Sippli, M. Rieger, and F. Huettig, GPs' and dentists' experiences and expectations of interprofessional collaboration: Findings from a qualitative study in Germany, *BMC Health Serv Res*. **17** (2017) 179.

[5] H. Singh, C. Spitzmueller, N. Petersen, M. Sawhney, and D. Sittig, Information overload and missed test results in electronic health record-based settings, *JAMA Intern Med*. **173** (2013) 702–704.

[6] C. Kingsley, and S. Patel, Patient-reported outcome measures and patient-reported experience measures, *BJA Educ*. **17** (2017) 137–144.

[7] C. Haux, S. Listl, M. Ganzinger, I. Schubert, and P. Knaup, Towards an evidence-informed and patient centered decision support system for intersectoral care, In: Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie. 62. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V. (GMDS). Oldenburg, 17.-21.09.2017. Düsseldorf: *German Medical Science GMS Publishing House*; 2017. DocAbstr. 233.

[8] M. van der Zande, and S. Listl, Patient-reported measures at the intersect of chronic-systemic and dental diseases: A scoping study, *Forthcoming*. (2018).

[9] D. Levac, H. Colquhoun, and K. K. O'Brien, Scoping studies: Advancing the methodology, *Implement Sci*. **5** (2010) 1–9.

[10] A. Cooper, R. Reimann, and D. Cronin, Modeling users: Personas and goals, *About Face*. **2** (2003) *55*–74.

[11] Axure Software Solutions, Axure RP. https://www.axure.com (accessed November 9, 2018).

[12] J. Brooke, SUS-A quick and dirty usability scale, *Usability Evaluation in Industry*. **189** (1996) 4–7.

[13] Apple Inc., ResearchKit. http://researchkit.org (accessed November 9, 2018).

[14] ResearchStack.http://researchstack.org (accessed November 9, 2018).

[15] J. Lindström, and J. Tuomilehto, The diabetes risk score: A practical tool to predict type 2 diabetes risk., *Diabetes Care*. **26** (2003) 725–731.

[16] Deutsche Gesellschaft für Parodontologie, Der Parodontitis Risikoscore - ein Selbsttest. https://www.dgparo.de/media/download-5acc9cf7d7b48 (accessed November 9, 2018).

[17] R. Mazza, and A. Berre, Focus group methodology for evaluating information visualization techniques and tools, in: 11th International Conference Information Visualization, 2007.

Address for correspondence

Christian Haux, Institute for Medical Biometry and Informatics, Heidelberg University, Im Neuenheimer Feld 130.3, 69120 Heidelberg, Germany; Christian.Haux@med.uni-heidelberg.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190418

How Do General-Purpose Sentiment Analyzers Perform when Applied to Health-Related Online Social Media Data?

Lu He, Kai Zheng

Department of Informatics, University of California, Irvine, City Irvine, CA, USA

Abstract

Sentiment analysis has been increasingly used to analyze online social media data such as tweets and health forum posts. However, previous studies often adopted existing, general-purpose sentiment analyzers developed in non-healthcare domains, without assessing their validity and without customizing them for the specific study context. In this work, we empirically evaluated three general-purpose sentiment analyzers popularly used in previous studies (Stanford Core NLP Sentiment Analysis, TextBlob, and VADER), based on two online health datasets and a general-purpose dataset as the baseline. We illustrate that none of these general-purpose sentiment analyzers were able to produce satisfactory classifications of sentiment polarity. Further, these sentiment analyzers generated inconsistent results when applied to the same dataset, and their performance varies to a great extent across the two health datasets. Significant future work is therefore needed to develop context-specific sentiment analysis tools for analyzing online health data.

Keywords:

Social Media; Computing Methodologies

Introduction

Increasingly, patients use the internet to ask questions related to their health conditions and write about their experience coping with diseases. According to a survey conducted by the Pew Internet Project, in the U.S., one in five patients' living with chronic diseases participated in creating online content about their health or medical issues through social media websites [1]. As such content is readily available and contains rich information and insights, researchers have started to utilize it as a new source of data to conduct novel research studies. In this paper, we refer to such content hereafter as "online health data."

Sentiment analysis, also known as opinion mining, is a branch of natural language processing (NLP) for describing emotions from text. It provides a computational means to automatically classify positive, negative, or neutral attitude toward a subject (e.g. a movie) based on the opinions expressed in a piece of text (e.g. a movie review). In health-related disciplines, sentiment analysis has been widely applied to analyze online health data to investigate in topics such as public opinions of health policies (e.g. the Affordable Care Act) [2], patient attitudes toward a medical treatment or intervention (e.g. vaccination) [3,4], consumer rating of healthcare services or products (e.g. hospital services; drugs and cosmetic products) [5,6] and patient journeys (e.g. stigmatization related to Alzheimer's Disease) [7].

The proliferation of research work that has applied sentiment analysis to online health data is attributable in part to the availability of several general-purpose sentiment analyzers (e.g. Stanford Core NLP Sentiment Analysis). However, all of these analyzers were initially developed in non-healthcare contexts; and were trained on non-healthcare datasets. It is therefore imperative to evaluate their validity before applying them to analyze online health data. Unfortunately, in reviewing the relevant literature, we found that previous studies used these general-purpose sentiment analyzers in a rather arbitrary manner. Most of the studies did not provide adequate justifications as to why a particular sentiment analyzer was chosen; whether it was appropriate for a study context; and whether other analyzers might produce better results.

In this paper, we aimed to address this gap by empirically evaluating the performance of three general-purpose sentiment analyzers that have been most popularly used in previous studies: Stanford Core NLP Sentiment Analysis, TextBlob, and VADER. We applied these sentiment analyzers on two datasets, representing two typical analytical scenarios with online health data: public opinions regarding health interventions and healthcare policy. We also used a non-domain specific twitter dataset to serve as the baseline and compared the performances of the analyzers on the baseline dataset with the other two health-related datasets. Through the empirical evaluation, we aimed to answer the following research questions:

1. 1. Do different sentiment analyzers produce consistent results when applied to the same online health dataset?
2. 2. Does the same general-purpose sentiment analyzer perform differently when applied to different online health datasets concerning different health topics?
3. 3. Are these general-purpose sentiment analyzers adequate enough to generate useful results without retuning for online health data?

Answering these questions may help the research community establish an evidence base as regards the validity of these general-purpose analyzers when applied in studies that involve online health data. The results may also provide insights into how to properly select the right sentiment analyzer for a particular study context, and how to improve their performance in future research.

Related Work

Most of the existing sentiment analyzers and lexicons were developed based on movie reviews or product reviews, possibly

because of the availability of large amounts of labeled data for training. The Hu&Liu sentiment lexicon, one of the earliest tools for sentiment analysis, was curated by manually grouping words contained in e-Commerce product reviews into different sentiment categories [8]. Similarly, the most popularly used sentiment analyzer, the Stanford CoreNLP Sentiment Analysis tool, based its sentiment treebank and model training on a movie review dataset [9,10].

In health-related studies, Korkontzelos et al. used sentiment analysis to detect adverse drug reaction (ADR) based on twitter data, and demonstrated a higher level of accuracy as compared to conventional approaches [11]. Davis et al. used a sentiment lexicon named labMT15 to assess public opinions expressed in tweets regarding the Affordable Care Act (ACA), and showed that the results were highly consistent with what could only be obtainable previously through expensive polls [2]. In another study, Du et al. used the supervised machine-learning method to automatically classify tweets based on the sentiments toward HPV, and to study the evolution of the sentiments over time using a time series analysis [3]. More recently, Burnap et al. incorporated sentiment scores into their feature sets to develop supervised machine learning models that automatically detect suicide-related tweets [12]; and Ji et al. extracted negative sentiments from online social media data to understand public health concerns regarding disease epidemics [13].

While researchers may opt to train machine-learning classifiers on their own datasets, doing so requires a significant amount of manually annotated data, in addition to sophisticated skills in developing, training, and testing machine learning models. As a result, most of the previous studies leveraged existing, general-purpose sentiment analyzers that are readily available and are relatively easy to adopt. However, as it has been previously demonstrated, sentiment analysis models trained in one domain may perform poorly when applied in another domain without adaptation. In a study on Ebola-related social media discussions, Lu et al. noticed a significant level of disagreement between the results generated by different sentiment analyzers that they experimented with [14]. In reviewing the previous studies, we also found that there was generally a lack of discussions on the rationale of choosing a particular sentiment analyzer; and very few studies validated the analyzer chosen before applying it to their datasets. It thus remains unknown whether the results reported in these studies, based on general-purpose sentiment analyzers developed in non-healthcare domains, are reliable and repeatable.

Methods

Sentiment Analyzers Studied

In this paper, we evaluated three sentiment analyzers that had been most popularly used in previous studies: Stanford Core NLP Sentiment Analysis, TextBlob, and VADER.

The Stanford Core NLP Sentiment Analysis tool was developed by the Stanford NLP group as a module in the Stanford Core NLP toolkit [9]. Its sentiment analysis model was trained using a recursive neural tensor network on a movie review dataset made available by Pang and Lee [10]. Unlike earlier sentiment analyzers, in which bag of words was used and word orders were ignored, the Stanford Sentiment Analysis tool parses input text into sentiment trees, where each leaf node refers to a word; every word in the input text is thus represented as a vector. Instead of producing numerical sentiment score, the results generated by the Stanford Sentiment Analysis tool are in the form of discrete categories, namely “very negative,” “neutral,” “positive,” and “very positive.”

TextBlob is a widely distributed Python Library with an easy-to-use API that can be called upon by other programs to analyze sentiments of text, in addition to performing other common NLP tasks such as part-of-speech tagging and tokenization [15]. Its sentiment analyzer has two implementations: one based on a collection of semantic patterns; and the other based on a Naïve Bayes learning module. TextBlob returns sentiment analysis results in the form of numerical polarity ranging from -1 (most negative) to 1 (most positive). It also produces a companion subjectivity score in the range of 0 (very objective) to 1 (very subjective).

VADER, which stands for Valence Aware Dictionary and sEntiment Reasoner, is an open-source sentiment analyzer developed and maintained by Hutto and Gilbert. VADER is a lexicon and rule-based tool optimized for classifying sentiments expressed in user-generated text in social media [16]. The rules it utilizes are heuristics derived from a manual review of a set of tweets by multiple independent human judges. The authors have also incorporated many features that are not found in other sentiment analyzers, such as punctuation, capitalization, slangs, emoticons, and degree modifiers. The results produced by VADER are in the form of sentiment polarity (positive vs. negative), and a numeric sentiment intensity score on a scale from -4 (extremely negative) to +4 (extremely positive).

Annotated Datasets for Evaluation

To evaluate the performance of these general-purpose sentiment analyzers, we leveraged two publicly available online health datasets that were annotated in previous studies for the purposes of understanding public opinions of the Health Care Reform (hereafter referred to as the “HCR” dataset) and public opinions of HPV (the “HPV” dataset), respectively. We also used a general-purpose twitter dataset to serve as the baseline.

The first dataset, or the HCR dataset, contains tweets that include the hashtag “#hcr” in March 2010. The original annotated dataset also includes 8 different targets, for instance, Obama, HCR, Liberals, etc. The annotated sentiments include 5 categories, positive, negative, neutral, irrelevant and unsure. For our study purpose, we only included the tweets that targeted on HCR and those that expressed positive, negative or neutral sentiments. This resulted in a total of 961 tweets, with 290 of them (30%) being labeled as positive, 422 (44%) being labeled as negative and 249 (26%) being labeled as neutral.

The second dataset used in this study was curated and made available by Du et al. This dataset was created by three human annotators in an attempt to discover hierarchical sentiment categories of public opinions regarding the HPV vaccination [3]. The manually annotated results consist of three major categories: “negative,” “positive/neutral,” and “unrelated.” The original dataset contained a total of 6,000 tweet IDs. After removing invalid tweet IDs (e.g. those that were deleted, or are no longer publicly accessible), 4,616 tweets are available to use in this study. As our study focused on sentiment analysis, we further excluded those tweets that were labeled as “unrelated,” which left us 3,211 tweets to work with. Among these tweets, 1,084 (34%) were annotated as negative, and 998 (31%) were positive and 1129 (35%) were neutral.

The baseline dataset contains non-domain specific tweets that have been annotated. It includes tweets regarding products (e.g., “am loving new malcolm gladwell book - outliers”), personal experiences or feelings (e.g., “so tired. I didn't sleep well at all last night.”), opinions (e.g., “If Google's self-driving

car is the future, I don't want to be a part of it. #savethemanuals"). It contains 381 positive tweets (38.9%), 387 negative tweets (39.5%) and 212 neutral tweets (21.6%).

Study Design

We separately applied the three sentiment analyzers to each of the annotated datasets. To answer the first research question ("do different sentiment analyzers produce consistent results when applied to the same online health dataset"), we calculated the inter-rater agreement rate between the results generated by different sentiment analyzers when applied to the same dataset. To answer the second research question ("does the same general-purpose sentiment analyzer perform differently when applied to different online health datasets concerning different health topics"), we computed the precision-recall-F1 metrics of the sentiment analyzers across the two annotated datasets. We then compared the classification results (i.e. sentiment polarity) to the ground truth—human annotations, to answer the third research question ("are these general-purpose sentiment analyzers adequate enough to generate useful results without retuning for online health data"). Last, we performed term frequency-inverse document frequency (TF-IDF) analysis to extract and compare the important words in each sentiment category across the three datasets.

Results

The overall results - the precision, recall, F1-measure metrics of each sentiment analyzer on each dataset are reported in Table 1, Table2 and Table 3. The recall measure should be interpreted with caution, as with skewed datasets it is easy to achieve a high recall by incorrectly labeling the data. Stanford NLP sentiment analyzer performed poorly on both the HCR and the HPV dataset, compared to the baseline dataset, as it has extremely low precision (e.g., 11%, 7.3%) in detecting positive sentiments on the HCR dataset and the HPV dataset. VADER performed fine on the HCR and the HPV dataset with around 42 % to 51% F1-measures in each sentiment category, however, the performance on the baseline dataset (65%~74%) still beats these numbers. TextBlob performed the worst in distinguishing sentiment classes, especially that it assigned most labels as neutral. It performed equally poorly on the HCR and the HPV dataset with extremely low precision for the positive and negative sentiment class, but had decent performance on the baseline dataset. Therefore, the three sentiment analyzers all performed poorly on the two health-related datasets, but had satisfying performance (i.e., most having a precision or a recall higher than 60% in detecting the three sentiment categories) on the baseline dataset that is non-domain specific.

Table 1- Precision, recall, F1-measure, Stanford NLP

		Dataset		
Class	**Metric**	**Baseline**	**HCR**	**HPV**
Positive	Precision	54%	11%	7.3%
	Recall	73%	62.7%	47%
	F1-measure	62%	18.8%	12.7%
Negative	Precision	39%	47.2%	58.2%
	Recall	71%	46.5%	35.8%
	F1-measure	51%	46.8%	44.3%
Neutral	Precision	74%	38.6%	39%
	Recall	34%	20%	34%
	F1-measure	47%	26.3%	36.6%

Table 2-Precision, recall, F1-measure, VADER

		Dataset		
Class	**Metric**	**Baseline**	**HCR**	**HPV**
Positive	Precision	82.6%	54.8%	39%
	Recall	67.6%	38.2%	45%
	F1-measure	74%	45%	42%
Negative	Precision	64.7%	35.4%	57%
	Recall	90.7%	56.4%	49%
	F1-measure	75.5%	43.5%	53%
Neutral	Precision	68.8%	46.6%	50%
	Recall	62%	41.4%	51%
	F1-measure	65.3%	43.9%	51%

Table 3-Precision, Recall, F1-measure, TextBlob

		Dataset		
Class	**Metric**	**Baseline**	**HCR**	**HPV**
Positive	Precision	59%	7.6%	3%
	Recall	72%	40.7%	49%
	F1-measure	65%	12.8%	5.9%
Negative	Precision	40%	2.6%	3.8%
	Recall	90%	68.8%	65%
	F1-measure	55%	5%	7.1%
Neutral	Precision	82%	96.8%	97.6%
	Recall	36.8%	27%	36%
	F1-measure	51%	42.2%	52%

Table 4 reports the inter-rater agreement rates when applying these sentiment analyzers to the first annotated dataset (HCR). Among the three sentiment analyzers, Stanford NLP and TextBlob exhibit a high degree of agreement with each other, 46.68%; while the results obtained by the VADER Sentiment Analysis tool correlate poorly with the other two analyzers, for instance, Stanford NLP in particular (29.88%). Shown in Table 5 are inter-rater agreement rates when applying each of the three sentiment analyzers to the second annotated dataset (HPV). Similarly, Stanford NLP and TextBlob have the highest agreement (39.3%). However, the results produced by VADER and by the Stanford NLP Sentiment Analysis tool are poorly correlated; the inter-rater agreement ratio is only 38.0%. In Table 6 we present the inter-rater agreement rates of the three sentiment analyzers on the baseline dataset. While the agreement rates are higher than those of the HPV and the HCR datasets, they are still unsatisfying with only around 50% agreement ratios of each pair of the three sentiment analyzers.

Table 4-Inter-rate agreement rates, the HCR dataset

Analyzer A	**Analyzer B**	**Agreement Ratio**
Stanford	TextBlob	46.68%
Stanford	VADER	29.88%
VADER	TextBlob	34.95%

Table 5-Inter-rate agreement rates, the HPV dataset

Analyzer A	**Analyzer B**	**Agreement Ratio**
Stanford	TextBlob	39.3%
Stanford	VADER	38.0%
VADER	TextBlob	38.5%

Table 6-Inter-rate agreement rates, the Baseline dataset

Analyzer A	Analyzer B	Agreement Ratio
Stanford	TextBlob	51.0 %
Stanford	VADER	42.6%
VADER	TextBlob	42.7%

To further explore possible reasons why general-purpose sentiment analyzers perform significantly worse on health-related datasets than on baseline, non-domain specific dataset, and how different health-related datasets are from the non-domain specific dataset, we used term frequency-inverse document frequency (TF-IDF) analysis. Table 7 to Table 9 present the top-5 weighted TF-IDF unigrams of the three datasets for each sentiment category.

Table 7: Top-5 positive unigrams

Baseline	HCR	HPV
love	Passthedamnbill	Act2015
happy	Affordable	availability
Lebron	Human	post2015
earlier	Petition	saving
seats	stupakpitts	literally

Table 8: Top-5 negative unigrams

Baseline	HCR	HPV
fucking	handsoff	neutral
warner	takeover	trigger
fuck	killthebill	victim
driverless	tax	injury
cable	codered	danish

Table 9: Top-5 neutral unigrams

Baseline	HCR	HPV
Driving	defazio	slightly
deflategate	holding	stance
Google	schedule	callaghan
Check	breaking	heather
flight	association	project

The TF-IDF analysis may in part explain why the sentiment analyzers tend to assign too many neutral labels and fail to recognize negative and positive sentiment classes. For instance, words and phrases such as "danish" itself is neutral, but tweets that express negative sentiments toward HPV often include such words to form their arguments. However, the three sentiment analyzers failed to recognize them as negative, because without a proper context, they are neutral. A sample tweet "Vaccines trigger genetically modified diseases" is labeled as neutral by VADER and TextBlob. While we do not have space to include the top-10 unigrams and bigrams of the baseline dataset, we observed that the top-10 weighted unigrams are sentiment words such as "love", "happy", "amazing", "better", "thank", etc, which do not appear in the unigrams in both the HCR and the HPV datasets. Therefore, the significant differences in performance of the three sentiment analyzers on health-related datasets and general datasets that are not in health-domain may be explained in part by the different ways of sentiment expression – in health-related social media data, people use words that are neutral to form arguments and express negative or positive sentiments, and such expressions cannot be accurately captured by general-purpose sentiment analyzers. We also noticed that for the HCR dataset, many of the combined word phrases such as "passthedamnbill" and "killthebill" are used to express sentiments, however current NLP tools may have difficulties parsing and splitting those combined words, and therefore makes it hard to sentiment analyzers to detect the underlying sentiments.

Discussion

Our results indicate that the three general-purpose sentiment analyzers popularly used in previous studies produced inconsistent classification results when applied to the same online health dataset, and their performance varies to a great extent across different datasets. Among these three analyzers, VADER and the Stanford Sentiment Analysis tool have the lowest degree of inter-rater agreement. This may be due to the fact that these two sentiment analyzers were developed from distinct domains: VADER drew its lexicon and rules primarily from social media data (tweets); while the Stanford Sentiment Analysis tool was developed based on movie reviews. This finding indicates that in sentiment analysis, the utility of a high-performing sentiment analyzer trained in one domain may not be transferable to other domains. Thus, when deciding what sentiment lexicons or sentiment analyzers to use, researchers should be aware of the contexts in which they were originally developed, in addition to its underlying classification mechanisms.

Overall, the performance of these three general-purpose sentiment analyzers is unsatisfactory (e.g., having extremely low precision in detecting positive sentiments or falsely labeling tweets as neutral) when applied to online health data, while they have decent performance on non-health related social media data. The Stanford Core NLP Sentiment Analysis tool may fall short because of the nature of user-generated content online, especially tweets, that contain an excessive number of anonyms, abbreviations, hashtags, and URLs; as compared to movie reviews that are generally well structured. However, this does not explain why tools such as VADER, which was specifically designed to process social media data, missed a vast majority of the text containing negative sentiments in the HPV dataset and HCR dataset. It is possible that "negativity," "positivity," and "neutrality" of sentiments expressed in the context of health-related discussions may be interpreted differently by human annotators, in contrast to sentiments conveyed in other types of social media exchanges, demonstrated by the TF-IDF analysis above. Thus, it is critical for health sciences researchers to be mindful of the limitation of the sentiment analyzer(s) that they choose to use.

The findings from this study provide insights into future work on how to improve the utility of sentiment analysis in studies that involve online health data. First, based on our review of previous work, it appears that a lack of comprehensive understanding of the state-of-the-art of sentiment analysis tools impedes researchers from picking the right tool for their studies, and from providing adequate rationale to justify their choices. Therefore, it will be very valuable to have a "road map" of the history and recent development of sentiment analyzers and lexicons, especially their context of development, working mechanisms, and intended use. Second, the existing general-purpose sentiment analyzers need to be significantly adapted when used to analyze online health data. This requires a thorough understanding of the nature of health-related social media discussions, and of the specific health-related topic being studied; for example, the algorithm and lexicon appropriate for determining sentiment polarity in detection of drug side effects

can be very different from what is appropriate for use in understanding the public's opinions toward a health policy. Future work is thus called for to explore how to develop context-specific lexicons and sentiment analyzers that are optimized for analyzing different types of online health data. Third, pre-study validation and post- study error analysis are essential to understand the utility and limitation of an existing sentiment analyzer, which should be performed and reported in every study that involves sentiment analysis. Unfortunately, most of the previous studies that we reviewed simply reported the results produced by a general-purpose sentiment analyzer, without conducting any assessment of the validity of the results. Lastly, knowing that sentiment analyzers usually do not perform well when applied across domains, the research community may consider developing domain adaptation techniques that can be readily applied to extend the capability of existing general-purpose sentiment analyzers, e.g., by means of re-training sentiment analysis models, adjusting heuristics and rules, or swapping lexicons.

Conclusion

In this study, we empirically evaluated the performance of three general-purpose sentiment analyzers on two different online health datasets. The results show that these general-purpose sentiment analyzers were unable to produce consistent results when applied to the same dataset, and their performance varies when applied to different datasets. These findings suggest that general-purpose sentiment analyzers developed in non-healthcare domains may perform poorly on online health data. Future work is thus needed to identify ways to tailor them, or develop new sentiment analyzers optimized for the health context.

References

[1] Social Media and Health | Pew Research Center, (2010). https://www.pewinternet.org/2010/03/24/social-media-and-health/ (accessed March 29, 2019).

[2] M.A. Davis, K. Zheng, Y. Liu, and H. Levy, Public Response to Obamacare on Twitter, *J. Med. Internet Res.* **19** (2017) e167. doi:10.2196/jmir.6946.

[3] J. Du, J. Xu, H. Song, X. Liu, and C. Tao, Optimization on machine learning based approaches for sentiment analysis on HPV vaccines related tweets, *J. Biomed. Semant.* **8** (2017). doi:10.1186/s13326-017-0120-6.

[4] J. Du, J. Xu, H.-Y. Song, and C. Tao, Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data, *BMC Med. Inform. Decis. Mak.* **17** (2017). doi:10.1186/s12911-017-0469-6.

[5] J.W. Huppertz, and P. Otto, Predicting HCAHPS scores from hospitals' social media pages: A sentiment analysis, *Health Care Manage. Rev.* (2017) 1. doi:10.1097/HMR.0000000000000154.

[6] H. Isah, P. Trundle, and D. Neagu, Social media analysis for product safety using text mining and sentiment analysis, in: 2014 14th UK Workshop Comput. Intell. UKCI, 2014: pp. 1–7. doi:10.1109/UKCI.2014.6930158.

[7] N. Oscar, P.A. Fox, R. Croucher, R. Wernick, J. Keune, and K. Hooker, Machine Learning, Sentiment Analysis, and Tweets: An Examination of Alzheimer's Disease Stigma on Twitter, *J. Gerontol. Ser. B*. **72** (2017) 742–751. doi:10.1093/geronb/gbx014.

[8] M. Hu, and B. Liu, Mining Opinion Features in Customer Reviews, (n.d.) 6.

[9] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky, The Stanford CoreNLP Natural Language Processing Toolkit, in: Proc. 52nd Annu. Meet. Assoc. Comput. Linguist. Syst. Demonstr., Association for Computational Linguistics, Baltimore, Maryland, 2014: pp. 55–60. doi:10.3115/v1/P14-5010.

[10] B. Pang, and L. Lee, Opinion mining and sentiment analysis, (n.d.) 94.

[11] I. Korkontzelos, A. Nikfarjam, M. Shardlow, A. Sarker, S. Ananiadou, and G.H. Gonzalez, Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts, *J. Biomed. Inform.* **62** (2016) 148–158. doi:10.1016/j.jbi.2016.06.007.

[12] P. Burnap, W. Colombo, and J. Scourfield, Machine Classification and Analysis of Suicide-Related Communication on Twitter, in: Proc. 26th ACM Conf. Hypertext Soc. Media - HT 15, ACM Press, Guzelyurt, Northern Cyprus, 2015: pp. 75–84. doi:10.1145/2700171.2791023.

[13] X. Ji, S.A. Chun, and J. Geller, Monitoring Public Health Concerns Using Twitter Sentiment Classifications, in: 2013 IEEE Int. Conf. Healthc. Inform., 2013: pp. 335–344. doi:10.1109/ICHI.2013.47.

[14] Y. Lu, X. Hu, F. Wang, S. Kumar, H. Liu, and R. Maciejewski, Visualizing Social Media Sentiment in Disaster Scenarios, in: Proc. 24th Int. Conf. World Wide Web - WWW 15 Companion, ACM Press, Florence, Italy, 2015: pp. 1211–1215. doi:10.1145/2740908.2741720.

[15] TextBlob: Simplified Text Processing — TextBlob 0.15.2 documentation, (n.d.). https://textblob.readthedocs.io/en/dev/index.html (accessed March 30, 2019).

[16] C.J. Hutto, and E. Gilbert, VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text, (n.d.) 10.

Address for correspondence

Lu He

E-mail: lhe11@uci.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190419

Prioritizing Features to Redesign in an EMR System

Samar Helou[a], Victoria Abou-Khalil[a], Goshiro Yamamoto[b], Eiji Kondoh[c], Hiroshi Tamura[d], Shusuke Hiragi[ab], Osamu Sugiyama[e], Kazuya Okamoto[ab], Masayuki Nambu[e], Tomohiro Kuroda[ab]

[a] *Department of Social Informatics, Graduate School of Informatics, Kyoto University, Kyoto, Japan*
[b] *Division of Medical IT and Administration planning, Kyoto University Hospital, Kyoto, Japan*
[c] *Department of Gynecology and Obstetrics, Graduate School of Medicine, Kyoto University, Kyoto, Japan*
[d] *Center for Innovative Research and Education in Data Science, Kyoto University, Kyoto, Japan*
[e] *Preemptive Medicine and Lifestyle Related Diseases Research Center, Kyoto University, Kyoto, Japan*

Abstract

Redesigning Electronic Medical Record (EMR) systems is needed to improve their usefulness and usability. For user-centered redesign, designers should consider which EMR features are the most important to the users. However, prioritizing the EMR features is complicated because: (i) EMR systems involve multiple users with different, and sometimes conflicting, priorities and (ii) targeting one feature will affect other features of the EMR system. In this work, we propose a method for prioritizing the features to target when redesigning an EMR system. The method takes into consideration the different priorities of the users and the relationships between the different features. We illustrate the method through a case study on redesigning EMR systems in Japanese antenatal care settings. Our results show the importance of considering the different types of EMR users and the relationships between different EMR features. Designers could use the proposed method as a decision-aid tool in EMR redesign projects.

Keywords:

Electronic Medical Records, Design, Decision Aids

Introduction

Electronic Medical Record (EMR) systems are increasingly implemented in clinics. Their main purpose is to support healthcare provision processes by making them more efficient, effective, and data driven [1,2]. However, healthcare providers had voiced concerns over EMR systems disrupting their workflows [3] and affecting their communication with their patients [4]. These concerns may be due to the low levels of usability and usefulness of existing EMR systems [5].

Previous studies have proposed multiple redesign opportunities to improve EMR systems such as: integrating them better into work practices [6–8], making it easier to navigate and find information inside of them [9–11], extending their functionalities to support all the tasks of the healthcare providers [12–14], improving their interoperability [15], and improving their usability and learnability [16], among others.

When redesigning an EMR system, multiple redesign opportunities exist. Usually, designers cannot act upon all of them due to their limited available resources. Therefore, designers have to make a choice over which EMR features to target. To make that choice, designers have to prioritize EMR features i.e., know which EMR features are the most important to work on.

Following a user-centered design approach, the designers would prioritize the EMR features based on the priorities of their users [17]. However, EMR systems have different users with different, and sometimes contradicting, priorities. For example: some patients may think it is important to have access to their EMRs while their providers may think it is not important at all. Taking into consideration the priorities of all the users is necessary to provide optimal designs.

Moreover, targeting one feature could affect other features of the EMR system. For example: adding more functionalities to the EMR system might make it less usable. As another example, making the EMR system easier to learn might make it easier to use. Thus, the overall effect of targeting one EMR feature consists of a *direct effect* i.e., the effect it has on the targeted feature and a *cascade effect* i.e., the effect it has on the other features.

In this paper, we propose a method to prioritize EMR features to target during redesign projects. The method takes into consideration the priorities of the different users and the relationships between the different EMR features. The method allows EMR designers to rank different EMR features based on their importance to the users and the overall effect of targeting them.

To illustrate the method, we conduct a case study where we prioritize EMR features to redesign in Japanese antenatal care settings.

Methods

Our method consists of four steps, as shown in Figure 1.

In *Step 1,* the designers identify the level of importance of the different EMR features to the users.

In *Step 2*, the designers analyze the relationships between the different EMR features.

In *Step 3*, the designers compute the overall effect of modifying each EMR feature.

In *Step 4*, the priority score for redesigning the different EMR features is calculated. The priority scores are used to identify the most strategic EMR features to redesign.

1- Identifying the importance of different EMR features to the users

Previous work identified multiple features of EMR systems in Japanese antenatal care settings [16]. To identify the importance of the different EMR features, we surveyed five obstetricians, ten midwives, and 413 pregnant women in Japan. We asked the respondents about the importance of each EMR feature to them. The respondents reported the importance of the different EMR features on a 4-point Likert scale of "Very Important (4)," "Important (3)," "Slightly Important (2)," "Not important at all (1)."

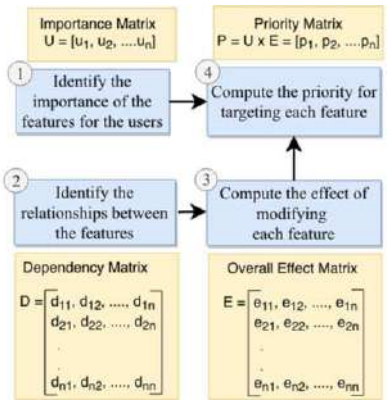


Figure 1- Four steps to prioritize EMR features to redesign

Using the survey results, we created the features' *importance matrix* as follows:

Suppose we have *n* features F = {$f_1, f_2, ..., f_n$} and *l* respondents R ={$r_1, r_2, ..., r_l$} who indicated the importance of each feature using the 4-point Likert scale. By aggregating the *l* respondents' answers, we obtain the features' *importance matrix* $U = [u_i]_n$ using:

$$u_i = \frac{1}{l}\sum_{k=1}^{l} u_i^k \quad i = 1,2,\dots,n$$

where u_i^k represents the answer of respondent r_k on the importance of feature f_i and u_i represents the importance score of feature f_i to the group of users.

We created the feature *importance matrix* separately for three different user groups: Obstetricians (U_o), Midwives (U_m), and Pregnant women (U_p).

Certain EMR features could only be experienced by the obstetricians and midwives, for example: the automatic collection of data from the medical devices used in the clinic. For these features, we did not ask the opinion of the pregnant women. Their importance scores were assigned as the average importance score given by the pregnant women for the other EMR features. The reasons and the consequences of this decision are discussed in the Discussion section.

By giving equal weight to the opinions of the three user groups, the feature *importance matrix* $U = [u_i]_n$ was calculated as:

$$U = \frac{1}{3}(U_o + U_m + U_p)$$

where u_i is the importance score of feature f_i.

2- Identifying the relationship between the EMR features

Multiple EMR features could be redesigned. One way to prioritize the EMR features to redesign is based on the users' priorities as identified in *Step 1*. However, the different EMR features are interdependent. In other terms, modifying one EMR feature could affect other features. Therefore, taking into consideration the interdependency of the features is needed to strategically prioritize them for redesign.

To identify the relationships between the EMR features, an analysis was conducted by three of the authors of this paper. We assumed that feature f_i has a negative influence on feature f_j when an improvement of f_i leads to a deterioration of f_j. Similarly, we assumed that feature f_i has a positive influence on feature f_j when an improvement of f_i leads to an improvement of f_j.

In the analysis, we rated the degree of influence between the different features. The direct influence of feature f_i on feature f_j was rated using a scale of "Very strong negative influence (-0.9)," "Moderate negative influence (-0.6)," "Weak negative influence (-0.3)," "Weak positive influence (+0.3)," "Moderate positive influence (+0.6)," "Strong positive influence (+0.9)."

Once the authors agreed over the influence ratings, the *feature dependence matrix* was formed as $D = [d_{ij}]_{n\times n}$ where all principal diagonal elements are equal to zero and d_{ij} represents the degree to which feature f_i has as influence on feature f_j.

3- Computing the overall effect of modifying each feature

In this step, we aim to calculate the overall effect of redesigning each EMR feature. The overall effect of modifying a feature f_i consists of the direct effect on f_i and the indirect effect on the other EMR features.

Let $A = [a_{ij}]_{n*n}$ be the *direct effect matrix* where a_{ij} is the direct effect that modifying feature f_i has on feature f_j. All the principal diagonal elements of A are equal to 1 and $a_{ij} = 0$ *where* $i \neq j$.

The *overall effect matrix* E is computed by summing the direct effects and the indirect effects of modifying the different EMR features as follows:

$$E = A + A \times D + A \times D^2 + \cdots + A \times D^h \,, \qquad h \to \infty$$

$E = [e_{ij}]_{n*n}$ where e_{ij} is the overall effect that modifying feature f_i has on feature f_j.

4- Computing the priority for targeting each EMR feature

To identify which features are more important to target, we take into consideration two factors: (i) the importance of the EMR features based on the users' opinions, and (ii) the overall effect of targeting the different EMR features.

Therefore, the *priority matrix* is computed by multiplying the *importance matrix* by the *overall effect matrix* : $P = U \times E$. The computation will result in $P = [p_i]_n$ where p_i represents the priority score of redesigning feature f_i.

Example to illustrate the method

Step 1.

Supposing we have two EMR features f_1 and f_2 that we could target in a redesign project.

Supposing that $\boldsymbol{U} = [\mathbf{2} \quad \mathbf{3}]$ is the *importance matrix* for features f_1 and f_2. This importance matrix indicates that the users think that f_2 is more important than f_1. Following a user-centered design approach, redesigning f_2 has a higher priority than redesigning f_1.

Step 2.

Supposing that $\boldsymbol{D} = \begin{bmatrix} \mathbf{0} & \mathbf{+0.6} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ is the *dependency matrix* of f_1 and f_2 based on the designers' analysis. In this case, f_1 has a moderate positive influence of f_2, and f_2 has no influence on f_1.

Step 3.

$\boldsymbol{A} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}$ is the *direct effect matrix*.

Consequently, $\boldsymbol{E} = \begin{bmatrix} \mathbf{1} & \mathbf{0.6} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}$ is the *overall effect matrix*. This means that modifying f_1 has a moderate indirect influence on f_2, while modifying f_2 has no influence on f_1.

Step 4.

Consequently, we obtain $\boldsymbol{P} = \boldsymbol{U} \times \boldsymbol{E} = [\mathbf{3.8} \quad \mathbf{3}]$ as the final *priority matrix*.

The priority scores in the final *priority matrix* indicate that redesigning f_1 has a higher priority than redesigning f_2. This

change in priorities between *Step 1* and *Step 4* shows the importance of considering the interdependency of the features when redesigning EMR systems.

Results

Importance of the different EMR features to the users

From previous work conducted on EMR systems in Japanese antenatal care settings [16], we identified fourteen EMR features. Based on the responses to the user surveys, we identified the importance of the different EMR features to the users.

Table 1 shows the EMR features, their assigned code names, and their importance to the users. The features are initially ranked based on their importance scores (**IR**) from highest to lowest.

U_o , U_m, and **U_p** represent the importance scores given by the obstetricians, midwives, and pregnant women, respectively. **U** is the aggregation of the scores for the three user groups.

The users agreed on importance of twelve out of fourteen features. However, the pregnant women reported that viewing their EMRs during and after the check-ups was important to them. On the other hand, the antenatal care providers did not report that feature to be important.

The overall top priorities for EMR users in Japanese antenatal care settings were: (1) *the use of the EMR system to exchange information between the antenatal care providers*, (2) *the EMR system being easy to use*, and (3) *the EMR system providing them with a quick summary of the pregnancy course*.

The overall lowest priorities for EMR users were (13) *the pregnant women viewing the EMR screen during the check-ups* and (14) *the pregnant women viewing their EMRs online*.

Relationships between the EMR features

Through the analysis conducted in *Step 2*, we identified the interdependency between the different EMR features. Figure 2 shows the map of interdependencies.

The nodes in the map represent the different EMR features. Inside the nodes, the EMR features are named using their corresponding code name shown in Table 1.

The weighted arrows connecting the nodes represent our judgement of the degree to which feature f_i influences feature f_j. The interdependence matrix $D = [d_{ij}]_{n\times n}$ was formed according to the map shown in Figure 2. Matrix D is not shown due to space constraints.

Priority for targeting each EMR feature

Through the computations conducted in *Step 3 and Step 4,* the final priority matrix was calculated. The final priority scores converged at h=10, as shown in Figure 3.

Table 1- The importance of different EMR features to the users

Code name	EMR feature	IR	U_o	U_m	U_p	U
EXCHANGE	EMR system is used to exchange information between the providers	**1**	3.8	4.0	3.46	3.75
EASY USE	EMR system is easy to use	**2**	4.0	4.0	2.67*	3.56
SUMMARY	EMR system can provide a quick summary of the pregnancy	**3**	3.8	4.0	2.76	3.52
EASY LEARN	EMR system is easy to learn	**4**	3.8	3.9	2.67*	3.46
INTEGRT	EMR system is well integrated with the medical devices in the clinic	**5**	3.8	3.6	2.67*	3.36
PSYCHO-SOCIAL	Sensitive psycho-social information is documented in detail in the EMR	**6**	3.4	3.6	2.75	3.25
NOT INT COMM	EMR system does not interrupt the communication	**7**	3.2	3.8	2.73	3.24
PREP	EMR system is used to prepare for the check-ups	**8**	3.4	3.5	2.67*	3.19
SCHEDULE	EMR system is used to manage the antenatal care appointments	**9**	3.4	3.8	2.37	3.19
NOT INT PROCESS	EMR system does not interrupt the clinical process	**10**	3.2	3.6	2.67*	3.16
EXPLN	EMR system supports the explanation	**11**	2.8	2.9	2.40	2.7
PAUSE SCREEN	EMR system is used to pause the communication when needed	**12**	2.4	2.4	2.67*	2.49
SCREEN ACCESS	EMR screen can be seen by the pregnant women during check-ups	**13**	1.6	2.3	2.61	2.17
WEB ACCESS	EMR can be accessed online by the pregnant women	**14**	1.0	2.3	2.30	1.87

* Importance scores were calculated by averaging the other importance scores reported by the pregnant women

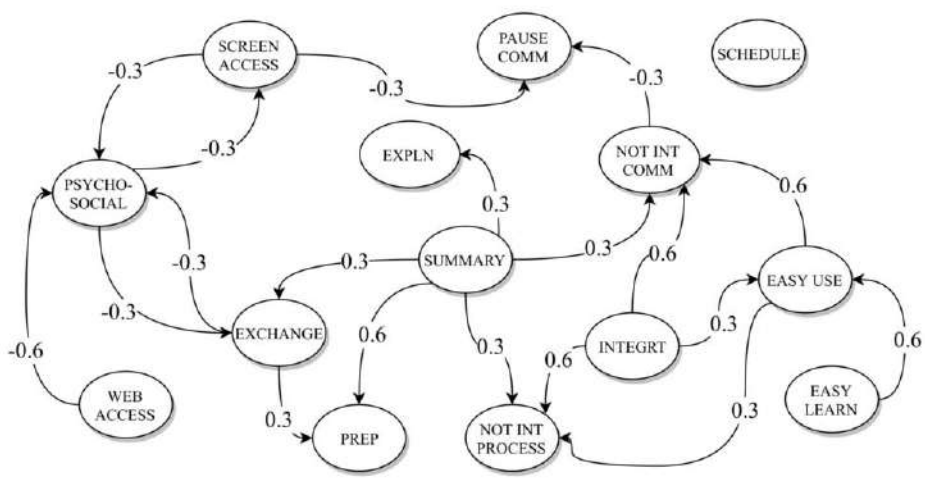


Figure 2- The map of interdependencies between the EMR feature

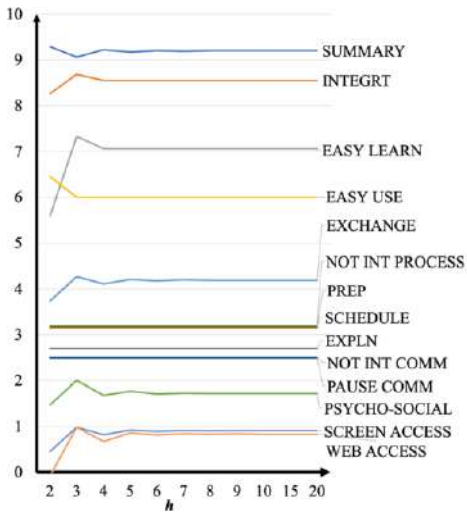


Figure 3- Convergence of the final importance scores

Table 2 shows the final priority scores (**P**) for the different EMR features after considering the feature interdependency. **IR** represents the initial ranking of the features based on the importance scores of the users. **FR** represents the final ranking of the features based on the final priority scores. **Change** represents the change in feature ranks after taking into consideration the features interdependency.

Table 2- Final ranking of EMR feature to redesign

Feature code	P	IR	FR	Change
EXCHANGE	4.19	**1**	**5**	**-4**
EASY USE	6	**2**	**4**	**-2**
SUMMARY	9.19	**3**	**1**	**+2**
EASY LEARN	7.06	**4**	**3**	**+1**
INTEGRT	8.55	**5**	**2**	**+3**
PREP	1.72	**6**	**12**	**-6**
NOT INT PROCESS	2.5	**7**	**10**	**-3**
PSYCHO-SOCIAL	3.19	**8**	**6**	**+2**
NOT INT COMM	3.19	**9**	**7**	**+2**
SCHEDULE	3.16	**10**	**8**	**+2**
EXPLN	2.7	**11**	**9**	**+2**
PAUSE COMM	2.49	**12**	**11**	**+1**
SCREEN ACCESS	0.91	**13**	**13**	**0**
WEB ACCESS	0.84	**14**	**14**	**0**

After taking into consideration the interdependency of the features, the priority ranks of the EMR features changed. The top priority EMR features to redesign in Japanese antenatal care settings was identified as (1) *the implementation of automatic summary generation in EMR systems.* The second, third, and fourth priorities were (2) *the integration of EMR systems with the used medical devices,* (3) *the improvement of the learnability of the EMR systems* and (4) *the improvement of the usability of EMR systems.*

The lowest priority EMR features to redesign were (13) *granting the pregnant viewing access to the EMR screen during the check-ups* and (14) *granting the pregnant women online viewing access to their EMR.*

Discussion

In this work, we proposed a method that allows the designers to prioritize EMR features to redesign. The method takes into consideration the priorities of the different users. By doing so, the method fits into the user-centered design paradigm where design activities respect the preferences and aspirations of the users. The method goes beyond the users' priorities and takes into consideration the relationships between the different EMR features. By doing so, the method regards EMR systems as complex dynamic systems.

We applied the method in a case study of redesigning EMR systems in Japanese antenatal care settings. In our case study, we considered three different types of users: the obstetricians, the midwives, and the pregnant women. The results of the survey showed that the three user groups have mostly overlapping priorities. However, the *ability of the pregnant women to access their EMRs* was viewed differently by the healthcare providers and the pregnant women. This shows the importance of considering the opinions of the different users when redesigning EMR systems.

The survey results showed that *the exchange of information between the different antenatal care providers* was the highest priority EMR feature for the three user groups. However, after considering the relationships between the different features, the highest priority EMR feature to redesign was found to be *the generation of summaries from EMR notes*. This change in priorities shows the importance of considering the relationships between the different features when redesigning EMR systems.

In the following paragraphs, we discuss the different design choices that we made in our case study and how these choices affected the end results.

Giving weights to the opinions of the different user groups

In our case study, we gave equal weights to the opinions of the three user groups. Some may argue that this choice is not patient-centered and that the opinions of the pregnant women should have a larger weight. For example, we could have considered the obstetricians and the midwives as one group of users: the antenatal care providers. Consequently, we could have given equal weights to the opinions of two user groups: the antenatal care providers, and the pregnant women. In this case, another question arises: how should the weights be distributed within the antenatal care providers' group? Therefore, the question of the weight for the different user groups remains open. We believe that the designers can decide on the weights depending on what they see best in their target situations.

Dealing with the missing data

As mentioned in the methods section, some EMR features cannot be experienced by the pregnant women. For these features, we did not ask for the opinions of the pregnant women in the survey. The importance score was assigned as the average of the other importance scores provided by the pregnant women for the other features. Other approaches could have been employed to address the missing scores issue. For example: we could have given those features the lowest possible importance score or only considered the opinions of the obstetricians and the midwives for these features. However, we decided to consider these features as neutral for the pregnant women by assigning them the average importance score that they have given for the other features. This decision was mainly driven by the fact that antenatal care providers gave higher importance scores on average. The other possible approaches in this case would have majorly biased the final priority scores.

Deciding on the relationships between the EMR features

Three of the authors of this paper conducted an analysis of the relationships between the features and reached an agreement. Based on their analysis, the final priorities were computed. Clearly, different designers could have come up with different

relationships, resulting in different priorities. The subjectivity of the designers is a limitation of this method. However, we argue that any design process is usually tainted with the subjectivity of the designers. Therefore, their subjectivity cannot be completely avoided, but could be reduced by defining stricter criteria on how to model and rate the influences between the different features. Further research is needed to develop a set of guidelines for designers to follow when modeling the interdependency of the features.

In other respects, the users may have considered some features' dependencies when answering the survey questions. For example, some respondents may have considered the effect of integrating the devices with the EMR system when rating its importance. Therefore, some interdependencies may be accounted for twice. Further consideration is needed to counter this effect when applying the method.

In other respects, deciding on which features to redesign could include other important criteria such as: the cost of the redesign activity, the difficulty of the redesign activity, and the alignment of the redesign activity with other pre-existing strategies. This method provides the designers with information about which feature redesign could have the largest user-desired effect on the system. Further work is needed to incorporate this method into a larger scope method that allows the designers to take into consideration other decision-making criteria when redesigning EMR systems.

Through the case study, we showed the importance of considering the different users and the feature interdependency. To understand the practical value of this method, it is important to further evaluate the method as a decision-aid tool for designers in redesign projects.

Finally, even though this method was proposed to prioritize EMR features, it can be used by designers of other complex systems with multiple user groups and interdependent features.

Conclusion

We presented a method to prioritize EMR features to redesign. The method takes into consideration the priorities of the different users and the relationships between the different EMR features. The method allows the designers to rank features based on the direct and indirect effects of targeting them. We applied the method in Japanese antenatal care settings. The top priority features to redesign were (i) the implementation of automatic summary generators in EMR systems and (ii) the integration of EMR systems with the used medical devices. The results showed the importance of taking into consideration the priorities of the different users and the interdependency of the EMR features. The proposed method could be used as a tool to support designers in making strategic decisions during the initial system redesign phase.

Acknowledgements

This work was supported by the Leading Graduate Schools Program "Collaborative Graduate Program in Design", Japan.

References

[1] B. Chaudhry, J. Wang, S. Wu, M. Maglione, W. Mojica, E. Roth, S.C. Morton, and P.G. Shekelle, - Systematic review: impact of health information technology on quality, efficiency, and costs of medical care, *Annals of Internal Medicine*. (2006) 742.

[2] A. Cornwall, Electronic health records: an international perspective, *Health Issues*. **73** (2002) 19–23.

[3] W. Hersh, Health care information technology: progress and barriers, *Jama*. **292** (2004) 2273–2274.

[4] J.S. Irani, J.L. Middleton, R. Marfatia, E.T. Omana, and F. D'amico, The use of electronic health records in the exam room and patient satisfaction: a systematic review, *The Journal of the American Board of Family Medicine*. **22** (2009) 553–562.

[5] M.L. Graber, D. Johnston, and R. Bailey, Report of the Evidence on Health IT Safety and Interventions, (2016).

[6] J. Aarts, J. Ash, and M. Berg, Extending the understanding of computerized physician order entry: implications for professional collaboration, workflow and quality of care, *International Journal of Medical Informatics*. **76** (2007) S4–S13.

[7] P. Carayon, B.-T. Karsh, R. Cartmill, P. Hoonakker, A.S. Hundt, D. Krueger, T. Thuemling, T. Wetterneck, P.F. Brennan, J. Walker, and others, Incorporating Health IT Into Workflow Redesign: Request for Information Summary Report, *AHRQ Publication*. (2010).

[8] K.M. Cresswell, A. Worth, and A. Sheikh, Integration of a nationally procured electronic health record system into user work practices, *BMC Medical Informatics and Decision Making*. **12** (2012) 15.

[9] J.S. Ash, P.N. Gorman, M. Lavelle, T.H. Payne, T.A. Massaro, G.L. Frantz, and J.A. Lyman, A cross-site qualitative study of physician order entry, *Journal of the American Medical Informatics Association*. **10** (2003) 188–200.

[10] R.H. Miller, and I. Sim, Physicians' use of electronic medical records: barriers and solutions, *Health Affairs*. **23** (2004) 116–126.

[11] A.F. Rose, J.L. Schnipper, E.R. Park, E.G. Poon, Q. Li, and B. Middleton, Using qualitative studies to improve the usability of an EMR, *Journal of Biomedical Informatics*. **38** (2005) 51–60.

[12] V. Blijleven, K. Koelemeijer, M. Wetzels, and M. Jaspers, Workarounds Emerging From Electronic Health Record System Usage: Consequences for Patient Safety, Effectiveness of Care, and Efficiency of Care, *JMIR Human Factors*. **4** (2017).

[13] M.E. Flanagan, J.J. Saleem, L.G. Millitello, A.L. Russ, and B.N. Doebbeling, Paper- and computer-based workarounds to electronic health record use at three benchmark institutions, *Journal of the American Medical Informatics Association*. **20** (2013) e59–e66.

[14] G. Marcu, A.K. Dey, and S. Kiesler, Designing for collaborative reflection, in: Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2014: pp. 9–16.

[15] S. Garde, P. Knaup, E. Hovenga, and S. Heard, Towards semantic interoperability for electronic health records, *Methods of Information in Medicine*. **46** (2007) 332–343. doi:10.1160/me5001.

[16] S. Helou, G. Yamamoto, E. Kondoh, H. Tamura, S. Hiragi, O. Sugiyama, K. Okamoto, M. Nambu, and T. Kuroda, Understanding the Roles of EMR Systems in Japanese Antenatal Care Settings., *Studies in Health Technology and Informatics*. **251** (2018) 257–260.

[17] C.M. Johnson, T.R. Johnson, and J. Zhang, A user-centered framework for redesigning health care interfaces, *Journal of Biomedical Informatics*. **38** (2005) 75–87.

Address for correspondence

Helou.samar@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190420

Towards the TIGER International Framework for Recommendations of Core Competencies in Health Informatics 2.0: Extending the Scope and the Roles

Ursula Hübner[a], Johannes Thye[a], Toria Shaw[b], Beth Elias[c], Nicole Egbert[a], Kaija Saranto[d], Birgit Babitsch[e], Paula Procter[f], Marion J. Ball[g]

[a] *Health Informatics Research Group, University of Applied Sciences Osnabrück, Osnabrück, Germany*
[b] *Healthcare Information and Management Systems Society (HIMSS), TIGER Initiative, Chicago, IL, USA*
[c] *University of Scranton, Scranton, Pennsylvania, USA*
[d] *Dept. of Health and Social Management, University of Eastern Finland, Kuopio, Finland*
[e] *Institute of Health and Education, New Public Health, Osnabrück University, Osnabrück, Germany*
[f] *Sheffield Hallam University, Sheffield, South Yorkshire, United Kingdom*
[g] *IBM Research, Baltimore, Maryland, USA*

Abstract

This paper describes the methodology and developments towards the TIGER International Recommendation Framework of Core Competencies in Health Informatics 2.0. This Framework is meant to augment the scope from nursing towards a series of six other professional roles, i.e. direct patient care, health information management, executives, chief information officers, engineers and health IT specialists and researchers and educators. Health informatics core competency areas were compiled from various sources that had integrated the literature and were grouped into consistent clusters. The relevance of these core competency areas was rated in a survey by 718 professional experts from 51 countries. Furthermore, 22 local case studies illustrated the competencies and gave insight into examples of local educational practice. The Framework contributes to the overall discourse on how to shape health informatics education to improve quality and safety of care by enabling useful and successful health information systems.

Keywords:

medical informatics, competency-based education, health professions

Introduction

The notion that health information systems success does not only depend on good technology but also on capable users has been accepted for many years in the context of participatory systems design [1] and the task-technology-individual models [2]. However, it has received momentum only in the last decade with increasing health IT adoption rates world wide [3]. Thus, realistic scenarios for using IT on a large scale emerged and necessitate a competent workforce. While clinical end users and their practical skills are widely discussed, health informatics competencies of IT decision makers, often at the board level, have only recently received attention [4]. Hence, health informatics covers a wide field of competencies for all different types of professionals in the healthcare arena, at different levels, and is connected with the need for life long learning. Against this backdrop, national and international professional and scientific associations are undertaking great efforts to develop [5-7] and update educational recommendations [8]. TIGER (Technology Informatics Guiding Education Reform) started issuing recommendations for basic IT competencies for nurses [9] and moved on to develop a comprehensive framework of recommendations of health informatics. This framework accounts for the increasing complexity and sophistication of competencies needed for different roles in nursing [10]. However, it is restricted to nursing and does not include other professions. The primary goal of this study, therefore, was to extend the scope of the framework of recommendations of health informatics core competencies beyond nursing and to include further professional roles that contribute to the success of health information systems. Hereby, the relevance of pertinent core competency areas for the different roles should be designated and illustrated. The secondary goal was to reconcile the global perspective on educational recommendations with the local perspective reflecting an in-depth view and practical experience. As the approach for this framework should pursue a very similar rationale and methodology we decided to call it "International Recommendation Framework of Core Competencies in Health Informatics 2.0". This paper describes major milestones towards this framework.

Methods

The development of the International Recommendation Framework of Core Competencies in Health Informatics 2.0 was performed under the umbrella of the EU Horizon 2020 project EU*US_eHealth_Work addressing workforce development and of TIGER (Technology Informatics Guiding Education Reform). TIGER is a grassroots initiative formalised in 2006 within the nursing community before transitioning to the Healthcare Information and Management Systems Society (HIMSS) in 2014 with members in 26 countries worldwide. TIGER now embraces an interprofessional focus that covers a great field of different health care professionals. In order to address the **first goal**, six professional roles were identified that belong to the communities of either IT users, IT decision makers, IT technologists or IT researchers & educators. Knowing that these communities overlap we more specifically defined the roles as (1) direct patient care (mainly physicians, nurses, therapists) and (2) health information managers belonging to the users, (3) executives (clinical and administrative) representing the decision makers, (4) chief information officers (technical and clinical) in their dual position as decision makers and technologists, (5) engineers and health IT specialists as members of technologists, and

finally (6) science & education forming the group of researchers and educators. The core competencies that should be assigned to the professional roles were compiled from the Recommendation Framework for nurses, which integrated existing recommendations of well-known scientific and professional associations, e.g. IMIA [5] and AMIA [7], and from the HITCOMP tool [11]. In comparison to the Recommendation Framework for nurses the following new competency areas were included: public health informatics, consumer health informatics and learning techniques. In addition, other competency areas were now marked as separate areas, i.e. communication, healthcare processes & IT integration, legal issues, interoperability & integration, and life cycle mangement. Previously these areas had been subsumed under other areas in the Framework for nurses. Information and communication technology was split into two areas: applications and architectures. Finally, three areas were rephrased to be more general. These changes were made to adapt the list to the broader professional scope. This adaptation was performed by four experts that mapped both competency lists, i.e. the nursing one [10] and the one from the HITCOMP tool [11], and finally agreed on utilising 33 core competency areas for the Framework 2.0 (see Table 1).

Table 1 – Core Competency Areas in Alphabetical Order

Applied computer science	Interoperability and integration
Assistive technology	IT risk management
Change/stakeholder management	Leadership
Clinical decision support by IT	Learning techniques
Communication	Legal issues in health IT
Consumer health informatics	Medical technology
Data analytics	Principles of health informatics
Data protection and security	Principles of management
Documentation	Process management
e/mHealth, telematics, telehealth	Project management
Ethics in health IT	Public health informatics
Financial management	Quality and safety management
Care processes and IT integration	Resource planning & management
ICT / systems (applications)	Strategic management
ICT / systems (architectures)	System lifecycle management
Information management research	Teaching, training, education
Information and knowledge management in patient care	

In order to obtain the relevance ratings for the core competence areas, a questionnaire was developed that amongst others included a section on competencies. Initially, twelve different roles were distingushed that were then grouped to match the six different roles of healthcare professionals mentioned above. Similar to the Framework for nurses, relevance ratings could range from 0 to 100. The survey participants rated the relevance of competencies for those professional roles they were competent to speak to. Explanations of what the core competency areas embraced were included in the questionnaire. After pretesting, the survey was finally made available online from the middle of February to the end of June 2017. As the relevance rating should yield a global picture, the survey link was deployed via 60 global listservs comprised of individuals and international, European and North American organisations that represent healthcare professionals. The organizations were asked to share it amongst their members. Due to this deployment policy the people who were invited could not be exactly specified by number. As the same data had been used to analyse options for interprofessional education, the survey methodology had been described also in [12] with a slightly different focus and with clusters of professions that partly varied from this approach.

The **second goal** of this study was to exemplify and illustrate the global findings with a local perspective. To this end, case studies from different countries were identified to illustrate the core competence areas by practical and detailed descriptions of individual competencies. As it was necessary to obtain comparable descriptions with a similar focus and structure, a template for reporting the case was developed that incorporated the principles of case studies [13]. The template was divided into sub-sections for: author, organisation, background, status of current developments, activities and measures, changes, results and outlook and lessons learnt. There was a checklist of eHealth topics alligned with the list of core competency areas of the questionnaire, e.g. process and workflow management, consumers and populations, research, data science, ethics, legal and data protection. The template also included a checklist to aid the case study authors referring to crucial areas, e.g. teaching the teachers, integrating health informatics into traditional curricula and motivating clinicans and managers.

In order to obtain authentic and first hand information, it was decided that the case descriptions should be provided by the persons who were actually involved in this case, e.g. as developer of the educational programme and/or as teacher in this programme. The recruitment of case study authors was initialised by an open call that was launched in July 2017 and closed in January 2018. The call was made public via HIMSS community listservs, the European Health Telematics Association (EHTEL) and other EU*US_eHealth_Work project members. A total of 214 individuals from all around the world were personally invited aiming at experts affiliated with major leading institutions in their field. In addition, general invitations via the HIMSS listservs, national and international conferences, e.g. MEDINFO 2017, were made public. Upon receipt of the case study manuscript the descriptions were edited by authors of this paper (TS, BE, UH) in cooperation with the case study authors.

Results

Goal 1: Relevance of core competency areas. A total of 718 experts from 51 countries provided answers to the questions on health informatics competencies. The 51 countries were composed of 28 European countries, 10 Asian countries, 8 countries from Middle and South America, 2 African countries, the USA, Canada and Australia (see [4] for further demographics). These answers corresponded with 1,571 relevance ratings for professional roles. Out of these answers, 27 were excluded either because they addressed health professions not meant to be focused on in this Recommendation Framework or received not enough answers, e.g. pharmacists. Table 2 gives an overview of the top 10 core competency areas and their mean relevance for the six professional roles. Relevance means were high among the top 10 across all roles ranging from 96.4 to 81.1. Each role was characterised by a unique pattern of top 10 competency areas out of which the first three were separately marked (see Table 2). Communication appeared in all six roles among the top 3. Other core competency areas in the top 3 were more role specific. Direct patient care for example was further featured by documentation and information & knowledge management while the executive role was characterised by leadership and quality & safety management. Beyond the rather distinct role profiles it is noteworthy that some core competency areas were shared by a majority of the roles: Among the top 10, leadership and ethics in health IT appeared in all six roles, quality & safety management, documentation and care processes & IT integration in four out of the six roles. All other competency areas were more specific and described only three or fewer roles. As the absolute relevance ratings ranged in a rather small interval from about 10 points per role (e.g. for direct patient care from 92.4. to 81.1) among the top 10, there were sometimes only minimal

differences between one rank and the next one. For example in the role of engineers and health IT specialists, documentation on rank 9 received a mean relevance of 82.1 and process management on rank 10 had a mean value of 82.0.

Table 2- Top 10 Core Competency Areas in the Six Roles and Related Mean Relevance (REL - 0…100)

	Direct patient care (DPC) (nurses/physicians/therapists)	
	Core competencies	***REL ± SD***
1	*Communication* [n=335]	92.4 ± 14.5
2	*Documentation* [n=337]	91.7 ± 17.2
3	*Information & knowledge management in patient care* [n=335]	90.0 ± 17.5
4	Quality & safety management [n=333]	87.5 ± 18.9
5	Leadership [n=336]	86.2 ± 19.0
6	Learning techniques [n=334]	85.6 ± 18.8
7	Teaching, training & education in healthcare [n=333]	84.4 ± 21.0
8	Ethics in health IT [n=334]	83.8 ± 23.0
9	Information & communication technology (applications) [n=332]	81.6 ± 20.5
10	Care processes & IT integration [n=333]	81.1 ± 21.3
	Health information management (HIM)	
	Core competency area	***REL ± SD***
1	*Communication* [n=184]	90.1 ± 19.0
2	*Documentation* [n=184]	87.7 ± 18.0
3	*Data analytics* [n=183]	87.7 ± 17.9
4	Leadership [n=184]	87.0 ± 19.0
5	Data protection & security [n=184]	86.9 ± 19.3
6	Information & knowledge management in patient care [n=182]	86.2 ± 19.4
7	Ethics in health IT [n=184]	85.6 ± 20.2
8	Principles of health informatics [n=182]	85.1 ± 18.4
9	Care processes & IT integration [n=183]	84.8 ± 19.1
10	Learning techniques [n=184]	84.2 ± 20.2
	Executives (EXC) (clinical and administrative)	
	Core competency area	***REL ± SD***
1	*Leadership* [n=55]	96.4 ± 7.8
2	*Communication* [n=55]	95.8 ± 8.3
3	*Quality & safety management* [n=55]	90.4 ± 16.1
4	Information & knowledge management in patient care [n=55]	89.2 ± 16.9
5	Strategic management [n=55]	89.1 ± 21.0
6	Principles of management [n=55]	88.6 ± 20.ß
7	Legal issues in health IT [n=55]	87.6 ± 16.3
8	Process management [n=55]	87.5 ± 16.4
9	Resource planning & management [n=55]	87.3 ± 21.7
10	Ethics in health IT [n=55]	87.0 ± 18.3
	Chief information officers (CIO) (clinical and technical)	
	Core competency area	***REL ± SD***
1	*Leadership* [n=62]	93.8 ± 9.6
2	*Communication* [n=62]	93.2 ± 10.7
3	*Care processes & IT integration* [n=62]	91.8 ± 13.7
4	Principles of management [n=61]	90.8 ± 12.2
5	Quality & safety management [n=61]	90.5 ± 12.7
6	Strategic management [n=61]	90.0 ± 13.4
7	Process management [n=62]	89.6 ± 13.6
8	Change & stakeholder management [n=61]	89.6 ± 12.6
9	Ethics in health IT [n=61]	88.7 ± 18.0
10	Resource planning & management [n=61]	88.4 ± 18.7
	Engineering or health IT specialist (ENG)	
	Core competency area	***REL ± SD***
1	*Communication* [n=172]	91.3 ± 14.3
2	*Care processes & IT integration* [n=171]	87.5 ± 18.9
3	*Information & communication technology (applications)* [n=171]	87.2 ± 18.0
4	Leadership [n=172]	86.1 ± 17.8
5	Project management [n=172]	85.4 ± 19.7
6	Data protection & security [n=171]	84.3 ± 22.6
7	Ethics in health IT [n=170]	83.4 ± 22.2
8	Interoperability & integration [n=172]	83.0 ± 21.7
9	Documentation [n=172]	82.1 ± 22.6
10	Process management [n=172]	82.0 ± 21.7
	Science and education (S&E)	
	Core competency area	***REL ± SD***
1	*Communication* [n=218]	91.6 ± 16.1
2	*Teaching, training & education in health care* [n=220]	89.2 ± 17.9
3	*Leadership* [n=218]	88.2 ± 17.3
4	Learning techniques [n=218]	88.1 ± 18.8
5	Ethics in health IT [n=219]	86.5 ± 21.3
6	Documentation [n=222]	86.3 ± 21.2
7	Information & knowledge management in patient care [n=221]	86.3 ± 20.2
8	Principles of health informatics [n=218]	83.3 ± 23.2
9	Quality & safety management [n=220]	83.1 ± 22.9
10	Data analytics [n=218]	81.9 ± 23.6

In order to further group the core competency areas, clusters used in the Recommendation Framework for nurses were tested for consistency and adapted if needed. Table 3 shows the Cronbach's alpha values for the six roles and seven clusters, i.e. data/information/knowledge (DIK), information exchange/ information sharing (IEIS), ethical/legal issues (EL), systems/ system principles (SYS), management (MAN), technology (TECH) and teaching/learning (LRN). Only data analytics (STAT) was not assigned to any of the clusters and stands on its own. Thus, a consistency check was not necessary. The great majority of alphas received acceptable values, well above or very close to 0.7, hinting at consistent clusters. This held not true only for three role cluster combinations. The clusters were constructed with some overlap [10], e.g. public health informatics was assigned to DIK and IEIS.

Table 3 - Cronbach's Alpha Values for the Roles and Clusters (No. Core Competency Areas)

Clusters	**Roles**					
	DPC	ENG	HIM	EXC	CIO	S&E
DIK (8)	0.86	0.88	0.90	0.86	0.82	0.92
n	322	161	174	54	61	211
IEIS (8)	0.88	0.88	0.91	0.91	0.88	0.92
n	321	160	171	54	59	207
EL (3)	0.82	0.87	0.90	0.79	0.87	0.89
n	330	169	182	55	61	217
SYS (4)	0.85	0.85	0.88	0.90	0.85	0.91
n	324	167	176	54	61	212
MAN (10)	0.92	0.92	0.95	0.92	0.92	0.95
n	326	166	175	54	61	212
TECH (2)	0.49	0.71	0.65	0.68	0.73	0.76
n	325	163	175	55	59	211
LRN (2)	0.68	0.57	0.83	0.63	0.81	0.80
n	332	166	181	54	62	218

Goal 2: Illustration of the core competency areas. So far a total of 22 case studies from 19 countries were obtained which covered the views from universities (15), from hospitals, (3) from the perspective of countries (3) and one from an educational IT system (decision support). The university courses described offered education at the level of Bachelor, Master and continuing education programmes. The educational activities of hospitals targeted workforce development while the country perspectives reflected needs and national programmes to establish and deepen health informatics education. Due to the nature of health informatics, all case studies blended technical and health topics, however, with various foci and targeting different roles. The majority (17) addressed students and professionals in Direct Patient Care (DPC) either as the only role or in combination with a different role. Six case studies covered the Executive role (EXC) and four the Chief Information Officer role (CIO) either alone or in

combination. Two cases exemplified a curriculum focusing on engineers/health IT specialists together with other roles. Finally, one case study described the need for health informatics in general irrespectively of a dedicated role. Currently all 22 case studies are available from [14]. There were 7 dedicated interprofessional cases and others also stressed the importance of mutual exchange between the students and professionals. In the following some examples are given to show how the case studies illustrate the core competency areas and break them down into individual competencies.

Case Study 1: Indiana University School of Informatics and Computing, Indianapolis, Indiana, United States (Josette Jones)

Case study 1 describes a module-based flexible workforce training program with 21 one-credit modules for anyone who needs training in health informatics, in particular students from health professional programs (e.g. physicians, nurses, public health), professional health care staff members (e.g. from patient centred medical homes, community health centres). It thus addresses all professional roles of this Framework. The following competencies belong to the areas systems/system principles (SYS) and data/information/ knowledge (DIK):

EHR systems development & implementation: Identify the range of clinical decision support (CDS) tools within the EHR; determine which tool is appropriate for specific situations; analyze how to develop and implement CDS tools to adhere to meaningful use criteria. Describe the processes of developing or selecting an EHR system, preparing and supporting clinicians for system implementation and evaluating system effectiveness. Clinical data and clinical process modeling; Technical security applications and issues; Systems testing and evaluation.

Case Study 2: Laurea University of Applied Sciences and Arcada University of Applied Sciences, Finland; Tartu Health Care College, Estonia; Red Cross Medical College of Rīga Stradiņš University, Latvia (Outi Ahonen, Jonas Tana, Gun-Britt Lejonqvist, Marge Mahla, Sanita Marnauza, Elina Rajalahti)

The curriculum, whose development was funded by the EU Central Baltic Program 2014-2020, is multi-professional and combines health and welfare with IT and service design. In the three study units (15 credit points), future professionals from different fields of study (IT, social care, economics and health care) are developing their own unique competencies according to the pedagogical principle "learning by developing". The following example of competencies is taken from unit 2 that focuses on the cluster ethical/legal issues (EL):

Understand ethical theories, safety procedures, principles and laws affecting digital health and welfare as well as customer privacy. Have the skills to practice ethical and high quality customer service taking responsibility for the safety and integrity of the client.

Case Study 3: Assuta Medical Centers, Israel (Rachelle Kaye)

The main drivers at Assuta Medical Centers, the largest private hospital system in Israel, for process changes and associated skills and capabilities, including eHealth competencies, is the striving to steadily improve the quality of care. Continuing education, hereby, is devided into developing basic, intermediate and advanced skills and competencies and is meant to reach all professionals within Assuta. Assuta, which publishes a professional journal, therefore emphasises analytical competencies (STAT) as the following example taken from advanced skills and capabilities shows:

Research and Data Analytics: Perform digitally supported research and database research, or data analytics, design database for research purposes, on-going management and patient care improvement.

The three case studies were chosen against the background to illustrate and reflect ongoing activities to increase the health informatics competencies of all healthcare professionals in a process of life long learning. The case studies were also selected on the basis to represent countries with a high adoption rate of health IT.

Discussion

The current state of the "International Recommendation Framework of Core Competencies in Health Informatics 2.0" developed by TIGER within the EU*US_eHealth_Work project utilizes a robust methodology of surveying healthcare stakeholders across countries worldwide about the relevance of core competency areas and is grounded on a rigorous method to obtain comparable local exemplar case study descriptions. The methodolody was rooted in the approach pursued by the Recommendation Framework for nurses [10] and was further developed regarding the breadth of core competency areas included, the outreach to obtain views from all around the world including Africa and the highly systematised manner of case study descriptions. The recommendation framework is meant to serve as a compass for teachers, students and healthcare organisations to identify patterns of core competency areas and practical advice how the competencies are embedded in a curriculum and realised in a local setting.

The relevance findings point to the paramount importance of communication as the connecting link between different stakeholders with various interests (silo mentality), different settings (primary, secondary vs. tertiary care) and other types of fragmentation. Communication is coupled with leadership, another competency area that runs like a golden thread through the relevance ratings across the roles. It is noteworthy that leadership is not only esteemed relevant at the board level but at all levels and goes along with different professional scopes. This finding matches the increasing awareness of intrapreneurship [15] as a key factor for health IT success. It describes the capability of individuals to assume responsibility, initiate projects and become an innovation champion. Communication and leadership as drivers for health IT correspond with the knowledge about the ethical constraints and skills how to balance diverging interests. Ethics in health IT was thus also found relevant for all roles and is illustrated by the Finish, Estonian and Latvian curriculum that dedicated one out of 3 modules to this topic.

At a more aggregated level, these results concur with four out of the ten foundational domains identified by the latest AMIA white paper on core competencies at master's degree level [8] that revolve around social and behavioural science/aspects and leadership. The IMIA recommendations from 2010 [5] mentioned socio-organizational and socio-technical issues and ethical and security issues as two areas from a list of 19 biomedical/health informatics core knowledge and skill areas for IT users and biomedical and health informatics specialists.

Among the information systems core competency areas care processes & IT integration was found to be essential not only for technically oriented roles but also for direct patient care. This demonstrates that the stakeholders must possess knowledge that crosses the health – technology boarder. Data analytics seems to be an emerging field in the age of Big Data,

however, not yet found central for all roles. In countries with a complete adoption of electronic health records, data analytics is just the next step in digitisation. The comprehensive utilisation of the data is very well illustrated by case study 3 from Israel.

A question that often arises in the context of education is whether interprofessional courses are meaningful. Judging by the core competency areas shared across the professional roles, interprofessional approaches seem feasible and are current practice as a series of the case studies demonstrated. A specific analysis on this topic based on the same data [12] had shown that the relevance ratings between nurses and physicians did not differ significantly, thus supporting this option. Also many of the case studies addressing direct patient care did not specifically distinguish between the professions working directly with patients. This discussion is further fueled by the demand of joining health and social care [16] in particular for the elderly and other vulnerable groups.

There are some limitations that need to be deliberated. Although the survey findings embraced the voice of experts from 51 countries, the very large majority came from North America and Europe. Thus, a bias towards industrialised countries cannot be excluded. This bias can be partly mitigated by including case studies from as many countries as possible. Indeed, it was possible to garner case descriptions from China, India, Saudi Arabia and Nigeria amonst others. However, more insight into local educational practice is required to complete the Framework. These case studies should not only represent more countries but also cover all professional roles in an even manner. Similar to the Framework for nurses an expert workshop with discussions on the roles, the core competency areas and the relevance ratings is desirable. These activities constitute the next steps towards finalising the Framework.

Conclusions

The TIGER International Recommendation Framework of Core Competencies in Health Informatics 2.0 is based on a proven methodology and well on its way with global findings and local exemplar case studies. It contributes to the overall discourse how to shape health informatics education. Furthermore, these findings should help stimulating the discussion within IMIA's work on educational recommendations.

Acknowledgements

This study was funded by the European Union's Horizon 2020 (grant: 727552 EUUSEHEALTHWORK) and by the German BMBF (grant: 16OH21026 KeGL). We wish to thank the case study authors for their invaluable contributions in particular Outi Ahonen, Elske Ammenwerth, Juris Bārzdiņš, Alexandrina Maria Ramos Cardoso, Jan Florin, Theofanis Fotis, Josette Jones, Taghreed Justina, Rachelle Kaye, Ulla-Mari Kinnunen, Gun-Britt Lejonqvist, Jessica Liston, Inge Madsen, Marge Mahla, Sanita Marnauza, Anne Moen, Lynn Nagle, Siobhán O'Connor, Omotayo Omojola, Elina Rajalahti, Ann Kristin Rotegård, Sabu K M, Paulino Souza, Jonas Tana, Helena Blažun Vošner and Zhuang Yiyu. In addition, we thank Rachelle Blake for integrating the HITCOMP perspective into this work.

References

[1] C. Sjöberg and T. Timpka, Participatory design of information systems in health care, *J Am Med Inform Assoc* **5** (1998), 177-83.

[2] E. Ammenwerth, C. Iller, and C. Mahler, IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study, *BMC Med Inform Decis Mak* **6** (2006), 3.

[3] J. Zelmer, E. Ronchi, H. Hyppönen, F. Lupiáñez-Villanueva, C. Codagnone, C. Nøhr, U. Huebner, A. Fazzalari, and J. Adler-Milstein, International health IT benchmarking: learning from cross-country comparisons, *J Am Med Inform Assoc* **24** (2017), 371-379.

[4] J. Thye, U. Hübner, J. Hüsers, and B. Babitsch, IT Decision Making in German Hospitals - Do CEOs Open the Black Box? *Stud Health Technol Inform* **243** (2017),112-116.

[5] J. Mantas, E. Ammenwerth, G. Demiris, A. Hasman, R. Haux, W. Hersh, E. Hovenga, K.C. Lun, H. Marin, F. Martin-Sanchez, G. Wright, Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics, *Methods Inf Med* **49** (2010), 105–120.

[6] Global Health Workforce Council (GHWC), Global Academic Curricula Competencies for Health Information Professionals, Chicago: The AHIMA Foundation, http://www.ahima.org/about/global/global-curricula [cited Nov 18th, 2018].

[7] C.A. Kulikowski, E.H. Shortliffe, L.M. Currie, P.L. Elkin, L.E. Hunter, T.R. Johnson, et al. AMIA Board white paper: definition of biomedical informatics and specification of core competencies for graduate education in the discipline, *J Am Med Inform Assoc* **19** (2012), 931–938.

[8] A. L. Valenta, E.S. Berner, S.A. Boren, et al. AMIA Board White Paper: AMIA 2017 core competencies for applied health informatics education at the master's degree level, *J Am Med Inform Assoc* **25** (2018), 1657-1668.

[9] B. Gugerty and C.W. Delaney, TIGER Informatics Competencies Collaborative (TICC) Final Report, 2009, http://tigercompetencies.pbworks.com/f/TICC_Final.pdf [cited Nov 18th, 2018].

[10] U. Hübner, T. Shaw, J. Thye, N. Egbert, H.F. Marin, P. Chang, et al. Technology Informatics Guiding Education Reform – TIGER, *Methods Inf Med* **57** (2018), e30-e42.

[11] HITCOMP – Health IT Competencies, http://hitcomp.org/ [cited Nov 18th, 2018].

[12] J. Thye, T. Shaw, J. Hüsers, M. Esdar, M. Ball, B. Babitsch, and U. Hübner, What Are Inter-Professional eHealth Competencies? *Stud Health Technol Inform* **253** (2018), 201-205.

[13] S. Crowe, K. Cresswell, A. Robertson, G. Huby, A. Avery, and A. Sheikh, The case study approach, *BMC Med Res Methodol* **11** (2011), 100.

[14] EU*US eHealth Work Project's Global Case Studies, https://www.himss.org/professional-development/tiger-case-studies [cited Nov 18th, 2018].

[15] J.D. Liebe, M. Esdar, J. Thye, and U. Hübner, Antecedents of CIOs' Innovation Capability in Hospitals: Results of an Empirical Study, *Stud Health Technol Inform* **243** (2017), 142-146.

[16] P.M. Procter, Ubiquitous Adoption of Innovative and Supportive Information and Communications Technology Across Health and Social Care Needs Education for Clinicians, *Stud Health Technol Inform* **235** (2017), 358-362.

Address for Correspondence

Ursula Hübner, PhD, University AS Osnabrück, Caprivistraße 30A, 49076 Osnabrück, Germany, E-mail: u.huebner@hs-osnabrueck.de (prefered), phone +49 541 969-2012.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190421

E-Consent for Data Privacy: Consent Management for Mobile Health Technologies in Public Health Surveys and Disease Surveillance

Leonardo H. Iwaya[ab], Jane Li[b], Simone Fischer-Hübner[a], Rose-Mharie Åhlfeldt[c], Leonardo A. Martucci[a]

[a] *Department of Mathematics and Computer Science, Karlstad University, Karlstad, Sweden,*
[b] *Australian e-Health Research Centre, CSIRO, Marsfield, NSW, Australia,*
[c] *School of Informatics, University of Skövde, Skövde, Sweden*

Abstract

Community health workers in primary care programs increasingly use Mobile Health Data Collection Systems (MDCSs) to report their activities and conduct health surveys, replacing paper-based approaches. The mHealth systems are inherently privacy invasive, thus informing individuals and obtaining their consent is important to protect their rights to privacy. In this paper, we introduce an e-Consent tool tailored for MDCSs. It is developed based on the requirement analysis of consent management for data privacy and built upon the solutions of Participant-Centered Consent toolkit and Consent Receipt specification. The e-Consent solution has been evaluated in a usability study. The study results show that the design is useful for informing individuals on the nature of data processing, allowing them to make informed decisions.

Keywords:

Mobile health, Privacy, Public surveillance

Introduction

Privacy protection of personal data (or "data privacy") is fundamental for developing sustainable mobile health (mHealth) technologies – the use of mobile devices to support the delivery of healthcare. This is particularly important for the initiatives in low- and middle-income countries where hundreds of mHealth projects have been developed to support the care of HIV, malaria, tuberculosis, diabetes and antenatal care [1]. As these projects involve the processing of highly sensitive personal data at a large scale, data privacy becomes one of the main challenges for gaining public trust, deploying and scaling-up systems [2].

The research field of privacy and data protection has evolved dramatically in the past decades, in both legal and technical terms, and has drawn attention to healthcare applications. Most recently, the European General Data Protection Regulation (GDPR) [3] has advanced the safeguards of people's right to privacy in the digital world. Privacy-enhancing technologies have been developed to improve transparency, intervenability, and security of the digital systems.

While the adoption of mHealth systems in low- and middle-income countries is growing, researchers have pointed out that a knowledge gap exists in the realms of privacy and mHealth [2][4]. What seems to be missing are more concrete examples on *how* to integrate best practices of data privacy in the existing mHealth initiatives. That is, providing mHealth practitioners with realistic recommendations on how to adhere to privacy laws, as well as how to engineer privacy and use existing privacy-enhancing technologies into their systems.

The research in this paper addresses the needs to support consent for data privacy in mHealth. Consent is traditionally a legal requirement for healthcare interventions and clinical trials, embodying the respect for patients' autonomy and dignity. This has been extended to the digital world to safeguard people's right to privacy. The GDPR uses the concept of "informed consent" as one of the main legal grounds for processing of personal data, i.e., data subjects should be able to determine when, how, and to what extent their information is communicated to others. Informed consent enables people to make decisions before any personal data is collected. More specifically, we have investigated an electronic consent (e-Consent) solution to support the consent management for Mobile Health Data Collection Systems (MDCSs). MDCSs have been widely used for conducting health surveys and surveillance in the primary care [5]. They are used by Community Health Workers (CHWs) to gather and report data as part of their care activities. Tablet- and smartphone-based systems (e.g., GeoHealth MDCS, Open Data Kit, Open Smart Registration platform) [5] have been developed to streamline the data collection to replace the paper-based approaches in many countries. This move has raised a particular challenge in technology design to support the consent process.

In this paper, we present our work in the design of an e-Consent solution to support informed consent in MDCSs. We first review related work and describe our research methods. We then analyse existing legal and technical requirements. We focus on presenting the e-Consent design which addresses the strategies on how to inform the data subject and how to handle and store the consent. We then present the usability study and results. Our work contributes to the research of protection of personal data with an e-Consent design which can be integrated into MDCSs and used by CHWs in primary care.

Related Work

Innovative solutions have been proposed to safeguard people's privacy when giving consent in digital applications. One of them is the Participant-Centered Consent (PCC) toolkit [6] which is designed to obtain informed consent from research participants. The interface guides the users through the consent process to allow them to make an informed decision. The toolkit has been incorporated in the Apple's ResearchKit and used by various application developers. The consent also needs to be captured in an appropriate data structure and managed by the data collection platform (e.g., MDCS). To do so, the Consent Receipt specification [7] defines a record of consent granted by an authority. This record can be portrayed in a human-readable and machine-readable format. The Consent Receipt includes a link to the existing privacy policy as well as a description of what information has been or will be col-

lected. It also states the purposes for data collection and relevant information about how that information will be processed or disclosed, and how long it is valid (i.e., expiration date). As a result, the Consent Receipt specification promotes interoperability with a data structure for representing consent in compliance with current privacy and data protection laws.

To the best of our knowledge, the development of e-Consent for MDCSs is relatively unexplored. Our work leverages on the PCC toolkit and the Consent Receipt specification. However, the PCC toolkit is designed primarily for app-mediated research and MDCSs require a more sophisticated consent interface which the existing PCC toolkit does not provide. That is especially in terms of selective consent and withdrawal, allowing data subjects to agree only to specific purposes. The Consent Receipt specification also needs to be extended to allow consent on behalf of minors (e.g., by a parental figure) or on behalf of people unable to give consent due to mental or physical limitations (e.g., by a guardian or health worker).

Methods

Our research had two main phases: the design of the e-Consent, and the usability evaluation of the e-Consent. During the first phase, the e-Consent interface was designed through an iterative approach. Based on prior work on MDCSs [5] and current privacy law (GDPR [3]), the legal and technical requirements for the e-Consent were first defined. This was followed by an investigation of how to use Consent Receipts [7] as a data structure to handle consent in the system. The e-Consent mock-up interfaces were then developed using the Mockplus prototyping tool. During the second phase, the e-Consent interface was evaluated in a usability study (approved by CSIRO's Human Research Ethics Committee Nr: 117/18). The study used a mixed method of cognitive walkthrough [8], questionnaire and interview in an experimental setting. Participants were researchers in a lab and had related experience in digital technologies in healthcare. Potential participants were invited to participate via email. Each experimental session involved one participant. During the cognitive walkthrough, the participants interacted with the interface by playing a role of patient or family member being enrolled in the primary care program and giving consent. After the walkthrough, each participant completed a paper-based questionnaire and attended a debrief interview session to talk about their experience with the interface. The questions asked in the questionnaire and interviews focused on assessing the e-Consent interface regarding the principles of informed consent as defined by Friedman *et al* [9], including information disclosure, comprehension, voluntariness and agreement. Table 1 summarizes a series of statements in the questionnaire which used 5-point Likert scale. The questionnaire also included a knowledge comprehension quiz in which the participants were asked 5 questions (Table 2) to reflect their understanding on the information provided in the e-Consent interface.

Results

Legal and Technical Requirements

The analysis of the legal and technical requirements for consent has served as the first step and foundation for our design, outlined in this section. The GDPR [3] introduces a higher standard for consent. Its Article 4(11) defines it as: *'consent' of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her*. If compared to the Friedman's principles of informed consent [9], *freely given* refers to *voluntariness* and *agreement*, while *specific*, *informed* and *unambiguous* relates to *information disclosure* and *comprehension*. In addition, according to Art. 7 GDPR [3], the conditions for consent are: the controller has to demonstrate that the data subject has consented (i.e., burden of proof); the request for consent must be clearly distinguishable from other matters; the data subject shall have the right to withdraw his or her consent at any time; the performance of a contract may not be made dependent upon the consent to process further personal data, which is not needed for the performance of that contract (e.g., "prohibition of coupling" of consent). The GDPR also includes other recitals and articles, covering: conditions for consent for scientific research, child's consent, freely given consent, and burden of proof (GDPR Recitals (32, 33, 38, 42, 43) [3]).

Table 1 – Questionnaire Part 1: Principles

Information disclosure	**[Data]** I know what personal information will be collected by the system. **[Use]** I know how my personal information will be used. **[Access]** I know who will have access to my personal information. **[Retention]** I know for how long my information will be stored or archived in the systems. **[Protection]** I know how they will protect my information.
Voluntariness	**[Forced]** I felt like forced or coerced to provide consent. **[Inf. Manip.]** I could identify some forms of manipulation in the options and information provided. **[Psy. Manip.]** I could identify some forms of psychological manipulation during the process.
Agreement	**[Accept/Decline]** It is clear whether I can accept or decline the consent. **[Withdraw]** It is clear that I can revoke or withdraw consent any time.

Table 2 – Questionnaire Part 2: Comprehension Quiz

Comprehension	What is (are) the purpose(s) of collecting my personal health data? (**3 points**) **[P1]** Support primary care teams in the provision of health care (Yes or No) **[P2]** Support public health programs (Yes or No) **[P3]** Support research in public health and clinical practice (Yes or No) **[Access]** My personal data can only be accessed by authorised personnel. (**1 point**, Yes or No) **[Sharing]** I can choose to share my data with qualified researchers. (**1 point**, Yes or No) **[Skip]** I will be able to skip any question during the surveys. (**1 point**, Yes or No) **[Stop]** I will be able to stop participating at any time. (**1 point**, Yes or No)

Besides legal requirements, we also elicited technical requirements based on our experiences with MDCSs [4, 5]. The current e-Consent tool was designed considering the Brazilian Community-Based Primary Health Care (CBPHC) program, Family Health Strategy [10], in which MDCSs have been increasingly adopted. CBPHC programs often rely on a broad and implicit consent for the processing of personal data. By receiving the CHWs at their homes, the families implicitly accept the health service and agree with the data collection. However, Art. 9 2. (a) requires explicit consent for the pro-

cessing of special categories of personal data (e.g., health data). This is a specific consent for data processing of personal data which differs from the traditional informed consent for receiving a medical treatment. Based on the existing literature and group discussions, we generated the requirement analysis for e-Consent in MDCS as shown in Table 3. These requirements not only address the current consent issue of the CBPHC program but also the challenges for consent under the GDPR and in the context of MDCSs.

Table 3 – Main Legal and Technical Requirements

No threat of disadvantage – Corresponds to freely given. Consent should not be required for the provision of care.
Provide information – Information about the personal data processing should be provided orally or in writing. It is important to consider the cases of illiterate people.
Information easy to understand – The information that is provided should be concise and easy to understand.
Support CHWs – The application should support the CHWs with all the information necessary regarding consent.
No additional hardware – Consent should be received without requiring the use of any computer technology on the part of the data subjects (i.e., may not have or afford a device).
Registration and consent – Allow consent to be obtained while the CHW is visiting the family or if a family member comes to the health unit and enrols in the program.
Selective consent and selective withdraw – Data subjects should be able to give or withdraw consent to specific purposes of data processing.
Consent withdrawing – Data subjects should have an interface to review and/or to withdraw consent (e.g., website portal, talk to CHWs, call Basic Health Unit).
Consent signing – Data subjects should be able to sign the consent (e.g., digital or wet signature) as otherwise consent would not be explicit.
Child's consent – Child's consent should be given or authorised by "the holder of parental responsibility".
Consent witnessing – If data subject is unable to sign, the health worker should sign the consent as a witness; and the consent should be marked as 'unable to sign'.
Consent receipt – A copy of the consent should be available in the system and sent to the data subject (e.g., via email).
Managing consents – Modified or revoked consents should be archived for the duration necessary for verification or provenance purposes.
Auditing and compliance – A process for paper trail (a written record, history, or collection of evidence) should be designed and implemented to demonstrate compliance.
Secure storage and transmission – Since the consent may contain sensitive personal information, the consent receipts should be securely stored and transmitted.

Design Considerations

Although the existing solutions (e.g., PCC toolkit) offer a strong starting point, adaptions were required to fit the e-Consent design into the particular context of the MDCSs. The PCC toolkit is designed mainly for app-mediated research, i.e., participate in research through smartphones, and share data with researchers. MDCSs are used primarily for the health care purposes (i.e., offering care and treatment, and meaningful use of health data), although they might also be used for secondary purposes (e.g., research and statistics). The range of data collected with MDCSs is also much larger, requiring better information about the data processing.

MDCSs also have a much broader range of purposes for data processing, so that, data subjects should be able to selectively consent (whenever possible) to the purposes that they agree and to selectively withdraw. For instance, because MDCSs are used to support public health, the data controller is normally the government and they can carry out processing activities without consent. That is, there are other lawful bases for data processing, such as the performance of a public tasks, to fulfil a contract, or on the legitimate interest of the data subjects (i.e., for their own health benefits). However, some MDCSs can also be used for secondary purposes, which should be made optional to data subjects, for instance, if linking their personal data to other electronic health records or disclosing it for research and statistics outside the public health sphere. In summary, the specification of purposes is more complex for MDCSs and the interface for consent must reflect such conditions, particularly with respect to selective consent and withdraw per purpose.

In addition, a particular context in the CBPHC consent setting is that the Tablet devices used for MDCSs are carried by the CHWs. This is different to PCC which is designed to run at the individuals' mobile phones. CHWs can help individuals to use the consent application, walking them through the consent process. Information can be provided orally and in writing, and data subjects can ask questions directly to CHWs before signing the consent. That also enables options for dynamic consent, i.e., asking individuals again for consent in case data should later be used for another purpose, or allowing them to change their consent over time. The design should allow the consent to be changed or withdrawn during the CHWs visits or by contacting the basic health unit, considering that consent revocation should be as easy as giving consent.

Consent Interface

Based on the elicited requirements, we defined the steps and information that should be conveyed by the e-Consent tool. The resulting consent interface (Figure 1) provides the data subjects (e.g., family members) with appropriate information about the primary care program and the MDCSs. Each page contains a short explanation about main aspects of the privacy policy. As recommended in [6], the interface uses two layers of information together with appropriate icons to reinforce its content. Users can access the second layer of information by clicking in *"learn more"* link. The consent interface includes the following steps and information:

Description about Primary Care: explains the primary care program offered by the government and the CHWs' tasks.

Data handling and use: explains the categories of personal data collected, emphasizing the highly sensitive data.

Selective consent and withdrawal: lists purposes of personal data processing, whether they are compulsory (e.g., public task) or optional (e.g., data linkage and sharing for research).

Overview on privacy and data protection: describes the data protection mechanisms yet also stressing privacy risks.

Your rights and choices: reminds data subjects that they can skip survey questions, revoke consent and determine to what extent they want to share their data.

Review and consent: informs about the data controller and gives a summary of the consent to be agreed and signed.

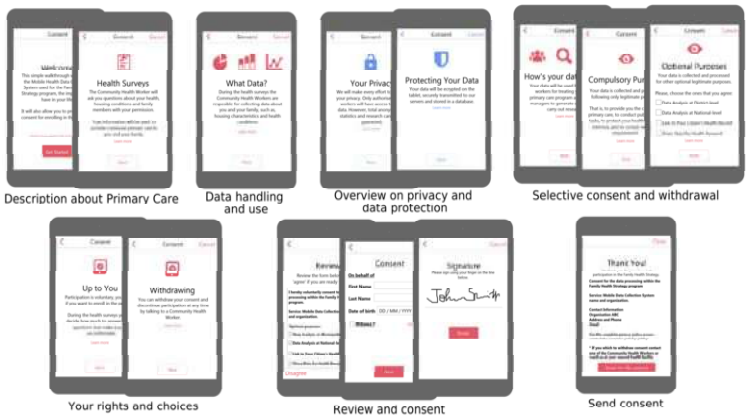


Figure 1 – Main Interfaces for the Consent Application

Send consent: allowing data subjects to send a copy of the consent in a human-readable format to their email address.

Generating Consent Receipts

To generate the Consent Receipt, the e-Consent application should create a signed data structure in a JSON Web Token format with all necessary attributes. This string can be represented in human- and machine-readable formats. Consent Receipts can also have a date; triggering an automatic revocation and requiring re-consent (e.g., every two years). However, the Consent Receipt does not support in its data structures the consent "on behalf of" other people that are unable to give consent. The data structure should therefore be extended to include not only the data subject but also a second individual that consents on behalf of the data subject. Developers can also decide how to handle the Consent Receipts, e.g., a JSON object can be associated to the data subject's record and stored in the database. Consent Receipts may however contain personal data. Thus, it is assumed that the MDCSs use security mechanisms (e.g., [11]) to protect the receipts as well.

Usability Evaluation

A total of 10 participants participated in the usability evaluation. They were researchers working in human-computer interaction, health informatics, and computer security. Their years of experience ranged from 3 to 35 years. Participants were positive on the design related to the principles of informed consent in their questionnaire responses (Figure 2). They reflected positively about their experience, such as recalling what data was being collected and how it would be used; did not feel forced, coerced or manipulated to give consent; options for accepting, declining, and withdrawing consent were clear to them. Nonetheless, the participants' satisfaction on some aspects of information disclosure, such as data *Access*, *Retention* and *Protection*, was lower than the satisfaction on voluntariness and agreement. Questionnaire results also showed that all participants understood what they were consenting to (Figure 3). On average, participants marked 5, 9 out of 7, 0 points in this quiz. Only the question regarding the right to *Skip* questions during health surveys seemed to be misunderstood.

Key findings of the interviews are summarized below.

Information disclosure – Participants (n=9) understood what data is disclosed, the purposes of data processing, and who has access to their data. They stated that *"[the e-Consent] gives*

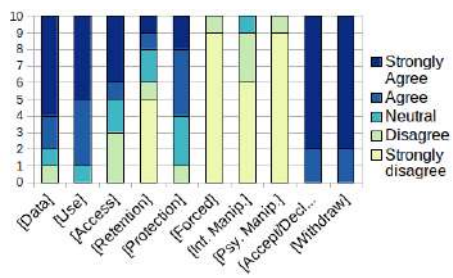


Figure 2 – Results from the First Part of the Questionnaire (information disclosure, voluntariness and agreement). Note that, disagreeing is beneficial for columns 6-8 (Forced, Inf Manip, Psy Manip), i.e., did not feel forced/manipulated.

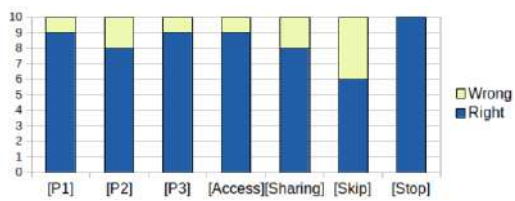


Figure 3 – Results for the Comprehension Test

me all the information I care about", and *"[y]eah, I think that is quite clear, laid out pretty simply, it's very concise"*.

However, one of the participants said that he/she did not *"remember reading how to withdraw the consent"*, and another participant also admitted: *"I'd say that I haven't read all the 'learn more' things. […] Maybe if people have time to read it properly, they can get all the information"*.

Comprehension – Although the participants' average score in the comprehension test was positive, some participants commented that specifically on data *Retention* the e-Consent needed to provide more information about what happened to their data when consent was withdrawn or changed (e.g., disable data sharing for health research).

Voluntariness – Most participants did not feel forced, coerced or manipulated to provide consent (n=9). They mentioned, *"[n]o, I didn't get any feeling of this kind of thing [i.e., manipulation]. There's no impression that something is dodgy.*

It's clear to me". Just one participant, stated that he/she *"felt [coerced] a small amount. If I were to declined consent that I would not receive healthcare [...]"*. Another participant added, *"I feel that the information maybe too much for me. That would've stopped me rather than coerced"*. Regarding manipulation, specifically, one participant stressed that he/she did not feel any kind of manipulation *"because there was not a lot of talking from you [e.g., the CHW]"*. That is, the e-Consent may not contain any manipulative content, but health workers should be trained to not attempt to talk people into something that they do not want to do.

Agreement – All the participants (n=10) stated that the options to accept or decline consent were clear, and they understood that it was possible to withdraw the consent. They mentioned, that they *"never felt that I didn't have a choice"* and *"[y]es, it was pretty clear whether I should accept or decline, even after completing the consent form I realized that I can withdraw"*.

Participants also provided general feedback and suggestions. Examples included: the second layer of information was text-heavy (n=7); more attractive/engaging buttons and links could be used (n=6); some terms used were too technical (n=6); some patients or family members might want more information on privacy aspects (n=5); the pressure of giving consent "on spot" could be a concern (n=4); and, a "progress bar" would be useful to show the steps and progress (n=2).

Discussions

Data privacy is challenging in the context of primary care programs (e.g., CBPHC) which often involve vulnerable populations that can be susceptible to stigmatization and discrimination caused by privacy violations. This, combined with the increase use of mHealth, has led to our exploration of providing appropriate solutions to the e-Consent process. Although preliminary results are positive, further design considerations and improvements can be discussed.

Privacy and protection of personal information needs to be well described in the e-Consent. This is particularly important to inform the aspects regarding data *Access* and *Retention*. We also found that participants were not fully aware that during the health surveys carried out by CHWs, that they have the option of not answering some questions unless it is a mandatory question. This needs to be more clearly stressed in the interface using clear and short statements and details need to be emphasized in the "learn more" link. Furthermore, the "quiz" can be potentially introduced as part of the e-Consent as an additional step before reaching the agreement and signature steps. We found that some participants missed important information either because they felt pressure to finish or thought they already knew it. Depending on the application and scenario, data subjects could be required to pass the comprehension test before providing consent.

The proposed e-Consent offers a simple solution with flexibility for technology refinements. The Consent Receipt specification already offers the minimum viable data structure. It is left to developers to provide more sophisticated solutions for authorisation mechanism (e.g., OAuth, UMA, and XACML). Ideally, data subjects would have access to a personalised web portal where they can access their information and see all records of consent, but providing such system interface can increase the costs of development and infrastructure. Finally, and importantly, the usefulness of the e-Consent designs will need to be assessed with the actual users (e.g., family members and CHWs) and in the real-world setting.

Conclusions

Addressing privacy is a priority for mHealth practitioners [2]. Informed consent is one of the grounds to safeguard individuals' privacy rights, giving back their autonomy and enabling informed decisions before starting the data collection. We have presented the design and evaluation of an e-Consent tool tailored to MDCSs in the context of public health surveys and disease surveillance. Early findings suggest that it has the potential to enhance system's transparency and give individuals more control over their data. Moreover, our requirement analysis, design considerations and usability study findings have implications for other mHealth applications in which data privacy and informed consent are crucial.

Acknowledgements

Funded by DigitalWell Research, a partnership between Region Värmland and Karlstad University.

References

[1] C. B. Aranda-Jan, N. Mohutsiwa-Dibe, S. Loukanova, Systematic review on what works, what does not work and why of implementation of mobile health (mhealth) projects in Africa, *BMC Public Health* **14(1)** (2014), 188.

[2] TrustLaw, *Patient privacy in a mobile world: A framework addresses privacy law issues in mobile health*, TrustLaw Connect, a Thomson Reuters Foundation Service, 2013.

[3] European Commission, Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation), *Official Journal of the European Union*, 2016.

[4] L. H. Iwaya, S. Fischer-Hübner, R. Åhlfeldt, L. A. Martucci, mhealth: A privacy threat analysis for public health surveillance systems, in 2018 *IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, (2018), 42–47.

[5] J. H. Sá, M. S. Rebelo, A. Brentani, S. J. Grisi, L. H. Iwaya, M. A. Simplício Jr., T. C. Carvalho, M. A. Gutierrez, Georeferenced and secure mobile health system for large scale data collection in primary care, *Int J of Med Inform* **94** (2016), 91 – 99.

[6] J. Wilbanks, Design issues in e-consent, *The Journal of Law, Medicine & Ethics* **46(1)** (2018), 110–118.

[7] M. Lizar, D. Turner, *Consent Receipt Specification*, Kantara Initiative, 2018.

[8] R. Spencer, The streamlined cognitive walkthrough method, working around social constraints encountered in a software development company, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI'00), ACM, (2000), 353–359.

[9] B. Friedman, P. Lin, J. K. Miller, Informed consent by design, *Security and Usability* (2005), 503–530.

[10] J. Macinko and M. J. Harris, Brazil's family health strategy delivering community-based primary care in a universal health system, *New England Journal of Medicine* **372(23)** (2015), 2177–2181.

[11] M. A. Simplício, L. H. Iwaya, B. M. Barros, T. C. M. B. Carvalho, M. Näslund, Secourhealth: A delay-tolerant security framework for mobile health data collection, *IEEE J Biomed Health Inform* **19(2)** (2015), 761–772.

Address for correspondence

L. H. Iwaya, Karlstads universitet, Datavetenskap, Karlstad, SE 651-88, +46 54-700 11 33, leonardo.iwaya@kau.se

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190422

Characterization of Behavioral Transitions Through Social Media Analysis: A Mixed-Methods Approach

Tavleen Singh[a], Carlos A Perez[a], Kirk Roberts[a], Nathan Cobb[b], Amy Franklin[a], Sahiti Myneni[a]

[a] *University of Texas School of Biomedical Informatics, Houston, TX, USA*
[b] *Georgetown University Medical Center, Washington, DC, USA*

Abstract

Unhealthy behaviors are a socioeconomic burden and lead to the development of chronic diseases. Relapse is a common issue that most individuals deal with as they adopt and sustain a positive healthy lifestyle. Proper identification of behavioral transitions can help design agile, adaptive, and just-in-time interventions. In this paper, we present a methodology that integrates qualitative coding, machine learning, and formal data analysis using stage transition probabilities and linguistics-based text analysis to track shifts in stages of behavior change as embedded in journal entries recorded by users in an online community for tobacco cessation. Results indicate that our semi-automated stage identification method has an accuracy of 90%. Further analysis revealed stage-specific language features and transition probabilities. Implications for targeted social interventions are discussed.

Keywords:
Health behavior , social media, machine learning

Introduction

Poor lifestyle choices and unhealthy behaviors such as smoking, alcohol consumption, poor diet, and physical inactivity affects the development and management of chronic diseases, like obesity, Type 2 diabetes mellitus, hypertension, cardiovascular diseases, and several types of cancer [1]. Chronic diseases kill 38 million people every year and hence their prevention is important [1]. Efficient self-management of chronic illnesses and adoption of positive health behaviors has shown to be related to overall better physical and psychological health outcomes [2]. However, such self-management is not driven by individuals solely but rather in a social context which includes formal health care providers, informal social peers, and their physical surroundings [3]. Behavior change can be a daunting task, and oftentimes individuals relapse as they attempt to embrace and sustain a positive health change.

Behavioral transitions are important to sustain lasting and improved health outcomes [4]. Several theories have been postulated to define and model these transitions and changes in behavior. Specifically, the Transtheoretical Model (TTM) of Change has been used to conceptualize the process of intentional behavior change by its two main components: stages and processes of change [5]. Stages explore the temporality of behavior change, while processes encompass cognitive and behavioral concepts such as decisional balance, self-efficacy, and rewards management [5]. It is important to identify these stages of change so that appropriate lifestyle interventions can be prescribed to individuals who are willing to quit an unhealthy habit and modify their behaviors. Many mobile interventions have been modeled after TTM. Adapting theory-driven interventions to the changing needs of individuals can help sustain positive health changes long-term [5].

Emerging research has shown that agile and adaptive interventions that respond to stage transitions can be effective tools of behavior change in real-time settings [6]. However, it is difficult to monitor stage transitions using traditional behavior modeling approaches at scale. Online communication forums, which form the dominant means of communication in the digital era, provide an opportunity to understand human behavior change at nuanced levels. According to the Global Digital Report 2018 the number of social media users worldwide is about 3.2 billion [7]. The electronic archival of digital interactions provide us with rich data sources that afford the opportunity to understand intricate processes and stages of behavior change. Emerging research suggests that online social media analysis can help us understand patterns of social factors underlying behavior change and help in the development of network interventions for health behavior changes [8]. Thus far, social media analysis has focused on (a) understanding the structure of social ties [9], (b) assigning peer-to-peer communication events to various social support categories [10]. However, prior research on health-related social networks has not merged theoretically-driven content-based approaches with machine learning methods to facilitate stage-based user behavior and transition classification. Such understanding can have important implications for the design of interventions that aim to enhance self-management behavior.

In this paper, we present a new methodology to understand an individual's behavior transitions using recorded journal entries in online communities. Our study has three major components: (a) qualitative analysis to manually label the user-generated data for different stages of behavior change, (b) automated categorization using machine learning models to scale the labeling of stages of behavior change to a large social media dataset, and (c) characterization of stage-specific language features and transition rates to enable personalized intervention refinement and technology development in scalable online settings. Our research described in this paper will help answer the following questions: 1) How can we adapt existing informatics approaches to automate stage identification in digital health platforms? 2) What are the relationship characteristics between user engagement on these platforms, behavior transitions, and language features?

Such an understanding can help in identifying the triggers for relapse and designing appropriate interventions for the users at risk for relapse or in need of more social support for long-term sustenance [11]. We apply our methodology to characterizing behavioral transitions among the users of an online community for tobacco cessation.

Methods

QuitNet is an online social communication forum that promotes smoking cessation amongst its members thereby leading to behavior change and health promotion [12]. It has been in existence from the past 16 years and has over 100,000 new registrants per year. The members of this platform are usually smokers who are willing to quit or ex-smokers who are willing to stay abstinent. Graham et al. [13] have shown a strong correlation of individual's participation in online community with abstinence compared to individuals who do not participate in such communities. The dataset used in this study consisted of de-identified journal text entries, spanning from 1999-2015 including 26,441 individuals, and 111,004 journal text entries. The journal text entries that were marked as public by the QuitNet users were used for this analysis. This study is exempted from human subjects review by the Institutional Review Board at the University of Texas Health Science Center at Houston.

Qualitative analysis

Firstly, we developed the annotation guidelines to inform the coding process and ensure objectivity in stage assignment. Table 1 shows the annotation guidelines that were adopted in our analysis. We used the TTM model to label every journal text entry with its associated stage of change. The TTM model defines Precontemplation, Contemplation, Preparation, Action, Maintenance, and Termination as the six stages of change, where each stage involves a process of progress (Figure 1). Fivehundred out of the 111,004 journal text entries were randomly selected and coded manually by two independent researchers by performing a line by-line analysis on the text entries to derive the stages of change codes from the data. The codes they assigned to the text entries had a Cohen's Kappa measure of 97%. Disagreement between the researchers was resolved through discussion.

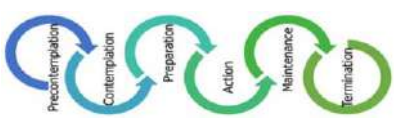


Figure 1– Stages of behavior change

Automated Text Analysis

Using the labeled dataset of 500 journal text entries, we first performed a feature inspection to get a better understanding of our textual data. The dataset was then cleaned using the following techniques - (a) tokenizing the text to split sentences into individual words, (b) removing the stop words to get rid of words that occur too frequently or infrequently as well as punctuation, (c) stemming the text to address word variants, and (d) vectorizing words since we needed to convert words into mathematical representations to feed them into various machine learning models. Latent Dirichlet Allocation (LDA) and word2vec models [14, 15] were trained using default parameters to create feature vectors and extract sensible feature data from the journal text entries. LDA is a probalistic topic model that creates probabilities on the word level, while word2vec is a deep learning method that creates feature vectors for the words in the text corpus [15]. After this we applied a few supervised classifiers to be able to predict what stage belongs to each journal text entry. The labeled dataset was split into a train and test set and a ten fold cross validation was performed. We used the features that were previously created to train three machine learning models: Logistic Regression (LR), Support Vector Machine (SVM) (linear), and Random Forest (RF). These models were chosen, given their performance on similar classification tasks [15]. We used recall, precision, and F-measure as our accuracy metrics to evaluate the machine learning algorithms used in this study. We selected the model with the highest accuracy to make final predictions on the unlabeled dataset. Scikit-Learn package was used for analysis [16].

Data Analysis

User level analysis was performed to map transitions between stages of behavior change. Pearson Correlation coefficient was calculated to identify if there was any correlation between user engagement and stage transition rates (the numbers of messages posted and the number of transitions from one stage to another). A Markov model was used to estimate the stage membership probabilities and the stage transition probabilities of movement from one stage to another. Stage membership probabilities indicate the prevalence of each stage and transition probabilities indicate the probability of stage movements conditional on the stage membership at the previous time point [17]. Further, we applied The Linguistic Inquiry and Word Count (LIWC) [18] dictionary that comprises of psychologically meaningful word categories, and whose output includes the percentage of words within a given text that belong to each stage of change. Applying this technique allows for a direct comparison of text features across various stage-specific journal entries.

Table 1– Annotation guidelines

Stage in TTM model	Characteristics of this stage as per TTM model	Patient's View / State of mind
Pre-contemplation (Not Ready)	People in this stage are not ready to change in the foreseeable future like for another 6 months.	Not thinking about change, Feeling of no control, Denial: does not believe it applies to self
Contemplation (Getting Ready)	People in this stage have the intention to change in the next 6 months.	Weighing benefits and costs of behavior, proposed change
Preparation (Ready)	People in this stage are ready to make a change immediately, they have a plan of action in place.	Experimenting with small changes
Action	In this stage people have made specific overt modifications in their lifestyles within the past six months.	Not all modifications count as Action in this model. Total abstinence is what counts as an action compared to switching to low nicotine cigarettes.
Maintenance	In this stage people have made specific overt modifications in their lifestyles and are working to prevent a relapse.	Maintaining a new behavior over time, has quit for over 6 months
Relapse	People in this stage have started smoking again following a quit attempt.	Experiencing a normal part of the process of change, usually feels demoralized

Results

On an average, a user posted four text entries. Table 2 below shows some of the textual characteristics of the unlabeled journal text entry corpus.

Table 2– Characteristics of an unlabeled text corpus

Characteristics of the corpus	Frequency
Total number of journal text entries	111,004
Total number of unique users	26,441
Mean age of the users	40.3 years
Total number of females	18,614 (~71%)
Total number of males	7,614 (~29%)
Average length of the journal entry	150 words

Qualitative Analysis

Table 3 shows examples of some of journal text entries and their associated stage of change as coded manually by the researchers. In the first example - as the individual was in the process of deciding to undertake the quit process and still thinking about it- the entry was coded as 'Contemplation'. The second example entry was coded as 'Preparation' since the individual specifically mentioned that they plan to quit smoking in 5 days. The last example entry was coded as 'Action' since the individual mentioned being smoke free for a certain amount of time, with a past quit date.

Table 3– Journal text entries labeled by manual coding

Stage of Change	Journal text entry example
Contemplation	I had originally set my quit date at Feb. 7 2003. I got to thinking about it and decided why wait?
Preparation	I was going to quit early but something upset me and I ran to the store. Ugh! Will I ever quit. I'm Quitting in 5 days!
Action	Tough day! Made it through! No slips today.4 days, 5 hours, 55 minutes and 17 seconds smoke free. 85 cigarettes not smoked. $33.60 and 15 hours of my life saved! My quit date: 8/15/2011 2:00:00 PM

Our dataset lacked examples of the 'Pre-contemplation', 'Maintenance' and 'Termination' stage and had limited examples of the 'Contemplation' stage (n=21). There were 210 text entries which were coded as 'Preparation' and 268 text entries which were coded as 'Action'. There was one text entry which was coded as 'NA' (Not Applicable).

Automated Text Analysis

Table 4 shows the accuracy of various machine learning models used for making predictions on the unlabeled dataset using 10-fold cross validation. Since the Random Forest Model gave the highest accuracy, this model was chosen to make the final predictions on unlabeled dataset.

Table 4– Classification report for machine learning models

Machine Learning Models	Precision	Recall	F1-score
LR	0.69	0.70	0.69
SVM (Linear)	0.71	0.72	0.72
RF	0.91	0.90	0.90

62% of the total entries were labeled as 'Action' (n=68,947) and 38% were labeled as 'Preparation' (n=42,041). Very few entries were labeled as 'Contemplation' (n=15) and 'NA' (n=1). Table 5 shows some example journal text entries that were labeled using the RF machine learning model.

Table 5– Journal text entries labeled by the Random Forest model

Stage of Change	Journal text entry example
Contemplation	just hooked up on the quitnet today... checking the site out... seeing what's available... so far so good... let my younger brother know i'd joined up by sending him a q-card... maybe he'll join up too...
Preparation	Reasonable nights sleep. crazy day. headache, a few cravings. no exercise. just work until 6 p.m. 1 nic gum.
Action	last night was bad; no sleep; up every freakin hour...hope day 2 is better. 1 day, 14 hours, 6 minutes and 31 seconds smoke free. 24 cigarettes not smoked. $5.62 and 4 hours of my life saved! my quit date: 7/26/2010

Data Analysis

Because a limited number of samples were coded to be in the 'Contemplation' stage, we conducted further data analysis with only those journal text entries that were labeled as either 'Action' or 'Preparation' by the machine learning model. The highest number of journal text entries posted by an individual was 418 and the number of transitions made by this particular individual was 95 from one to stage of behavior change to another. The Pearson Correlation coefficient between the number of journal text entries posted by the users and their transitions between the stages of change calculated was 0.857 (Figure 2).

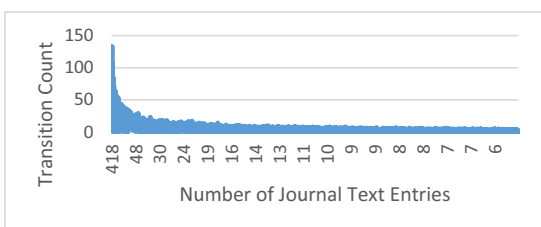


Figure 2– Line graph showing the relationship between the journal text entries posted and transitions between stages of change

This value showed that the higher the frequency of journaling by an individual, the higher the number of behavioral transitions for that individual, which may occur in either a positive direction (Preparation to Action) or a negative

direction (Action to Preparation). This engagement phenomena may indicate that the denser the digital footprint of an individual in a digital platform like QuitNet, the higher the probability to identify behavioral transitions.

Figure 3 shows a Markov chain model representing the movement pattern between the two stages of change – 'Preparation' and 'Action'. The membership probabilities for 'Preparation' is 0.37 and for 'Action' is 0.32. The transition probabilities from 'Preparation to Action' is 0.27 and from 'Action to Preparation' is 0.21.

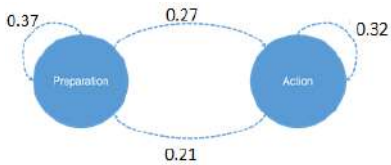


Figure 3– A two-state Markov chain representing the transition between stages of behavior change

As can be seen from the figure above, the probability of an individual staying in the 'Preparation' stage is higher compared to an individual staying in the 'Action' stage. The probability of transitioning from 'Preparation' to 'Action' stage is higher compared to transitioning from 'Action' to 'Preparation' stage.

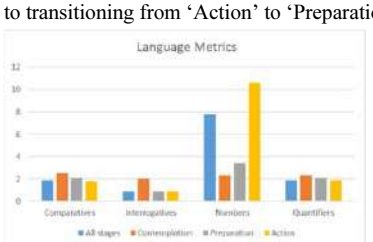


Figure 4- Comparison of means for language metrics

Exploratory text analysis using LIWC revealed specific language features that were prominent within the 'Preparation' and 'Action' stages. As can be seen in Figure 4, interrogatives were prevalent in the 'Contemplation' stage (e.g seeking information), while numbers were highly used in the 'Action' stage, which is probably expected due to expressing quantities of both time of abstinence and cigarettes not smoked (e.g. demonstrating progress). Figure 5 indicates specific foci of journal entries within each stage of change. For example, in the 'Contemplation' stage language with individual drives and needs, achievement, power, and reward were shown to be common. Sense of achievement and work- induced stress (obstacles to quitting) were emphasized in the 'Action' stage.

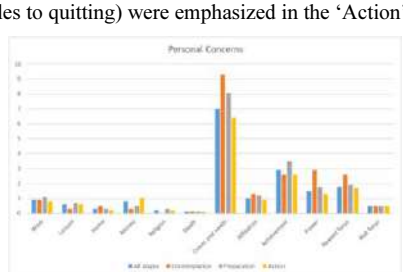


Figure 5- Comparison of means for personal communication topics in different stages of behavior change

Discussion

Technologies like online communities provide support for individuals to modify their behavior for better health outcomes. On the other hand, there are theory-driven principles and processes that work best at every stage of behavior change to reduce resistance, facilitate progress, and prevent relapse. It is important to develop advanced analytic tools to integrate theory and technology, which can result in enhanced just-in-time support infrastructure for sustained behavior change. Such an ecosystem will ultimately lead to a greater understanding of the ways in which such social phenomena mediate behavior change at individual and community levels, thereby providing us with the opportunity to develop superior and scalable interventions. This to further assist the development of multilevel systems that harness behavior change mechanisms aimed at augmenting the support for individuals attempting to achieve their health goals [19].

To the best of our knowledge, the research reported in this paper is among the few studies that have attempted to apply various machine learning models in the context of online communities and social journaling to understand and characterize stages of behavior change in tobacco cessation. Our work provides new techniques to analyze user generated unstructured health data to understand an individual's behavior transition over a period of time. Our model was able to label the stages of behavior change with an accuracy of 90% in journal text entries. The word2vec based approach described in this paper allows for the extension of human-intensive qualitative analysis to large social media datasets.

We found that people who engaged more in journaling had higher transition flags, which may indicate (a) desired or undesired stage shifts, and (b) higher chances of being flagged for stage transitions. In either case, better support infrastructure through adaptive and personalized means aimed at increasing user engagement may ultimately lead to better health outcomes. Further, we have seen that there is a higher probability of an individual staying in the 'Preparation' stage and targeted interventions should be developed to encourage such individuals to move to the 'Action' stage and attain a positive behavior change. The individuals undergoing relapse can be identified based on their stage transitions and appropriate interventions can be designed for them – encouraging them to pick a new quit date or helping them form new relationships online so that they can stay on the path of quitting.

Specific language traits revealed using LIWC analysis, while preliminary in nature, still highlight the authenticity and individual drive embedded in journal entries highlighting the need to understand emotional tone, speech intentions, and cognitive focus within each stage of behavior change. Our study can provide new directions for developing network interventions [20] for tobacco cessation and health promotion by focusing on content-based, targeted behavior change strategies while addressing stage-specific constraints and associations simultaneously.

One of the limitations of our work was the limited number journal text entries that were coded manually to identify stages of change. To improve the generalizability of our results, it is important to have higher number of journal text entries in the qualitative sample so that higher training accuracy of machine learning models can be achieved. It is possible,given the low fraction of journal text entries coded for stages of behavior change, that the distribution of various stages may not have been accurately represented. It is important that a larger number of journal text entries are coded to reach stage

saturation. Also, the content of the journal text entries plays an important role during the predictions made by machine learning models. For example, an individual who has just started the 'Action' but has provided lots of detail about the 'Preparation' could have led the classifier to make a wrong stage prediction. Individual-specific linguistic features and demographics should also be considered when analyzing journal entries. There are also some limitations inherent to the TTM such as a lack of consideration of the social context, biological, and environmental issues related to changes in health behaviors [21], that should be considered when understanding behavior transitions. Another limitation of our study is focusing only on tobacco cessation. Future work should extend these methodologies to datasets in other areas such as diabetes, cancer prevention and survivorship, etc. A more formal social network analysis that utilizes peer interactions with journal entries needs to be performed so that content specific network-patterns can be identified as they have implications in the design of behavioral support systems to promote public health and wellness.

Conclusions

Risky health behaviors contribute to a large number of preventable deaths around the world. The ubiquity of online social platforms allows us to examine inter- and intra-personal processes and stages of health behaviors among individuals. In this paper, we described a mixed methods approach that combines qualitative coding and automated text analysis to provide deeper insight into the mechanisms underlying behavior change through the utilization of digital footprints in the form of online journaling. It is very important to develop scalable methods to help health researchers and professionals analyze large amounts of textual data generated from online communities in today's digital era. Such techniques can help design personalized and targeted interventions that persuade people to initiate or adhere to a positive behavior change. This can help in establishing novel digital and translational interventions in public health and behavioral sciences.

Acknowledgements

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Number R21CA220670. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] Noncommunicable Diseases And Their Risk Factors, World Health Organization. (n.d.). https://www.who.int/ncds/en/ (accessed April 1, 2019).
[2] M.P. Gallant, The Influence of Social Support on Chronic Illness Self-Management: A Review and Directions for Research, Health Educ Behav. 30 (2003) 170-195.
[3] Å. Audulv, K. Asplund and K.-G. Norbergh, The integration of chronic illness self-management, Qual Health Res 22 (2012), 332-345.
[4] C. Abraham and S. Michie, A taxonomy of behavior change techniques used in interventions., Health Psychology. 27 (n.d.) 379-387.
[5] J.O. Prochaska and W.F. Velicer, The Transtheoretical Model of Health Behavior Change, Am J Health Promot. 12 (1997) 38-48.
[6] D. Centola, Social Media and the Science of Health Behavior, Circulation. 127 (2013) 2135-2144.
[7] Digital In 2018: World's Internet Users Pass The 4 Billion Mark - We Are Social, We Are Social. (2018). https://wearesocial.com/blog/2018/01/global-digital-report-2018 (accessed April 1, 2019).
[8] S. Myneni, K. Fujimoto, N. Cobb and T. Cohen, Content-Driven Analysis of an Online Community for Smoking Cessation: Integration of Qualitative Techniques, Automated Text Analysis, and Affiliation Networks, Am J Public Health. 105 (2015) 1206-1212.
[9] D. Centola, The Spread of Behavior in an Online Social Network Experiment, Science. 329 (2010) 1194-1197.
[10] K. Glanz, Health Behavior and Health Education, John Wiley & Sons, 2008.
[11] S. Myneni, N. Cobb and T. Cohen, In Pursuit of Theoretical Ground in Behavior Change Support Systems: Analysis of Peer-to-Peer Communication in a Health-Related Online Community, J Med Internet Res. 18 (n.d.) e28.
[12] QuitNet, (n.d.). https://quitnet.meyouhealth.com/#/ (accessed April 1, 2019).
[13] A.L. Graham, N.K. Cobb, L. Raymond, S. Sill and J. Young, Effectiveness of an Internet-Based Worksite Smoking Cessation Intervention at 12 Months, Journal of Occupational and Environmental Medicine. 49 (2007) 821-828.
[14] T. Mikolov, K. Chen, G. Corrado, and J.J.a.p.a. Dean, Efficient estimation of word representations in vector space, (2013).
[15] Classification Combining LDA And Word2Vec | Kaggle, (n.d.). https://www.kaggle.com/vukglisovic/classification-combining-lda-and-word2vec (accessed March 31, 2019).
[16] Model selection and evaluation, (n.d.). https://scikit-learn.org/stable/model_selection.html#model-selection (accessed April 1, 2019)
[17] L.A. Brick, C.A. Redding, A.L. Paiva and W.F. Velicer, Intervention effects on stage transitions for adolescent smoking and alcohol use acquisition., Psychology of Addictive Behaviors. 31 (n.d.) 614-624.
[18] J.W. Pennebaker, M.E. Francis, and R.J. Booth. Linguistic inquiry and word count: LIWC 2001. Mahway: Lawrence Erlbaum Associates. 2001;71(2001):2001.
[19] S. Myneni, K. Fujimoto and T. Cohen, Leveraging Social Media for Health Promotion and Behavior Change: Methods of Analysis and Opportunities for Intervention, (2017) 315-345.
[20] T.W. Valente, Network Interventions, Science. 337 (2012) 49-53.
[21] C.J. Armitage, Is there utility in the transtheoretical model?, 14 (2009) 195-210.

Address for Correspondence

Tavleen Kaur Ranjit Singh, MS. The University of Texas School of Biomedical Informatics at Houston, 7000 Fannin, Houston, TX 77030. tavleen.kaur.ranjit.singh@uth.tmc.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190423

Would Geriatric Patients Accept Using a Telemedicine Platform for Post ICU-Discharge Follow-Up Visits?

Saif Khairat[a], Katie Tirtanadi[a], Paige Ottmar[a], Ritika Gudhe[b], Charles Adrian Austin[b]

[a] *School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, NC, US,*
[b] *School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, US*

Abstract

For those specialized in geriatric medicine, telemedicine innovations provide a new alternative to in-person follow-up care, allowing clinicians to connect and treat patients with more convenience. Telemedicine will likely play a vital role in reaching underserved populations in remote areas. This study investigates first impressions of a telemedicine-based delirium assessment tool. The overall response from participants is positive, supporting the theory that these types of tools will be welcome within the geriatric patient population. Feedback surrounding interactions with the interface are also positive, showing that while many elderly patients may refrain from working with tablets daily, they can successful interact with the tool when needed for care reasons. While this study and sample size are not all-inclusive regarding the diversity of patients and distinct challenges, it serves as a preliminary step towards future research exploring the feasibility and acceptability of such tools within this specific population.

Keywords:

Telemedicine, Critical Care, Aged

Introduction

Challenges in caring for geriatric populations, individuals older than 60 years of age who are receiving care [1], persist in the US health system. Similar to other vulnerable populations, the population of geriatric patients is growing in number as the baby boomer generation ages, often bringing with them complex diagnoses requiring the cooperation of many different follow-up plans and treatment specialists [2]. Primary care has struggled to keep up with the complex nature of geriatric care, as well as the growing number of patients requiring attention [2]. This is further compounded by the increasing shortage of medical professionals specializing in geriatric medicine [3,4]. Attributes of geriatric care often hinder traditional in-person care methods The development of telemedicine, particularly telecommunication platforms with patient-centered features, provide a possible solution to the unique challenges faced within the demand of this medical specialty. Telemedicine allows clinicians to connect and evaluate patients remotely and is ideally suited to help improve overall care outcomes within this population as well as address access issues.

A pilot study gathered preliminary evidence of geriatric patient satisfaction while using computer systems for follow-up care in the home [5]. This small pilot referred to these visits as "electronic house calls" and demonstrated that most geriatric patients are comfortable with computer-assisted follow-up care. It also revealed that patients did not believe the computer to negatively impact their clinician relationship with their physician [5,6]. While this study employed personal computers, as telemedicine technology has advanced, it has opened opportunities for health informatics research to further study the satisfaction and feasibility of other remote tools. These remote telemedicine-focused platforms can enhance the interaction between patient and clinician, offering a "new kind of service relationship" by providing direct and personal care with more convenience [6].

Telemedicine has been incorporated into ICU patient care and education. These new technologies allow for critical care patients to be treated by health professionals remotely [7,8]. ICU telemedicine interventions have demonstrated a reduction in hospital cost, patient mortality and patient length of stay [8,9,10]. While many benefits of telemedicine in the ICU are financially related, telemedicine has most prominently allowed for more efficient care to critically ill patients. Faster response time to alarms and the capturing of performance data for review and education has helped this improvement [8]. For ICU settings in rural or underserved areas, telemedicine has been a welcome addition, helping staff to monitor patients during their stay and to provide relief to physicians and nurses, in the event of workforce shortages [11]. ICU telemedicine interventions have been instrumental in clinical improvement and this continues to be an arguing force for the adoption of telemedicine platforms in hospital ICU settings across the country [9].

As opposed to previous generations of geriatric patients, individuals currently aging into this population group now have had some exposure to telecommunication or connected devices at some point in their lifetime [12]. Videoconferencing with patients as a form of follow-up care has already shown to be successful with this population [2]. These sessions are comparable to face-to-face encounters in terms of satisfaction, reliability and usability [12]. Furthermore, telemedicine can also potentially alleviate clinical access issues for geriatric patients in rural and remote locations. Barriers related to technology, such as information overload, lack of devices and or infrastructure, and cost, are still real concerns for geriatric patients [13]. Evidence to support telemedicine's effective deployment and sustainability in such areas is still being researched [14]. However, there will likely be an increase in the number of jobs that focus specifically on the coordination of telemedicine technology for patient use. This will potentially assist diffusion of telemedicine practices across clinical systems and regions, especially as service areas continue to grow in scope [6]. For our study, we investigate the feasibility of using a remote telemedicine tool for follow-up post-acute care within the geriatric population.

Methods

Thirty (n=30) participants from an inpatient geriatric specialty unit in North Carolina were selected for this study. Daily patient schedules were reviewed to identify potential study participants. Participants were English-speaking and 65 years of age

or older. This study was completed in partnership with another observational study gathering preliminary data on a new telemedicine delirium diagnostic tool. Participants were initially evaluated to determine their baseline cognitive state [15]. Those who screened positive for dementia as defined by an abnormal Mini-Cog test or had a documented history of dementia and/or brain abnormalities, as well as those unable to give informed consent, were excluded. Participants were given a tablet and instructions to complete the new delirium assessment. Once this assessment concluded, participants completed a usability questionnaire to detail their impressions of the new tool. The usability data, detailed in this paper, will be used to support hypothesis of a positive potential feasibility and patient acceptability of remote tablet devices in care settings outside of the hospital.

To evaluate the new assessment tool, participants completed the Questionnaire for User Interface Satisfaction (QUIS), a short form survey designed to uncover useful impressions from initial participant interactions. Participants answered questions about their overall reactions, the design, terminology used, and understanding of the system - both overall and when directed to complete a task. The final section of the questionnaire, system capabilities, was not applicable to this study.

Participants' responses for each question of the QUIS were determined from a bipolar Likert scale ranging from 0 to 9. These individual data points were averaged into four overarching categories of the QUIS for each participant: overall reaction, overall screen, terminology and system information, and meaning. A one-way ANOVA was performed in order to test for significant differences in the mean scores for these four categories by age group (60-69 years, 70-79 years, and 80+ years).

Results

Of the 30 participants recruited for the study, 26 (87%) were female. 100% of the participants were white and non-Hispanic. Participants ranged from 67 to 92 years old, with an average age of 77.80 years. Four participants (13.3%) were 60-69 years old, 14 participants (46.7%) were 70-79 years old, and 12 participants (40%) were 80 years old or older.

Overall Satisfaction Results

Questions focused on four different factors of the tablet experience, which were collapsed into four average scores for each of the 30 participants. The domains were overall reaction to the software, screen, terminology and system information, and learning. Questions to which participants responded "NA" were not included in the analysis.

Participants' responses—as determined from a bipolar Likert scale ranging from 0 to 9—were averaged across the four main categories to create a mean score for each participant for each category. These individual means were then averaged into overall means for each category, which ranged from 8.16 to 8.72. See Table 1 for descriptive statistics for all four domains.

Age and Satisfaction

An alpha level of .05 was used to assess statistical significance. A one-way ANOVA was used to assess whether responses varied by age category. There was no significant difference in mean responses found in any domain by age category. See Figure 1 for the mean average scores across age categories.

The 60-69-year-old age group had the highest mean score for the Overall Reaction, with a mean value of 8.90. The 80+ year-old age group had the lowest mean score with 7.98. The 60-69 age group also had the highest mean score for the Screen domain, with a mean value of 8.94. In this domain, the 70-79-year-old age group had the lowest mean score of 8.45. Conversely, the 80+ age group had the highest mean scores for the Terminology and System Information and Learning domains, with mean values of 8.72 and 8.85, respectively. The 60-69-year-old age group had the lowest scores for these two domains, with means of 8.42 and 8.25, respectively. However, it should be noted that the range between the highest- and lowest-scoring age categories for each domain is small, ranging from 0.30 to 0.92. See Table 2 for the mean scores of the four domains by age category.

Table 1—Summary of Mean Scores for the Four Domains of Interface Satisfaction Survey

Mean Score	N	Minimum	Maximum	Mean	SD
Overall Reaction Mean Score	30	6.00	9.00	8.16	1.66
Overall Screen Mean Score	30	6.25	9.00	8.60	0.72
Terminology and System Information Mean Score	30	6.66	9.00	8.58	0.85
Learning Mean Score	30	6.00	9.00	8.72	0.69

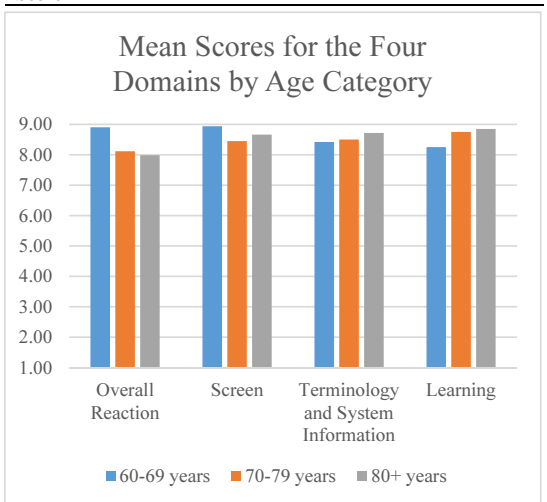


Figure 1—Mean Scores for Four Domains by Age Category

Additionally, we broke down the four domains into their various indicators. Though a one-way ANOVA of the indicators by age group found no significant differences in score by age category, the mean scores for the different indicators in each domain are highlighted in the rest of this section to note overall trends between age categories.

The Overall Reaction domain was broken down into its indicators, which can be seen in Table 3. Overall Reaction scores were high, ranging from 7.67 to 9.00. For most of the indicators, the 60-69-year-old age group scored the highest, followed by the 70-79 group and then the 80+ group. The "Rigid or Flexible" indicator is the only one which does not follow this pattern, with the 80+ group having a higher mean score than the 70-79 group, with mean scores of 8.08 and 7.79, respectively.

Table 2—Summary of Mean Scores for the Four Domains of Interface Satisfaction Survey by Age Group

Domain	60-69 Years Mean Score	70-79 Years Mean Score	80+ Years Mean Score
Overall Reaction	8.90	8.11	7.98
Overall Screen	8.94	8.45	8.66
Terminology and System Information	8.42	8.50	8.72
Learning	8.25	8.75	8.85

Table 3—Summary of Mean Scores for Overall Reaction Domain by Age Group

Domain	60-69 Years Mean Score	70-79 Years Mean Score	80+ Years Mean Score
Terrible or Wonderful	8.50	8.07	7.83
Difficult or Easy	9.00	8.36	7.67
Frustrating or Satisfying	9.00	8.14	8.17
Inadequate or Adequate	9.00	8.21	8.17
Rigid or Flexible	9.00	7.79	8.08

The Overall Screen domain was broken down into its indicators, which can be seen in Table 4. Scores for these indicators mirrored those in the Overall Reaction domain, ranging from 7.86 to 9.00. Contrary to the Overall Reaction domain, while the 60-69 group scored the highest across all indicators, here the 80+ group scored higher than the 70-79 group. The exception to this is the "Reading Characters on Screen" indicator, where the 70-79 group scored higher than the 80+.

Table 4—Summary of Mean Scores for Overall Screen Domain by Age Group

Domain	60-69 Years Mean Score	70-79 Years Mean Score	80+ Years Mean Score
Reading Characters on Screen	9.00	8.90	8.83
Organization of Information	8.67	8.54	8.73
Sequence of Screens	9.00	8.54	9.00
Help Messages on the Screen	9.00	7.86	8.17

The Terminology and System Information domain was broken down into its indicators, which can be seen in Table 5. For three of these indicators—Use of Terms Throughout the System, Position of Messages on Screen, and Prompts for Input—again the 60-69 age group scores the highest. However, for Error Messages and Information Accessibility, this group scored the lowest. This is mostly due to the small N for the 60-69 group for these indicators (N = 3), with most respondents in this age category choosing "Not Applicable". One participant rated the system a 2 for both of these indicators, decreasing the respective mean scores significantly. For the other indicators, the 60-69 group is followed by the 80+ group and then the 70-79, mirroring that of the Overall Screen domain.

Table 5—Summary of Mean Scores for Terminology and System Information Domain by Age Group

Domain	60-69 Years Mean Score	70-79 Years Mean Score	80+ Years Mean Score
Use of Terms Throughout the System	9.00	8.78	9.00
Position of Messages on Screen	9.00	8.78	8.89
Prompts for Input	9.00	8.31	8.64
Error Messages	5.50	8.20	8.33
Information Accessibility	7.25	8.50	8.83

The Learning Domain was broken down into its indicators, which can be seen in Table 6. Analyses of this domain were hindered by the high number of participants who gave a "Not Applicable" rating for various indicators, most notably the Exploring New Features by Trial and Error and Reference Materials. This is likely because many participants did not opt to explore features on the telemedicine tool and were not instructed to by researchers. Additionally, reference materials were not readily provided for participants. For the other indicators, Remembering Commands and Straightforwardness of Tasks were scored the highest by the 60-69 group followed by the 70-79 group and then the 80+ group. Most notably, Learning to Operate the System exhibited the opposite pattern, with the 80+ group rating this the highest and the 60-69 group rating it the lowest, with mean scores of 9.00 and 8.50, respectively.

Table 6—Summary of Mean Scores for Learning Domain by Age Group

Domain	60-69 Years Mean Score	70-79 Years Mean Score	80+ Years Mean Score
Learning to Operate the System	8.50	8.75	9.00
Exploring New Features by Trial and Error	N/A	7.67	8.33
Remembering Commands	9.00	8.61	8.58
Straightforwardness of Tasks	9.00	8.85	8.83
Reference Materials	N/A	8.20	8.50

Discussion

Overall, participants were very satisfied with the interface, with an overall reaction mean score of 8.16. The other three domains exhibited similarly high scores, ranging from means of 8.58 to 8.72. A one-way ANOVA failed to find a statistically significant difference in scores by age category, suggesting that even the eldest of this geriatric population did not react differently

from the youngest age category. We demonstrate that for three of the four domains, the domain means actually increased between the 60-69 group and the 80+ group. Together, these results suggest that geriatric populations are highly satisfied with the telemedicine software, indicating that it could be feasible and acceptable for this population to use.

The development of telemedicine, particularly telecommunication platforms with patient-centered features, provide a possible solution to some of the challenges associated with providing geriatric care, especially as the proportion of the population who are older grows. Telemedicine platforms can be utilized to improve primary care by allowing providers to follow-up with their geriatric patients in a time and place that is most convenient for both groups. This could be especially convenient for scheduling follow-ups with rural patients or for those without access to reliable transportation to and from their appointments. Though these individuals have generally had some exposure to connected devices [12], previous studies found that older participants exhibited a lower overall reaction mean score. Though this study did not focus on patients and instead solicited results from community members [16].

Our study contradicts some of the results from a previous study that found that older participants were less satisfied with telemedicine platforms. However, this is the first study to look at the experience of in-patient geriatric populations in relation to telemedicine platforms [16]. Future research is needed to truly assess the acceptability and feasibility of telemedicine platforms for the geriatric population. Most ICU specific research regarding telemedicine interventions detail support for adoption of these platforms via financial incentives [9,10]. While many show greater contribution margins once this technology is implemented within ICU processes, other research demonstrates better clinical outcomes, such as decreased mortality, as a result of more acute monitoring and intensivist involvement via telemedicine platforms [8]. Conversely, our study looks at the patient side of telemedicine in the ICU. Our preliminary findings show that from a patient perspective, telemedicine is a welcome and useful tool during their treatment.

Over time, rural counties have received a net influx of people over the age of 50, suggesting that many geriatric people are opting to move into rural locations [17]. In light of this, future studies are needed to focus on the satisfaction and feasibility of telemedicine software for rural, geriatric populations who may face different barriers than their peers in urban areas. As this is a potentially vulnerable population, researchers must ensure that the usage of telemedicine software as a form of follow-up care does not also interfere with medical comprehension.

At this time, this research suggests that telemedicine software is a viable solution to overcoming barriers to reaching geriatric populations. As the population of the United States ages and this population tends to require more medical care, telemedicine allows primary care physicians the ability to follow-up with these patients in a manner that is convenient for both parties. Offering telemedicine platforms as an option for geriatric populations can reduce the burden on the patient and the physician and provide greater access to patients who may otherwise have difficulty traveling to appointments.

Limitations

The main limitation of this study is the sample size of 30 participants, all of whom are English speakers. Participants did vary in age, however all participants were white and not Hispanic/Latino. Additionally, the vast majority of patients were women. As this group is not representative of the demographics of the area, it is likely that this is not a representative sample. Finally, patients who had tested positive for dementia or had a documented history of dementia and/or brain abnormalities were excluded, so our sample is not representative of the entire geriatric population.

Conclusions

Telemedicine can potentially alleviate access and availability issues for geriatric patients. However, previous research found that older populations were less likely to be satisfied with telemedicine software [5]. This study focused on how geriatric patients reacted to the telemedicine software in terms of receiving follow-up care outside of the hospital setting.

Results found that telemedicine software is an effective tool for receiving follow-up care, with no differences in mean satisfaction between age categories. Future research is needed to study how rural geriatric patients respond to telemedicine software.

Acknowledgements

This study was funded by the National Center for Advancing Translational Sciences (NCATS) #R21TR002088-01.

References

[1] P. Thane, History and the sociology of ageing. *Soc. His. of Med*. **2** (1989) 93-96. doi:https://doi.org/10.1093/shm/2.1.93

[2] R.C. Merrell, Geriatric telemedicine: Background and evidence for telemedicine as a way to address the challenges of geriatrics, *Healthc. Inform. Res.* **21** (2015) 223–229. doi:10.4258/hir.2015.21.4.223.

[3] G.A. Warshaw, E.J. Bragg, D.E. Brewer, K. Meganathan, and M. Ho, The development of academic geriatric medicine: Progress toward preparing the nation's physicians to care for an aging population, *J. Am. Geriatr. Soc.* **55** (2007) 2075–2082. doi:10.1111/j.1532-5415.2007.01519.x.

[4] M.N. Shah, S.M. Gillespie, N. Wood, E.B. Wasserman, D.L. Nelson, A. Dozier, and K.M. McConnochie, High-intensity telemedicine-enhanced acute care for older adults: An innovative healthcare delivery model, *J. Am. Geriatr. Soc.* **61** (2013) 2000–2007. doi:10.1111/jgs.12523.

[5] R.L. Bratton, and C. Cody, Telemedicine applications in primary care, *Mayo Clin. Proc.* **75** (2000) 365–368. doi:10.4065/75.4.365.

[6] C. LeRouge, M.J. Garfield, and R.W. Collins, Telemedicine: Technology mediated service relationship, encounter, or something else?, *Int. J. Med. Inform.* **81** (2012) 622–636. doi:10.1016/j.ijmedinf.2012.04.001.

[7] H.N. Reynolds, H. Rogove, J. Bander, M. McCambridge, E. Cowboy, and M. Niemeier, A Working Lexicon for the Tele-Intensive Care Unit: We Need to Define Tele-Intensive Care Unit to Grow and Understand It, *Telemed. e-Health*. **17** (2011) 773–783. doi:10.1089/tmj.2011.0045.

[8] C.M. Lilly, J.M. McLaughlin, H. Zhao, S.P. Baker, and S. Cody, A multicenter study of ICU telemedicine reengineering of adult critical care, *Chest*. **145** (2014) 500–507. doi:10.1378/chest.13-1973.

[9] W.J. Binder, J.L. Cook, N. Gramze, and S. Airhart, Telemedicine in the Intensive Care Unit: Improved Access to Care at What Cost?, *Crit. Care Nurs. Clin. North Am.* **30** (2018) 289–296. doi:10.1016/j.cnc.2018.02.010.

[10] C.M. Lilly, C. Motzkus, T. Rincon, S.E. Cody, K. Landry, and R.S. Irwin, ICU Telemedicine Program Financial Outcomes, *Chest*. **151** (2017) 286–297. doi:10.1016/j.chest.2016.11.029.

[11] M.M. Ward, F. Ullrich, A.J. Potter, A.C. MacKinney, S. Kappel, and K.J. Mueller, Factors Affecting Staff Perceptions of Tele-ICU Service in Rural Hospitals, *Telemed. e-Health*. **21** (2015) 459–466. doi:10.1089/tmj.2014.0137.

[12] S. Narasimha, K.C. Madathil, S. Agnisarman, H. Rogers, B. Welch, A. Ashok, A. Nair, and J. McElligott, Designing Telemedicine Systems for Geriatric Patients: A Review of the Usability Studies, *Telemed. e-Health*. **23** (2017) 459–472. doi:10.1089/tmj.2016.0178.

[13] M. Levine, J.E. Richardson, E. Granieri, and M.C. Reid, PAIN & AGING SECTION Original Research Articles Novel Telemedicine Technologies in Geriatric Chronic Non-Cancer Pain : Primary Care Providers ' Perspectives practices serving older adults with CNCP in New, (2014) 206–213.

[14] M. Brignell, R. Wootton, and L. Gray, The application of telemedicine to geriatric medicine, *Age Ageing*. **36** (2007) 369–374. doi:10.1093/ageing/afm045.

[15] K.K.F. Tsoi, J.Y.C. Chan, H.W. Hirai, S.Y.S. Wong, and T.C.Y. Kwok, Cognitive tests to detect dementia a systematic review and meta-analysis, *JAMA Intern. Med.* **175** (2015) 1450–1458. doi:10.1001/jamainternmed.2015.2152.

[16] T. Newlin, T. McCall, P. Ottmar, B. Welch, and S. Khairat, Assessing the satisfaction of citizens using teleconsent in clinical research, *Stud. Health Technol. Inform.* **247** (2018) 685–689. doi:10.3233/978-1-61499-852-5-685.

[17] K. Johnson, R E P O RT S O N Demographic Trends in Rural and Small Town America, n.d.

Address for correspondence

Dr. Saif Khariat, 428 Carrington Hall, S. Columbia St., Chapel Hill, NC 27599

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190424

Feasibility of Three Head Mounted Eye-Tracker in Anesthesia: A Feasibility Study

Andreas Klausen[a], Rainer Röhrig[b], Myriam Lipprandt[b]

[a] *Department for Anesthesiology, Intensive Care-, Emergency Medicine, Pain Therapy, Klinikum Oldenburg, Oldenburg, Germany,*
[b] *Department of Medical Informatics, Carl von Ossietzky University of Oldenburg, Oldenburg, Germany*

Abstract

Many studies use eye-tracker to analyse the socio-technical system, also in medical research. Only a few articles describe the use of eye-tracker to examine human-computer interaction in a critical care environment, especially in the field of anaesthesia or surgery. Therefore, we have tested in a feasibility study head-mounted eye-tracker of three different manufactures in a simulated anesthesia surrounding with mankind patient simulators. The research question was to analyse whether the field scene camera of the eye-tracker can be used in the light conditions and changes in brightness of the operating room. In addition, it was tested whether the eye-tracker was still calibrated and held on the subject's head during the resuscitation movement. All eye-trackers tested had a good adaptation on changing light or changing distances.

Keywords:

Patient Simulation, Anesthesia, Patient Safety

Introduction

Anesthesia as a medical discipline in hospitals include anaesthesiology, which includes all medical procedures such as intensive care, pain therapy and emergency medicine. Within intensive care, procedures such as intubation in the operating theatre are omnipresent. The anesthesiologist's work is characterized by the use of various technical devices such as ventilators and a high stress level due to medical emergencies. Errors at the level of human-computer interaction or human-human communication can lead to errors that impair patient safety. For this reason, the anaesthetic area is also a safety-critical area, which is the subject of research.

The main aspects of safety-critical systems are characterized by interactions and rules between people (teamwork), (medical-) devices and are influenced by organizational conditions. In case of an adverse event, the cause can usually occur on several levels of the socio-technical system. An analysis of this system is therefore necessary to increase patient safety. It depends also on human-machine-interaction. Especially in emergency medicine, a risk for patient safety arises from situations with a lot of stress, changing teams and complex human-machine interaction. The perception, attention and education level of each team member is important for patient safety. Loss of attention, insufficient teamwork or problems in human-human-interaction can be a reason of making wrong decisions or insufficient troubleshooting[1].

The identification of risk-based human-machine interactions and usability are a core component of the research project in which this pilot study is embedded. For this purpose, this pilot study determines whether eye-tracker can be used as a measuring system for analysing part of the socio-technical system and its limitations. So far, eye-tracker has been used as a measuring instrument for the determination of visual attention, perception and decision making [2]. Currently, they are used in emergency and intensive care medicine to analyze the surgeon's vigilance by measuring their gazes [3]. The anesthetist's distribution of visual attention [4], the mental workload [5], the anesthetists' experience [6] and the eye behaviour between novice and expert are being investigated with eye-tracker [7–9]. However, the applicability of eye-tracking as a measuring instrument for the identification of risks in human-machine interaction in the anesthesiological area is still missing [10].

The aim of this pilot study is to test the feasibility and effectiveness of eye-tracker in emergency and intensive care medicine. The suitability of eye-tracker is tested under real-world conditions, which in particular address the context factors such as light and brightness differences. This study is part of a larger future research project to analyse the entire socio-technical system in medicine.

The following research questions are answered in this study:

- Is the eye-tracker applicable in anesthesia workspaces?
- Do the video data of the eye-tracker allow an analysis afterwards of the scene?
- Do the video data of the eye-tracker represent the scene and its objects in a realistic way corresponding to the perception of the person wearing the eye-tracker glasses?
- Is the eye-tracker still attached to the head and calibrated after rapid head movements (resuscitation)?
- Are there any limitations in the use of eye-tracker in these environments?

Methods

Study Design

In this study, three eye-trackers were tested for use in the operating room and in intensive care medicine. The test referred to the practical applicability of Eye-Tracking technology in intensive care unit environments. For this purpose, three scenarios were selected in which the research question was answered.

The operation rooms are characterized by large brightness und distance differences. It should be tested if the video material provided by the field scene camera of the eye-tracker depicts adequately and correctly the crucial areas such as monitor, glottis and patient simulator. The correct representation of the objects is crucial for the subsequent qualitative analysis of the risky actions in emergency medicine. The detection of vital signs on the monitor or the perception of the glottis during intubation must be realistically reproduced by the eye-tracker to ensure a successional analysis of the video material. For future use of the eye-tracker as a measuring instrument in the socio-technical system, the comparison between the perception of the subject and the recorded video material must take place

in order to be able to use qualitative analysis methods. In addition, it must be found out if the glottis lying in a dark hole can be recognized as a Region of Interest (ROI) by the eye-tracker. The recognition of the glottis is an important criterion for correct intubation. Therefore, it must be checked whether the glottis can be taken as Region of Interest (ROI).

The domain-specific environment has yielded in to three research questions (see the research questions assigned to the scenario Table 1.

Table 1 – Research Goals Assigned to the Scenarios

Research Goal	Scenario and Requirement
Detection of different objects (ROI) with different luminous intensity	Scenario1 Changing Gaze: Accurate changing gaze e.g. to monitor with different backlights in greater distances
Recognition of the glottis	Scenario 2 Intubation: Accurate detection with very low light intensity and low distance
The calibration of the eye-tracker remains stable after heavy movement of the head (resuscitation)	Scenario 3 Resuscitation: Freedom to move your head and continue to recognize the objects in the room

The study was performed in an operating room with a patient simulator. An anesthesiologist performed the scenarios with the three eye-trackers according to a given protocol. A study leader was present and directed the process of the study. Before each test, the eye-tracker was calibrated. Once the eye-tracker worked, each scenario was performed and the video data were recorded. First, the scenario "changing gaze" was performed on all three eye-trackers. Thereafter, the intubation scenario and the resuscitation scenario were performed. The analysis of the video data was carried out by two raters (anesthetist and computer scientist) who defined the evaluation criteria. The analysis process took place by consensus.

Material

Table 2 – Used Equipment and Material

eye-tracker	Ergoneers Dikablis Eye Tracking Glasses	SMI Eyetracking Glasses 2.0	Tobii Pro Glasses 2
Software	Avidemux 2.6.18	BeGaze 3.6 build 52	Tobii Pro Lab 55.5126

Evaluation Criteria

This evaluation is based on an analysis of the video material. Two raters evaluate the recording due to the time of recognizability (sharpness) of the objects (ROI) in the video. By using the timer and frame-to-frame-function of the video software we measured the time from beginning of fixation to the end of adjustment (mean values). We choose the frame where the objects were very cleary depicted. The raters assess the recognizability of objects subjectively.

The applicability of eye-tracker in intensive care medicine must meet the following requirements:

- Rapid focus adaptation and rapid adaptation when brightness changed.
  - Anaesthesiologists always have to change their gazes. Changing of gaze is always combined with changing of brightness and changing of distance.
- Good correlation between the real scene and the recorded video frame
  - The look into the orifice of the body (e.g. laryngoscopy) is often combined with loss of brightness and cramped visibility conditions, a special point of view or a special angle of vision.
- The eye-tracker must stay fixed on the head and stay calibrated during the resuscitation
  - During the resuscitation intense movement of the head occurs.
- Are the values visible on the screens with different backlighting? In order to recognize the ROI, the objects must be clearly recorded. Monitors with different lighting systems are often used in anesthesia.
  - Here we have selected the categories "recognizable", "partially recognizable" and "not recognizable".

Simulation Environment

The study takes place in an operation theatre to have a realistic environment. We use a patient simulator (Laerdal, Resusci Anne Advanced SkillTrainer). The patient simulator and the monitoring machine were connected to display the simulated physiological parameters. This patient simulator was lying on the operation table. The operation lamp was focused on the thorax of this patient simulator. We simulated different scenarios to test the properties of the following products: Ergoneers Dikablis Eye-tracking Glasses Professional (Ergoneers GmbH), SMI Eyetracking Glasses 2.0 (SensoMotoric Instruments GmbH) Tobii Pro Glasses 2 (Tobii AB).

The operating theatre was 5m x 5m with an operation table placed in the middle of the room (Figure 1). At the end of the head the anaesthesia machine was on the right side with a light background (Primus, Dräger). The distance was nearly 70 cm. On the right side of the anaesthesia machine was the patient monitor with a black background (Datex Ohmeda).

Scenario 1: Changing Gaze

We defined four different Regions of Interest (ROI) reflecting the tasks during the medical procedure.

- ROI 1: Face of Patient Simulator,
- ROI 2: Display of Patient Monitoring
- ROI 3: Display of the Aaesthesia Machine

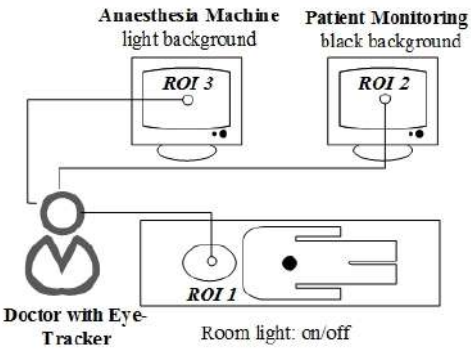


Figure 1 – This figure shows a model of the operation theater (5m x 5m) to depict the scenario more easily.

The anesthesiologist focussed all of the ROI two times (displays) up to 4 times (patient simulator and through the room). After each change of the gaze, he had to fix the ROI for at least 10 seconds, so we could measure the time for the adjustment of the eye-trackers. This was performed with the light switched on and off.

Scenario 2: Oral Intubation

The anesthesiologist did the laryngoscopy. If he could see the epiglottis, arytenoid cartilage, vocal chords and subglottic area he gave a sign to find this scene on the video later. We looked for the typical anatomical structures on the video at the time the anaesthesiologist saw all structures in real field of vision.

Scenario 3: Resuscitation

The anesthesiologist performed a heart pressure massage for 1 minute. Calibration of the eye-tracker was measured by using a target disc (Figure 2) before and after the heart pressure massage. The centre of this target disc was placed at a height of 170 cm, the distance to the disc was 100 cm. The deviation was analysed on the recorded video.

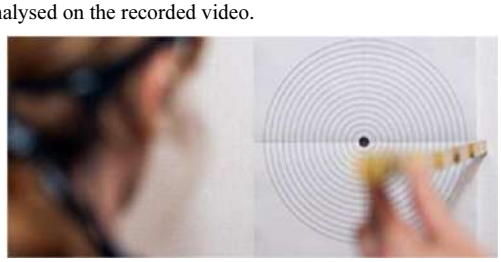


Figure 2 – After 1 minute resuscitation the deviation of calibrated eye-tracker was tested with the help of a target disc.

Results

Scenario 1: Changing Gaze

The face of the patient simulator was depicted clearly in the video material of all three eye-trackers in under one minute (

Table 3). When the light was turned off, it took a little longer for the eye-tracker to be in focus or to see something on the video. Nevertheless, the eye-tracker took less than a second to focus.

Table 3 –ROI 1-Face of Patient Simulator with Room Light Switched On and Off. It depicted the time of adaption till the video of the eye-tracker was focused.

	Ergoneers	Tobii	SMI
Light On	0,11 sec	0,70 sec	0,40 sec
Light Off	0,39 sec	0,90 sec	0,53 sec

In ROI 2, the eye-trackers should record the monitor to determine what can still be seen when the light is switched on and off. The monitor background colour is black. The ability to see details on the monitor strongly depends on the colour background of the monitor and the ambient light. With the light on, it takes less than a second to focus the eye-trackers. Only with the Ergoneer eye-tracker the contents of the monitor are fully recognizable (Table 4 and Figure 3).

With the light switched off, all eye-tracker needs more time to focus. Nevertheless, the highest setting time is just 1,01 second (Ergoneers eye-tracker). In this experiment, SMI's Eye Tracker had the shortest set-up time and the objects on the patient monitor screen were all clearly visible (Table 5 and Figure 4).

Table 4 – ROI 2: Display of Patient Monitoring with Room Light Switched On. .It depicts the time of adaption

	Ergoneers	Tobii	SMI
Light On	0,48 sec	0,96 sec	0,34 sec
	everything recognizable	partly recognizable	partly recognizable

Figure 3 – Screenshot of the Patient Monitor with Room Light On. The first monitor is seen with the Ergoneer eye-tracker, the second with Tobii and the third with SMI.

Table 5 – ROI 2-Display of Patient Monitoring with Room Light Switched Off. It depicted the time of adaption

	Ergoneers	Tobii	SMI
Light Off	1,01 sec	0,84 sec	0,78 sec
	partly recognizable	partly recognizable	everything recognizable

Figure 4 – Screenshot of the Patient Monitor with Room Light Off. The first monitor is seen with the Ergoneer eye-tracker, the second with Tobii and the third with SMI.

In ROI 3 the eye-trackers should record the ventilator (anesthesia machine) with background light to determine what can still be seen when the light is switched on and off. The ability to see details on the ventilator strongly depends on the background colour of the ventilator and the ambient light (Figure 5 and Figure 6).

In total, all eye-trackers needed the longest time to focus the scene with the light on and a backlight monitor. The eye-tracker from Tobii took with 1,42 sec. the longest time. With the Ergoneers and SMI eye-tracker the contents of the screen have been displayed in detail and can be recognized. With the light switched off, it took less than a second to focus again. However, the objects on the monitor are either barely visible or not at all (Table 6 and Table 7).

Table 6 – ROI 3-Display of the Anesthesia Machine with Room Light Switched On.

	Ergoneers	Tobii	SMI
Light On	0,52 sec	1,42 sec	0,58 sec
	everything recognizable	partly recognizable	everything recognizable

Figure 5 – Screenshot of the Anaesthesia Machine Display with Room Light On. The first monitor is seen with the Ergoneer eye-tracker, the second with Tobii and the third with SMI.

Table 7 – ROI 3-Display of the Anesthesia Machine with Room Light Switched Off.

	Ergoneers	Tobii	SMI
Light Off	0,75 sec	0,90 sec	0,58 sec
	nothing recognizable	partly recognizable	nothing recognizable

Figure 6 – Screenshot of the Anaesthesia Machine Display with Room Light Off. The first monitor is seen with the Ergoneer eye-tracker, the second with Tobii and the third with SMI.

Scenario 2: Oral Intubation

During a laryngoscopy the arytenoid is fully visible on all eye-tracker video data. The vocal cord is partially recognized by the Tobii eye-tracker but the other eye-trackers do not display it sufficiently. The glottis is completely recognized by Tobbii and SMI but only partially by Tobii. With all eye-trackers the sections of the scene camera and the real view of the test person were different (Figure 7 and Table 8).

Figure 7 – Screenshots of the Direct Laryngoscopy. The first monitor is seen with the Ergoneer eye-tracker, the second with Tobii and the third with SMI.

Table 8 – Scenario 2 Oral Intubation.

	Ergoneers	Tobii	SMI
Arytenoid cartilage	recognizable	recognizable	recognizable
Vocal chords	not recognizable	partly recognizable	not recognizable
Subglottic area	partly recognizable	recognizable	recognizable

Scenario 3: Resuscitation

Rapid head movements during a cardiac compression changed the calibration of the eye-tracker of Ergoneers (2 rings, 16mm) and SMI (7 rings, 56 mm). The eye-tracker of Tobii had the best fixation in our test without any changing of calibration (Table 9).

Table 9 – Scenario 3 - Deviation After Cardiac Compression

	Ergoneers	Tobii	SMI
Mm	16 mm	0 mm	56 mm
α-Angle	0,92 deg		3,21 deg

Discussion

The aim of the "Changing Gaze" experiment was to analyze the image and video quality of eye-tracker cameras in different light conditions and with different screen backgrounds. Only if the ROI is also recognizable, a subsequent video analysis can be carried out. On average, all Eye Trackers have adapted to the environment and lighting conditions in less than a second. On average it took them a little longer to focus in the dark. Regardless of how the camera is focused, object detection is not always possible. On the display with a dark background, the contents are partially or completely visible when the light is switched off and on. In contrast, the objects on the screen with backlight and the light switched on are almost all recognizable. However, when the light is switched off, the background lighting outshines the contents of the display, so that almost nothing is visible.

During oral intubation, the light conditions between the surroundings and the oral cavity are very high in contrast. The angle at which the subject looks into the oral cavity is small, so that the subject's view is different from that of the scene camera. The exact adjustment of the scene camera must be ensured without disturbing the subject during intubation.

The position at the head and the calibration of the eye-tracker were measured during the resuscitation. The strong head movement led to a slight shift in the calibration of two eye-trackers (Ergoneers and SMI). One eye-tracker was still calibrated afterwards (Tobii). All the eye-trackers sat tight on the subject's head. A change of the calibration can make a subsequent analysis impossible, so that data sets cannot be used [7]. In a qualitative feedback, the subjects said that they would wear the eye trackers during resuscitation [11]. On the other hand, wearing the eye tracker and the restriction of mobility by the eye tracker was perceived as marginal disturbing [5].

There are many reasons for data loss or poor data quality with eye-trackers. The data loss for eye-trackers was approx. 27%, whereby the causes were calibration, poor focus and hard contact lenses in the test person [12]. The quality of recording depends on one hand on the subject that can lead to noisy video material [3]. Poor recordings can also be caused by hard- and software issues of the eye-tracker like equipment failure and poor tracking quality [13] or technical problems [4]. The reasons for poor data quality have not yet been sufficiently investigated, since data quality has never been the goal of the

studies, but always only one aspect of limitation. In addition to the technology, the environment such as lighting conditions and monitor background are also responsible for the quality of the recorded material.

Limitations

Activities and procedures in emergency medicine and anesthetics are crucial for patient safety. To analyze aspects of the sociotechnical systems, the measuring systems have to be proved. Extremely changing light conditions caused by surgical lights, also procedures that look at "dark holes" e.g. the glottis and almost extreme head movements such as resuscitation, necessitate a feasibility analysis of the eye-trackers with regard to efficiency and effectiveness. The limitation of this pilot study is the limited number of rooms in which the study took place. Other operating theatres may have different light conditions and other arrangements of the devices, which may have an effect on the recordings of the eye-tracker.

Conclusions

The aim of this study was the applicability of eye-tracker in anesthesia, which is characterized by changing light conditions, displays with many values and different background illumination. Three eye-trackers are used, but they are not in competition. The measured values and qualitative results do not allow a decision on a winner. All eye-trackers are suitable for use in anesthesia and have their advantages and disadvantages. The Tobii Eye Tracker was the only one that didn't have a shift of the calibration for tight head movements, but took the longest time for the focus position. The Ergoneers and SMI needed the same average time to capture the videos, but with strong head movement the calibration was slightly shifted. All in all, the eye-trackers differ in the display of the objects on the screens. SMI and Ergoneers displayed objects very clearly or not at all, while Tobi Eye Tracker always displayed something, but often a little blurred.

References

[1] St. Pierre M and Hofinger G 2014 *Human Factors und Patientensicherheit in der Akutmedizin* (Berlin, Heidelberg: Springer Berlin Heidelberg)

[2] Law B H Y, Cheung P-Y, Wagner M, van Os S, Zheng B and Schmölzer G 2018 Analysis of neonatal resuscitation using eye tracking: A pilot study *Archives of disease in childhood. Fetal and neonatal edition* **103** F82-F84

[3] Atkins M S, Tien G, Khan R S A, Meneghetti A and Zheng B 2013 What do surgeons see: capturing and synchronizing eye gaze for surgery applications *Surgical innovation* **20** 241–8

[4] Schulz C M, Schneider E, Fritz L, Vockeroth J, Hapfelmeier A, Brandt T, Kochs E F and Schneider G 2011 Visual attention of anaesthetists during simulated critical incidents *British Journal of Anaesthesia* **106** 807–13

[5] Schulz C M, Schneider E, Fritz L, Vockeroth J, Hapfelmeier A, Wasmaier M, Kochs E F and Schneider G 2010 Eye tracking for assessment of workload: A pilot study in an anaesthesia simulator environment *British Journal of Anaesthesia* **106** 44–50

[6] Schulz C M, Schneider E, Kohlbecher S, Hapfelmeier A, Heuser F, Wagner K J, Kochs E F and Schneider G 2014 The influence of anaesthetists' experience on workload, performance and visual attention during simulated critical incidents *J Clin Monit Comput* **28** 475–80

[7] McCormack C, Wiggins M W, Loveday T and Festa M 2014 Expert and competent non-expert visual cues during simulated diagnosis in intensive care *Front. Psychol.* **5** 17

[8] Tien T, Pucher P H, Sodergren M H, Sriskandarajah K, Yang G-Z and Darzi A 2014 Eye tracking for skills assessment and training: A systematic review *Journal of Surgical Research* **191** 169–78

[9] Tomizawa Y, Aoki H, Suzuki S, Matayoshi T and Yozu R 2012 Eye-tracking analysis of skilled performance in clinical extracorporeal circulation *J Artif Organs* **15** 146–57

[10] Klausen A, Röhrig R and Lipprandt M 2016 Feasability of Eyetracking in Critical Care Environments - A Systematic Review *HEC 2016 Health - Exploring Complexity: An Interdisciplinary Systems Approach*

[11] Katz T A, Weinberg D D, Fishman C E, Nadkarni V, Tremoulet P, Te Pas A B, Sarcevic A and Foglia E E 2018 Visual attention on a respiratory function monitor during simulated neonatal resuscitation: An eye-tracking study *Archives of disease in childhood. Fetal and neonatal edition*

[12] Henneman E A, Marquard J L, Fisher D L and Gawlinski A 2017 Eye Tracking: A Novel Approach for Evaluating and Improving the Safety of Healthcare Processes in the Simulated Setting *Simulation in healthcare : journal of the Society for Simulation in Healthcare* **12** 51–6

[13] Tien T, Pucher P H, Sodergren M H, Sriskandarajah K, Yang G-Z and Darzi A 2015 Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair *Surgical endoscopy* **29** 405–13

Address for Correspondence

Dr.-Ing. Myriam Lipprandt
Email: myriam.lipprandt@uol.de,
Department of Medical Informatics
Carl von Ossietzky University of Oldenburg
Oldenburg, Germany, 26129

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190425

Augmenting Analytics Software for Clinical Microbiology by Man-Machine Interaction

Walter Koller[a], Gabriel Kleinoscheg[b], Birgit Willinger[c], Andrea Rappelsberger[d], Klaus-Peter Adlassnig[b,d]

[a] *Department for Hospital Epidemiology and Infection Control, Vienna General Hospital and Medical University of Vienna, Vienna, Austria*
[b] *Medexter Healthcare GmbH, Vienna, Austria*
[c] *Division of Clinical Microbiology, Vienna General Hospital and Medical University of Vienna, Vienna, Austria*
[d] *Section for Artificial Intelligence and Decision Support, Medical University of Vienna, Vienna, Austria*

Abstract

In the present study, we intended to solve identification problems in analyzing the results of microbiology by proactive man-machine interaction. We modified the analytics software MOMO so that it flags laboratory results containing textual elements unknown to the thesaurus, and a human expert assigns the elements to the respective existing thesaurus elements or creates new ones. In 773,309 laboratory results, roughly 2.6% contained unassigned elements and would have been ignored in thesaurus-based analyses for purposes other than simply reporting microbiological findings to physicians. In current use, the thesaurus is kept up to date with synonyms, syntactic deviations, misspellings, and entries not contained earlier, with man-machine interaction of 2–3 hours per week. This approach helps to accommodate both up-to-date clinical reporting for immediate patient care as well as up-to-date queries for infection surveillance and epidemiology, outbreak management, quality control and benchmarking, and antimicrobial stewardship.

Keywords:
Data Analytics; Software; Microbiology

Introduction

Why are clinical information technology (IT) solutions – despite high sophistication and the latest IT standards [1] – sometimes not well accepted by users? This question has been addressed in extensive clinical informatics studies [e.g., 2-4] as well as studies focused on our topic [5, 6], and is still impeding the optimal use of IT tools in clinical and laboratory medicine. A common feature of all studies cited above is the attempt to overcome shortcomings in daily routine outputs which could be allocated to insufficient involvement of human experts during the execution of IT tasks.

In the course of our long-standing involvement in fully automated IT-assisted analysis of microbiological and clinical data, we learned that strict adherence to coded data alone is not enough to avoid deficits in IT analysis [7]. We looked into the discrepancies that emerged when comparing the results of automated IT analyses with the respective gold standards. The search for the root cause led us to elements and terms in our clinical reporting scheme which were not allocated to the thesaurus (terminology coding of the laboratory information system (LIS)), and therefore could not be recognized by our software. In many instances, simple misspellings or orthographic variants were the cause. In other cases, a number, or even whole arrays, of different entities were allocated to a single code, which is why a distinction by code was no longer feasible. So far, such deficits – provided they are detected at all – could explain mistrust in automated IT tools and call for a scrupulous check of each data entry and manual thesaurus allocation of missing terms by human experts. This hardship may be accepted for a research study but is not a realistic approach for a reliable routine IT clinical reporting tool.

Two aspects should be mentioned here: First, free-text entries have been introduced by the users of the microbiology LIS (with arguments discussed later in this paper), thus "invading" the LIS which originally had been focused on merely coded entries and results. Second, microbiology findings play an important role not only in immediate patient treatment but also in contributory disciplines, such as infection surveillance, outbreak management, and antimicrobial stewardship. Hence, microbiology reports and the respective meta-analyses/queries are significant and must therefore be concise, correct, and – last but not least – timely.

The aim of this study was to enhance the precision of our automated analytics and clinical software by adding and augmenting man-machine interaction. In detail, we aimed to solve ambiguity and identification problems in digital reports of clinical microbiology by proactive man-machine intervention.

Methods

General

We modified our automated analytics software so that it flagged laboratory reports which contained elements unknown to the thesaurus and forwarded these to a human expert on a regular basis. The expert then assigned the textual elements to the respective existing entities or created new ones. Thus, we "trained" the thesaurus on a regular basis to recognize possible synonyms, syntactic deviations and misspellings, and thus be extended by new entries.

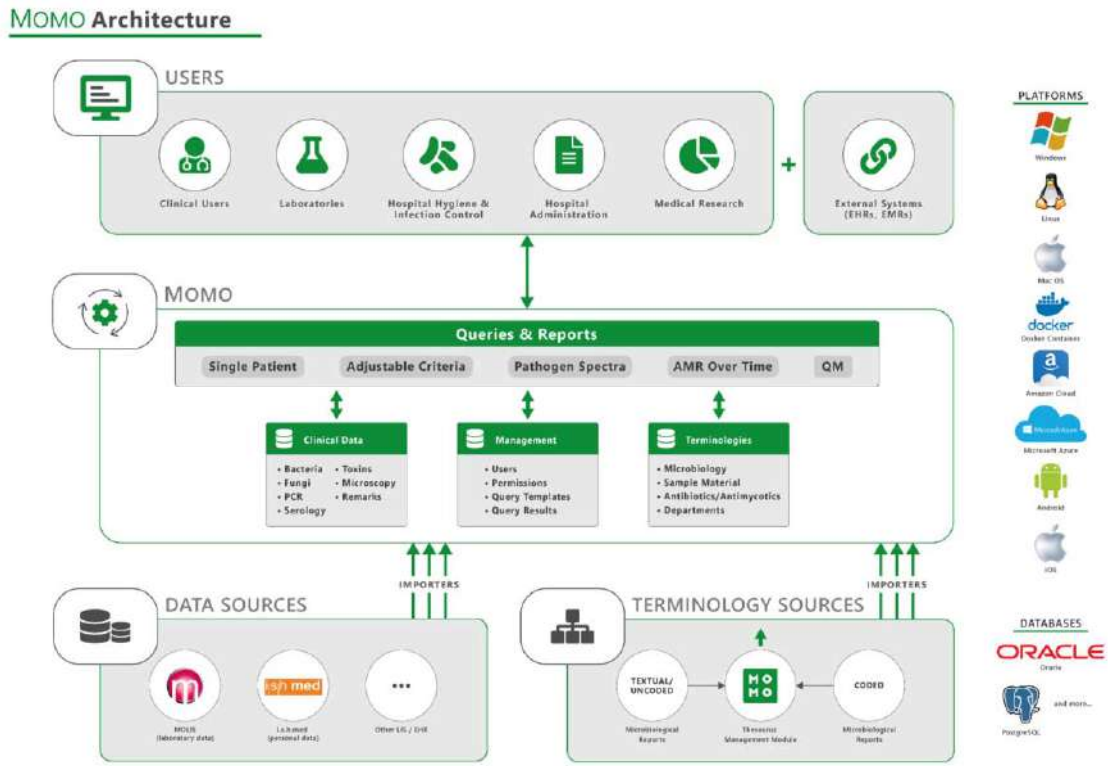


Figure 1 – MOMO is a multifunctional tool for analyzing, monitoring, and reporting pathogens and antimicrobial resistance. It receives 58 different parameters from the microbiology laboratory; four of them are based on terminologies. Most data are structured or coded, some are textual. The latter include not only several comments or report additions, but also microbiological terminology such as bacteria, fungi, PCR, serology, toxins, and microscopy.

Study Setting and Design

We performed a retrospective single-center analysis of validated clinical microbiology results from Vienna General Hospital (VGH), Austria – a 1,900-bed tertiary-care and teaching hospital. Laboratory data were obtained through systematic interrogation of MOLIS (Modular Open Laboratory Information System, Compu Group Medical (CGM) LAB Belgium S.A., Barchon, Belgium) [8] from patients of all VGH clinics from July 4, 2013 to February 16, 2018. MOLIS is designed for IT support of laboratory processes and issues laboratory findings in digital reports, which are delivered as pdf files to the requesters.

Microbiology laboratory data were imported and analyzed by the use of MOMO (Monitoring of Microorganisms, Medexter Healthcare, Vienna, Austria) [9], a microbiology analytics tool for generating analyses of pathogens, spectra, and antimicrobial resistances from routine microbiology results. In addition, MOMO provides immediate answers to questions related to microbiology results for single patients. This feature is now routinely employed by clinicians at their offices and at the bedside. Figure 1 shows the principal architecture of MOMO. MOMO automatically checks incoming textual identifiers (e.g., specimen, detection method, microbes, antibiotics) for compatibility with existing thesaurus entries, and provides different analysis options. It employs four thesaurus categories: requester/department, specimen type, microbiology, and antibiotics/antimycotics.

Thesaurus Management

MOMO uses software elements which check all incoming entities against the thesaurus they belong to. Importantly, each thesaurus may identify entities either by code or by text depending on its configuration. Due to the structure of MOLIS, MOMO's thesauri for requester/department, specimen type, and antibiotics/antimycotics are based on the entities' respective codes. On the other hand, MOMO's thesaurus for microbiology identifies all entities by their texts because the respective codes in MOLIS were frequently inconclusive or missing.

Regarding the microbiology thesaurus, each of its entries (concepts) consists of an internally generated number and the corresponding textual label/name. Externally provided codes can be attached. MOMO's thesaurus management permits the definition of synonyms for each concept or the creation of new concepts (see Figure 2). The concepts may be organized hierarchically into superordinate and subordinate concepts across several levels (e.g., family, genus, species of bacteria). Concepts may possess more than one parent element. *Staphylococcus aureus*, for instance, is a species of the genus *Staphylococcus* within the family of Staphylococcaceae (parent 1); under the distinction of Gram staining, *Staphylococcus aureus* possesses

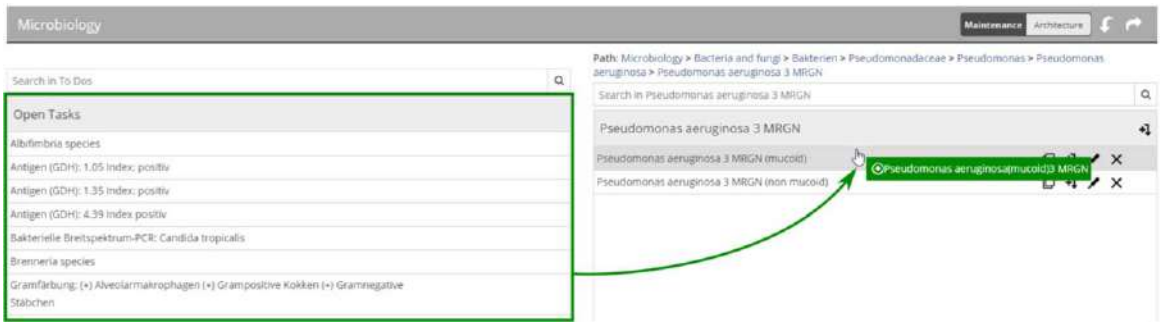


Figure 2 – MOMO thesaurus management – Within MOMO, concepts that cannot be automatically allocated to existing concepts are collected as "open tasks". During thesaurus maintenance these concepts become distinct concepts within the thesaurus hierarchy or will be allocated to an existing concept as a new synonym.

a Gram stain retaining cell wall and is therefore attributed to the Gram-positive group of bacteria (parent 2).

A particular advantage of identifying microbiological results by their texts instead of their codes is that they may be incorporated into the thesaurus immediately, although they may not yet have an officially defined and assigned code [3].

Thus, any microbiology result will be allocated to the respective concept in the thesaurus, or a new concept will be introduced, which is from then on available in clinical as well as secondary query results. Concepts representing misspellings, orthographic variants, or completely new elements must initially be allocated manually; thereafter these texts will be found automatically.

Results

In the investigation period from July 4, 2013 (the day on which the data transfer from MOLIS to MOMO commenced) to February 16, 2018, altogether 773,309 laboratory results were available for analysis. As shown in Table 1, the yearly yield ranged from 154,079 to 209,290, indicating a rising trend. The 2013 and 2018 batches of results were truncated by the respective start and ending dates of the study period.

Table 1 – Imported and analyzed microbiology results by year

Year	Number
2013	25,645
2014	157,627
2015	154,079
2016	196,203
2017	209,290
2018	30,465
Total	773,309

Fractions of Laboratory Results Containing Elements Unknown to the Thesaurus

We started counting the elements unknown to the microbiology thesaurus after the study period had started and can only give an estimate of their total numbers (Table 2).

Roughly 2.6% of all results contained unknown elements and would therefore have been ignored in MOMO analyses based on correctly sorted concepts alone. The types of incompatibility with existing thesaurus concepts were manifold and ranged from typos to completely new elements. Two-thirds required allocation to new sub-concepts under existing concepts. One third of the required actions were the allocation of synonyms or textual variants.

Table 2 – Required thesaurus adaptations after import of 773,309 microbiology results

Type of incompatibility	N*	Fraction
Entries requiring manual allocation	20,000	2.6% of total number
Thereof		
– New species	100	0.5% of N
– New synonym or textual variants	6,900	34.5% of N
– Allocation of new sub-concepts under already existing concepts (especially for serology and microscopy)	13,000	65% of N

* estimated

A prospectively conducted analysis of unassigned microbiology concepts gave a deeper view of the number of concepts allocated to one of the categories within the microbiology thesaurus (Table 3). This analysis includes 89,973 microbiology results from October 30, 2018 to March 12, 2019. Of these, 1,663 (approximately 1.9%) concepts had to be assigned manually. As presented in Table 3, the majority fell into the categories culture, serology, and microscopy.

Table 3 – Current figures for 1,663 manually allocated microbiology concepts between October 30, 2018 and March 12, 2019

Category	N	Fraction
Culture	311	18.7%
PCR	130	7.8%
Serology	525	31,6%
Toxins	72	4.3%
Microscopy	422	25.4%
Miscellaneous	203	12.2%
Total	1,663	100.0%

Effect of the Intervention

On completion of the allocation of unknown elements, 100% of the microbiological results in the trial period became accessible for MOMO analyses.

Manpower Expenditure

In the current use of MOMO, two to three hours per week are required to keep up with changes in laboratory as well as clinical routine. We call this the "man-machine terminology interface". The work is done by persons who are familiar with the MOMO thesaurus as well as clinical and microbiological terms. Top-level clinical or microbiological experts need to be contacted only on rare occasions.

Discussion

Free text entries are known sources of typing errors, misspellings, and unwanted inaccuracies. For this reason, they are "banned" from many modern clinical IT applications, and health care workers are familiar with mandatory coded entries. However, textual descriptions can be more meaningful and may fit individual characteristics better. User-friendly "colloquial" terminologies [2, 4] are indispensable in attracting the use of IT tools by medical professionals.

In the microbiology LIS at the VGH, we were confronted with user demands for free-text elements, which in the course of years of use had become significantly inaccurate. An extraordinary example was a two-digit code connected with 83 – mostly unrelated – entities in the thesaurus. Discrepancies between codes and text were one of the reasons, which had led us to the decision that MOMO should access text rather than codes in the analysis of microbiology terms.

As it turned out, the ambiguity of terms was not an immediate problem for the clinical recipients of microbiology reports, but rather for users relying on MOMO query results, as they would have had to manually select all hidden variations. Hence, the quota of coding incompatibilities observed in our study was more or less irrelevant from the immediate clinical standpoint, but a crucial factor in creating analyses for infection surveillance or outbreak management, where missing results are unacceptable.

As described by de Quirós et al. [2], we had to deal with different acronyms and synonyms for the same clinical finding. For the generation of concise query results, it was necessary to allocate those different terms to the same concept. In contrast to [2], however, we did not have to provide standardized codes based on our thesaurus. Nevertheless, it would be possible to incorporate codes and terms of official terminologies (e.g., SNOMED) as synonyms for those terms that are now available in the microbiology thesaurus.

For this study, we decided not to create a thesaurus based primarily on a so-called reference terminology (e.g., SNOMED as proposed by Rosenbloom et al. [4]) which provides the users with a set of terms as complete as possible. On the contrary, we included only those terms in the thesaurus for which at least one microbiological result was available. Thus, users implicitly know what they may query and do not have to create "test" queries to see whether results are available for certain entries of the thesaurus.

Compliant with, and supplementary to, FAIR principles (code sets are required to be Findable, Accessible, Interoperable, and Reusable) [10], the observations of our study trigger the following microbiology-specific discussion:

Why admit free text entries in lab requests or in (microbiological) lab reports at all?

1. In clinical reporting, we need procedures to deal with inevitable and unanticipated advents of new communication and knowledge elements, which are not yet included in the ordering schemata of the system. This concerns input (emerging clinical demands and laboratory methods) as well as output (new report details or messages). Even brand-new knowledge must be reported precisely and on time, irrespective of the status of the reporting system. This not only applies to microbiology reports for the ordering clinician (which may contain free text information or even handwritten information), but equally to IT analyses and meta-analyses built on microbiology reports, especially if they serve (hospital) epidemiology and outbreak management.
2. Therapeutic imperative: Lab reports must be released a.s.a.p. to facilitate the earliest possible start of appropriate therapy (or adaptation of current therapy) – ideally in a matter of hours. In contrast, as stated by [3], the construction of clinical code sets is usually a time-consuming activity. Thus, the required codes may not be available at the time they are requested.
3. In the man-machine terminology interface, concepts based on free text are organized in a structured manner. This makes them accessible to automated analysis without additional coding.

Following this, the second question pertaining to the analyses of content in microbiology reports is whether they should be based on codes or on free text entries. An analysis based on codes is a plain approach to fully automated analyses; non-coded entries are more difficult to handle and need not be excluded from automated analysis. In free-text-based analyses, each data entry is available for analysis, but the procedure may be time consuming because it requires the scrutiny of hitherto unknown entries.

Neither of the two approaches meets all demands. And, if we agree that non-coded entries have a place in clinical reporting, we ought to provide IT supplements which capture this type of "wild characters" as well.

Our results support two basic recommendations:

1. Optimal and timely thesaurus maintenance is an indispensable provision for all fully automated and rapid IT analyses, the results of which must be trustworthy in terms of their conciseness and correctness.
2. IT systems which draw on codes when appropriate and on free text when appropriate – as MOMO does – provide comfort, speed, and conciseness.

As a result, in thesaurus management high priority is given to capturing synonyms, syntactic deviations and misspellings, which in our analyses were predominant causes of missed entities or misinterpreted reports.

Of similar importance is the ability to capture clinical and microbiological free text comments in laboratory reports which provide special knowledge aimed at new/improved diagnosis and therapy, and which address epidemiology. The ability of a system to communicate such high-level information supports the proficiency/professionality of the daily dialog between laboratory, clinicians and hospital epidemiologists.

Finally, this type of IT tool may serve as an internal knowledge engine for the clinical microbiology laboratory striving for continuous knowledge acquisition and its provision in routine microbiological work.

This topic also encompasses the notorious question: can analyses of routine clinical reporting be accepted for research purposes, especially in epidemiology and public health matters?

This question was discussed extensively in the process of establishing international benchmarking networks (e.g., HELICS [11], IPSE [12], and EARSS [13], which were forerunners of the present European Centre for Disease Prevention and Control networks HAI-net [14] and EARS-net [15]). In the end, it was agreed that routine laboratory data are indispensable for this purpose, despite the fact that they may be generated under missing or unknown scientific standards. There is no other way of obtaining the required information, because impeccable scientific studies of appropriate size and duration are not feasible. In our study, we present a new focus as well as a software solution to bring analysis data from routine clinical reporting closer to the desired degree of conciseness, reliability, and relevance.

Conclusions

In our opinion, clinical IT solutions must focus on a good balance between full automation and man-machine interactivity for successful clinician-laboratory dialog, which in turn supports patient care and infection control.

Given this balance, we may expect considerable progress from such IT solutions in microbiology-related "Good Clinical Practice" as well as in infection prevention and control. Other areas that will benefit from such progress are research in clinical microbiology, healthcare-associated infection prevention and control, epidemiology, and public health!

"When used properly, informatics tools can help the clinical microbiology laboratory to do more with less while improving the quality of patient and public health care" [1].

Acknowledgments

The authors are indebted to Wolfgang Barousch for his fruitful help and discussions.

References

[1] D.D. Rhoads, V. Sintchenko, C.A. Rauch, and L. Pantanowitz, Clinical Microbiology Informatics, *Clin Microbiol Rev* **27**(4) (2014), 1025-1047.

[2] F.G.B. de Quirós, C. Otero, D. Luna, Terminology Services: Standard Terminologies to Control Health Vocabulary, *Yearb Med Inform* **27**(1) (2018), 227-233.

[3] R. Williams, B. Brown, E. Kontopantelis, T. van Staa, N. Peek, Term Sets: A Transparent and Reproducible Representation of Clinical Code Sets, *PloS On*e **14**(2) (2019) :e0212291.

[4] S.T. Rosenbloom, R.A. Miller, K.B. Johnson, P.L. Elkin, S.H. Brown, Interface Terminologies: Facilitating Direct Entry of Clinical Data into Electronic Health Record Systems, *J Am Med Inform Assoc* **13**(3) (2006), 277-288

[5] R. Freeman, L.S.P. Moore, L. García Álvarez, A. Charlett, A. Holmes, Advances in Electronic Surveillance for Healthcare-Associated Infections in the 21st Century: A Systematic Review, *J Hosp Infect* **84**(2) (2013), 106-119.

[6] W. Koller, K.-P. Adlassnig, A. Rappelsberger, A. Blacky, Plea for Use of Intelligent Information and Communication Technologies in Infection Surveillance and Benchmarking by Healthcare Institutions, in M. Bienkiewicz, C. Verdier, G. Plantier, T. Schultz, A. Fred, H. Gamboa (Eds.) *Proceedings of HEALTHINF 2014*, Scitepress, Portugal, (2014) 399-404.

[7] W. Koller, J. de Bruin, W. Barousch, B. Willinger, K.-P. Adlassnig, Man-Machine Dialog: How to Optimize Results of Intelligent IT Tools in Infection Surveillance and in Clinical Decision Support, *Int J Infect Control* **14**(Supp 1: Abstracts of IFIC 2018, 25–27 April 2018, Krakow, Poland) (2018), 10.

[8] CompuGroup Medical (CGM) SE, *MOLIS – Laboratory Information System,* 2017. Available at: https://www.cgm.com/media/cgm_de/documents/labor_1/2017-01-24_MOLIS_Brosch.pdf, last access: 21 October 2018.

[9] Medexter Healthcare GmbH, *Microbiology Analytics with MOMO – Monitoring and Reporting Pathogens & AMR,* 2018. Available at: https://www.medexter.com/products-and-services/clinical-solutions/microbiology-and-amr, last access: 28 March 2019.

[10] M.D. Wilkinson, M. Dumontier, U. Aalbersberg I.J., G. Appleton, M. Axton, A. Baak, et al., The FAIR Guiding. Principles for Scientific Data Management and Stewardship, *Sci Data* **3** (2016) :160018.

[11] HELICS – Hospital in Europe Link for Infection Control through Surveillance, *Surveillance of Nosocomial Infections in Intensive Care Units. Protocol Version 6.1 (Based on Version 5.0 including technical amendments) September 2004*, 2004. Available at: http://www.sicsag.scot.nhs.uk/hai/helics_protocol.pdf, last access: 28 March 2019.

[12] European Centre for Disease Prevention and Control, *Improving Patient Safety in Europe. Technical Implementation Report 2005-2008. November 2008, Vol. I*, 2008. Available at: https://ecdc.europa.eu/sites/portal/files/media/en/healthtopics/Healthcare-associated_infections/HAI-Net/Documents/healthcare-associated-infections-IPSE-Technical-Report.pdf, last access: 28 March 2019.

[13] National Institute for Public Health and the Environment, *AMR Surveillance Infrastructure*, 2015. Available at: https://www.rivm.nl/en/Topics/W/WHO_Collaborating_Centre_Antimicrobial_Resistance_Epidemiology_and_Surveillance/Fields_of_expertise/AMR_surveillance_infrastructure, last access: 28 March 2019.

[14] European Centre for Disease Prevention and Control, *About the Network*, 2018. Available at: https://ecdc.europa.eu/en/about-us/networks/disease-networks-and-laboratory-networks/hai-net-about, last access: 28 March 2019.

[15] European Centre for Disease Prevention and Control, *European Antimicrobial Resistance Surveillance Network (EARS-Net),* 2018. Available at: https://ecdc.europa.eu/en/about-us/partnerships-and-networks/disease-and-laboratory-networks/ears-net, last access: 28 March 2019.

Address for Correspondence

Univ.-Prof. Dr. Walter Koller, Medical University of Vienna; Mail c/o Martinstrasse 26, 1180 Vienna, Austria; Tel.: +43-676-5008106; E-Mail: walter.koller@meduniwien.ac.at.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190426

Identification of Influencing Factors Regarding the Decision for or Against an Open Access Publication of Scientists of Medical Informatics: Description and First Results of Group Discussions and Interviews

Stefanie Kuballa[a], Mareike Schulze[b], Corinna Mielke[a], Monika Taddicken[c], Reinhold Haux[a]

[a] *Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig – Institute of Technology and Hannover Medical School, Braunschweig, Germany,*
[b] *Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig – Institute of Technology and Hannover Medical School, Hannover, Germany,*
[c] *Department of Communication and Media Studies, Institute of Social Sciences, University of Braunschweig, Germany*

Abstract

Open Access provides researchers another opportunity of publishing, besides the traditional publication in subscription-based journals. Providing higher dissemination and therefore visibility as well as better accessibility, among others, Open Access helps to fulfil changed needs of authors and readers in our information and communication society of today. Though this publication model provides a lot of advantages both for readers and authors, there are also some obstacles. In order to identify the incentives that can lead scientists of medical informatics to an Open-Access-publication, we conducted a study consisting of group discussions, interviews, and surveys. This tripartite evaluation starts in its first part with group discussions and interviews. First results of them show that, among others, the higher visibility, indexing, Impact Factor and better accessibility are factors for an Open-Access-publication.

Keywords:

Open Access Publishing, Motivation, Focus Groups

Introduction

Within the last decade, a creeping alteration in the publication landscape can be observed. Due to the changed needs of researchers, in the role of readers as well as authors, in regard to the modified information and communication society, a new publication model has been developed. Open Access (OA) seems to fulfill these needs better than the traditional subscription-based model [1-3]. The internet allows the publication of journal articles, whole journals, or books online and within the OA model without any barriers such as paywalls. This means that everyone with an internet connection can freely access content and depending on the license, distribute, use, and modify it. In recent years many governments, research funding organizations and universities have changed their policies demanding to publish funded research outcomes in OA [4, 5]. For financing this new publication model, a shift of costs from readers to authors can often be found. For publication of accepted manuscripts, they have to pay a so-called "Article Publication Charge" (APC).

Today nearly all journals offer their authors the option to publish their article in OA. There is also a growing number of genuine OA journals, though this depends on the discipline, in which Medical Informatics authors can only choose from a small number of pure OA journals [6]. Another issue is that only a few authors make use of this publication model so far and only about 13% of research papers are published OA [7]. This is effected by concerns connected to OA. Besides the assumed positive characteristics of the publication model as, for example, better visibility and increased citation rates, there are also some significant barriers. An often-higher effort for authors, for example in terms of administrative tasks and the above-mentioned publication charges, is to refrain from OA publishing [8, 9].

The factors that influence researchers in their decision on where or how to publish their research results have not been sufficiently investigated yet. In order to identify them, among others, we are conducting the project "Trans-O-MIM" (full title "Strategies, models and evaluation metrics for the goal-oriented, stepwise, sustainable and fair transformation of established subscription-based scientific journals into open-access-based journals with *Methods of Information in Medicine as example*") [10, 11], funded by the German Research Foundation (DFG). The results are intended to help traditional subscription-based journals transform successfully into OA-based journals. Within the framework of this project, we applied them on the concrete transformation of *Methods of Information in Medicine*.

Therefore, the goal of this study is the exploration of factors influencing researchers in the role of authors both positively and negatively in regards to OA publishing. For the construction of a successful and sustainable business model, it is essential to understand and consider the needs and wishes of authors. So the overall aim of the Trans-O-MIM-Incentives-Study (full title "Study on the Identification of Influencing Factors Regarding the Decision For or Against an Open Access Publication of Scientists of Biomedical and Health Informatics") is to determine what lead researchers to publish their research results to be freely available in OA. The sub-goals are:

- The identification of incentives and/or incentive systems for scientists with regard to a publication in OA (motivators of the use of OA).
- Also of interest is the determination of obstacles and barriers for scientists relating to an OA publication (motivators of the disuse of OA).

- In addition, also the superordinate level needs to be discussed, namely how the scientific system should be shaped for OA and what or how publication service providers can contribute to the promotion of OA from the scientists' point of view.

In the subsequent sections, we described the research method for this study and its composition, its conduction and preliminary results with a special focus on the first phase of the Trans-O-MIM-Incentives-Study. Afterward, we discussed these first findings and future perspectives of the study followed by a conclusion in the last section.

Methods

The main focus of this paper is on the first phase of the Trans-O-MIM-Incentives-Study, though altogether it consists of three parts:

1. Qualitative phase with guide-based group discussions and interviews with selected scientists in the field of biomedical and health informatics.
2. Quantitative phase with a standardized survey of representatives of scientific organizations in the field of biomedical and health informatics.
3. Quantitative phase with a standardized survey of members of scientific organizations in the field of biomedical and health informatics.

All three phases of the study are run in close collaboration with IMIA, the International Medical Informatics Association.

In preparation of the study we designed a questionnaire regarding the incentivizing of OA publications for a short survey that was conducted at HEC 2016 (Health Exploring Complexity: An Interdisciplinary Systems Approach; GMDS & DGEpi & IEA-EEF annual meeting, Medical Informatics Europe – MIE2016; 28 August – 2 September 2016, Munich, Germany), one of the main international conferences in the field of medical informatics, biometry and epidemiology [12]. Based on these not very productive and satisfactory results we decided not to choose mixed methods for the intended study in order to gather more detailed and in-depth information. This leads us to the previous explained study design consisting out of three phases with qualitative and quantitative methods.

Phase 1: Guide-based group discussions and interviews

The guide-based group discussions and interviews are intended to serve as a qualitative pre-stage of the subsequent surveys of scientific organizations and their members. The goal is to make participants narrate and create a discussion that develops its own dynamic in order to gain new and disregarded aspects pertaining to the wishes and needs of scientists in respect of OA publishing. In addition, it is to be reasoned which aspects should be taken into account in the subsequent survey of scientific organizations regarding country differences in OA. This first phase of the Trans-O-MIM-Incentives-Study started in July 2018 and was completed by the last interviews in February 2019.

Taking the results of the preparational survey into account, we created a guide for semi-structured group discussions. The guide serves as a framework and does not have to be followed rigidly, but rather provides the topics that need to be covered. It consists of six theme groups each containing one central question and several in-depth questions (see Table 1). The central questions serve as an introduction to different areas of interest and allow participants an unconfined statement of thoughts and opinions. In-depth questions are used if certain aspects need to be inquired, if answers are not sufficient or if the discussion has haltered. The theme groups are the attitude towards OA, reception behavior at OA articles, publication behavior of scientists, motivators of the use or disuse of OA, positive/negative external influence and promotion of OA.

Table 1– Questions included in the guide

Central questions	Most used in-depth questions
What are the participants thinking about OA?	• How do participants define OA? • Do participants advocate a change of the publication system to OA – and why or why not?
How do the participants read OA articles?	• Where do participants inform themselves about new articles and do they differ between subscription-based articles and OA articles? • How do participants experience new functionalities and evaluation tools, as Altmetrics?
What is the participant's procedure for publishing their research results?	• According to what criteria do participants choose a journal in which they want to publish? • Have the participants already published OA and how was their experience compared to a subscription-based publication? If not: why?
What leads the participants to publish their research results in OA and what deters them from it?	• What do participants think about the licenses that are used for OA? Are they well versed in the various Creative Commons licenses? • To what extent do participants feel constricted by the OA policy of many funding organizations in terms of their publication freedom?
Whereby do publication service providers and the scientific system promote or impede OA in the participant's point of view?	• What offers of publication service providers have a positive/negative influence on the decision for/against OA? • In what way does the scientific system have any structures that support/deter participants in an OA publication?
With what or rather whereby can OA publications be promoted in the participant's opinion?	• What changes pertaining to the scientists are necessary to reach more OA publications? • How will OA develop within the next five years?

For the group discussions, the sample consists of single members of IMIA who are suggested by IMIA for participation. It was planned to include 36 scientists categorized in the IMIA regionalities (North America, Latin America and the Caribbean, Europe, Middle East, Asia and the Pacific, Africa) and three seniority levels (scientists with long experience (e.g. department chairs), scientists with intermediate experience (e.g. postdocs), early scientists (e.g. Ph.D. students)), divided into six groups for discussions via video conference. Due to the difficulty of finding enough participants for group discussions we decided to conduct

additional interviews based on the same guide. Therefore corresponding authors of papers in *Methods of Information in Medicine*'s OA track *Methods Open* were invited for interviews. Additional interviews were held at conferences as for example on APAMI 2018 (APAMI 2018 - 10th Biennial Conference of the Asia Pacific Association for Medical Informatics; 09 – 11 November 2018, Colombo, Sri Lanka).

For the video conferences and non-face-to-face interviews, we used the video conference software of DFN, the German National Research and Education Network, based on Adobe Connect. The discussions and interviews were scheduled to last for an hour each. For an exact evaluation, the soundtrack and video (where available) were recorded. A method of analysis for the group discussions and interviews, the qualitative structuring content analysis of Mayring is chosen. Their data is categorized in accordance to previously determined criteria by use and preparation of a coding scheme (containing e.g. values, variables, and dimensions) and a coding guideline (containing e.g. coding rules and a collection of anchor examples for orientation). After determination of the content-analytical analysis unit, i.e. the coding unit and context unit, the coding of all group discussions and interviews is conducted by marking the discovery points according to the variables [13]. Therefore a transcription of all data in a pure verbatim protocol in connection with special characters is conducted. As not all data is collected at the time of preparing this paper, a preliminary evaluation is based on a comprehensive protocol.

Phase 2: Survey of scientific organizations

This first quantitative phase of the study conduces to capture the various conditions of OA in individual countries. Though prerequisites for publications in OA are obviously heterogeneous, the differences are not yet known in detail. Therefore we want to ascertain them by means of a standardized survey of the respective scientific organizations.

The questionnaire asked for various aspects regarding OA in the country of the respective scientific organization with the main focus on the current situation. Taking into account the findings of the group discussions and interviews, the survey will mainly cover the following aspects:

- In which countries the scientific organization operates
- If they mainly follow the green or gold road to OA and how the sentiment of scientists is
- If the government has provided guidance or recommendations concerning OA and if funding organizations have provided regulations or requirements
- If research institutions (e.g. universities) have OA policies
- How the development of OA is predicted in the next five years

The sampling frame for the survey consists of all scientific organizations that are member societies of IMIA. The target group has already been informed about the upcoming survey during a presentation at APAMI 2018. For convenience, we will conduct an online survey by using the software tool "eSurvey Creator" [https://www.esurveycreator.com]. The online survey is in preparation (see Figure 1) and will take place in April and May 2019. For participation, an e-mail containing a link to the survey will be sent to representatives of several scientific organizations. Upon the data received a quantitative evaluation will be conducted.

Phase 3: Survey of members of scientific organizations

This third part corresponds to the second quantitative phase of the study and builds on the guide-based group discussions and interviews. Herewith it is intended to complete the picture of wishes and needs of scientific authors with regard to OA publishing. While the first phase provides comprehensive and deep insights into the subject, the survey will reach the broad mass of scientists of biomedical and health informatics.

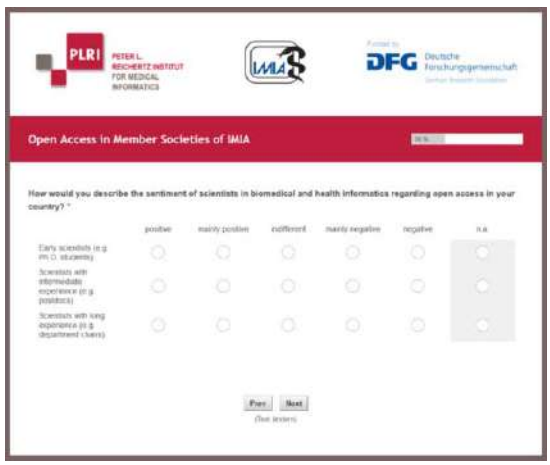


Figure 1– Exemplary page of the online survey

The questionnaire will partially use conjoint analysis to implement various scenarios from which respondents can choose the preferred ones. The questions are divided into various question complexes, according to the guide for group discussions and interviews.

The sampling frame for the standardized survey consists of all members of scientific organizations, who are member societies of IMIA, which currently corresponds to approximately 60,000 people. For greater reach, it will also be conducted as an online survey in May 2019.

Results

This analysis includes the data from 34 participants and was gathered in

- six group discussions with 18 participants suggested by IMIA representatives
- eleven interviews with participants also suggested by IMIA representatives
- five interviews with corresponding authors of a paper in *Methods Open*

Altogether 93 scientists of biomedical and health informatics have been invited for group discussions and interviews, so the participation quote is 36,56 %.

Attitude towards OA

The knowledge regarding OA is very heterogeneous but the whole sample has already heard from it. Most of the participants have a positive and open meaning regarding OA. The broad majority also welcomes a change of the publication system to OA. Thereby many mentioned that a complete

change seems unlikely and that they prefer the coexistence of the traditional system and OA. In this context also the freedom of choice between the publication models was stated as important.

Reception behavior at OA articles

The general tenor of responses to the question about possible differences in the way the participants inform themselves about new print or OA articles is that they do not exist. Some only annotated that OA articles are mostly not included in the subscribed table of contents or alerts, that is a popular information source. Others also outlined that they generally prefer OA articles as they are easier to access even if institutional subscriptions are available. In the way, articles are read no differences have been discovered. The opinion on new functionalities and evaluation metrics is divided and often these tools are still unknown. While some welcome it as a necessary innovation others describe it as insignificant.

Publication behavior of scientists

Asking the participants according to what criteria they choose a journal to publish in following key determinants have been discovered: a suitable scope of the journal to reach the targeted audience, a fast processing time, a high-quality review process and the Impact Factor. Also often mentioned was the cost of publication. The copyright issue by the use of Creative Commons is of little relevance and knowledge to scientists.

The question if participants in the role of authors have already published a paper in OA was affirmed by about half of the sample. Though it needs to be mentioned that only very few differentiated between Green and Gold OA, while about a third knew about the hybrid model. A previous non-publication in OA was often explained with missing funds or experience in publishing.

Motivators of the use or disuse of OA

The answers with respect to the motivators of the use and also the disuse of OA as a publication model show the previously anticipated factors but also new and unforeseen aspects, as presented in Table 2.

Relating to a constriction in their publication freedom due to OA policies of many funding organizations and universities the participants consistently stated that they do not feel restricted. They endorse these policies if existent and do not see any disadvantages.

Table 2– Factors for or against OA publications

Motivators for OA	Motivators against OA
• Faster processing time and publishing of articles	• High costs, for example, Article Publication Charges (APC)
• Higher visibility and therefore more citations	• Lower or initially missing impact factor
• Free accessibility and availability	• Missing reputation of (new) journals
• Indexation in renowned registers	• The distinction between serious and untrustworthy or predatory journals is sometimes difficult
• Better rights on personal use for authors	
• Fulfillment of requirements of funding organizations	• Financing problems, esp. in structurally weaker countries
• Better publication opportunities for scientists from structurally weaker countries	• Fixed/traditional publication structures in the departments
	• There are only a few pure OA journals in the subject area
• Strengthening of the competition for traditional journals	

Positive/negative external influence on OA

Most of the participants had difficulties in answering the questions what offer of publication service providers would have a positive/negative influence on OA and how the scientific system could enhance or impede it. With regard to publication service providers, no concrete ideas for desired offers have been mentioned while the financing policy has been broadly criticized. In this context, it was discovered that nearly no author has made use of waiver systems or even knew about them, though often provided by publication service providers. Identified demanding framework conditions are: facilitation of the financing of articles in OA journals, initiation of new ways/methods for peer review, orientation away from the Impact Factor, production of scientific journals away from commercial publishing houses, and guide to choosing the right OA journal.

Promotion of OA

Regarding the further promotion of OA, no more findings have been gathered. Asking what changes pertaining to the scientists would be necessary most participants responded that they think most researchers are ready for the OA publication model. Some mentioned that senior researchers tend to have more reservations compared to the younger generation.

On the question of how the near future for OA will look like the whole sample predicts a positive development. Most are convinced that the proportion of OA articles will rise steadily but until the transformation is complete (if ever possible) it will still take a long time.

Discussion

The results show that OA has not yet reached the awareness of the majority of researchers of biomedical and health informatics. The findings regarding the attitude towards OA demonstrate accordingly that the topic is not considered in-depth and differentiated by many participants. Moreover, a missing distinction between OA and electronic publications is widely spread. Though a difference between young and senior researchers can be observed. While younger scientists often have only cursory knowledge and experience they are very open-minded on this still young publication model. Senior scientists, in turn, are often more focussed on political aspects of OA publishing.

The reception and publication behavior seems barely influenced by the underlying publication model. This corresponds with the findings that researchers in most cases are not very interested in publication issues and in consequence, are also not aware of them. Interestingly OA articles are not only preferred in structurally weak regions lacking of journal subscriptions but also by well-equipped researchers due to the generally easier accessibility and sometimes also due to the literature search methods. With

regard to the criteria for choosing a journal for publication, it is noteworthy that the scope of the journal is mentioned first and not the Impact Factor. Thereby young researchers often have or want to adhere to the specifications of their principals regarding the choice of journal, so finally, it is their stipulation where and how a paper will be published.

The most mentioned positive and negative motivators for OA are better visibility and accessibility/availability, indexing, more citations, faster processing time and on the downside costs, reputation, Impact Factor and lack of funding. Besides these factors, many participants also provided more detailed factors that can often be attributed to the respondent's seniority or origin. This can also be observed at answers on questions relating to the external influence on OA. On the other hand, only a few findings on further promotion could have been collected, which also seems to correlate with the often found disinterest in publication issues as thoughts regarding this topic have not been made yet.

Future perspectives

As already described in the preceding sections the group discussions and interviews will be followed by two online surveys to gain even deeper insights. The results show that a survey of scientific organizations is necessary in order to clear what influence the origin of participants has on their answers and opinions. A survey of members of the scientific organizations is important to gain a greater sample. This allows a check if the gained findings can be transferred to all scientists of biomedical and health informatics.

Limitations

A limitation to the in this paper focussed the first phase is that not all interviews are conducted yet and therefore a complete analysis and relative frequencies in the results section are missing. However, the sample size has nearly been reached what leads us to the conclusion that a comprehensive picture of factors regarding the decision for or against an OA publication can already be presented. A further limitation to the Trans-O-MIM-Incentives-Study is that this research can not or only limited be transferred to other fields as there are huge differences between several disciplines. Therefore further research for example in disciplines of the social sciences is necessary for comparison.

Conclusions

Through group discussions and interviews provided us with various new insights with regard to the incentives and obstacles of an OA publication, the answers also contain many well-known influence factors. The upcoming surveys will show if there are more, so far unconsidered, advantages and disadvantages of OA. It remains a question of why researchers seem prevalently disinterested in the conscious choice of a publication model. As factors for or against OA are now broadly discovered within this survey, in the next step it might be interesting to find out how researchers can be made aware of the OA publication model in general. This might round up the factors that need to be taken into account for a successful transformation of well-established subscription-based journals into OA-based journals.

Acknowledgments

This research has been supported by Deutsche Forschungsgemeinschaft under grant HA 1438/17-1.

We would like to thank IMIA and its members for supporting this study in various ways.

References

[1] M. Laakso, P. Welling, H. Bukvova, L. Nyman, B.C. Björk, and T. Hedlund, The development of open access journal publishing from 1993 to 2009, *PLoS One* **6** *(*2011), e20961.

[2] K. Kurata, T. Morioka, K. Yokoi, and M. Matsubayashi, Remarkable growth of open access in the biomedical field: analysis of PubMed articles from 2006 to 2010, *PLoS One* **8** *(*2013), e60925.

[3] A.J. Wolpert, For the sake of inquiry and knowledge – the inevitability of open access, *N Engl J Med* **368** (2013), 785-787.

[4] Knowledge Exchange. Bristol: Knowledge Exchanged. open access. [cited 2018 Nov 5] Available from: http://www.knowledge-exchange.info/.

[5] European Commission, *open innovation, open science, open to the world – a vision for Europe,* European Commission, Brussels, 2016. [cited 2018 Nov 5] Available from: https://publications.europa.eu/en/publication-detail/-/publication/3213b335-1cbc-11e6-ba9a-01aa75ed71a1.

[6] S. Kuballa, Open Access Publishing in the Field of Medical Informatics, *J Med Syst* **41** (2017), 82.

[7] R. Schimmer, K.K. Geschuhn, and A. Vogler, *Disrupting the subscription journals' business model for the necessary large-scale transformation to open access*, 2015. [cited 2018 Nov 5] Available from https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_2148961.

[8] C. Haug, The downside of open-access publishing, *N Engl J Med* **368** (2013), 791-793.

[9] R. van Noorden, OA: The true cost of science publishing, *Nature* **495** (2013), 426-429. Errata in **496,** 151 and **499,** 19.

[10] R. Haux, S. Kuballa, M. Schulze, C. Böhm, O. Gefeller, J. Haaf, P. Henning, C. Mielke, F. Niggemann, A. Schürg, and D. Bergemann, Exploring possibilities for transforming established subscription-based scientific journals into open access journals. Present situation, transformation criteria, and exemplary implementation within Trans-O-MIM, *Methods Inf Med* **55** (2016), 481-487.

[11] S. Koch, R. Haux, O. Gefeller, I.N. Sarkar and D. Bergemann, Methods Open – A New Journal Track Starting in 2017, *Methods Inf Med* **55** (2016), 478-480.

[12] S. Kuballa, M. Schulze, C. Mielke, O. Gefeller and R. Haux, On Incentives for Open Access Publishing: A Survey at IMIA's Annual General Assembly During HEC 2016, *Stud Health Technol Inform* **245** (2017), 901-904.

[13] P. Mayring, *Qualitative Content Analysis: Theoretical Foundation, Basic Procedures and Software Solution*, 2014. [cited 2019 Mar 21] Available from https:// qualitative-content-analysis.org/_files/200000075-82241831d6/Mayring(2014)QualitativeContentAnalysis.pdf.

Address for correspondence

Stefanie Kuballa, M. Sc.

Peter L. Reichertz Institute for Medical Informatics

University of Braunschweig - Institute of Technology and Hannover Medical School

Muehlenpfordtstr. 23

38106 Braunschweig

Germany

E-mail: stefanie.kuballa@plri.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190427

Clinical Leaders' Self-Perceived eHealth Competencies in the Implementation of New eHealth Services

Sari Kujala[a], Tarja Heponiemi[b], Pirjo Hilama[c]

[a] *Aalto University, Department of Computer Science, Finland*
[b] *National Institute for Health and Welfare, Finland*
[c] *The South Savo Social and Health Care Authority, Finland*

Abstract

The training and competence of healthcare professionals are key factors in adopting new eHealth services. The scope of eHealth is broadening from information systems to eHealth services such as patient portals supporting self-management, which create a need for new competencies. In this study, we evaluated clinical leaders' eHealth competencies and training needs in two public healthcare organizations in Finland. The target organizations' goal was to increase the number of patients' eHealth services and clinical leaders were seen as critical in this change. Data were collected through an online survey of 98 clinical leaders working in two healthcare organizations. The results imply that managing change and planning implementation are challenging to clinical leaders. They need more information about eHealth services, their possibilities, and benefits in order to support their subordinates and patients. The clinical leaders seem to be in a critical role in supporting healthcare professionals and avoiding resistance to change.

Keywords:
eHealth, Competence, Implementation

Introduction

The training and competence of healthcare professionals have been identified as key factors in adopting Health Information Technologies [1]. Healthcare professionals are end-users operating applications and services, and thus, they need varying competencies for applying the technologies in their work or otherwise the benefits of new technology are not realized.

Previous research has focused especially on students' and nurses' ICT skills and informatics competence [2,3]. The Technology Informatics Guiding Educational Reform (TIGER) [4] identifies three categories of competencies in nursing informatics: basic computer competencies, information literacy, and information management.

While the level of healthcare professionals' basic computer skills is improving, the scope of eHealth is broadening from information systems to eHealth services. By eHealth services, we mean online services such as patient portals that support patients with self-managing their health. Increasingly, eHealth services are used to invite patients to take a more active role in the maintenance of their health [5]. For example, patients may have access to personal health information, educational materials, appointment scheduling, and patient-provider communication tools. Consequently, new eHealth competencies of healthcare professionals have been identified [6]. For example, professionals are in a critical role in motivating and guiding patients to use eHealth services and they need to be able to communicate with patients via a computer.

In addition to training, a champion or a superuser who is an enthusiastic visioneer willing to initiate and drive services forward is good in motivating and supporting staff [7,8]. On the other hand, common barriers to successful adoption of eHealth are negative effects on health professionals' workload, workflow, and roles [9]. Thus, it is not only about individual healthcare professionals' eHealth competencies to use eHealth services that are important, but professionals need to be also motivated and supported in adopting new eHealth services. In addition, work organizations and teams need to plan and adopt new workflows, roles, responsibilities, and care processes.

Leaders have a major role in motivating and supporting healthcare professionals and providing favorable conditions for them when new eHealth services are implemented to their organization. Indeed, there is plenty of existing evidence that effective leadership and change management contribute to the successful implementation of eHealth [10–14]. For example, according to Ingebrigtsen et al. [15] leaders should communicate a clear vision and goals for new eHealth services, provide leadership support, arrange training to healthcare professionals, identify and appoint champions, address work process change, and follow up the implementation.

The leadership activities are important as healthcare professionals and their attitudes are critical in preventing resistance and ensuring the active use of new innovations [16–18]. The good implementation practices that are related to successful implementation of eHealth services are well identified [11,17,19–22]. However, these good practices do not seem to be well known or used in healthcare organizations [23,24]

It should not be considered self-evident that leaders know how to support healthcare professionals. Even if they had good traditional leadership skills, the number of new technologies and eHealth services is constantly increasing and changing the care work. Especially patient empowerment and self-management related to the new eHealth services are changing the traditional expert role of healthcare professionals and creating tension as the control needs to be shared with patients [25,26]. Furthermore, as the new services are also used by patients, they should be informed, motivated and supported.

Patient work needs to be also redesigned in a way that new eHealth services support traditional health care and services efficiently.

In this study, we examine the current state of eHealth competencies of clinical leaders in two public healthcare organizations in Finland. Ingebrigtsen et al. [15] define clinical leaders as leaders responsible for leadership within an organization that delivers care. Clinical leaders include division and department directors, and personnel in designated leader positions in frontline units (e.g., wards, outpatient clinics, primary care practices, etc.) [15].

The identification of deficits in eHealth competencies helps in developing training and support for clinical leaders. We consider eHealth competencies broadly including use, implementation and the organizational change related to the new eHealth services. Thus, we are interested in what kind of challenges clinical leader experience in new eHealth services and their perceptions of the quality of implementation practices used in their units.

The strategy of the target health organizations is to use eHealth services widely, and organizations have already provided a set of self-management services to patients. The organizations' goal is to increase the number of eHealth services provided to patients so that patients initiate their care through eHealth services. In the organizations, clinical leaders' role was perceived as important in supporting healthcare professionals, and the identification of training and support needs was considered essential.

Methods

An online questionnaire was developed to assess the eHealth competencies of both healthcare professionals, including nurses, social workers, and physicians, and their supervisors and leaders. The leaders had their own version of the questionnaire and in this paper, we focus on their responses. The survey contained questions about demographics, self-perceived eHealth competencies, training, and support needs, challenges related to eHealth services, and the use of the recommended implementation practices.

Self-perceived eHealth competencies were assessed as in previous studies [6,27] using multiple-choice questions with competence statements with a five-point Likert scale. The scale ranged from 1 (fully disagree) to 4 (fully agree) and included a fifth option, 5 (I don't know), that was removed from the analysis. Seven statements were selected to broadly represent different eHealth competence areas identified in the literature and found important for leaders (see Table 1).

Also, the use of the recommended implementation practices was assessed using the multiple-choice questions as in our previous study [23,24] and with the same five-point Likert scale. In addition, respondents were asked open-ended questions with regard to training and support needs and challenges related to eHealth services. The quantitative data were analyzed using descriptive statistics and the responses to the open questions were content-analyzed.

The survey study was carried out between June and August 2017 in the South Savo and between October and November in 2017 in the East Savo Social and Health Care Authority. Both authorities are public healthcare organizations in Finland. The healthcare organizations already had eHealth services in use and the goal was to use the services widely.

The survey study was introduced in the intranet news page of the organization and an invitation was sent to participants by their work email. The invitation letter included a description of the target group and the purpose of the survey. Although participation was anonymous, respondents could submit their email address at the end of the survey to participate in a draw of movie tickets and two wireless computer mice.

Results

A total of 98 clinical leaders filled in the questionnaire; the response rates were 24% and 20% of the total healthcare personnel in the studied organizations. The mean age of the respondents was 50.1 years (*SD* = 8.5), 86.5% of them were females and they had worked for an average of 7.7 years (*SD*=7.2) in similar tasks. Most of the respondents were nurse leaders (32%). In addition, respondents included social worker leaders (25%), physician and dentist leaders (11%), other leaders (13%), and managers (19%).

Table 1 shows the self-perceived level of competencies. The healthcare organizations had recently offered training on the security and privacy protection and this competence was evaluated to be the strongest one, 98.8% of the respondents agreed or fully agreed that they are able to work based on the principles. Most of the respondents were also confident that they can use eHealth applications and services and do traditional leading tasks such as leading a change and supporting their subordinates' competence. The respondents had the least confidence in communicating about the new eHealth services to patients, developing services in a customer-centered way and planning an implementation of a new eHealth service.

Table 1– Percentage of clinical leaders agreeing with the competence statements. Scale from 1 (Fully disagree) to 4 (Fully agree).

Competence statement	M	S.D.	Respondents agree (%)
I can work based on the principles of information security and privacy protection.	3.7	.63	98.9
I can use eHealth applications and services.	3.3	.73	92.9
I can support the competence of my subordinates	3.0	.77	89.5
I can lead the change when new services are implemented	2.9	.70	84.2
I can communicate about the eHealth services to patients	3.0	.73	81.5
I can develop services in a customer-centered way	2.8	.78	80.0
I can plan the implementation of a new eHealth service	2.4	.88	52.2

Of the respondents, 42 took the opportunity to answer the open-ended question about their training and support needs. Most frequently, the respondents wanted more information about eHealth services, their possibilities and benefits (10 mentions). In addition, they wished to learn technical skills (4 mentions) and receive information about the contact and responsible

persons (3 mentions). The respondents also realized that they need skills to guide colleagues and customers (2 mentions). The need to learn the customer-centered development work (3 mentions) and implementation skills (2 mentions) were also mentioned.

74 respondents described the challenges of eHealth services. Most commonly, they were concerned about the competence of the healthcare professionals (12 mentions) and how to train, engage and commit the healthcare professionals to use new eHealth services (31 mentions).

Many pointed out challenges related to patients (18), how to guide them and how all different patient groups, especially elderly people, are able to use the services. The shortage of resources, concerning both time and equipment were perceived as challenging (16 mentions) as well as the usability, interoperability, and functioning of the services (14 mentions). The respondents were also unsure how to plan and renew the operations and services (8 mentions) and afraid of the negative attitudes of the professionals and their resistance to change (6 mentions).

Table 2– Percentage of clinical leaders agreeing with the statements of good implementation practices. Scale from 1 (Fully disagree) to 4 (Fully agree).

Competence statement	M	S.D.	Respondents agree (%)
New services are planned to be part of the work process	2.9	.89	69.2
The personnel will be informed enough about the new eHealth services	2.7	.87	67.3
The benefits of new eHealth services are told to the personnel	2.7	.91	62.1
My unit has a person who is responsible for the implementation of the eHealth services	2.7	1.1	62.0
There is enough training for using new eHealth services	2.7	.93	62.0
In our unit, we set goals for the implementation of new eHealth services	2.7	.86	58.7
Our unit has a person who encourages others to use eHealth services	2.7	1.0	58.3
The impact of the new eHealth services to our work is evaluated	2.6	.99	55.7
The personnel has the possibility to participate in the planning of new services	2.6	.97	54.9
We set measures for monitoring the implementation of new services in our unit	2.4	.86	46.9
The personnel is given allocated work time for adopting services	2.3	1.1	40.3

Table 2 presents the respondents' evaluations of how good implementation practices were used in their unit. The results resonate with the fact that almost half of the respondents experienced that they do not know how to plan the implementation of a new eHealth service. One-third of the respondents did not believe that the personnel is not informed enough about the new eHealth services or the benefits of the services. Even fewer respondents agreed that there is a person who is responsible for the implementation or would encourage others to use eHealth services. Almost half of the respondents did not agree that the personnel has the possibility to participate in the planning of new services. Less than half of the respondents' agreed that the implementation of new services is monitored using measures or the personnel is given allocated work time for adopting services as recommended.

Discussion

The results of the clinical leaders' eHealth competence survey in two healthcare organizations show that clinical leaders are in a pressure of many demands and changing environment. They need to support healthcare professionals in the change, organize workflows, make sure that patients are informed and guided, and that the quality of new eHealth services is good enough. They also perceived that the shortage of resources, regarding both time and devices, is challenging.

Many clinical leaders did not feel confident in their tasks indicating that they need more support and training. Especially, they requested more information about eHealth services, their possibilities, and benefits. They can't support their subordinates and patients without knowing about the new eHealth services by themselves.

Majority of the clinical leaders evaluated that they were able to use the new eHealth services, but one-fifth of them felt less confident in new tasks such as communicating about the new eHealth services to patients and developing services in a customer-centered way. Almost half of the leaders perceived that they are not able to plan the implementation of a new eHealth service. It is probable that leaders have not received training of these activities during their basic education and the organizations do not yet have traditions and practices for handling these tasks.

Consistent with the identified lack of implementation skills, the survey results imply that good implementation practices were not systematically used in the organizations. All leaders did not even know who is responsible for the implementation of the eHealth services in their organizations. The benefits of the new services and the goals of the implementation were not clear to all either.

The results suggest that clinical leaders should be better informed about the new eHealth services and their implementations. Many clinical leaders did not know whether the new implementations are well planned and good implementation practices are used. Still, the successful implementation does not only depend on the strategic level of management, but also the operational and frontline levels should participate [15,28].

All leadership levels have an important role in informing, motivating and engaging healthcare professionals. The responses to the open question about the challenges related to eHealth services reveal that the leaders are especially concerned about how to do these tasks for avoiding health professionals' resistance to change.

The results imply that managing change and planning implementations are particularly challenging to clinical leaders. We recommend healthcare organizations to improve the training of clinical leaders on these tasks. In addition, the identified training needs should be considered in the design of continuing education and degree programs.

Informing and engaging numerous health professionals is a challenging task and clinical leaders are in a critical role in this work. Trained clinical champions have been found to be successful in motivating and supporting healthcare professionals [8,11,15] and one potential strategy to healthcare organizations is to train frontline leaders to act as champions in their work units.

The limitation of the study is a small number of respondents in only two healthcare organizations in one country. Thus, the results are not well generalizable, but they are more preliminary in nature. The findings give insights about the potential competence needs and pain points of eHealth leaders. Future studies are needed to test the results in wider settings. However, the studied health organizations can be considered as forerunners as they already have used new kinds of eHealth services supporting self-management and they had the willingness to train and support both leaders and professionals.

Conclusions

In this study, we identified clinical leaders' eHealth competencies focusing on new eHealth services supporting patients self-care and self-management. The results provide insights into the clinical leaders' training and support needs. The new training needs are especially related to managing change, planning implementations of new eHealth services. The clinical leaders are in a critical role in supporting healthcare professionals in order to avoid the resistance to change and ensure that the adoption of new eHealth services is successful.

Acknowledgements

The authors are grateful to the DigiSote project and the South and East Savo Social and Health Care Authorities. This work was supported by the Strategic Research Council at the Academy of Finland, decision number 303606.

References

[1] M. Lluch, Healthcare professionals' organisational barriers to health information technologies—A literature review, *International Journal of Medical Informatics*. **80** (2011) 849–862. doi:10.1016/j.ijmedinf.2011.09.005.

[2] M.K. Lam, M. Hines, R. Lowe, S. Nagarajan, M. Keep, M. Penman, and E. Power, Preparedness for eHealth: Health Sciences Students' Knowledge, Skills, and Confidence., *Journal of Information Technology Education*. **15** (2016).

[3] S. Collins, P.-Y. Yen, A. Phillips, and M.K. Kennedy, Nursing Informatics Competency Assessment for the Nurse Leader: The Delphi Study, *JONA: The Journal of Nursing Administration*. **47** (2017) 212–218. doi:10.1097/NNA.0000000000000467.

[4] S. O'Connor, U. Hubner, T. Shaw, R. Blake, and M. Ball, Time for TIGER to ROAR! Technology Informatics Guiding Education Reform, *Nurse Education Today*. **58** (2017) 78–81. doi:10.1016/j.nedt.2017.07.014.

[5] D. Flynn, P. Gregory, H. Makki, and M. Gabbay, Expectations and experiences of eHealth in primary care: A qualitative practice-based investigation, *International Journal of Medical Informatics*. **78** (2009) 588–604. doi:10.1016/j.ijmedinf.2009.03.008.

[6] S. Kujala, E. Rajalahti, and T. Heponiemi, Health Professionals' Expanding eHealth competencies for Supporting Patients' Self-Management, *Studies in Health Technology and Informatics*. **247** (2018) 181–185.

[7] A. MacFarlane, A.W. Murphy, and P. Clerkin, Telemedicine services in the Republic of Ireland: An evolving policy context, *Health Policy*. **76** (2006) 245–258. doi:10.1016/j.healthpol.2005.06.006.

[8] J.R.B. Halbesleben, D.S. Wakefield, M.M. Ward, J. Brokel, and D. Crandall, The Relationship Between Super Users' Attitudes and Employee Experiences With Clinical Information Systems, *Medical Care Research and Review*. **66** (2009) 82–96.

[9] C. Granja, W. Janssen, and M.A. Johansen, Factors Determining the Success and Failure of eHealth Interventions: Systematic Review of the Literature, *J Med Internet Res*. **20** (2018). doi:10.2196/10235.

[10] J. Sligo, R. Gauld, V. Roberts, and L. Villa, A literature review for large-scale health information system project planning, implementation and evaluation, *International Journal of Medical Informatics*. **97** (2017) 86–97. doi:10.1016/j.ijmedinf.2016.09.007.

[11] K. Cresswell, and A. Sheikh, Organizational issues in the implementation and adoption of health information technology innovations: An interpretative review, *International Journal of Medical Informatics*. **82** (2013) e73–e86. doi:10.1016/j.ijmedinf.2012.10.007.

[12] C.L. Damberg, M.S. Ridgely, R. Shaw, R.C. Meili, M.E.S. Sorbero, L.A. Bradley, and D.O. Farley, Adopting Information Technology to Drive Improvements in Patient Safety: Lessons from the Agency for Healthcare Research and Quality Health Information Technology Grantees, *Health Services Research*. **44** (2009) 684–700. doi:10.1111/j.1475-6773.2008.00928.x.

[13] J.-L. Hsiao, H.-C. Chang, and R.-F. Chen, A Study of Factors Affecting Acceptance of Hospital Information Systems: A Nursing Perspective, *Journal of Nursing Research*. **19** (2011) 150. doi:10.1097/JNR.0b013e31821cbb25.

[14] R.-F. Chen, and J.-L. Hsiao, An investigation on physicians' acceptance of hospital information systems: A case study, *International Journal of Medical Informatics*. **81** (2012) 810–820. doi:10.1016/j.ijmedinf.2012.05.003.

[15] T. Ingebrigtsen, A. Georgiou, R. Clay-Williams, F. Magrabi, A. Hordern, M. Prgomet, J. Li, J. Westbrook, and J. Braithwaite, The impact of clinical leadership on health information technology adoption: Systematic review, *International Journal of Medical Informatics*. **83** (2014) 393–405. doi:10.1016/j.ijmedinf.2014.02.005.

[16] L. Lapointe, and S. Rivard, Getting physicians to accept new information technology: insights from case studies, *CMAJ*. **174** (2006) 1573–1578. doi:10.1503/cmaj.050281.

[17] C.S. Kruse, C. Kristof, B. Jones, E. Mitchell, and A. Martinez, Barriers to Electronic Health Record Adoption: a Systematic Literature Review, *J Med Syst*. **40** (2016) 252. doi:10.1007/s10916-016-0628-9.

[18] S. Kujala, I. Hörhammer, T. Heponiemi, and K. Josefsson, The role of frontline leaders in building health professionals' support for a new patient portal: Survey study, *Journal of Medical Internet Research*. **21** (2019).

[19] D.A. Ludwick, and J. Doucette, Adopting electronic medical records in primary care: Lessons learned from health information systems implementation experience in seven

countries, *International Journal of Medical Informatics*. **78** (2009) 22–31. doi:10.1016/j.ijmedinf.2008.06.005.
[20] P. Esmaeilzadeh, and M. Sambasivan, Health Information Exchange (HIE): A literature review, assimilation pattern and a proposed classification for a new policy approach, *Journal of Biomedical Informatics*. **64** (2016) 74–86. doi:10.1016/j.jbi.2016.09.011.
[21] L.W. Peute, J. Aarts, P.J.M. Bakker, and M.W.M. Jaspers, Anatomy of a failure: A sociotechnical evaluation of a laboratory physician order entry system implementation, *International Journal of Medical Informatics*. **79** (2010) e58–e70. doi:10.1016/j.ijmedinf.2009.06.008.
[22] C. Sicotte, and G. Paré, Success in health information exchange projects: Solving the implementation puzzle, *Social Science & Medicine*. **70** (2010) 1159–1165. doi:10.1016/j.socscimed.2009.11.041.
[23] S. Kujala, I. Hörhammer, M. Ervast, H. Kolanen, and M. Rauhala, Johtamisen hyvät käytännöt sähköisten omahoitopalveluiden käyttöönotossa, *1*. **10** (2018) 221–235–221–235. doi:10.23996/fjhw.69140.
[24] S. Kujala, I. Hörhammer, J. Kaipio, and T. Heponiemi, Health professionals' expectations of a national patient portal for self-management, *International Journal of Medical Informatics*. **117** (2018) 82–87. doi:10.1016/j.ijmedinf.2018.06.005.
[25] S. Mudge, N. Kayes, and K. McPherson, Who is in control? Clinicians' view on their role in self-management approaches: a qualitative metasynthesis, *BMJ Open*. **5** (2015) e007413. doi:10.1136/bmjopen-2014-007413.
[26] R.M. Anderson, and M.M. Funnell, Patient empowerment: Myths and misconceptions, *Patient Education and Counseling*. **79** (2010) 277–282. doi:10.1016/j.pec.2009.07.025.
[27] A. Jauhiainen, P. Sihvo, and H. Ikonen, Terveydenhuollon henkilöstön osaaminen ja valmiudet ottaa käyttöön sähköiset terveyspalvelut, in: A. Jauhiainen, and P. Sihvo (Eds.), Sähköiset Terveyspalvelut Asiakkaan Käyttöön Terveydenhuollossa - Teoriasta Käytäntöön, Karelia-ammattikorkeakoulu, Joensuu, 2014: pp. 26–39.
[28] A.C. Avgar, A.S. Litwin, and P.J. Pronovost, Drivers and Barriers in Health IT Adoption: A Proposed Framework, *Applied Clinical Informatics*. **3** (2012) 488–500. doi:10.4338/ACI-2012-07-R-0029.

Address for correspondence

Sari Kujala, sari.kujala@aalto.fi, +358 50 3862768.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190428

It Needs More Than Just User Participation: Combining Perspectives of Clinical Leaders and Chief Information Officers on Determinants of Hospitals' IT Innovativeness

Jan-David Liebe[a], Moritz Esdar[a], Jens Rauch[a] and Ursula Hübner[a]

[a]*Health Informatics Research Group, Osnabrück University AS, Germany*

Abstract

Although user participation may facilitate the realisation of IT innovations, various literature analyses show only minimal to moderate evidence for such effects possibly due to disregard of mediating factors. Against this background, this study examines the extent to which joint intrapreneurship of clinical leaders and IT leaders as well as a distinct innovation culture mediate the effect of user participation on hospitals' IT innovativeness. IT innovativeness was measured by the availability and usability of IT functions and by the perceived 'innovative power' of a hospital. An empirical model was developed and tested with data from 168 clinical leaders and IT leaders who participated pairwise in a survey representing 84 German hospitals. Three parallel mediation analyses indicated that the participation of users could only lead to IT innovativeness if they were accompanied by intrapreneurial leadership on the part of clinical directors and IT leaders and if a pronounced innovation culture prevailed.

Keywords:

User participation, intrapreneurship, innovation culture, diffusion of innovation, organisational culture

Introduction

Research has repeatedly shown that user participation can facilitate the realisation of IT innovations by blending the technical expertise of the information management department with the system-related functional expertise of users [1]. User participation can refer to various information management activities that take place during IT related innovation processes, ranging from identification of user requirements, implementation and deployment of new applications, to training of colleagues and regular system evaluations [2].

The basic idea behind user participation is to increase the acceptance and motivation of the users and thus contribute to maintaining long-term relationships between the system and its users. In addition to these rather psychological benefits, user participation is expected to generate a number of management-related, methodological and cultural benefits that, taken together, increase an organisations ability to innovate also regarding health IT [3].

Despite this theoretical potential, various literature analyses on the benefits of user participation provide only limited evidence of such effects. Bano and Zowghi [4], for example, conclude in their review of 87 studies that user participation is a double-edged sword which, if not handled properly, can cause more problems than benefits. Frequently cited difficulties that led to misalignments of participatory IT projects included disagreement with project objectives and conflicts about the extent to which users, IT staff and the top management should be empowered to make decisions [4]. Furthermore, He & King [3] showed in a review of 82 studies that the direct effects of user participation on the successful implementation of IT projects are rather minimal to moderate [2]. One reason for the lack of directly measured effects is seen in the fact that mediation factors are widely ignored. A second reason is seen in the different outcome measures with which the effect of user participation is investigated [2-4].

In hospitals, which often are highly specialised, fragmented expert organisations with complex hierarchies, two determinants in particular may mediate the effect of user participation on the ability to innovate regarding health IT: On the one hand, there is the extent to which clinical leaders and IT leaders jointly value IT innovations and therefore promote and demand user participation (top-down mediation). On the other hand, it is the degree of an innovation friendly organisational culture, which enables change through flexible and agile organisational processes and which is characterised by a clear vision of the future, defining the path for corresponding changes and thus facilitating the realisation of user participation (bottom-up mediation) [5].

Against this background, our study explores the following research questions: (1.) To what extent is the effect of user participation on the hospitals' IT innovativeness mediated by the attiudes of the clinical leaders and IT leaders and (2.) to what extent is the effect mediated by the organisational culture.

In theory, these so called top-down and bottom-up mediators can be described with the concepts of intrapreneurship and innovation culture. Intrapreneurship refers to acting on one's own responsibility on behalf of the organisation or part of the organisation, taking risks and anticipating the impact of one's actions, whereby corresponding initiatives and actions mainly concern the development of new products or the reorganisation and optimisation of existing practices [6]. In hospitals, intrapreneurship among executives is seen as an essential precondition of innovation [7]. Corresponding leadership types are characterised as "boundary spanners", constantly looking for innovative optimisation approaches within the hospital (across professional boundaries) and outside the hospital incorporating these approaches into their strategic actions [5,8]. At the same time, intrapreneurial leaders, especially in medium-sized and larger organisations such as hospitals, cannot drive innovative optimisations on their own. They rather rely on interdisciplinary teams that can work together in a purposeful manner to implement innovative concepts [7]. Against this background, it can be assumed that clinical leaders and IT leaders with strong intrapreneurship personalities encourage their employees to work together closely, constructively and therefore innovatively. Thus, we posit:

H1. *The more pronounced intrapreneurship is at the level of clinical leaders and IT leaders the more likely is a positive effect of* user *on a hospital's ability to innovate ~~in~~ health IT.*

Innovation culture can be defined as the extent to which an organisation's values and norms focus on the steady

introduction of new and improved practices and products [8]. Hence, a distinctive innovation culture is characterised by the fact that corresponding optimisation approaches are facilitated by a versatile organisational environment and that they are at the same time guided by an organisation-wide vision of the hospital's future [5]. Previous research suggests that organisational culture in hospitals also has a strong influence on the degree to which IT innovations are disseminated [9]. On the basis of prior studies, that already touched on the association between innovation culture and user participation [10], it is assumed that the nature of the organisational culture (innovation-oriented or not) also determines the direction of user participation (innovation-oriented or not). Thus, we posit:

H2. *The more pronounced the innovation culture of a hospital is, the more likely it is that* user participation *will have a positive effect on a hospital's ability to innovate ~~in~~ health IT.*

Figure 1 displays the research model in which the indirect effects of user participation on the IT innovativeness of a hospital are to be tested. It is assumed that joint intrapreneurship and innovation culture interact in parallel. In order to verify the indirect effect of these mediators, a total of three outcome variables were tested, which together characterise the different facets of IT innovativeness of a hospital. These were a) the availability of a selected set of IT functions, b) the usability of these IT functions and c) the perceived IT related innovation power. The hypothesis H1 is represented by the indirect paths a_1 and b_1 (top-down mediation). Complementarily, hypothesis H2 is represented by the indirect paths a_2 and b_2 (bottom-up mediation). The mediated, direct effect of user participation on innovation power is represented by the path c`.

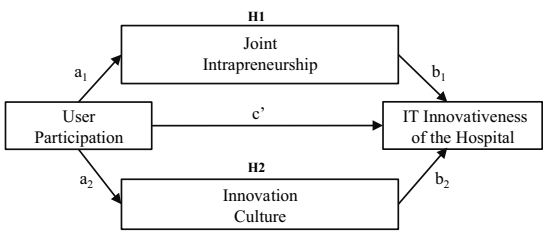


Figure 1 – Research model

Method

The research model was tested with data from 84 German hospitals. One clinical leader and one IT leader of each hospital took part jointly in a survey conducted by the IT Report Healthcare [11]. The IT Report Healthcare consists of a series of surveys that regularly invites IT stakeholders and other key players in all German hospitals to answer questions on topics such as IT maturity and information management.

To test the research model, we conducted three parallel multiple mediator analyses. Parallel mediation analysis allows the simultaneous testing for more than one mediator while accounting for the shared association between them. Model 1 examined the mediating effect of joint intrapreneurship and innovation culture on IT-related innovation power. Model 2 assessed the same mediating effects on the availability of IT functions and Model 3 gauged the mediating effects on the usability of these IT functions.

User participation was measured by asking the clinical leaders to assess the extent to which clinical staff (e.g. physicians / nurses) were involved in IT issues in the hospital on a scale ranging from 1 = "no participation at all" to 10 = intensive participation". The user participation scale comprised seven items that refer to different activities along a typical IT related innovation process in hospitals (starting with participation in IT strategy development, requirements analysis and system selection followed by user training and system evaluation).

Joint intrapreneurship was assessed by both, the clinical leader and the IT leader with regard to three aspects: (1.) to what extent they regularly take the time to think about IT-supported optimisations of hospital operations, (2.) to what extent they regularly exchange information about new IT solutions with external parties (e.g. suppliers, other IT managers, researchers), and (3.) to what extent they stimulate IT-related innovation. In a fourth item, the clinicial leaders were asked to assess the extent to which they regularly seek talks with the IT leader in order to discuss strategic IT issues. Correspondingly, the IT leaders were asked to what extent they seek regular talks on strategic IT issues with clinical leaders. The rating was made on a five-point Likert scale ranging from 1 = "do not agree at all" to 5 = "totally agree".

Innovation culture was also measured by a combined assessment of IT leaders and clinical leaders. Two items were identical in their wording and assessed (1.) the degree to which the hospital, in the participants' opinion, was agile and flexible when it came to the use of new IT solutions and (2.) the extent to which the hospital had a vision of the future that explicitly included the use of IT. In a third item, clinicial leaders should assess the scope to which the IT department was visible in their hospital. In a complementary manner, the IT leaders should rate the degree to which they regularly exchange information with clinical departments. The rating was made on a five-point Likert Scale ranging from 1 = "do not agree at all" to 5 = "totally agree".

Availability of IT functions was assessed solely by the IT leaders. The scale covered eight IT functions in four functional classes that primarily cover advanced IT functions: clinical decision support functions (i.e. medication therapy, alerting, clinical reminders), functions that address patient safety (i.e. electronic tracking of medication loop from ordering to administration), decision support functions (i.e. access to clinical databases) and clinical documentation functions (i.e. discharge letter, electronic nursing records, electronic ICU records). The availability was calculated on a five-point Likert scale ranging from 1 = "No, the IT function does not yet exist and an implementation is not planned" to 5 = "The IT function is completely implemented in all units".

Usability of the IT functions was assessed by the clinical leaders: They were asked to rate how well each of the eight abovementioned functions supported the corresponding documentation tasks and clinical processes. The rating was based on a 10-point Likert scale ranging from 1 = "not satisfied at all" to 10 = "completely satisfied".

Finally, IT-related innovative power was measured by a combined assessment of the clinical leaers and the IT leaders. Therefore, the participants of both groups evaluated how innovative they generally considered the hospital to be in terms of the use of IT on a scale from 1 = "not at all innovative" to 10 = "very innovative". The scales used had already been validated in other studies [9-12]. However, since the answers from two groups were combined for three of the six main variables (joint intrapreneurship, innovation culture and innovation power) and since the item sets deviated slightly from the original scales, we recalculated Cronbach's α to check their internal consistency.

We used PROCESS Version 3.1 for the calculations of the mediation models which utilises an ordinary least squares based path-analytical framework to test for both direct and indirect effects [13]. Since PROCESS does not allow testing the statistical assumptions of the data (i.e. linearity, homoscedasticity, normality of the estimation error and

multicolliniarity), we additionally calculated eleven simple regression models (one for each path in the three mediation models) and three multiple regression models (taking into account all main variables in the three mediation models). Based on these calculations, we furthermore tested the basic requirements for mediation modelling following Baron and Kenny [14]. According to this test procedure, four criteria must be met to determine mediation effects. The first condition is that there is a significant influence of the independent variable on the dependent variable (path c', see Fig. 1). The second condition is that there is a significant influence of the independent variable on the mediator(s) (path a_1 and a_2). The third condition is that there is a significant influence of the mediator(s) on the dependent variable, taking into account the independent variable (path b_1 and b_2). The final condition is that, taking into account the mediator(s), the effect of the independent variable on the dependent variable (path c') disappears or at least decreases.

After testing all model assumptions and the basic requirements for mediation modelling, we calculated the three mediation models with PROCESS. Statistical significance of indirect effects was assessed using bootstrapped bias-corrected percentile based confidence intervals, based on 10.000 bootstrap samples as recommended by Hayes [13]. Lastly, we calculated 95% confidence intervalls for the indirect and total regression coefficents and for the differences between the indirect effects.

Results

Table 1 shows the mean values, standard deviations and Cronbach's α of the six main variables (all values were scaled to range between 1 and 10 to improve comparability) as well as their intercorrelations. Cronbach's α indicated a satisfactory internal consistency for all scales, whereby the value for innovation power was slightly below .6 and was therefore marginally accaptable.

Table 1– Descriptive statistics, α and intercorrelations *(*p < .01) UP - user participation, JIP - joint intrapreneurship, IC - innovation culture, IP - innovation power, AIF - availability of IT functions, UIF - usability of IT functions*

	$\bar{x}$	$\sigma_{\bar{x}}$	α	1	2	3	4	5
1UP	6.2	1.6	.92					.
2JIP	7.3	1.2	.82	.57*				
3IC	7.1	1.4	.79	.49*	.60*			
4IP	6.4	1.8	.59	.43*	.59*	.72*		
5AIF	7.7	1.5	.76	.33*	.54*	.53*	.60*	
6UIF	7.2	1.8	.87	.34*	.48*	.28*	.36*	.26*

The examination of the model assumptions showed that the relationship of the independent and dependent variables could be regarded as linear in all regression models and thus also in the indirect pathes. Estimation errors were distributed relatively evenly over the predicted Y-values in all models and thus no heteroscedasticity was found. Q-Q plots yielded that the estimation errors of the calculated models were normally distributed. The calculated VIF values were all below 2.0, indicating no multicollinearity. On the basis of the four test criteria according to Baron and Kenny [16], it could be assumed that the basic requirements for mediation modelling were met.

The results of the three parallel mediation analyses together generally confirmed the hypothesized model. In all models the effect of user participation was mediated by joint intrapreneurship and in two models additionally mediated by the innovation culture. The direct effect of the user participation on the dependent variables (path c') disappeared in all models, indicating complete mediations. Together, user participation, joint intrapreneurship and innovation culture were able to explain 53.5 % of the variance of innovation power and 33.6 % of the variance of the availability of IT functions. Explained variance of the usability of IT functions amounted to 21.1%. Figure 2 shows the standardised regression coefficients for the relationship between user participation and IT-related innovation power, as it was mediated by joint intrapreneurship and innovation culture with the corrected R^2 displayed at the top right. The standardised regression coefficient between user participation and IT-related innovation power, controlling for the two mediators, is presented in parentheses. As the model demonstrates, user participation was significantly associated with joint intrapreneurship (a_1), and a high degree of joint intrapreneurship was significantly associated with a higher IT-related innovation power (b_1). The standardised indirect effect mediated by joint intrapreneurship (a_1b_1) was (.57)(.25) = .14 (see also indirect effects Tab. 2). Figure 2 furthermore shows that user participation was also significantly associated with innovation culture (a_2), and a high degree of innovation culture was significantly associated with a higher IT-related innovation power (b_2). The standardised indirect effect mediated via innovation culture (a_2b_2) was (.49)(.56) = .27 (Tab. 2). Figures 3 and 4 can be read in the same way by presenting the standardised regression coefficients for the direct and indirect effects of user participation on the availability respectively the usability of IT functions.

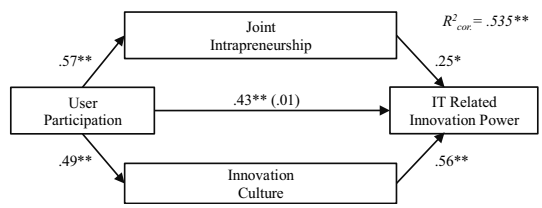


*Figure 2 – Mediation model 1 (*p < .05, **p < .01)*

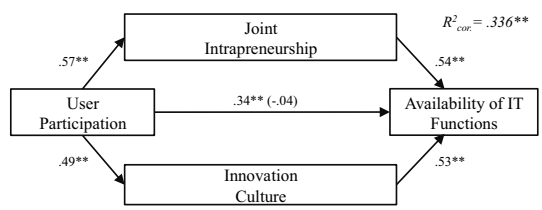


*Figure 3 – Mediation model 2 (**p < .01)*

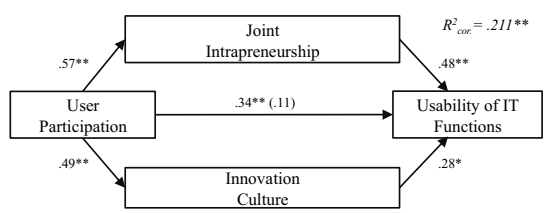


*Figure 4 – Mediation model 3 (**p < .01)*

Table 2 presents the total and indirect effects of user participation on the IT-related innovation power, availability of IT functions and usability of IT functions. In addition, the 95% bias-corrected confidence intervals based on 10.000 bootstrap samples are displayed. The lower bound of the 95% confidence interval (LCI) via joint intrapreneurship was above zero in all models and therefore significant. The same was true for the

indirect effect via innovation culture in models 1 and 2. The difference between the coefficients was not significant in all models (LCI to UCI) in Model 1: -.41 to .06; LCI to UCI in Model 2: -.16 to .23.

Table 2– Total and indirect effects with bias corrected 95% confidence intervals (CI) from bootstrapping. UP - user participation, JIP - joint intrapreneurship, IC - innovation culture, IP - innovation power, AIF - availability of IT functions, UIF - usability of IT functions

		95% CI	
Path	**Coefficient**	**Lower**	**Upper**
Total effects			
UP → IP	.43	.31	.70
UP → AIF	.53	.21	.53
UP → UIF	.52	.08	.45
Indirect Effects			
UP → JIP → IP	.14	.02	.32
UP → IC → IP	.27	.19	.50
UP → JIP → AIF	.31	.06	.35
UP → IC → AIF	.26	.06	.27
UP → JIP → UIF	.27	.13	.46
UP → IC → UIF	.14	-.16	.13

Discussion

With the increasing potential of information technologies to meet the general challenges of efficent and safe health care delivery, the importance of management practices that make IT adoption processes in hospitals smoother and sustainable increases as well. Against this background, user participation is regarded as a promising imperative in information management. Despite this, only moderate empirical evidence of corresponding effects have been provided to date [2-4].

In order to explain the discrepancy between expected and actually proven effects, various mediation factors are discussed in the literature. These range from the complexity and scope of the workflows at hand and the number of IT stakeholders involved in the change process to the complexity of the technology itself and the competencies and attitudes of the users [3]. In addition, some studies point to a mediating effect of leadership behaviour and organisational culture [2-4]. For the hospital environment, in which the latter aspects could be of particular importance due to partially inflexible hierarchies and fragmented organisational structures [5,8], these mediator effects on user participation have not yet been tested on a broad empirical basis.

The present study is a first approach to close this knowledge gap by investigating the mediating effect of intrapreneural leadership and innovation culture on hospitals' IT innovativeness. The chosen study design had two major advantages: on the one hand, it combined the perspective of clinical leaders and IT leaders, so that the multi-professional character of user participation was taken into account. On the other hand, the assumed mediation effects are examined not only in relation to one, but to three outcome variables. This made possible to test whether the mediation effects considered occur independently of the choosen outcome variable.

The results of the mediation analysis confirmed the hypothesis model. It could be shown that there is less of a direct effect of user participation on the IT innovativeness of hospitals, but that this effect mainly occurs when clinical leaders and IT leaders have a strong intrapreneurship personality and when a distinct innovation culture prevails. It could also be shown that the indirect effects explain up to half of the variance of IT innovation, depending on how it is operationalised.

In order to further explain the critical role of intrapreneurial leadership, the findings can be linked to the results of previous studies and existing leadership theories. Two aspects seem to be particularly important. The first one concerns the connection between intrapreneurship and transformative leadership. The results indicate that intrapreneurial leaders are able to achieve a high degree of IT innovativeness on their own by regularly and proactively thinking about optimisation possibilities and searching for suitable ideas inside and outside the organisation. At the same time, intrapreneurial leaders seem to have the ability to transfer this intrapreneurial thinking and acting to their employees. Corresponding indications are already given by Lega [7], who argues that intrapreneurial leaders must have a team behind them in order to implement their innovative approaches. This mechanism of an interpreneurial feedback loop from the leadership level to the employees can be explained by the concept of transformative leadership. Transformative leaders are characterised by the desire to intrinsically motivate their employees. They therefore provide promising visions of the future, offer suitable ways to achieve corresponding goals, act as role models and promote individual employee developments [15]. As such, they are able to forge alliances and propel a sense of communal spirit so that creative momentum from user participation can be transformed for the good of the organisation.

The second important aspect that emerges from the results relates to the joint appearance of intrapreneurship at the leadership level. Since the operationalization of joint intrapreneurship was designed in such a way that high values can only be achieved if both the clinical leader and the IT leader have a pronounced intrapreneurship personality, it can be assumed that positive effects of user participation occur to a lower degree if only one of the two leaders thinks and acts intrapreneurially. Conversely, this also means that participation should not take place at the user level alone if innovations are to be successfully implemented. Rather, participation must reach into the management level, where clinical leaders seek exchange with IT leaders and vice versa. These findings go hand in hand with previous studies that examined the effectiveness of hospital CIOs and concluded that a close connection to the hospital board goes hand in hand with IT performance and IT innovation [5]. Above all, however, the results regarding intrapreneurship highlight the decisive role that clinical leaders can play for the implementation of IT innovations. They act as agents and representatives of the users and close the gap to IT. To put it simply: If clinical leaders are removed from the equation, the greatest organisational efforts, including user participation, will not lead to successful IT innovations [7].

In order to interpret the demonstrated indirect effect of innovation culture, a closer look at the way in which user participation was captured appears to be helpful. The user participation scale measures the extent to which clinical users are involved in information management tasks, which in turn relate to a typical IT innovation process in an organsation - ranging from the identification and selection of new IT solutions to the participation in implementation projects and the conduct of user training and evaluation studys. The results of this study now indicate that these tasks can be carried out more effectively if they are supported by agile and flexible organisational structures. This finding is largely consistent with previous studies which have identified a low level of bureaucracy, low and permeable hierarchies and a low level of formalisation as basic prerequisites for innovative employee activities [8]. The great importance of a common vision of the

future for innovative employee activities has also been identified in other studies [16]. Last but not least, user participation seems to unfold positive effects on the hospital's innovativeness, if it is accompanied by a strong motivation for interprofessional exchange on the part of the clinical units and the IT department.

Finally, the results indicate that the effect size of joint intrapreneurship and innovation culture does not substantially differ. At the same time, the indirect effect of user participation via innovation culture was significant in only two of the three models. Taken together, it can therefore be concluded that without the observed top-down and bottom-up mediators, user participation will have little effect.

The present study has some limitations which also point to further research needs. Although the sample size can be considered sufficiently big taking into account the calculated regression coefficients [17], the examined effects should be retested in order to validate our results. Here, moderating factors such as hospital size or ownership could also be factored in. These characteristics are regarded as structural determinants of the hospital's IT innovativeness, but were not taken into account in the analyses due to the small group sizes of the individual characteristics. As the results suggest that transformational leadership may be the missing link between intrapreneurial leaders and user participation, this could be explicitly tested in future research. In the present study only the joint impact of intrapreneurship was considered, as it was assumed that both groups involved in participation (users and IT) need leadership. The study primarily used quantitative methods to answer the research questions, therefore the results should be further elaborated through a qualitative approach. Finally, the results demonsrate that joint intrapreneurship and innovation culture correlate. Even if no multicollinarity was found in the mediation models, this correlation could be investigated more closely. An innovative culture could, for example, promote the rise of intrapreneurial leaders, who would then be able to lead parts of the organisation to faster change.

Conclusions

This study investigated to what extent joint intrapreneurship and innovation culture mediate the effect of user participation on the IT innovativeness of hospitals. The results indicate that these so-called top-down and bottom-up mediators are not only sufficient but necessary conditions for successful user participation. In summary, the findings consolidate the network of possible preconditions for the hospital's IT innovativeness and - perhaps even more interestingly - contributes to the order of these preconditions.

Acknowledgements

This research was funded by the Federal State of Lower Saxony Germany (no. VWZN3062).

References

[1] N.Saleem, D. Steel, G. Gercek and A. Chandra, Significance of User Participation in a Hospital Information System Success: Insights From a Case Study. *Health Care Manag* **36**(2)(2017), 199–205.

[2] U. Abelein and B. Paech, Understanding the Influence of User Participation and Involvement on System Success - A Systematic Mapping Study, *Emp Softw Engin*, **20(1)** (2015), 28-81.

[3] J. He and W.R. King, The role of user participation in Information systems development: implications from a meta-analysis", J of Manag Inf Sy, **25 (1)** (2008), 301–331.

[4] M. Bano and D. Zowghi. A systematic review on the relationship between user involvement and system success. *Inform and Softw Techn* **58** (2015) 148–169.

[5] K. Cresswell and A. Sheikh, Organizational issues in the implementation and adoption of health information technology innovations: An interpretative review *Int J Med Inform* **82(5)** (2013) e73-86.

[6] B. Antoncic and R.D. Hisrich, Clarifying the intrapreneurship concept. *J Sm Bus Enterp Dev*, **10 (1)** (2003), 7-24.

[7] F. Lega, Strategic, organizational and managerial issues related to innovation, entrepreneurship and entrepreneurship in the hospital context: remarks form the Italian experience. *J Manag & Marke Healthcare* **2(1)** (2009) 77-93.

[8] T. Greenhalgh, G. Robert, F. MacFarlane, P. Bate and O. Kyriakidou, Diffusion of innovations in service organizations: Systematic review and recommendations. *Milbank Quarterly*. **2(4)** (2004), 581–629.

[9] M. Esdar, J.D. Liebe, J.P. Weiß and U. Hübner. Exploring Innovation Capabilities of Hospital CIOs: An Empirical Assessment. *Stud Health Techn Inform* **235** (2017), 383-7.

[10] M. Esdar, J.D. Liebe, B. Babitsch and U. Hübner, Determinants of Clinical Information Logistics: Tracing Socio-Organisational Factors and Country Differences from the Perspective of Clinical Directors. *Stud Health Techn Inform* **253** (2018), 143-147.

[11] IT-Report Healthcare [homepage on the Internet, cited 2018 Nov 25]. Available from: https://www.hs-osnabrueck.de/de/it-report-gesundheitswesen/publikationen/

[12] J. Hüsers, U. Hübner, M. Esdar, E. Ammenwerth, W.O. Hackl, L. Naumann and J.D. Liebe, Innovative Power of Health Care Organisations Affects IT Adoption: A bi-National Health IT Benchmark Comparing Austria and Germany. *J Med Syst* **41** (2017), 33.

[13] Hayes, A. F. (2013). An introduction to mediation, moderation, and conditional process analysis: A regression-based approach. New York, NY: Guilford Press.

[14] R.M Baron and D.A. Kenny, The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* **51** (1986), 1173-1182.

[15] N.K. Jaiswal, L.D. Rajib, Transformational leadership, innovation climate, creative self-efficacy and employee creativity: A multilevel study. *Int J of Hosp Manag* **51** (2015), 30-41.

[16] D.E. Leidner, D. Preston, D. Chen. An examination of the antecedents and consequences of organizational IT innovation in hospitals. *J Strateg. Inf. Syst.* **19** (2010), 154–70.

[17] M.S. Fritz and D.P. MacKinnon. Required sample size to detect the mediated effect. *Psychol Sci* **18** (2007) 233-239.

Address for correspondence

Jan-David Liebe, PhD, email: j.liebe@hs-osnabrueck.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190429

SNOMEDtxt: Natural Language Generation from SNOMED Ontology

Olga Lyudovyk[a], Chunhua Weng[a]

[a] *Department of Biomedical Informatics, Columbia University, New York, NY, USA*

Abstract

SNOMED Clinical Terms (SNOMED CT) defines over 70,000 diseases, including many rare ones. Meanwhile, descriptions of rare conditions are missing from online educational resources. SNOMEDtxt converts ontological concept definitions and relations contained in SNOMED CT into narrative disease descriptions using Natural Language Generation techniques. Generated text is evaluated using both computational methods and clinician and lay user feedback. User evaluations indicate that lay people prefer generated text to the original SNOMED content, find it more informative, and understand it significantly better. This method promises to improve access to clinical knowledge for patients and the medical community and to assist in ontology auditing through natural language descriptions.

Keywords:

Systematized Nomenclature of Medicine, Natural Language Processing, Access to Information

Introduction

SNOMED CT is the world's most comprehensive clinical terminology [1]. The March 2018 release of the US version contains 347,231 unique concepts, including 78,561 diseases, and defines 1,088,068 unique active relationships between these concepts [2]. In contrast, the largest professional medical reference source, Medscape (medscape.com), contains 7,600 diseases, representing less than 10% of the diseases defined in SNOMED CT, and the largest consumer health resource, Mayo Clinic (mayoclinic.org), describes 2,215 diseases. Disease descriptions in these resources are manually curated, thus limiting the number of diseases which can be covered. Topics may be chosen according to popularity in search results [3], thus rare diseases are often excluded from these resources. Counts of disease concepts in major medical information sources are shown in Figure 1. Google Knowledge Graph for diseases is not available, but since it is curated from the sources listed in Figure 1, it is likely on the same order of magnitude.

While extensive, SNOMED CT is not easily accessible to the public and is known to be difficult to use even for clinicians without training in ontologies [4,5]. Like other structured ontologies, SNOMED CT is not designed to be used directly by lay people. The US version of SNOMED CT contains only 4,372 text definitions easily interpretable by untrained personnel, covering 2,608 diseases, corresponding to 1.3% of all SNOMED CT concepts and 3.3% of disease concepts.

We propose a method called SNOMEDtxt to automatically generate disease descriptions from SNOMED CT in order to make available to both patients and the medical community the valuable clinical knowledge contained in SNOMED CT.

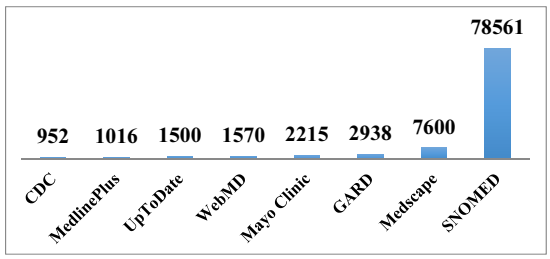


Figure 1– Counts of Diseases in cdc.gov, medlineplus.gov, uptodate.com, webmd.com, mayoclinic.org, rarediseases.info.nih.gov, medscape.com, SNOMED [2] (Nov. 10, 2018)

An additional use case for SNOMEDtxt is to enable clinicians and domain experts without specialised technical training or experience working with structured terminologies to review and critique clinical knowledge defined in SNOMED CT. This task is critically important as biomedical knowledge is growing exponentially with numerous data types and tools emerging rapidly on the daily basis. For example, Campbell et al. reported that the absence of a robust granular ontology represents a barrier to capturing and analyzing data in the field of cancer research and precision medicine [6], while Fung et al. made a similar observation in the area of rare diseases [7]. However, ontology auditing or quality ascertainment is largely performed by knowledge engineers with specialized training in ontology design and maintenance. The workforce of this profession is rare, hence creating a bottleneck for enabling scalable ontology expansion or for crowdsourcing ontology auditing. SNOMEDtxt allows representation of concepts and related information in natural text, thus expanding the group of potential reviewers to include any medical professionals who are not necessarily familiar with structured ontologies. Wider review of SNOMED CT by clinicians can be expected to improve accuracy, reduce missing information, and enable faster SNOMED CT evolution as the body of clinical knowledge expands.

Natural Language Generation (NLG) is a technology utilizing advanced computational methods to generate natural language descriptions from structured knowledge or data representation. Attempts to apply NLG to generate text from SNOMED CT have been reported by Liang et al. [8] and Kanhov et al. [9]. Liang and colleagues developed OntoVerbal, a generic tool for ontology verbalization that was then applied to SNOMED CT. While Kanhov and colleagues utilized an off-the-shelf natural language generator, they developed a methodology for user evaluation of the fluidity and readability of NLG texts in the Biomedical domain. OntoVerbal was developed as a Protégé 4.2 plugin and is not available in more recent Protégé versions

or as a standalone application. The NLG system developed by Kanhov et al. was not made available for download or use.

OntoVerbal implements a generic verbalization approach for ontologies, with an emphasis on the ability to handle any OWL ontology and generate natural language descriptions for any entity type in that ontology [8]. This approach restricts handling of relationships, or ontology axioms, to generic lexical choices and results in some redundant and inelegant phrases, such as "*chronic disease of the genitourinary system … has a finding site in a structure of the genitourinary system*." In contrast, our method trades off generalizability for improved readability and comprehensibility through more specific verbalizations of SNOMED CT axioms and simplifying structures tailored to SNOMED CT concepts, so that the same construct is simplified by SNOMEDtxt as "*… affects the genitourinary system*." Moreover, OntoVerbal takes the generic approach to ordering information from simpler sentences to more complex ones, whereas SNOMEDtxt follows the common flow of information found in disease descriptions in reference medical texts: definition is followed by possible causes, presentation, diagnosis, clinical course, and finally additional information.

SNOMEDtxt is a novel NLG engine and interface, intended to evolve and improve over time with user feedback. The current version focuses specifically on disease concepts and can be easily extended to summarize procedures, treatments, and other information contained within SNOMED CT and relevant to the wider audience.

Methods

SNOMEDtxt follows a 4-step framework outlined in Figure 2 to generate a disease description for a given disease.

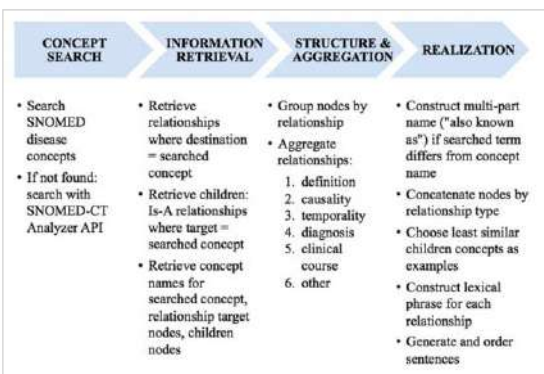


Figure 2 – Framework for Disease Description Generation

Concept Search And Information Retrieval

The current implementation of SNOMEDtxt is based on the 03/01/2018 release of SNOMED CT terminology, US edition, Snapshot version, available for download from the SNOMED CT website [2]. SNOMEDtxt uses a local copy of this database.

The system has the capability to randomly sample diseases from SNOMED CT and to search for disease names entered by a user. The search is undertaken in two steps: first, a simple match on concept names and synonyms in SNOMED CT database is attempted. If the search term is not found, the system then uses SNOMED CT Analyzer API (snomedct.t3as.org) to search for the term, provided that the API site is online.

Concepts are the key component of SNOMED CT. They are organized in a polyhierarchical structure with "Is-A" (parent-child) relationship and can be additionally defined or described through other relationships. Each relationship has a type, a source concept, and a destination concept. Once a disease Concept ID is found, relevant relationships are retrieved from SNOMED CT database:

- Relationships where the searched disease Concept ID is the source
- Is-A relationships where the searched disease Concept ID is the destination: the source concepts represent subtypes or examples of the disease and are included in the definition

Concept names are then retrieved for the corresponding target concepts. The generated text is the product of concept names arranged in lexical patterns corresponding to types of relationship between these concepts. Concept names undergo minimal string cleaning to remove non-informative structures such as "(Disorder)" and "(Body Structure)".

Structure And Aggregation

In order to produce fluid and coherent text and avoid redundancy wherever possible, SNOMEDtxt aggregates and structures information in three steps: Firstly, it groups all target nodes for the same relationship; secondly, it organizes relationships in broad logical groups; thirdly, it orders relationships within each group and the groups themselves following a typical flow of information in a disease description in medical reference texts. This stepwise grouping of relationships is a simplified application of the Rhetorical Structure Theory [10] that describes a recursive approach to organizing relationships in a text.

Table 1 – Organizing Relationships

Group	Relationship	Lexical Pattern
Definition	IS-A	"is a kind of"
	Finding site	"that affects the"
	Has definitional manifestation	"It manifests itself in"
	Associated morphology	"The associated morphology is"
	Pathological process	"Pathological process associated with … is"
	Children: IS-A, searched term=destination	"An example of … is" / "Examples of … are"
Causality	Causative agent	"is caused by"
	Due to	"occurs due to"
	Associated with	"is associated with"
Tempora-lity	Occurrence	"presents in" (period)
	During/Following/ After	"can occur during / following / after"
	Temporally related	"can be temporally related to"
Diagnosis	Finding method	"is discovered by"
	Finding informer	"<is discovered> through"
Clinical Course	Clinical course	"Clinical course is"
	Severity	"The severity of … is"
	Episodicity	"The episodicity of … is"
Other	Interprets	"interprets or evaluates"
	Has interpretation	"… as"
	Other	"Other related concepts include…"

Text Realization

The first task SNOMEDtxt undertakes in the Text Realization phase is constructing an informative disease name. If the search term is significantly different from the preferred term for the disease concept, as measured by Jaro-Winkler string distance [11], the disease description will combine both in the form of "<Preferred disease concept name> (also known as <searched term>)", e.g. "Influenza (also known as flu)".

Additionally, SNOMEDtxt concatenates all target nodes for the same relationship type which were aggregated in the previous step by following the "A and B", "A, B, and C" format. When concatenating examples of a given disease, SNOMEDtxt selects a maximum of three examples, based on the largest string dissimilarity with the given disease name, as a tradeoff between completeness and relevancy.

Finally, relationship types are converted into corresponding lexical patterns (see Table 1) and sentences are generated. For the sake of conciseness, relationships in the same group are combined into one sentence wherever this approach produces fluid text. For example, the Is-A and the Finding site relationships are combined into one sentence that forms the concise definition of the disease: "*Asthma is a kind of Respiratory disorder that affects the Airway*". Sentences are then ordered according to the order of relationships in Table 1.

Results

User Interface of SNOMEDtxt

A simple user interface is implemented in RShiny and is available online at https://sno2eng.shinyapps.io/sno2Eng.

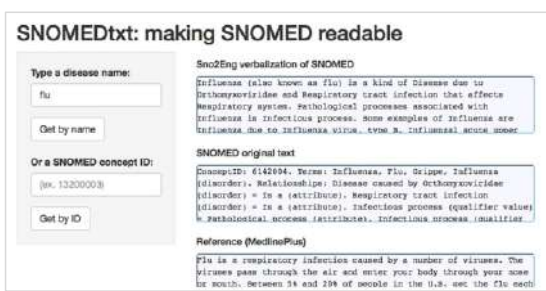


Figure 3 – Screenshot of SNOMEDtxt Interface

An example disease description generated by SNOMEDtxt and the corresponding concatenated SNOMED CT content are illustrated below.

SNOMEDtxt Disease Description

Lupus erythematosus (also known as Lupus) is a kind of Autoimmune disease and Connective tissue disease that affects Connective tissue. Some examples of Lupus erythematosus are Systemic lupus erythematosus, Drug-induced lupus erythematosus, and Neonatal lupus erythematosus. Pathological process associated with Lupus erythematosus is AI - autoimmune. Other related concepts are Cutaneous lupus erythematosus, Lupus erythematosus profundus, and Discoid lupus erythematosus of eyelid.

SNOMED CT Content

ConceptID: 200936003. Terms: Lupus erythematosus, LE - Lupus erythematosus, Lupus, Lupus erythematosus (disorder). Relationships: Disorder of connective tissue (disorder) = Is a (attribute). Connective tissue structure (body structure) = Finding site (attribute). Autoimmune disease (disorder) = Is a (attribute). Autoimmune (qualifier value) = Pathological process (attribute). Related concepts: Systemic lupus erythematosus (disorder) - Is a (attribute). Drug-induced lupus erythematosus (disorder) - Is a (attribute). Neonatal lupus erythematosus (disorder) - Is a (attribute). Discoid lupus erythematosus (disorder) - Due to (attribute).

We evaluated disease descriptions generated by SNOMEDtxt against the concatenated SNOMED CT content using computed metrics and user evaluations. Both sets of evaluations indicate that SNOMEDtxt succeeds in making SNOMED CT content more readable and comprehensible.

Computed Metrics

We computed readability and redundancy metrics for disease definitions of the top 20 most searched diseases in 2017 [13] and of 20 diseases randomly retrieved from SNOMED CT:

1. Readability: Flesch-Kincaid grade level (FK) and Automated Readability Index (ARI) estimate the number of years of education needed to understand a text. We calculated both with sylcount R package [12].
2. Redundancy: calculated as the ratio of unique word count to total word count after removing stop words.

Full summaries of health concepts retrieved from Medline Plus web service (MedlinePlus.org) were used as reference for the first set of disease concepts. Since only 4 out of 20 randomly sampled disease concepts had a reference health topic in Medline Plus, comparison with reference is not provided for the second set.

Table 2 – Evaluation with Computed Metrics

	Readability		Redundancy	
	FK	**ARI**	**Words**	**Unique/All**
Top 20 most searched diseases				
SNOMEDtxt	14.3	12.0	49.3	0.74
SNOMED CT	17.9	15.0	64.1	0.55
Reference	6.6	6.1	263	0.77
Random 20 SNOMED CT disease concepts				
SNOMEDtxt	11.7	9.7	47.3	0.69
SNOMED CT	15.7	13.8	69.7	0.56

For both measures of readability, a lower score indicates a lower grade of education needed to understand the text and therefore better readability. These metrics indicate that SNOMEDtxt texts are more readable than the original SNOMED CT content. For the 20 most searched diseases, the average FK score for SNOMEDtxt texts (14.3) is equivalent to the second year of undergraduate degree, and FK for SNOMED CT content (17.9) corresponds to the graduate school level. ARI score of 12.0 for SNOMEDtxt is equivalent to twelfth grade, while ARI of 15.0 for SNOMED CT content indicates that the text is appropriate for readers at the Professor level. Readability scores for the MedlinePlus reference texts are significantly lower, indicating that they can be read by a much wider audience than either SNOMEDtxt or the original SNOMED CT content.

SNOMEDtxt texts also improve on the redundancy metric compared to SNOMED CT content for the top 20 searched diseases (0.74 vs. 0.55) and for the 20 randomly sampled diseases (0.69 vs. 0.56).

User Evaluation

A survey evaluating results of SNOMEDtxt was conducted among 51 lay people recruited using Amazon Mechanical Turk (MTurk) and 6 clinicians from Columbia University Medical Center. MTurk is a crowdsourcing marketplace that enables outsourcing tasks like surveys to a distributed workforce for a small reward. Evaluations of all MTurk taskers that applied and did not self-identify as clinicians were included in the results. Evaluations of all 6 clinicians who responded to the survey were included in the results. All evaluators were provided with a basic description of the project, but were not aware of the study design or the research question.

We randomly selected a set of 20 disease concepts from SNOMED CT for evaluating readability, preference, accuracy, and completeness (set 1). Helpfulness was evaluated on a set of 20 disease concepts for which a medical reference text was available (set 2). Questions probing the degree of understanding were constructed for 10 diseases with sufficient information selected from 40 randomly sampled disease concepts (set 3). Comparison with OntoVerbal was restricted to 4 diseases for which OntoVerbal description was available in [8] (set 4). For all 4 sets, we generated a SNOMEDtxt disease description and a concatenation of SNOMED CT content. The survey was conducted using Qualtrics survey platform (www.qualtrics.com) and included randomization: each evaluator was presented with 3 randomly selected diseases from set 1, 2 from set 2, 2 from set 3, and 1 from set 4.

In order to assess readability and general preference, we presented evaluators with SNOMEDtxt disease descriptions and the SNOMED CT content for 3 diseases from set 1 and asked whether one or the other was more readable or generally preferred, or there was no difference. Evaluators were not informed which text represented SNOMEDtxt output. Lay people found 76.5% of SNOMEDtxt disease descriptions easier to read than the SNOMED CT content, and preferred 69% of SNOMEDtxt descriptions to SNOMED CT content. Clinicians found 83% of SNOMEDtxt descriptions easier to read and preferred 44% of them to the SNOMED CT content.

We tested understanding by presenting the evaluators with either the SNOMEDtxt description or the SNOMED CT content for a concept, followed by a multiple choice question designed to test whether the evaluator understood the text; we then compared the number of correct answers given when presented with SNOMEDtxt description or with the SNOMED content. SNOMEDtxt format appeared to be significantly easier to understand for lay users: they gave the correct answer 72% of the time when presented with SNOMEDtxt description and only 51% when presented with SNOMED CT original content. There was no difference for clinicians: they gave the correct answer 100% of the time regardless of what text they were presented with.

To evaluate helpfulness, we presented evaluators with the SNOMEDtxt description, the SNOMED CT content, and a description of the same concept from either Medline Plus or Google Knowledge Graph as a reference and asked "How helpful was the terminology content compared to" the reference, on a scale from 1 to 10. Lay people found SNOMEDtxt descriptions more helpful: the average helpfulness score for SNOMEDtxt texts was 5.7, compared to 4.8 for SNOMED CT content. On the other hand, clinicians found SNOMEDtxt descriptions on average minimally less helpful than SNOMED CT content (3.50 versus 3.58).

Clinician evaluators were also asked to assess accuracy and completeness for disease concepts from set 1. In most cases clinicians thought the SNOMEDtxt descriptions were as accurate (72%) and as complete (78%) as the original content, while they found 28% of descriptions to be somewhat less accurate, 6% somewhat less complete, and 17% significantly less complete.

Table 3 – User Evaluation: SNOMEDtxt vs. SNOMED

Readability and Preference

	SNOMEDtxt	SNOMED CT	No Difference
Lay Audience (n=51)			
Easier to read	76.5%	14.4%	9.2%
Preferred	68.6%	21.6%	9.8%
Clinicians (n=6)			
Easier to read	83%	11%	6%
Preferred	44%	28%	28%

Helpfulness and Understanding

	SNOMEDtxt	SNOMED CT
Lay Audience (n=51)		
Helpful (1-10)	5.7	4.8
Correctly understood	72.1%	51%
Clinicians (n=6)		
Helpful (1-10)	3.50	3.58
Correctly understood	100%	100%

Accuracy and Completeness

SNOMEDtxt vs. SNOMED CT	Significantly Worse	Somewhat Worse	Same
Clinicians (n=6)			
Accuracy	0%	28%	72%
Completeness	17%	6%	78%

A conclusive comparison between OntoVerbal and SNOMEDtxt was not feasible since only 4 disease descriptions were available for OntoVerbal. We conducted a limited comparison by presenting all evaluators with the SNOMED CT content and with a disease description from either OntoVerbal or SNOMEDtxt for the same disease (evaluators were unaware of the source of each text). All evaluators were asked which text they found easier to read and generally preferred; clinicians were additionally asked whether the text description was less accurate / complete than (denoted in Table 4 as "Worse") or as accurate / complete as (denoted as "Same") the SNOMED CT content. This limited comparative evaluation points to a preference for SNOMEDtxt disease descriptions with the same or better performance on readability, accuracy, and completeness.

Table 4 – User Evaluation: Comparison with OntoVerbal

	SNOMEDtxt	Onto Verbal	SNOMED CT	No Difference
Lay Audience (n=51)				
Easier to read	49%	43%	3.9%	3.9%
Preferred	52%	31%	11.8%	3.9%
Clinicians (n=6)				
Easier to read	50%	50%	0%	0%
Preferred	50%	17%	17%	17%
	SNOMEDtxt vs. SNOMED CT		**OntoVerbal vs. SNOMED CT**	
Clinicians (n=6)	**Worse**	**Same**	**Worse**	**Same**
Accuracy	0%	45%	27%	27%
Completeness	18%	37%	18%	27%

User evaluation demonstrates potential utility of SNOMEDtxt for lay users: they find the generated disease descriptions more readable and easier to understand than the structured SNOMED

CT content. The accuracy and completeness of SNOMEDtxt's natural language descriptions is close to the original SNOMED CT content. The use case of assisting in SNOMED CT content review would require some adjustments to SNOMEDtxt design in order to produce more faithful representations of the SNOMED CT content.

Discussion

We introduce a method to generate disease descriptions directly from the SNOMED CT ontology for two main applications: providing access to definitions of rare diseases or disease variants not described in clinical reference resources and enabling easier comprehension of SNOMED CT content for those reviewing, verifying, and extending the ontology.

In the design of SNOMEDtxt, we have made several choices that favor fluidity and ease of comprehension over faithful and complete representation of information, at the risk of possible loss of information. The human evaluation of results confirms that we achieved the goal. However, these choices may not be appropriate when SNOMEDtxt output is used to verify content of SNOMED CT. It may be desirable to provide users with configurations such as "more precise" and "easier to understand" when generating the natural language texts. Another tradeoff made in the design of SNOMEDtxt was readability at the expense of generalizability. In order to extend SNOMEDtxt to other types of concepts or to other terminologies, verbalizations of relationships and handling of aggregated sentence structures would need to be adjusted.

A significant limitation to the use of SNOMEDtxt for the wider audience is the amount of content available for each disease concept in SNOMED CT. Expanding, i.e. explaining, some related nodes, for example parent disease node or finding site, may add meaningful and relevant information to the generated disease descriptions. A navigable user interface where a user could click on confusing terms and see them explained would be an alternative approach to this challenge. Developing APIs to access SNOMEDtxt would enable integration of textual disease descriptions into other electronic resources and reference materials, such as EHR help function or patient portals. The search functionality in the current implementation is limited to exact string match with either the SNOMED term name or any of the term's synonyms and can be further improved with string search algorithms.

Results of the evaluation by lay people and clinicians presented in this paper are encouraging for the potential use of SNOMEDtxt in making SNOMED CT content more accessible and easier to review; however, a more rigorous evaluation with a larger audience and a greater number of tested concepts is recommended.

Finally, to allow the system to continuously learn and evolve, evaluation and feedback elicitation can be bulit into the user interface. Presenting users with different verbalization options at random and gathering user feedback would enable the system to learn verbalization patterns favored by users and evolve the NLG engine accordingly.

Conclusion

This work presents an ontology verbalizer for SNOMED CT disease concepts: a tool that generates natural language concept descriptions balancing completeness and accuracy with the ease of human comprehension. User evaluation shows that lay people prefer to read natural text instead of structured ontologies and understand textual descriptions better.

More broadly, natural langauge processing is growing in importance with many potential applications in Healthcare systems. NLG involves several important tradeoffs, which should be made with a specific application in mind. Two such tradeoffs are balancing completeness and accuracy on one hand with fluidity and comprehensibility on the other; and generalizability versus linguistic polish and expressiveness.

References

[1] J. Millar, The Need for a Global Language - SNOMED CT Introduction, *Stud Health Technol Inform* **225** (2016) 683-5.
[2] US National Library of Medicine, SNOMED CT United States Edition, www.nlm.nih.gov/healthit/snomedct/us_edition .html (Accessed April 28, 2018).
[3] N. Miller, E. M. Lacroix, J. E. Backus, MEDLINEplus: building and maintaining the National Library of Medicine's consumer health Web service, *BMLA* **88(1)** (2000), 11-7.
[4] S.Y. Kim et al., Comparison of Knowledge Levels Required for SNOMED CT Coding of Diagnosis and Operation Names in Clinical Records, *Healthc Inform Res* **18(3)** (2012), 186-90.
[5] J. E. Andrews, R. L. Richesson, J. Krischer, Variation of SNOMED CT Coding of Clinical Research Concepts among Coding Experts. *JAMIA* **14(4)** (2007), 497–506.
[6] W.S. Campbell et al. A computable pathology report for precision medicine: extending an observables ontology unifying SNOMED CT and LOINC, *JAMIA* **25(3)** (2018), 259–266
[7] K.W. Fung, Coverage of rare disease names in standard terminologies and implications for patients, providers, and research, *AMIA Annu Symp Proc* (2014 Nov 14) 564-72.
[8] S.F. Liang et al., OntoVerbal: A Generic Tool and Practical Application to SNOMED CT, *IJASCA* **4(6)** (2013) 227-239.
[9] M. Kanhov, X. Feng, H. Dalianis, H. Natural Language Generation from SNOMED Specifications, *CLEFeHealth 2012 Workshop* (2012).
[10] M. Taboada, W.C. Mann, Applications of Rhetorical Structure Theory, *Discourse Studies*, **8(4)** (2006), 567–588.
[11] W.E. Winkler, String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage, *American Statistical Association,* (1990), 354–359.
[12] D. Schmidt, 'sylcount' R package (2017).
[13] K. Sheridan, The 20 most Googled diseases, *StatNews* (June 6, 2017).

Address for correspondence

Chunhua Weng, Department of Biomedical Informatics, Columbia University, email: chunhua@columbia.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190430

Exploring the Social Structure of a Health-Related Online Community for Tobacco Cessation: A Two-Mode Network Approach

Shruthi Manas[a], Lindsay E. Young[b], Kayo Fujimoto[c], Amy Franklin[a], Sahiti Myneni[a]

[a] *Department of Biomedical Informatics, University of Texas, Houston, Texas, USA*
[b] *Department of Medicine, University of Chicago, Illinois, Chicago, USA*
[c] *Department of Public Health, University of Texas, Houston, Texas, USA*

Abstract

Unhealthy behaviors, such as tobacco use, increase individual health risk while also creating a global economic burden on the healthcare system. Social ties have been seen as an important, yet complex factor, to sustain abstinence from these modifiable risk behaviors. However, the underlying social mechanisms are still opaque and poorly understood. Digital health communities provide opportunities to understand social dependencies of behavior change because peer interactions in these platforms are digitized. In this paper, we present a novel approach that integrates theories of behavior change and Exponential Random Graph Models (ERGMs) to understand structural dependencies between users of an online community and the behavior change techniques that are manifested in their communication using an affiliation network. Results indicate population specific traits in terms of individuals' engagement in peer communication embed behavior change techniques in online social settings. Implications for personalized health promotion technologies are discussed.

Keywords:

Health Communities, Social networks, Tobacco cessation

Introduction and Background

According to the World Health Organization (WHO), tobacco use is one of the leading causes of preventable death, causing over seven million deaths worldwide annually. Further exposure to second-hand smoke alone leads to 600,000 deaths. Tobacco use causes various chronic conditions such as cancer, heart disease, stroke, lung diseases, diabetes and Chronic Obstructive Pulmonary Disease (COPD), which collectively account for an estimated $1.7 trillion a year leading to an economic burden on the global healthcare system [1]. Several interventions have been suggested and implemented to address tobacco use around the world – use of behavior counseling and nicotine replacement therapy [2]. Despite ongoing public health campaigns, tobacco use rates are still concerning. It is imperative for us to uncover mechanisms of behavior change to design scalable interventions that support cessation [3].

Studies indicate that social support and networks can have complex relationships with health behaviors and health outcomes [4]. For example, being married to a non-smoker and having a lower number of non-smokers in one's social network are associated with greater rates of smoking cessation. Christakis and Fowler demonstrated that quit attempts occurred within social clusters and that a variety of network variables were predictive of cessation, including the smoking cessation of a spouse, sibling, friend and co-worker [5]. Researchers have investigated these effects through a multitude of methods grounded in socio-behavioral theories and network science methods [3,6]. However, our understanding of the network dynamics of behavior change is still developing.

Advances in information and communication technologies have revolutionized the way in which individuals seek peer support to adopt and sustain positive health behaviors. Social media platforms have increasingly been used as behavior change venus where users reach out to their peers for support and guidance [7]. Analysis of peer interactions in these digital platforms can provide us with a deeper understanding of how social influence impacts users to quit tobacco use and stay abstinent. Consequently, this has raised interest in understanding how users of these platforms communicate with one another about health behavior change in these supportive settings. In this vein, previous work has primarily employed methods of content analysis including qualitative analysis [6], automated text analysis [8] and simulation-based effects [9] which enable identification of communication topics that could be linked to theoretical constructs and behavioral outcomes like abstinence and relapse rates. Although insightful, the problem with these approaches is that they remain agnostic to the structure of relationships that underlies the communication environment and interdependent nature of communicative choices [7]. For example, what one talks about the characteristics of those users, and their ongoing patterns of interaction is rarely integrated [6].

More recently, advances in social media analytics and social network modeling provide opportunities to extract the relational infrastructure of online communication forums like those supporting healthy behavior change, to uncover the social processes that bring structure to that communication, and to account for the interdependencies in those data. In this paper, we draw on these techniques to examine QuitNet, an online community for tobacco cessation. We conceptualize this environment as a bipartite (or two-mode) network, comprised of two sets of nodes–QuitNet users and the behavior change techniques they adopt and talk about in their communication threads –with edges between node sets representing a technique adoption tie. Then, using a class of statistical network models, the exponential random graph models (ERGMs), we investigate the degree to which the attributes of QuitNet users, namely their gender and age, organize the structure of their communication behavior change techniques. Findings from this analysis will reveal patterns of engagement with behavior change techniques that constitute online support structures for community members in online settings.

Methods and Materials

Data Site and Sample

QuitNet is a social media platform designed to help people quit smoking through peer support and pragmatic engagement. It is one of the first online social networks for health behavior change and has been in continuous existence since 1995 [10]. QuitNet offers various forms of social support. Users communicate with their peers in one of the following ways – one to many messages in threaded forums, through synchronous channels such as chat rooms or through asynchronous channels (e.g., private internal e-mail). Previous studies show that participation in QuitNet is associated with abstinence [3]. The primary mode of communication is through forum interactions. For the purpose of this study, we examined communication of 126 highly engaged QuitNet users in 2014 and 2015. In total, these 126 users exchanged 17,451 messages. This research project was reviewed and exempted by the Institutional Review Board at the University of Texas Health Science Center at Houston.

Qualitative Analysis

Two independent researchers coded 900 messages from this dataset, selected at random using the taxonomy of behavior change techniques (BCT taxonomy) [11] to ensure objectivity of the coding process and identify theoretical manifestations in QuitNet interactions. We used Cohen's Kappa to measure inter-rater reliability of the qualitative coding. The BCT taxonomy consists of 93 theory-linked techniques grouped into 16 classes. A single behavior change technique can be related to similar behavior change processes from multiple behavior change theories such as the Health Belief Model, Transtheoretical Model, Social Cognitive Theory and Social Change Theory [12]. Each message was coded appropriately to one or more techniques that were linked to a particular behavior (for example, smoking), focusing on a particular population (for example, QuitNet users). The definition of each technique can be found in [11] with illustrative examples.

Automated Classification System

Using methods from distributional semantics and machine learning, we developed an automated classification system to scale up the BCT annotation to the remaining messages in the dataset. To generate vector representations of messages, we used neural word embeddings, specifically the Skipgram-with-Negative-Sampling (SGNS) algorithm developed by Mikolov and colleagues [13], as implemented in the open source semantic vectors [14] package for distributional semantics. Additionally, Wikipedia was used as a background corpus. Our Wikipedia corpus contains 1.9 billion words in more than 4.4 million articles, and 500-dimensional Wikipedia-derived term vectors were obtained by applying the SGNS algorithm to the Wikipedia background corpus. This decision was motivated in part by the terse nature of the messages exchanged in QuitNet user forums, which often do not provide enough contextual information to train a distributional model [6, 8, 15]. We first superposed (added together) the Wikipedia term vectors for the terms that occur in each QuitNet message to obtain Wikipedia-based QuitNet message vectors. We then composed term vectors for the terms that occur in QuitNet by adding QuitNet message vectors for each message in which a given term in QuitNet occurred. As such, these term vectors encode distributional information from Wikipedia and from QuitNet-specific contextual use of terms. Finally, QuitNet message vectors were generated by superposing these term vectors. The components of the vectors generated in this way were used as feature vectors for supervised machine learning that was conducted using the widely used Waikato Environment for Knowledge Analysis (WEKA) open source package for machine learning [16]. Each of the techniques of behavior change taxonomy was used as a target for classification. We used only eight techniques due to sparse representation in qualitative coding. Ten-fold cross validation was applied using the random forest classifier to evaluate a binary classifier for each of the themes. Each of the trained validated classification models was then used to classify the entire dataset.

Two-Mode Network Analysis

We dichotomized the messages exchanged by a QuitNet user based on whether or not the message is assigned a specific behavior change technique during our text analysis (presence of a technique: 1, else: 0). From this, we were able to create a user to behavior change technique, binary network.

Data were thus analyzed using two-mode network analysis by creating an affiliation network composed of two sets (or nodes) of actors – users of the QuitNet platform (N=126) and the BCT techniques embedded in their QuitNet posts (N=8). Ties are defined only between members of each set but not within sets of nodes [17]. For this reason, the network is comprised only of user-to-technique ties; ties connecting QuitNet users to one another or ties connecting behavior change techniques to one another are not explored.

We then used a class of statistical models for social networks called Exponential Random Graph Models (ERGMs) [18, 19], which allow us to account for the interdependencies in mentions of behavior change techniques among users of the same platform. ERGMs have been applied to a wide array of fields as diverse as epidemiology [20], political science [21],communication studies [22], biological sciences [23] and archeology [24]. As described in previously published work [25], ERGMs permit inferences about how network connections emerge by estimating the likelihood of a tie being present (or absent) in the network as a function of small local tie-based configurations (or patterns). Each configuration represents a distinct social process, such as reciprocity or balance, and corresponds to a specific parameter in the model. From this perspective, ERGMs adopt a logic similar to logistic regression, in which a binary outcome like the adoption of behavior (or the presence of a tie in ERGMs) is modeled as a function of selected, predictor variables (or local configurations in ERGMs) that are thought to explain the observed outcome.The configurations modeled in ERGMs can emerge from endogenous processes – i.e., when actors form ties in response to the other ties being made in their social environment (e.g., when frequently mentioned behavior change techniques continue to be mentioned by more QuitNet users) or in response to exogenous properties that exist outside the network like the attributes of other QuitNet users (e.g., when male QuitNet users tend to adopt the same behavior change techniques).

In statistical terms, the effect of each parameter is estimated by determining the prevalence of the modeled configurations in the observed network and then assessing their statistical likelihood above what would occur by chance alone [19]. A parameter estimate that is positive and significant indicates that ties are more likely to occur within the configuration tested than by chance alone. Conversely, parameter estimates that are negative and significant suggest that ties are less likely to occur within the configuration tested than by chance alone. ERGMs are deemed acceptable when the parameters converge, which occurs when the convergence t-ratios for each parameter

reaches <0.10. Details about the specification, estimation, and simulation of ERGMs can be found here [26, 27].

Selecting which parameters to be included in an ERGM should be grounded in which distinct theories of interaction one wants to test or the research questions one wants to explore. Given that the focus of this analysis centers on the relationship between demographic attributes of QuitNet users and their patterns of communication about behavior change techniques, our estimation focuses on the effects of five attribute-based parameters to determine whether they were more likely to occur than expected by chance alone. Figure 1 illustrates the corresponding configurations of these effects

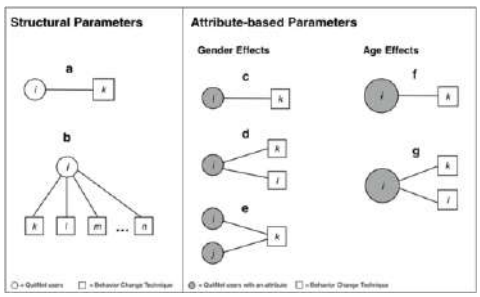


Figure 1. Parameterized Local Configurations among QuitNet Users a and j and Behavior Change Techniques k, l, m, and n. The parameters represent the following processes: (a) the likelihood of QuitNet users mentioning a behavior change technique, (b) the likelihood of QuitNet users mentioning multiple behavior change techniques, (c) the likelihood of male QuitNet users mentioning a behavior change technique, (d) the likelihood of male QuitNet users mentioning multiple behavior change techniques, (e) the likelihood of male QuitNet users mentioning the same behavior change technique, (f) the likelihood of older QuitNet users mentioning a behavior change technique, and (g) the likelihood of older QuitNet users mentioning multiple behavior change techniques.

First, to account for endogenous aspects of network emergence, we include two purely structural (or non-attribute) effects. The endogenous edge parameter (Figure 1a) is a required parameter and represents the overall propensity for QuitNet users to mention behavior change techniques in their posts. We also included the Alternating K-star parameter (Figure 1b), which represents the tendency for QuitNet users to mention multiple behavior change techniques and, therefore, corresponds to the variance in the distribution of technique mentions among QuitNet users.

With respect to our main interest -- attribute-based effects -- we parameterize several configurations that test the effects of gender and age on QuitNet users' patterns of communication about behavior change techniques. Specifically, we include parameters that represent the likelihood of male QuitNet users mentioning a behavior change technique (Figure 1c), the likelihood of male QuitNet users mentioning multiple behavior change techniques (Figure 1d), the likelihood of male QuitNet users mentioning the same behavior change technique (Figure 1e), the likelihood of older QuitNet users mentioning a behavior change technique (Figure 1f), and the likelihood of older QuitNet users mentioning multiple behavior change techniques (Figure 1g). Modeling results were implemented using MPNet, a network estimation program designed for one-mode, two-mode and multilevel network data [17-18].

Results

Qualitative Analysis

Table 1 provides sample QuitNet messages and the BCT techniques (and their definitions) assigned to them during manual coding. Of the manually coded messages, 32% were related to feedback and monitoring, 29% were related to social support, 21% were goals and planning, followed by rewards, natural consequences, associations, self belief and comparison of behavior and outcomes, repetition, and substitution.

Table 1. Illustration of QuitNet Messages and Assigned Behavior Change Technique

Quit message	Assigned technique
To tell you the truth, it's a new experience for me NOT to cough (I smoked for 38 or so years - YUK!). Good luck to you.	Natural consequences
Wow!!! xyxx is correct. You control your attitude. Deep breathing, chew gum, take a walk (or maybe a hike:-) Hang in there. This is not easy, your an addict.	Repetetion and substitution
At this point in your Quit it may be best to look at more immediate gains such as money saved or improved self esteem or better health. My dollar savings are $5,293 and that is real and for now. My life saved is an unrealized 15 weeks and 20 minutes that may never happen.	Comparison of outcomes

Automated Classification System

Due to insufficient positive examples in the training set, we disregarded eight of the 16 techniques of the taxonomy for final classification. For the remaining eight techniques, the precision, recall, and f-measure using Random forest classifier were 0.76, 0.74, and 0.78, respectively. The themes considered for further analysis were goals and planning, feedback and monitoring, social support, natural consequences, comparison of behavior, comparison of outcomes, rewards and threat, and self-belief. Based on the automated classification, we found that each of the users exchanged messages that embedded at least one behavior change technique and up to eight techniques. On average from the full sample, we see that far fewer users (13%) mention goals and planning. 55% of the users mention outcomes, 27% mention comparison of behavior, 53% users mention feedback and monitoring, 32% users mention natural consequence, self belief is mentioned by 33% of the users and social support is mentioned by 37% of users and finally 46% of users mention reward and threat. Table 2 shows the frequencies of each behavior change technique. Among the 126 users, 36 (29%) are men and 90 (71%) are women.

Table 2. Descriptive Statistics of Users and the Behavior Change Techniques within the QuitNet Forum

Behavior Change Techniques	Full user Sample (N=126)	Men (N=36)	Women (N=90)
Outcomes	70(55)	22(61)	48(53)
Feedback and monitoring	67(53)	19(53)	48(53)
Reward and threat	58(46)	17(47)	41(45)
Social support	47(37)	15(42)	32(35)
Self belief	42(33)	10(28)	32(35)
Natural consequence	40(32)	10(28)	30(33)
Comparison of behavior	34(27)	5(14)	29(32)
Goals and planning	17(13)	3(8)	14(15)

The bias towards females in the sample is a result of the inherent skewness in the utility of online social interventions such as QuitNet [6]. In the comparison of means tests not shown here, we find no significant differences between men and women with respect to the types of behavior change techniques they mention in their posts.

Figure 2. illustrates the two-mode affiliation network of 126 QuitNet users and eight behavior change techniques. QuitNet users are depicted as circles and the techniques are depicted as squares. The BCT technique nodes (squares) are sized by the number of mentions they received. Between node sets, there are a total of 375 ties, representing users' mentions of behavior change techniques in their QuitNet posts. Users mentioned on average 2.98 behavior change techniques and each technique had on average 46.88 mentions.

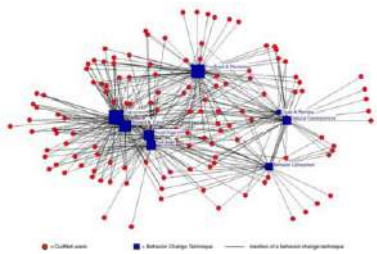


Figure 2. The network structure of mentions of behavior change techniques among 126 members of QuitNet as a 2-mode cessation technique affiliation network: 2014 and 2015.

Exponential Random Graph Model (ERGM)

Results of the two-mode ERGM are shown in Table 3. Reported in this table are parameter estimates, standard errors, and t-ratios, which reveal parameter convergence (t<0.1). The maximum likelihood (ML) estimate for a parameter indicates the direction of its propensity above what would be expected by chance alone. The significance is achieved when the ML estimate is greater than twice the SE.

Table 3. Exponential Random Graph Model of 2-Mode QuitNet User Affiliation Network with Behavior Change Techniques with Gender Modeled as Covariates

Parameter	ML Estimate	Standard Error(SE)	t-ratio
2-mode structure			
Edge density	-0.26	1.01	-0.01
Alternating k-star	-0.3	0.6	-0.02
User attribute-based edges			
Gender edge density	-1.23*	0.6	-0.02
Age edge density	0.004	0.01	-0.02
Gender based expansiveness	-0.09	0.2	-0.04
Gender homophily	0.10*	0.01	0.01
Age based expansiveness	0.0001	0.003	-0.04

Two-Mode Structural Effects

Here, the model estimates the effects of two structural or non-attribute parameters, which represent local structural patterns in the QuitNet user affiliation network. The required Edge parameter is negative, indicating that QuitNet users are less likely than expected to affiliate with behavior change techniques (ML Estimate=-0.26, SE=1.01). This reinforces the low density of the user affiliation network and suggests that adopting behavior change techniques is not random or haphazard. The negative estimate of alternating K-star parameter suggests that the likelihood of a user affiliating with multiple techniques is less likely than expected by chance (ML=-0.30, SE=-0.59).

Attribute - Based Edge Effects

The first attribute-based parameter, Gender density, is negative and significant (ML Estimate=-1.23, SE=0.60) indicating that men are less likely to adopt a behavior change technique than expected by chance. The attribute-based edge parameter for users' age (ML Estimate=0.004, SE=0.01) is positive indicating that older people are likely to adopt behavior change techniques. However, the effect is not significant. The gender-based expansiveness parameter estimate is not significant (ML Estimate=-0.09, SE=0.20), suggesting that men are no more or less likely than expected by chance to adopt multiple behavior change techniques. The gender homophily parameter estimate is positive and significant (ML Estimate=0.10, SE=0.02), indicating that men are more likely than expected by chance to adopt the same techniques. Finally, age-based expansiveness, which represents the likelihood that older QuitNet users are more likely to adopt multiple techniques is not significant (ML Estimate=0.0001, SE=0.003).

Discussion

Given that people increasingly turn to online social environments to find community and support, an opportunity arises to employ these platforms in health behavior interventions. Adopting a two-mode network analytic approach, we used a class of stochastic models called exponential random graph models (ERGMs) to determine the ways in which attributes of community users structure their communicative interactions about behavior change. Descriptively, we found that users more often discussed techniques that fell into the *Comparison of Outcomes*, *Feedback & Monitoring*, and *Reward & Threat* categories of behavior change. This finding can inform future interventions the need for a more adaptive and personalized, technology driven interventions such as a mobile app that includes effective quit plans, use of novel machine learning algorithms based on feature selection to obtain high classification rates to predict relapse (for e.g., smoking urges) and population-level interventions focusing on health effects of tobacco use.

A strength of this study lies in its analytic approach. Standard regression models are designed for data where independence among observations can be assumed. However, in a networked and open online forum like QuitNet, where users observe and respond to what others are posting about, we cannot assume that observation of what one user communicates about is independent of what others are talking about in those spaces. As such, an approach was needed that could account for those dependencies. To this end, we drew on the exponential random graph model (ERGM), which allows us to see how characteristics of QuitNet users condition their interactions and communication about behavior change techniques. The utility of automated text analysis approaches has allowed us to bridge two threads of socio-behavioral science driven by theoretical constructs and network models. Findings from our analysis provide insights about the underlying social dynamics of behavioral choices (e.g., adopting a behavior change technique) that would otherwise go undetected using standard regression models. These insights provide health researchers and interventionists with new directions for designing network based behavioral interventions [28] that target the social dynamics of behaviors like abstinence from tobacco use.

Our study has several limitations. Firstly, because these data are cross-sectional, inferences about cause and effect cannot be made. Secondly, our sample is limited to only those users who are highly engaged during 2014-2015. Moving forward, we should adopt the proposed techniques to identify patterns of relapse and quit sustenance. In addition, the accuracy of the automated system can further be improved through the use of advanced word representations [29, 30].

Conclusion

Tobacco use causes seven million deaths around the world each year. New modes of nicotine intake (JUULS, e-cigs) are now considered safer than traditional cigarettes and misinformation

is quite prevalent regarding these topics. In this new media environment, adaptive social interventions to help people abstain from tobacco use is a high-priority task. The major contribution of this study is its employment of integrative analytical framework to elucidate the structural dependencies between community users and their communication attributes. Insights from this work allow implementation of targeted, recommendation engines to promote meaningful affiliations with content and social ties based on user characteristics. Such translation can set the stage for scalable network-based interventions that can enable individuals and communities to engage in positive health behaviors.

Acknowledgements

Research reported in this publication was supported by the National Library of Medicine and National Cancer Institute of the National Institutes of Health under award numbers R21LM012271 and R21CA220670. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] World Health Organization. The World Health Report. Depression and Other common mental disorders. 2017.

[2] S. Zhu, M. Lee, Y. Zhuang, A. Gamst, and T. Wolfson, Interventions to increase smoking cessation at the population level: how much progress has been made in the last two decades? *Tobacco Control* **1(2)** (2012), 110–118.

[3] A. Graham, N. Cobb, L. Raymond, S. Sill and J. Young, Effectiveness of an internet-based worksite smoking cessation intervention at 12 months, *Journal of Occupational and Environmental Medicine* **49(8)** (2007), 821–828.

[4] D. Umberson, and J. K. Montez, Social Relationships and Health. A flashpoint for Health Policy, *SAGE Journals* (2011), S54-S66.

[5] N. A. Christakis, and J. H. Fowler, Quitting in Droves: Collective Dynamics of Smoking Behavior in a Large Social Network. *New England Journal of Medicine* **358(21)** (2008), 2249–2258.

[6] S. Myneni, K. Fujimoto, N. Cobb, and T. Cohen, Content-Driven Analysis of an Online Community for Smoking Cessation: Integration of Qualitative Techniques, Automated Text Analysis, and Affiliation Networks, *American Journal of Public Health* **105(6)** (2015), 1206–1212.

[7] C. A. Latkin, and A. R. Knowlton, Social Network Assessments and Interventions for Health Behavior Change: A Critical Review, *Journal of Behavioral Medicine* **41(3)** (2015), 90-97.

[8] V. Sridharan, T. Cohen, N. Cobb, and S. Myneni, Characterization of Temporal Semantic Shifts of Peer-to-Peer Communication in a Health-Related Online Community: Implications for Data-driven Health Promotion, *AMIA Annual Symposium Proceedings Archive* (2016), 1977–1986.

[9] D. R. Schaefer, J. Adams, and S. A. Haas, Social Networks and Smoking: Exploring the Effects of Influence and Smoker Popularity through Simulations. *Health Education and Behavior* **40(10)** (2015), 24S – 32S.

[10] N. K. Cobb, A. L. Graham, B. C. Bock, G. Papandonatos, Initial evaluation of real-world internet smoking cessation system, *American Journal of Public Health* **100(7)** (2010), 1282–1289.

[11] S. Michie, M. Richardson, M. Johnston, C. Abraham, J. Francis, W. Hardeman, and C.E. Wood, The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting change interventions, *Annals of Behavioral Medicine* **46(1)** (2013), 81–95.

[12] K. Glanz, B. K. Rimer, and K. Vishwanath, In: Health behavior and health education: theory, research, and practice, *5th Edition.* 2008.

[13] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, Distributed representations of words and phrases and their compositionality, *Advances in neural information processing systems*. 2013.

[14] D. Widdows, K. Ferraro, Semantic Vectors: A scalable open source package and online technology Management application. *Proceedings of LREC*'08 (2008), 1183-90.

[15] S. Myneni, N. K. Cobb, and T. Cohen. Content-specific network analysis of peer-to-peer communication in an online community for smoking cessation, *AMIA Annual Symposium Proceedings Archive* (2016), 934-943.

[16] G. Holmes, A. Donkin, and I. H. Witten, Weka: A machine learning workbench, 1994.

[17] P. Wang. ERGM extensions. Model for multiple networks and bipartite networks Exponential random graph models for social networks: theory, methods and applications. New York: Cambridge University Press. 2013.

[18] G. Robins, P. Pattison, Y. Kalish, D. Lusher. An introduction to exponential random graph (p*) models for social networks. 2007; 29(2):173 -91.

[19] S. Wasserman and K. Faust, Social network analysis: Methods and applications. Cambridge, England: Cambridge University Press. 1994.

[20] S. M. Goodreau, N. B. Carnegie, E. Vittinghoff, J. R. Lama, J. Sanchez, B. Grinsztejn, S. P. Buchbinder, What drives the US and Peruvian HIV epidemics in men who have sex with men (MSM)? *PloS one* **7(11)** (2012), e50522.

[21] M. T. Heaney, Multiplex networks and interest group influence reputation: An exponential random graph model. *Social Networks* (2014), 36, 66-81.

[22] M. Shumate, and E.T. Palazzolo, Exponential random graph (p*) models as a method for social network analysis in communication research. *Communication Methods and Measures* **4(4)** (2010), 341-371.

[23] Z. M. Saul and V. Filkov, Exploring biological network structure using exponential random graph models, *Bioinformatics* **23(19)** (2007), 2604-2611.

[24] V. Amati, T. Shafie, and U. Brandes, Reconstructing Archeological Networks the Structural Holes, *Journal of Archeological Method and Theory* **25(1)** (2017), 226-253.

[25] E. L. Young, K. Fujimoto, A. J. Schneider, HIV Prevention and Sex Behaviors as Organizing Mechanisms in a Face-book Group Affiliation Network Among Young Black Men Who Have Sex with Men, *AIDS and Behavior* **22** (2018), 3324-3334

[26] D. Lusher, J. Koskinen, and G. Robins, Exponential Random Graph Models for Social Networks. Cambridge University Press, 2013.

[27] P. Wang, G. L. Robins, P. E. Pattison, and J. H. Koskinen, MPNET: Program for the simulation and estimation of p*exponential random graph models. Melbourne, Australia: University of Melbourne, 2016.

[28] T. W. Valente, Network interventions, *Science* **337(6090)** (2012), 49-53.

[29] Y. LeCun, Y. Bengio, G. Hinton. Deep learning. *Nature*, **521(7553)** (2015), 436.

[30] S. Zhang, E. Grave, E. Sklar, N. Elhadad, Longitudinal analysis of discussion topics in an online breast cancer community using convolutional neural networks. *Journal of biomedical informatics* **69** (2017), 1-9.

Address for correspondence

Shruthi Manas, shruthi.manas@uth.tmc.edu

The University of Texas School of Biomedical Informatics at Houston, 7000 Fannin, UCT165, Houston, TX 77030

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190431

Global Workforce Trends in Health Informatics & Information Management

David Marc[a], Kerryn Butler-Henderson[b], Prerna Dua[c], Karima Lalani[d] , Susan H. Fenton[e]

[a] *Department of Health Informatics and Information Management, The College of St. Scholastica, Duluth, Minnesota, USA*
[b] *College of Business & Economics, University of Tasmania, Tasmania, Australia*
[c] *Department of Health Informatics and Information Management, Louisiana Tech University, Ruston, Louisiana, USA*
[d] *Department of Health Information Management, Texas State University, San Marcos, Texas, USA*
[e] *The University of Texas Health Science Center School of Biomedical Informatics, Houston, Texas, USA*

Abstract

In a data driven environment, healthcare has seen ongoing digital transformation to meet both clinical and business needs. But, have the educational and functional requirements of the health informatics and information management (HIIM) workforce also adapted? This study examined the current employment opportunities in HIIM globally. Using 11 keywords generated from a literature review, postings on the job advertisement website Indeed™ for all available countries were analyzed. The results show that job postings tend to fall within 4 discrete categories: 1) health information technology; 2) health research; 3) health leadership and project management; and 4) health compliance. Data indicated a higher prevalence for certain areas by country. The findings from this study can inform HIIM educational providers about future skill requirements.

Keywords:

Career Choice, Medical Informatics, Cluster Analysis

Introduction

One of the primary reasons to seek higher education is to prepare for a career with interesting work that pays a living wage. In an attempt to determine which skills and competencies are currently most desired by employers, a group of health informatics and information management (HIIM) educators undertook a qualitative analysis of health information management and health informatics-related job postings from Indeed.com.

In health informatics, many of the educational competencies required for training programs have been determined primarily by educators or professionals who have been in the field for years. While these stakeholders certainly have experience and a role to play, they may not be the main employers. Worse, these people are often not aware of emerging trends or knowledgeable about all of the new skills needed by the workforce. This is more of a concern for fields such as health informatics and information management, where significant technical change is occurring, as well as for any current job where machine learning has the potential to be employed at less expense than the human workers.

The purpose of this study is to describe the characteristics of careers related to health informatics and information management, as well as explain the geographical commonalities and differences in the skills and knowledge required for employment using international job posting data.

Background

With the growth of Information Technology (IT) and the advent of big data, the healthcare industry is dynamically adapting to the current need to harness big data for improving healthcare performance and decision-making [1]. The adoption of electronic health records (EHRs), has been a central focus in United States in the past few years with providers all around the world continuing to adapt to technology at a rapid pace. Several studies from Canada, Iran, Nigeria and Saudi Arabia have identified obstacles in IT implementation relating to a health information management workforce shortage [2-4]. A specific study by the Canadian Health Information Management Association in 2014 identified workers needed in the areas of standards, data quality and management, information governance, change and project management [5].

Globally, EHR adoption rates have been influenced by the availability of funding, governance, standardization, interoperability, and communication [6]. Recently, Black Book Research conducted a poll surveying 7,459 physicians, health administrators, technology managers and clinical leaders in both inpatient and ambulatory settings across 23 countries to identify gaps, challenges and successes in healthcare IT adoption and connectivity. Included in those surveyed were five countries with evidence of nearly 100% EHR adoption rates: Norway, Netherlands, United Kingdom, New Zealand, and Australia, as well as five countries with nearly 75% provider adoption rates: Germany, France, Canada, Switzerland, and Singapore. Regions from Europe, Middle East, and South Asia showed a drift from siloed EHRs to more integrated healthcare IT systems. With the increased adoption and use of the digital technology, new roles have emerged in e-health to support the implementation and the administration of the operations and maintenance of the technological infrastructure [7]. To augment these roles there is a need for enhancing the skillset to strengthen the health informatics and information management workforce. The International Federation of Health Information Management (IFHIMA) documented the Global Health Information Curricula Competencies in their comprehensive 2015 report for Health Informatics, Health Information Management and Health Information and Communication Technologies. The premise was that even though universities strive to provide relevant curricula to students, it may not represent the actual needs of a career-ready graduate.

Furthermore, with the adaptation of new technology, growing consumer engagement and changes in legislation and regulatory requirements there are changes to the healthcare workforce [8]. The workforce has seen a trend of new career opportunities emerging in the areas of data analytics, data governance, privacy and security and interoperability. Our objective through this study is to highlight trends in health informatics and information management careers that are most

prevalent globally and examine if there are differences across countries using data acquired from the job site Indeed.com.

Methods

Keyword Derivation

The research team generated a list of 11 categories to search for online job postings based on recent literature, suggesting trends of increased growth in these areas: Information Technology, Classification and Clinical Documentation Improvement, Consumer Engagement, Leadership, Research, Information Governance, Project Management, Health Informatics, Information Technology, Compliance, and Health Data Analysis. These categories were used as the keywords when searching for job postings.

Query of International Job Postings

In September 2018, job postings from 64 countries were queried from Indeed.com. Queries were generated for the 11 job categories described earlier. In addition to the job titles, the word "health" was used as a qualifier to exclude job postings not related to healthcare. The job title, date of posting, location, company, and a URL to the full job description was obtained using the Indeed.com API. The job description and requirements were derived from the text located on the URL from each job posting. Using the *urllib.request* python library only valid URLs were included in the analysis, as determined by an HTTP status code of 200 (i.e., indicating a successfully working URL). The *BeautifulSoup* Python library was used to query the contents of the HTML files. All of the content located on the specified URL landing page was used as the original resource text. The HTML files were then prepared to remove duplicate records and aberrant text.

Analysis of Text

The job posting URL landing page text was prepared using the R statistical software package. The *tm* R package was used to tokenize the text which was transformed to lowercase, punctuation was removed, the resulting white space was stripped, common English words (i.e., stopwords) were removed from the text, and word stems were discarded. The text corpus was normalized using Term Frequency-Inverse Document Frequency (TF-IDF) to emphasize terms that appear multiple times in a single job posting while decreasing the importance of terms that appear many times across all job postings. Sparse terms amongst the jobs posted were also removed.

The corpus of text was evaluated to cluster similar job postings with K-Means clustering using Euclidean distances. Cluster tags were generated based on the cluster centers to determine which postings belonged to particular clusters. The optimal number of clusters was determined using the average silhouette approach to obtain the number of clusters that yields the highest quality of separation. The clusters were characterized by the terms at cluster centers that had the highest and lowest means to better understand the differentiating features of the clusters.

To compare the frequency of job posting categories by country, a heatmap was used to compare the frequency of postings across all countries and the 11 job categories. Additionally, the relative differences of needed skills and knowledge emphasized across job categories in each country was depicted in a stacked bar plot based on the proportion of job postings in each country for a particular cluster across each of the 11 job categories.

Results

The results show that by using an automated method of clustering, job postings tend to fall within 4 discrete categories. Additionally, there are differences in the frequency of job postings related to the 11 key categories when comparing countries.

Table 1 - Characteristics of Clusters

Cluster	WSS*	Size	Key Terms
1	83.1	7350	data, engine, software, test, analyst, design, security, system, technology, solution
2	63.3	4894	clinic, patient, medic, care, hospitals, research, studies, healthcare, staff, site
3	194.7	18695	sale, market, custom, research, client, business, project, account, service, product
4	87.2	7039	safety, engine, site, maintenance, equip, project, regulatory, environment, manufacturer, compliance

*WSS= Within-Sum Squares

In Table 1, the clusters are distinguished by key characteristics. The total within-cluster sum of squares (WSS) was 428.3 and the between-cluster sum of squares was 20.1 resulting in an internal cohesion of 0.045. This demonstrates that there was a large variance of job postings within clusters and moderate to small separation of job postings between clusters (Figure 1). Cluster 1 is the 2nd largest and is distinguished by containing jobs postings related to more technical skills and knowledge related to data, software, design, and technology. Custer 2 was the smallest in size and the job postings within the cluster were the most highly related. Job postings in Cluster 2 were more clinically focused, as shown by the prevalence of terms related to clinic, patient, medicine, and care. Cluster 3 had the largest number of job postings and the highest degree of variation. The most common terms to occur in job postings in Cluster 3 included sale, market, research, client, business, and terms more related to healthcare sales, marketing, and project management. Lastly, job postings in Cluster 4 included terms related to compliance such as safety, regulatory, and maintenance.

The number of job postings in each cluster across the 11 categories shows that the majority of the jobs in Cluster 1 are related to health information technology; in Cluster 2 to health research; in Cluster 3 to health leadership and health project management, and in Cluster 4 to health compliance (Figure 2).

When comparing the job postings by country, the proportion of postings are predominantly assigned to Cluster 3 (Figure 3). When comparing the frequency of job postings by category for each country (Figure 4), the most common category is Health Project Management followed by Health Leadership, Health Information Technology, and Health Research. The least common categories include Health Informatics, Health Consumer Engagement, and Clinical Documentation Improvement. Interestingly, there are categories in the United States which occur frequently that do not occur often in other countries. These include Health Consumer Engagement, and Clinical Documentation Improvement. Additionally, Health Data Analysis is frequent in the United Kingdom, Canada, Australia, and India, but is less frequent in the United States.

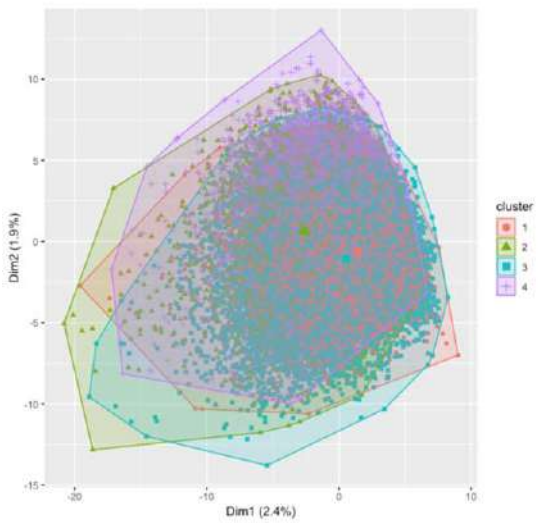


Figure 1 - K-Means Clustering of Job Postings

Discussion

This study demonstrated that jobs in health informatics and information management can be distinguished by key characteristics and there are differences across countries in the prevalence of jobs postings by these characteristics. In general, this study showed that job listings can be categorized into 4 distinct groups based on required skills and knowledge: technology focused; clinically focused; compliance focused; and sales/marketing/management focused. This approach is different than previously published initiatives to define needed competencies for the health informatics and information management workforce by the International Medical Informatics Association (IMIA), the Health Information Management and Systems Society (HIMSS) in cooperation with the European Union, the U.S., Australia, and Canada [5, 9-14]. The mapping of these can be found in Table 2.

	1	2	3	4
Clinical Documentation Improvement	100	627	236	138
Health Compliance	442	541	2249	2220
Health Consumer Engagement	110	64	763	30
Health Data Analysis	1119	184	1291	344
Health Informatics	249	268	155	11
Health Information Governance	325	232	674	135
Health Information Technology	2064	606	2651	460
Health Leadership	737	786	3941	1432
Health Project Management	1281	357	3538	1869
Health Research	923	1229	3197	400

Figure 2 - Heatmap of Job Postings Within Each Cluster Across Job Categories

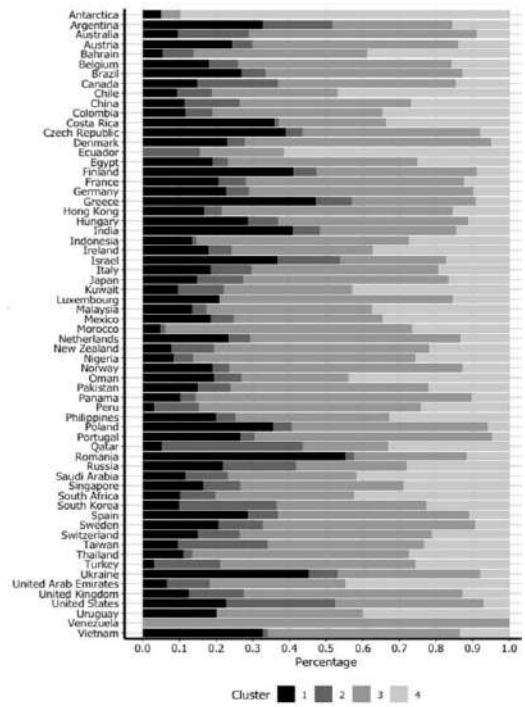


Figure 3 - Proportion of Job Postings in Each Cluster for Each Country

Global Comparison

In 2010, IMIA made recommendations for biomedical and health informatics education [9]. Interestingly, the IMIA domain area of informatics/computer science, mathematics, and biometry is closely related to technology cluster of this study, with the domain of medicine, health and biosciences, and health system organization related to clinical cluster. The third domain of the IMIA recommendations relates to foundational informatics knowledge and skills.

There has also been a cooperative effort between the EU and the U.S. to develop a comprehensive list of Health Information Technology Competencies (HITComp) [10]. There are 1,025 detailed competencies in the HITComp set, mapped to 15 domains. These 15 domains have 4 general categories of Administration, Direct Patient Care, Informatics, and Research/Biomedicine. This is consistent with the study's technology, clinical, and compliance clusters.

Table 2 – Mapping of Clusters to Competencies

Cluster & Focus	IMIA	HIT-Comp	CHIA	COACH	AMIA
1 – technology	XX	XX	XX	XX	XX
2 – clinical	XX	XX	XX	XX	XX
3 – business			x	x	
4 – compliance		XX	XX	XX	

Australasia

The Health Information Management Association of Australia (HIMAA) HIM Competency Standards (3rd edition) has three competency levels, from graduate entry through advanced practitioner [11]. The competencies are set within 9 domains, including clinical, management, terminologies and classifications, compliance, and technology. The Certified Health Informatician Australasia (CHIA) released the first edition of their Certified Health Informatics Competencies Framework in 2013 [12]. The CHIA framework is consistent with the technology, clinical, and compliance clusters, though CHIA includes project management in their management science domain, which is found in the business cluster of this analysis.

Canada

Canada's Health Informatics Association, COACH, has 7 categories of competencies [13]. As with the previous organizations, it includes content relevant to the technology, clinical, and compliance clusters. Consistent with the HIMAA and CHIA frameworks, it includes project management from the business cluster.

United States

The American Medical Informatics Association (AMIA) released a new set of core competencies in 2017 [14]. Very broadly drawn to include a variety of focus areas, the knowledge, skills, and abilities (KSAs) map most clearly to the technology and clinical clusters.
The American Health Information Management Association (AHIMA) also revised their core competencies as part of their HIM Reimagined strategy. The 2018 draft competencies expand capabilities across the areas of data analytics, auditing, and information governance [15].

Conclusions

Overall, the competencies currently articulated for educational programs by IMIA, the HITComp project, and associations in several countries, covers a majority of the terms found in job descriptions; however, sales and marketing are noticeably missing. Health informatics jobs are generally low, possibly due to a lack of standardized job titles in the informatics field.

When looking at the job categories, the majority of jobs are focused on project management, health leadership, and health information technology with fewest on health informatics, health consumer engagement, and clinical documentation improvement. The UK, Canada, Australia, and India have a greater prevalence of data analysis jobs compared with the U.S., which has a high number of consumer engagement, clinical documentation improvement, and information governance jobs. The findings from this study can inform HIIM educational providers about future skill requirements.

References

[1] Sandefer, R., et al. (2014). "Keeping Current in the Electronic Era: Data Age Transforming HIM's Mandatory Workforce Competencies." *J AHIMA* **85**(11): 38-44.

[2] Abdekhoda, M., et al. (2014). "Information technology acceptance in health information management." *Methods Inf Med* **53**(1): 14-20.

[3] Hasanain, R. A., et al. (2015). "Electronic Medical Record Systems in Saudi Arabia: Knowledge and Preferences of Healthcare Professionals." *Journal of Health Informatics in Developing Countries* **9**(1).

[4] Zayyad, Musa Ahmed and Mehmet Toycan. "Factors affecting sustainable adoption of e-health technology in developing countries: an exploratory survey of Nigerian hospitals from the perspective of healthcare professionals" *Peer J* vol. 6 e4436. 1 Mar. 2018, doi:10.7717/peerj.4436

[5] Gibson, C. J., et al. (2015). *"Health Information Management Workforce Transformation: New Roles, New Skills and Experiences in Canada*.

[6] A HIMSS Study: Electronic Health Records: A Global Perspective, Second Edition (August, 2010) Part I

[7] Gibson, C. J., et al. "Convergent evolution of health information management and health informatics: a perspective on the future of information professionals in health care" *Appl. Clin. Inform.* **6**(1) 163-84. 18 Mar. 2015, doi:10.4338/ACI-2014-09-RA-0077

[8] D. Fridsma, Developing the health informatics workforce of the future: academic and industry partners*, J Am Med Inform Assoc.* **24**(3), 2017, 677-678.

[9] Mantas, J., et al. "Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics." *Methods Inf Med 2010;* 49: 105-120; doi: 10.3414/ME5119.

[10] Health Information Technology Competencies (HITComp), EU*US eHealth Work, http://hitcomp.org/, accessed 11/17/2018.

[11] Health Information Management Association of Australia (HIMAA), *Health Information Management (HIM) competency standards*, 3rd edition, himaa2.org.au/sites/default/files/HIMAA_HIM_Competency_Standards_Version_3_FNL_June2017.pdf, accessed 11/19/2018.

[12] Certified Health Informatician Australasia, *Certified Health Informatics Competencies Framework*, 1st edition, December 2013.

[13] Canada's Health Informatics Association (COACH). *Health Informatics Professional Core Competencies, v. 3*, https://www.coachorg.com/en/resourcecentre/resources/Health-Informatics-Core-Competencies.pdf, accessed 11/17/2018.

[14] Valenta, A.L., et al. "AMIA Board White Paper: AMIA core competencies for applied health informatics education at the master's degree level." *J Am Med Inform Assoc.* **0**(0), 2018, 1–12 doi: 10.1093/jamia/ocy132.

[15] American Health Information Management Association (AHIMA) *Academic Curricula*, http://www.ahima.org/education/academic-affairs/academic-curricula, accessed 11/21/2018.

Address for correspondence

Corresponding author: David T. Marc, PhD
Assistant Professor and Chair of Department of Health Informatics and Information Management
The College of St. Scholastica
1200 Kenwood Ave | Duluth, Minnesota 55811
Tel: 1 1 218 625 4892; Email: dmarc@css.edu

Figure 4 - Heatmap of the prevalence of jobs by category across each country

	Health Informatics	Health Consumer Engagement	Clinical Documentation Improvement	Health Information Governance	Health Data Analysis	Health Compliance	Health Research	Health Information Technology	Health Leadership	Health Project Management	Total Postings
United States	369	410	413	399	378	402	401	420	411	408	4011
United Kingdom	65	139	167	375	311	359	379	413	399	358	2965
Canada	58	89	166	141	303	344	355	418	389	355	2618
Australia	14	56	80	148	263	328	351	410	396	325	2371
India	15	39	47	39	256	277	347	411	361	282	2074
China	12	19	30	17	143	299	342	230	395	320	1807
Ireland	1	19	28	19	85	286	176	251	344	310	1519
Netherlands	24	20	18	17	81	175	280	270	200	308	1393
Germany	20	27	10	11	75	177	350	210	260	245	1385
Singapore	5	18	9	6	77	230	215	175	276	302	1313
South Africa	1	1	1	40	41	265	108	100	310	257	1124
Poland	8	20	10	11	94	112	109	170	174	199	907
United Arab Emirates	0	1	15	7	35	198	100	70	180	202	808
New Zealand	1	1	1	12	23	105	78	90	288	194	793
Philippines	0	1	10	8	48	160	77	129	153	175	761
Hong Kong	6	10	1	8	62	72	132	100	127	232	750
Belgium	0	9	20	11	35	86	109	90	94	231	685
Switzerland	10	1	20	5	33	81	147	70	141	138	646
France	6	10	10	13	24	67	99	120	112	133	594
Romania	38	10	1	0	40	55	65	120	63	182	574
Malaysia	0	9	1	5	31	112	58	70	120	124	530
Nigeria	0	1	1	18	20	74	85	50	124	157	530
Vietnam	0	1	1	0	32	73	82	90	108	101	488
Spain	0	9	9	6	37	48	90	80	82	126	487
Japan	5	9	10	4	25	67	90	60	91	92	453
Thailand	0	1	1	8	23	86	48	50	134	69	420
Italy	1	9	1	6	25	56	56	90	79	68	391
Saudi Arabia	1	1	1	0	18	88	46	60	75	67	357
Mexico	0	1	1	1	22	70	35	60	85	76	351
Qatar	1	0	1	1	12	65	73	60	60	60	333
Denmark	1	9	1	8	17	22	102	50	56	65	331
Indonesia	1	1	1	1	17	49	37	30	97	74	308
Sweden	1	1	1	1	19	31	54	60	54	56	278
Czech Republic	1	1	1	5	16	59	32	70	43	49	277
Portugal	6	0	0	0	16	28	54	70	33	67	274
Hungary	6	0	1	0	16	42	63	40	37	52	257
Brazil	1	1	0	1	15	26	28	50	33	49	204
Egypt	0	1	0	0	10	45	24	30	47	38	195
South Korea	0	1	1	1	5	28	48	30	36	31	181
Ukraine	0	1	1	1	20	12	37	20	28	57	177
Greece	1	0	0	1	21	13	25	50	27	36	174
Austria	1	1	1	1	11	15	37	30	30	30	157
Norway	0	1	1	0	18	16	36	30	21	25	148
Israel	0	1	1	0	8	28	47	20	17	17	139
Argentina	1	1	0	0	6	20	20	30	23	34	135
Russia	0	0	1	1	7	12	25	20	29	34	129
Turkey	0	0	1	1	12	26	24	20	27	18	129
Pakistan	0	1	1	1	12	18	26	17	27	23	126
Kuwait	0	0	1	0	2	26	5	20	38	22	114
Finland	0	1	1	1	10	5	26	40	16	12	112
Taiwan	0	1	0	0	11	12	21	20	25	13	103
Luxembourg	0	1	0	1	5	13	13	20	16	22	91
Costa Rica	1	0	1	0	5	10	18	20	14	17	86
Colombia	1	0	0	1	2	17	17	10	9	12	69
Chile	0	0	0	1	0	8	9	10	17	19	64
Morocco	0	1	0	0	1	14	1	20	8	19	64
Panama	0	0	0	1	2	2	8	10	16	10	49
Oman	0	0	1	0	1	14	1	1	10	13	41
Antarctica	0	0	0	0	0	2	9	1	10	18	40
Bahrain	0	0	0	1	0	9	9	1	9	7	36
Peru	0	0	0	1	0	3	10	1	10	8	33
Ecuador	0	0	0	0	0	9	1	1	1	1	13
Uruguay	0	1	0	0	1	1	1	1	0	0	5
Venezuela	0	0	0	0	0	0	0	1	1	1	3
Total Postings	683	967	1101	1366	2938	5452	5751	5781	6896	7045	**37980**

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190432

User -Centered Design of a Patient Medication Reconciliation Module in an Integrated Personal Health Record

Santiago Márquez Fosser[a], Janine Sommer[a], Mariana Simón[a], Liliana Giraldo[a], Fernando Plazzotta[a], Daniel Luna[a]

[a] *Department of Health Informatics, Hospital Italiano de Buenos Aires, Ciudad de Buenos Aires, Argentina.*

Abstract

With the adoption of Personal Health Records (PHRs) integrated with Electronic Health Records (EHRs) and the increase of accessibility to data, institutions have the possibility of exchanging medical information with their patients. Involving the patient reported data has the potential to improve the quality of care and safety and create a feedback loop between patients and health professionals. The objective of this study is to describe a user-centered design of a module for medication list with reconciliation functionalities managed by the patients themselves, and connected to their EHR for supervision and medical validation. We conducted 42 interviews (31 patients and 11 general practitioners). From the interviews, we performed qualitative analysis and extracted the main findings from comments in both groups. Correctitude rate was 57 to 100%, and satisfaction of use (SUS) maximum was 96% and 92%. These findings may be relevant to patients, health care providers, and policymakers.

Keywords:

Electronic Health Record, Medication Reconciliation, Personal Health Record

Introduction

The medication reconciliation is the process of obtaining the best, complete, and accurate list of each patient's current home medications including name, dosage, frequency, and route of administration, and comparing admission, transfer, and/or discharge medication orders to that list. The reconciliation is done to avoid medication errors.

The patient's medication list represents one of the most important components in the Electronic Health Record (EHR). It describes the pharmacological and non-pharmacological indications that the doctor prescribes medications to the patient, and provides the best list of medical indications that the patient has at that time. In the EHR, this list serves as a record, for the same prescribing physician, and his colleagues. The data are used to conduct research studies, indicate electronic prescriptions, and evaluate the quality of patient care and support decision-making.

The literature reports the existence of more than 50% of a discrepancy between the medication that patients take and their medical record [1–3]. Patients can discontinue their medications, change regimens when their symptoms improve or when they suffer an adverse effect, and do not report these changes to their doctors [4–6]. Moreover, medical specialists prescribe new medications or change previous regimens without notifying the attending physician [5,6]. The existence of discrepancies between the list of ambulatory medications recorded in the EHR and what the patient takes is a source of potential problems in health care, which may induce inadequate decisions by the treating team, generate side effects reactions, and exacerbations of comorbidities.

The international organizations such as the Joint Commission make accreditation and certification and are recognized as a symbol of quality and commitment to meeting performance standards. The organizations recommend that: "*Find out what medicines the patient is taking. Compare those medicines to new medicines given to the patient. Make sure the patient knows which medicines to take when they are at home. Tell the patient, it is important to bring their up-to-date list of medicines every time they visit a doctor*" [7].

New strategies should be considered to mitigate this situation. The reconciliation of medication is the process by which the best list of a patient's usual medication is elaborated to use that information in different areas of care, and provide the appropriate treatment. If a patient changes the scope of care, for example at home without requiring assistance or in the outpatient setting, and consults the ward for an acute event, it is critical to establish the reconciliation of the medication. Reconciling the last record taken from the list of habitual medications with what the patient or the relatives confirm through a series of questions, generate the best possible list. In this model, the patient or family members play a passive role in which the healthcare team asks them and compiles the list.

There are experiences with different strategies that reduce discrepancies, especially during the admission interview at the hospital [2,8,9]. One of them is involving pharmacists at the admission of patients. Pharmacists have a lower error rate in the registration of the list of drugs in the EHR, compared with doctors and nurses. Another strategy is incorporating a permanent education to those who make the prescriptions [2]. Researchers investigated the usage of patient-reported data in which patients are asked to review their medications, corrected them [5]. Afterward, a general practitioner validates this information and updates the data in the EHR [5].

The objective of this study is to describe a user-centered design (UCD) of a medication reconciliation module managed by patients, connected to their EHR for supervision and medical validation managed by physicians.

Methods

We carried out this descriptive study, which is a UCD of a medication reconciliation framework involving patients and doctors, in Hospital Italiano de Buenos Aires. For the UCD, we conducted iterative cycles of designing and testing with patients

on the reconciliation functionalities in the Personal Health Portal and to doctors for the EHR functionalities.

Setting

Hospital Italiano de Buenos Aires (HIBA) is a non-profit organization with 165 years of history in Argentina. Its healthcare network includes a university hospital of high complexity that covers health care for outpatients, inpatients, emergencies, critical care, home care, chronic care, and medical and surgical specialties. It has its medical insurance service (health maintenance organization), with more than 160,000 affiliates, and provides health services to 1,500,000 people with other health insurances. Annually, more than 45,000 patients are admitted to their hospitals, and 45,000 surgical procedures and 3,000,000 outpatient visits take place.

Since 1998, the HIBA has its own health information system (in-house) that includes the management of clinical and administrative information. Its EHR is an integrated, modular, problem-oriented, and patient-centered system, and is used in the different clinical scenarios (ambulatory, hospitalization, emergency center, and home care) [10,11]. As part of the information system, an integrated Personal Health Records (PHR) called POPES is available to all outpatients since 2007. The PHR allows patients to receive medical care in the hospital network, to access and verify clinical and administrative information, and to interact with the health system. Among its main functionalities, the PHR allows users to update their personal information, share information, manage scheduled appointments, view test results, check, order and buy prescribed medication, and the possibility to consult with the healthcare team through Telemedicine tools. Moreover, it has a messaging service for communication with the general practitioner" At present, POPES has approximately 400,000 registered users [12].

Medication reconciliation by patient

PHR currently has a module for patient medications that display prescribed drugs, generic and brand name, dosage, and prescribing physician. It also allows the patient to ask to deliver medication to their homes (purchasing). The number of patients who used this module from January 1, 2018 to November 9, was 90,977, and the number of medications purchased in the same period was 145,188. In this context, the redesign of the software incorporated new features.

We built the patient's medication reconciliation in the drug module in the PHR and integrated it with the prescription order entry of the EHR. This functionality consists of a new list of actions that is self-management by the patient in the system (load, modify posology, or eliminate medications on the list). After generating the patient-provided list, doctors viewed it in the EHR and validated or rejected it. A UCD methodology was used for the development of new functionalities in the medication module for both, the PHR (patient reconciliation) and the EHR (doctor validation).

User Centered Design

Health professionals and patients value health care software, but they demand better products that satisfy their clinical information needs [3]. According to reviews, many of the issues that prevent adoption and satisfactory use are related to poor design and low usability [1, 2, 8]. Usability is a measure of efficacy, efficiency, and user satisfaction [9]. Usability has been established as a key factor for applications that provide the needed support while focusing on their tasks [6]. Usable applications are easy to learn, efficient to use, easy to remember, not prone to errors, and subjectively pleasing to use [5]. UCD is an approach to achieve usable products introduced by Norman and Draper [4]. UCD has evolved from the human-computer interaction field, with the contribution of cognitive psychology, software engineering, sociology, and other disciplines that influenced it and developed methodologies that embody their points of view [6]. Usability engineering presents an iterative cycle of design and evaluation until the fulfillment of established quantitative goals measured on lab tests [5]. The contextual design includes ethnographic methods for user research while participatory design involves users by setting goals and exploring design solutions, instead of just taking part in evaluation [6]. UCD has been applied successfully to enhance the adoption and success of products [13].

However, these applications presented usability challenges described in the literature mentioned above. While the users valued the availability of online health records and tools for administrative tasks, they also noted the lack of desired functionality and difficulties to learn and use the tools. Therefore, the department implemented UCD in 2011, applying usability engineering, participatory design, contextual design, and a user-centered framework for redesigning health care software [14].

Prototype development

The design consists of different stages, which we will describe in this section. We start the prototype development from the design of the already existing medication module. To achieve an efficient design, with lower error rate and greater value for the user, we divided the design into 4 phases: *initial prototype design, initial usability tests* (for both, patients and doctors), *modifications in the prototype* (2nd iteration), and *final usability tests.* We tested the prototypes, in different iterations, through similar clinical scenarios, for a total of 31 patients (7 on paper and 24 digital) and 11 doctors digitally.

Phase 1: Initial prototype design

We designed the prototype using Adobe Fireworks, Illustrator and Balsamiq.

Phase 2: Initial usability tests

We displayed the prototype to 5 users (patients and GP), aimed to detect usability problems [15]. The objectives were identifying the user's level of understanding about the complete process of each task and measuring the ease of use. For both, patients and doctors, we used a desktop version.

Patient's usability tasks were:

1. Identify the module of medications in the home page
2. Add a new medication (e.g., Levothyroxine)
3. Modify the dose of recently added medication (e.g., Esomeprazole)
4. Remove an existing medication from the list (e.g., Amoxicillin)
5. Request a prescription renewal (e.g., Ibuprofen)

All of them divided into different hypothetical scenarios for better understanding. We provided the doctors with clinical cases specially designed for the test.

GP usability tasks:

1. Validate new medication (Levothyroxine) and generate a prescription with new credits.
2. Validate the elimination of a drug (Amoxicillin)

3. Validate the modification of dose for an already existing drug (Esomeprazole)
4. Reject Clonazepam suggested by the patient and justify it.
5. Reject Tramadol suggested by the patient and justify it. Replace it with another medication (Ibuprofen).

The test is an exercise for practice use. The user must read the task aloud and try to solve it. At the same time, they were requested, as the test went forward, to express what happens and comment on its interpretation. The session included 3 participants: the user, the observer who paid attention to the user and registered on paper, and the facilitator who moderated the test. Finally, at the end of the test, we conducted a final System Usability Scale (SUS) questionnaire to evaluate the satisfaction of use. During the interviews, we consented the users orally and recorded notes and audio their voices. We also recorded the video of the screen with MORAE software. In this way, we could record their interaction with the application while performing the assigned tasks.

Phase 3. Prototypes' modifications

Based on the tests, we modified the prototypes with the detected errors and created an updated design.

Phase 4. Final usability tests

After thorough preparation of the scenarios scripts for the new test, between August 2017 and May 2018, we individually contacted patients in the waiting room of the Health Plan of the HIBA and invited them to participate in the test. They were performed in an empty room, which physician usually use to see patients.

Between February and May of 2018, we individually invite GP from the Family Medicine Department to participate. The tests were performed on a laboratory version, with interactive PDF and were registered in Morae.

Results

The first version (Figure 1) proposed an update of the drug labels with the new features. It contained the "Add Medication" button, and the actions to modify, delete each medication previously prescribed by the physician.

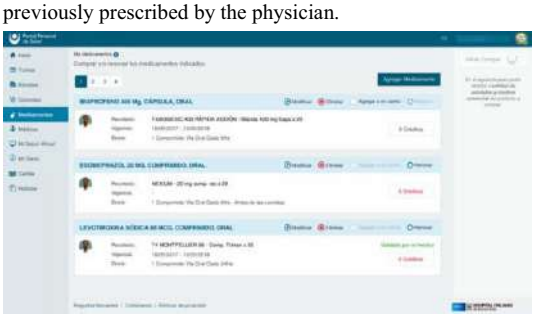


Figure 1– Initial prototype in the PHR

On the side of the GP, a new update prototype was displayed with different color under the title "Loaded by the patient", that was located at the top of the pre-existing prescribed medication list in the EHR. A "Validate" button was added to each new suggestion where the doctor had to enter it to decide what steps should be performed in each case (Figure 2).

Qualitative analysis

We performed a qualitative analysis for the registered notes and audio recordings from 31 patients and 11 physicians (end users). Characteristics of users tested are shown in Table 1.

Figure 2– Initial prototype in the EHR

Table 1– Characteristics of users tested

Variable	Value
Patients	31
Age average (years old)	52
Female	87%
Technology level*	77/100
General Practitioners	11
Age average (years old)	28
Female	
Specialty	72%
Technology level*	Family doctors (GP) 81/100
Total users	42

**Technology level was self-reported using a Likert scale.*

The main findings and comments during the testing were:

Patients:

- "It can be difficult for old people"
- "I find it useful but more for the doctor who does the control, not for me"
- "I do not know how long it will take the doctor to validate it, it should be automatically"
- "I think that could be of helpful, I would use it"
- "I do not know if I would use it, I'm not going to medicate myself"
- "I come from a different healthcare institution where everything is paper, this is wonderful"

Doctors:

- "Is going to add work to us"
- "There are doctors who do not want to prescribe medications indicated by another doctor"
- "It's good that they write it, it's easier and faster. That speeds up a lot"
- "I think it's good for patients to be able to promote autonomy, and contact more quickly with us"
- "It is useful. It saves the patient and the doctor time "

- "As a negative, you may lose the opportunity to explain to the patient why you would not indicate such medication that he suggests. e.g. He suggest clonazepam and I want to indicate something else"
- "I find it useful. It is good that there is a record of what the patient takes"
- "The change is very positive. Now, these things notify me by messages. In this way I have everything ordered"

The following findings were taken into account to make modifications on the initial prototypes:

- PHR: we included a medication search field and a print drug list button. We incorporated the possibility of uploading the medication photo for better identification. The main actions of the patient reconciliation functionality happened to have greater hierarchy, from being labels to buttons.

The fields of the medication identification were improved. We added brand name, dose, frequency, and prescribing physician. Moreover, we modified the main button of "add medication" by loading new medication. (Figure 3 and Figure 4). Table 2 presents the first 3 tests carried out with patients.

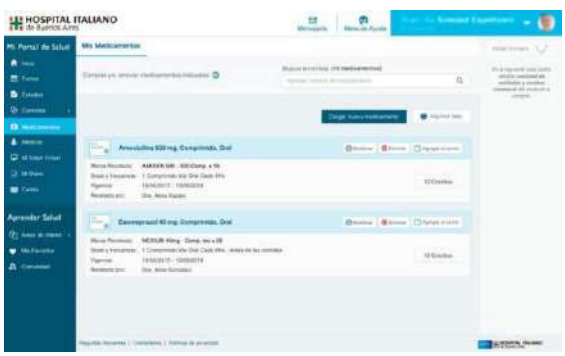


Figure 3– Final medication list. Patient view in the PHR

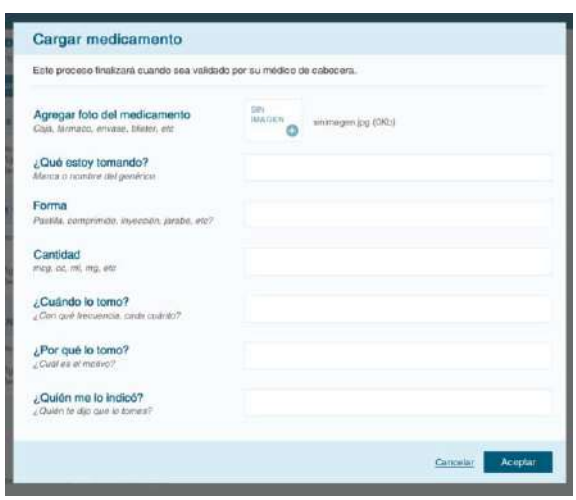


Figure 4– Patient view in the PHR (Recording a new medication):Patient should write the drug's generic or brand name, dosage, reason for prescription and prescribing physician (based on a FDA patient form to charge of a new medication) [16,17]

Discussion

Several articles have been carried out where patients were asked to review their medication list. Findings demonstrate that patients agree to correct their medication and review it effectively. Other studies concluded that the inclusion of a module with the medication list in PHRs improves the patient's medication list.

Figure 5– Prescription order entry on the EHR. In blue are listed the medication previously charged by the patient in the PHR and displayed here. Physicians have the ability to validate or dismiss changes

Table 2– Effectiveness of the exercise tasks

Task	Test 1 Correct %	Test 2 Correct%	Test 3 Correct%
Identify the module of Medications in the home page.	85%	88%	100%
Add a new medication.	71%	100%	100%
Modify the dose of recently added medication.	71%	88%	100%
Remove an existing medication from the list.	71%	100%	100%
Request a prescription renewal.	57%	100%	100%

The results of the SUS measurement of the medical and patients tests are shown in Table 3. For the first iteration, cycle we did not perform a satisfaction of use assessment.

Table 3– Patients and general practitioners satisfaction of use. Results of SUS

Number of iteration	Users	Number of users	SUS
1st	Patients	8	85%
2nd	Patients	3	93%
3rd	Patients	4	96%
4th	Patients	5	89%
1nd	Doctors	5	87%
2nd	Doctors	6	92%

Patients send an alert to updated medication prescription, but the doctors do not update the list. This problem could occur due to cultural, organizational, or relational problems, since professionals may have doubts when modifying the medications prescribed by another professional, or doubt about the information provided by patients [4,5].

In recent years, it has been proposed, as a strategy to improve the list of medication, that the patients themselves refine the list [5] since they know the medications they took in the past and the current medications. This model, with a vision of empowering patient in their care added to the correct use of the information systems, has a practical impact for the generation or modification of current computerized systems involving the active participation of patients in the medication list generation. The participation of the patient (or caregiver) is proposed as a fundamental element in the preparation of the best medication list.

We presented a user-centered design of a module of medication list with reconciliation functionalities and managed by the patients, connected to their EHR for supervision and medical validation. It is important to establish guidelines on these key activities to support the implementation in a hospital environment [18]. Our findings of SUS satisfaction of use in the final iteration (89% at 4th cycle, while 96% previously) demonstrate that we need to identify the probable cause and new modifications that will be necessary with re-testing of the final version.

Conclusions

This paper identifies the importance of involving primary care patients by integrating them into care processes and strengthening the doctor-patient relationship. A common issue on medication management arises when a patient is discharged from hospital. The study suggests possible modifications to the established models for medication reconciliation. The findings may be relevant to health care providers and policymakers who have an interest in encouraging patients to use PHR. This type of study could be replicated and extended in other settings to better understanding how to make the technology more useful for patients in their own health care.

References

[1] C. LeRouge, and N. Wickramasinghe, A Review of User-Centered Design for Diabetes-Related Consumer Health Informatics Technologies, *J. Diabetes Sci. Technol.* **7** (2013) 1039–1056.

[2] M. Zahabi, D.B. Kaber, and M. Swangnetr, Usability and Safety in Electronic Medical Records Interface Design: A Review of Recent Literature and Guideline Formulation, *Hum. Factors*. **57** (2015) 805–834.

[3] M.W. Friedberg, P.G. Chen, K.R. Van Busum, F. Aunon, C. Pham, J. Caloyeras, S. Mattke, E. Pitchforth, D.D. Quigley, R.H. Brook, F.J. Crosson, and M. Tutty, Factors Affecting Physician Professional Satisfaction and Their Implications for Patient Care, Health Systems, and Health Policy, *Rand Health Q* **3** (2014), 1.

[4] D.A. Norman and S.W. Draper, *User Centered System Design; New Perspectives on Human-Computer Interaction*, L. Erlbaum Associates Inc., (1986).

[5] D.J. Mayhew, The Usability Engineering Lifecycle: A Practitioner's Handbook for User Interface Design, Morgan Kaufmann Publishers Inc., (1999) 71–114.

[6] J.M. Carroll, Human-computer interaction: psychology as a science of design, *Annual review of psychology* **48** (1997), 61-83.

[7] T.J.C.A. Hospital, Hospital National Patient Safety Goals, in. https://www.jointcommission.org/assets/1/6/2019_HAP_NPSGs_final2.pdf (accessed November 21, 2018).

[8] S. Hermawati, and G. Lawson, Managing obesity through mobile phone applications: a state-of-the-art review from a user-centred design perspective, *Pers. Ubiquit. Comput.* **18** (2014) 2003–2023.

[9] ISO, Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability , (n.d.). https://www.sis.se/api/document/preview/611299/ (accessed November 2, 2018).

[10] F. Plazzotta, D. Luna, and F.B. de Quirós González, Health information systems: integrating clinical data in different scenarios and users, *Revista peruana de medicina experimental y salud publica* **32** (2015), 343-351.

[11] D. Luna, F. Plazzotta, C. Otero, F. González Bernaldo de Quirós, A. Baum, and S. Benítez, Incorporación de tecnologías de la información y de las comunicaciones en el Hospital Italiano de Buenos Aires, (2012). http://hdl.handle.net/11362/3959 (accessed November 23, 2018).

[12] M.I. Smith, G. Garcia, M. Simon, L. Bruchanski, J. Frangella, J.A. Sommer, M.V. Giussi, and D.R. Luna, Lessons Learned After Redesigning a Personal Health Record, *Stud. Health Technol. Inform.* **245** (2017) 216–220.

[13] G. Perlman, Book review: Human-Computer Interaction, by Jenny Preece, Yvonne Rogers, Helen Sharp, David Benyon, Simon Holland, and Tom Carey (Addison-Wesley, 1994), *ACM SIGCHI Bulletin*. **26** (1994) 82–85.

[14] C.M. Johnson, T.R. Johnson, and J. Zhang, A user-centered framework for redesigning health care interfaces, *J Biomed Inform* **38** (2005), 75-87.

[15] J. Nielsen, Why you only need to test with 5 users, in, Useit. com Alertbox, (2000).

[16] Center for Drug Evaluation, and Research, Resources for You - My Medicine Record, (n.d.). https://www.fda.gov/drugs/resourcesforyou/ucm079489.htm (accessed November 21, 2018).

[17] My Medicine List, (n.d.). http://www.safemedication.com/safemed/MyMedicineList.aspx (accessed November 21, 2018).

[18] AHRQ, Guide to Patient and Family Engagement: Environmental Scan Report, (n.d.). https://www.ahrq.gov/sites/default/files/publications/files/ptfamilyscan.pdf (accessed November 20, 2018).

Address for correspondence

Márquez Fosser Santiago. MD,

Health Informatician,

Hospital Italiano de Buenos Aires, Buenos Aires, Argentina.

email: santiago.marquez@hospitalitaliano.org.ar

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190433

Designing New Buildings to Accommodate Current Technologies

Ryan McNew[a], Elizabeth Weiner[a], Jeffry Gordon[a]

[a] *Frist Nursing Informatics Center, Vanderbilt University School of Nursing Nashville, TN, USA*

Abstract

Growing higher educational programs are always interested in adding physical facilities to their campus. Many times the initial plans do not include all the technology details needed for either a standard or a cutting-edge program. This paper describes planning and implementation efforts at the Vanderbilt University School of Nursing, which has resulted in a new 29,947-square-foot facility addition. The new space includes a complete floor designated as simulation laboratory space, an interactive classroom with five screens suitable for small group interaction, and a virtual classroom for cutting-edge multimedia production and transmission. Additional floors house both technology rich seminar meeting rooms as well as collaborative offices. All spaces are integrated using the innovative Crestron control system, including a multitude of their solutions that help to make this a smarter workplace that is both secure, smart, and worthy of completing educational goals.

Keywords:

Educational technology, distance education, organization and administration.

Introduction

Adding an additional facility into a school with a rich history brings challenges which are not always conceived in the early planning and design phases of the project. The Vanderbilt University School of Nursing (VUSN) has been providing quality education for nursing students since 1908 and distance education for over twenty years. The addition of a new building to the current facility brings the challenge of creating state-of-the-art spaces that not only provide a rich educational experience to our local students but also to our distance students. What we have found is that it is vital to have an informatics representative involved starting in the early planning stages and attend not only the Owners but also the sub-contractor's meetings. This representative needs to provide insight into what technologies (hardware, software, and networking) will need to be implemented into the new space as well as the requirements to make them function properly.

Prior to construction, VUSN consisted of three buildings, which were built at different times and housed all of the classrooms, simulation laboratory, conference rooms as well as faculty and staff offices. The total square footage for VUSN, prior to construction, was 79,304 square feet. The new building adds an additional 29,947 square feet for a grand total of 109,251 square feet. This presentation describes the lessons that we learned from not having an informatics representative involved from the beginning of the project, but also how we were able to ultimately achieve our goal of providing an educational experience, which utilizes the state-of-the-art technologies in the curriculum.

Throughout the years as technology changed so did the needs of our students, which meant that our educational spaces required upgrades. Technologies used in the everyday life of a student are expected to be used within the higher education environment to aid the student with developing their skills [1]. A major hindrance we discovered in bringing that technology is that due to the physical limits of the buildings, we were not always capable of implementing the latest technologies without significant costs. The new building was designed so that all of the buildings would connect, thus providing us with a means of upgrading several technologies used within the school.

Methods

Our biggest challenge was designing these new spaces without having any input into the architectural design of the facility. The informatics team at VUSN, led by Senior Associate Dean for Informatics Dr. Betsy Weiner, consists of twenty-eight individuals. Positions include classroom support, computer lab coordinator, network manager, web developer, graphic artists, general IT support, instructional designers, program coordinator, videographers, materials coordinator, videographers, media services, programmers, simulation staff, and informatics faculty. Due to the size of the informatics team, and because they are housed in Frist Hall, the name Frist Nursing Informatics Center has been given to them.

The first interaction with the new building was after months of designing and development had already taken place by the Owners and architects. There was zero input during the initial design phase from anyone within the Informatics Department. Our department received the completed floor plans and was given the directive to "make it work." When asked why our group was late to the planning, the leadership team responded that they didn't want "too many opinions at once." Unfortunately, that also meant that original budgets for fund raising purposes was not adequate for meeting the IT and simulation infrastructure needs.

Simulation Laboratory

The simulation laboratory, originally housed in a 1,290-square-foot space and located within Godchaux Hall, was to be relocated into a space within the new building which occupies an entire floor. The new space is over three times (3,932 square feet) the square footage of the original location. The original laboratory made use of a single, medium sized " L" shaped laboratory using rudimentary means with a large amount of imagination to run a simulation and zero control rooms in which the simulation is guided. Another classroom down the hall from the laboratory was being used for pre and post briefing sessions. During the post brief, the faculty could

only discuss the simulation using their recollection of events as well as that of students. Occasionally videotaped sessions were directed there so that group critique could take place, but there was no effective method to capture feedback for individual students.

The new simulation lab takes up an entire floor and consists of thirteen bed bays. Each bed has its own assigned control booth with direct line of sight to the bed. There is a large storage and prep room located in the center of the lab along with a complete computer setup allowing the simulation technicians to monitor the simulations in real time as well as aid faculty in the event they are needed. The facility looks like any hospital floor, complete with headwalls, hospital beds, simulated gasses, and computer monitors. Additional equipments can be rolled in from the storage room as required by the simulation. Curtains to block the sounds of the control rooms were also deemed important.

A barrier, which we encountered early on in the project, was that there was zero input from educators or technologists that would use or support the simulation laboratory. Another was that some of our floor space was taken up with main support structure for the building and could not be moved, providing some space constraints that we could not physically avoid. Our first priority was to ensure that it not only met the needs of the students, faculty, and staff, but also fit within in the confines of the new space we were given. After the initial design phase was complete, the informatics team was given the plans and tasked with developing a technology plan for a 13-bed simulation lab each with its own control room as well as two pre and post briefing rooms.

We quickly identified key members of the needed planning team. Participants included the Informatics Department (including the simulation staff who run the lab on a daily basis), and other faculty who are skilled in using advanced simulation within the Medical Center environment called the Center for Learning and Experiential Assessment (CELA). Initial discussion centered on revising the floor plan to better suit the running of complex simulations, complete with control rooms and simulation management software (B-Line Medical®). Additional considerations were noted so that faculty could best take advantage of all the factors the new facility would bring. Suggested revisions prompted several larger meetings with the Dean, architects and a smaller subset of the task force to discuss why the plans would not work as well as give our suggestions of change. This process repeated four times until we had a floor plan that would work for all parties involved. There was some initial resistance to taking this time due to the established opening date for the new facility, but the importance of a well-planned facility that met our needs was finally understood. Once the plans were finalized, we were able to focus on making the lab function properly.

The simulation lab utilizes a product from B-Line Medical®, called SimCapture which seamlessly captures audio and high quality video during the simulation encounter, streams live to faculty and students watching the simulation in real time, as well as records it for playback during the debrief and assessment sessions [2]. Students being recorded during their simulation experience then watched their experiences post-simulation and helped students learn by using active and contextual aspects from within the meaningful learning model [3]. Under optimal conditions, the equipment required to run the system should be on the same floor as the lab itself. Our simulation lab "server room," however, is located on the floor below the lab. Working closely with the audio/video (AV) integrator as well as the architects we designed the cable paths to run throughout the lab ceiling and then follow a conduit down into the "server room" below.

Each bed bay is equipped with a Stryker bed and over the bed table, two Axis cameras, two microphones, a functioning Amico headwall with simulated med gasses, vitals monitor, a computer with monitor and a B-Line SimCapture Node. Both monitors bedside are ELO Touch mounted to the headwall with custom Amico monitor arms. We installed a camera at the head of the bed facing the feet and another camera placed at the foot of the bed facing the head so that the entire simulation bed can be covered. The two ceiling mounted microphones use a cardioid polar pattern to pick up the audio bedside while helping to reduce the amount of ambient noise from other areas outside of the bed bay. Every bed has a functioning Amico headwall with simulated med gasses.

Figure 1- Bed bay

Each control room is equipped with one-way glass so that the simulation controller can operate the simulation while having a direct line of sight into the bed bay, a computer with dual monitors (with space permitting), headphones and speaker and a voice changer allowing the controller to alter his or her voice to fit the simulation happening bedside. For instance, a male faculty member could use the voice changer to sound like a female.

Figure 2 - Control room

We decided that it would be beneficial to build in redundancy. In order to ensure a simulation takes place, we designed the system so that any control station has the ability to operate any bed bay. For example, control room one could have a technical issue preventing the scheduled simulation from being performed; we now have the ability to utilize any unoccupied control room to monitor the simulation happening in bay one. The only downside is that the controller would lose direct line of sight.

In order to make this happen, we needed custom programming and additional equipment to ensure that the video and audio streams are routed from beds to control rooms. We contracted with a local integrator, Tristar Digital Connections, LLC, to provide all the equipment and programming. Each night at midnight, the system reboots and resets all of the streams to their default locations. In our "server room," we have the master controls for audio and video routing. The programming allows us to reroute the audio and video signals from any bed to any control room should a control room go down. The display uses a Crestron 10" panel and allows us to drag any bed from the top bank to any control room in the bottom bank. We also have the ability to reset all of the connections manually as a failsafe.

Figure 3- Server room

Wachtmeister Interactive Classroom

This classroom is unlike any classroom we have ever had as it flips the traditional classroom. A typical classroom has a lectern with a computer, feeding into a projector, which then projects the image, typically a PowerPoint presentation, onto a screen located behind the instructor. Students sit in rows of chairs normally bolted to the floor, they then take notes and watch the presentation. We chose to do none of that in this classroom in an effort to promote better student interactivity.

This classroom is equipped with five interactive ViewSonic, 4K displays each containing a blade PC, a PanaCast 2 panoramic-4K camera mounted on a fully articulating Triad-Orbit mount and a Crestron AirMedia 200; however, there is not a display located at the front of the classroom with the lectern. The Crestron AirMedia 200 allows our students to wirelessly present content from any of their devices such as laptops, tablets, and even cellphones [4]. The use of technology is what enables our students to become active learners and as such our nursing educators must also integrate technology into their curriculum [5]. Our main goal of using this technology was to embrace the interactive learning model by finding ways to have the faculty interact with the students and provide engaging content to the learning model.

During the design phase of this room, we developed multiple scenarios of potential use cases. The sole intent of this space was to develop learning environment where the faculty and students interact not only with the faculty, content but also amongst their peers during the class.

One such scenario developed was that this classroom would operate with the faculty starting class at the lectern by loading the lessons content onto the PC located within the lectern or onto a Microsoft Surface Pro. Using the controls located within the lectern touch panel control station or using a mirrored set of control on the Surface Pro, they are capable of sending the information to all 5 displays. The students can then cluster around any of the five displays and engage in the content with the instructor interacting with the student clusters.

Another scenario describes where up to five topics of discussion are identified. Students are divided into small breakout groups to discuss their specific topics and then to present their findings to the class during the class period.

Figure 4- 3 of the 5 displays in the Interactive classroom

For this scenario to work, we met with our integrator and discussed how we wanted the room to function: (1) each of the displays must be able to display the content from the lectern; (2) each display must be capable of displaying its own content; (3) we must be able to matrix the displays so that, as an example two of the five displays must have the same content, and the other three be able to have their own content; (4) students must be allowed to connect their devices to the display; (5) the faculty must be able to take any display and make it display onto any number of the displays. Our integrator came back with a solution that entailed over 60,000 lines of code to make the matrixing function using our guidelines. With the addition of the Crestron AirMedia we can allow our students to connect any device wirelessly to the display. Each display has a Crestron TouchPanel adjacent to the display; with a simple touch of a button the input for the display can easily be changed. The faculty have a master control, which allows them to control the input for each display, along with changing the matrixing.

Virtual Classroom

The concept behind the Virtual Classroom is for it to be more of a production studio than a typical classroom. We have found that there has been a growing demand for various types of recordings from faculty and students. The process to make a single recording happen was extremely labor intensive in regard to setup and tear down. Our team had no place to setup cameras, teleprompters as well as lights and leave the setup in place for any length of time due to the fact that we had to reserve classroom space for these recordings.

The space is a large corner room with windows facing both the street outside as well as the large open-air atrium on the right of the room. Our primary concerns were sound and light filtering into the space and causing detrimental effects on the production quality. Our first attempt to mitigate the effects was to request that the windows be blanked out (filled in). Our request was promptly declined due to it throwing off the aesthetics of the entire building. Our second attempt was to request blackout shades installed over every window that would be lowered while filming, that request was approved.

Due to being a LEED®- and WELL-certified building™, we were required to have regular offices lights in the space [6]. These lights do not provide satisfactory results for producing videos. We were able to discuss it with the architects and have the minimal amount of required lights as well as the ability to control them manually. Lighting was still problematic. Our solution was to remove the ceiling and paint it black at 9' above the finished floor (AFF). This exposed the deck and allowed us to install a grid system with electrical boxes located in numerous places around the grid to support the numerous LED lights for video production. This allows us to hang them in any configuration required for the video shoot. We also made the decision to hang the required office lights higher than the light grid to avoid the possibility of casting shadows.

Secondly, we were tasked with how to dampen the sound during recordings since all the walls were hard and flat causing sound waves to bounce off them. In discussing the situation with our AV vendor, we came up with the plan to add acoustic tiles around the peripheral of the room. These panels will help to absorb sound providing with the desired audio effect.

Thirdly, we had the possibility of having noisy equipment, such as light or sound boards, in the space during recording. Our solution to this issue was moving our Media Producer/Director's office adjacent to the classroom with a large pass-through conduit running from his office into the classroom. This solves the noisy board issue by having the equipment run from inside his office then passing the cables through the conduit into the classroom to the actual devices. The Media Producers office has one-way glass allowing him to see into the room but not distract the filming.

We also mounted a ViewSonic 4K Interactive display with blade PC on an articulating arm allowing us to angle the display to best suit the application. We also included a Crestron AirMedia 200 which allows any device to wirelessly connect to and display content.

This space can now be used for various scenarios; however, there will not be an audience in most cases. The question was how to get the content to our student population? We devised two different scenarios. Our first is that we would record and edit the video for dispersal at a later point in time, such as video on demand. Our second case was for doctoral defenses, which we could stream to larger classrooms. For this we used Crestron NVX, which provides streaming audio and video with no latency over a 1-Gigabit network connection to another NVX device located in another one of our classrooms [7].

Crestron NVX allows us to remove costly, large profile processors and adopt a more cost-effective means of sending and receiving data throughout the school. With one NVX node set to send and the other set to receive, we are able to send data HD (high definition) content across the network without latency; all while providing a large cost savings. Without NVX, we would have had to incur more costs by upgrading our switcher as well as a lot more AV wiring. To make use of the NVX environment, the user must give formal permissions for the content to be viewable by another group; our AV team then uses the NVX Director to establish a connection between the two or more rooms.

Seminar Meeting Rooms and Collaborative Offices

Both the Simulation lab and Interactive Classroom are technology laden and thus require some training in order to make full use of the space. We chose a unified and simplistic approach to the seminar spaces and collaborative offices. Each space makes use of the same ViewSonic 4K Interactive displays as well as the Crestron AirMedia 200 used within the Interactive Classroom. Each one of these rooms is also equipped with a Crestron NVX node allowing us to stream any of these rooms to any other NVX equipped room within the School of Nursing. Our objective was to create spaces using the latest technology while designing the space for ease of use to diminish the amount of tech support calls.

We accomplished this task by designing the spaces using the same technology in all collaborative spaces. This allows anyone within the Vanderbilt University School of Nursing to walk into any collaborative space and immediately feel comfortable knowing that the user will be able to operate the room as intended. Following the same layout design as the Interactive Classroom, we mounted one to two interactive displays, located next to each display is a Crestron TouchPanel controlling the inputs as well as a Crestron AirMedia 200 allowing the user the ability to connect wirelessly any device to share content. Numerous factors were taken into account to determine the number of displays required for a space as well as screen size and resolution. Factors considered were: size of room and content displayed (content such as detailed images maintains requirements much greater than that of a Word or Excel document). When two displays were required, we made the decision to install a Crestron AirMedia 200 with each display allowing both displays to show the same content or depending on the use case each display could be showing separate content.

Results

The various spaces discussed within this paper will be open for use early January 2019. The various, unforseen, delays during construction moved our opening from August 2018 until Janurary 2019. A short video will be shown during this presentation showing the Simulation Lab and Wachtmeister Interactive Classroom fucntioning and performing the tasks described within this paper, along with photos to show other spaces.

Evaluations will be given to both faculty and students in regards to ease of use and overall functionality of the systems and spaces. Results will be presented.

Discussion

It is anticipated that the results of the surveys will be positive from both the faculty and students. During the design phase we had input from many faculty within the VUSON, adjustments were made to the designs based off of their feedback. Our instructional designers were brought into the classroom space and asked to provide feedback on the design as well as to start the process of helping faculty to design course material utilizing the interactive of the space.

Feeback was taken into consideration and adjustments made based off of the feedback from all parties. There are numerous possibilities for altering how synchronous learning is administered within the school with a bleed over into asynchronous as well.

Conclusions

Technology is used to instruct our students; technology is constantly changing and so should the way the our learners are taught. Technology use, such as mobile technology, in education shows that it impacts the students learning objectives positively [8]. Interactive learning is happening on a daily basis, when parents hand their child a cellphone or tablet to play a game or watch an educational video, that child is learning using technology. We should not be satisfied with a learning model decades old and our students should not be either.

Nurses are known for their abilty to get things done when responding to situations, thinking quickly to find solutions to problems that are presented; technology can only enhance the abilily of nurses by provinding access to real-time data [9]. Why should we not use the latest technology to enhance the skillset of our students? In order to produce the highest quality students, they need to be exposed to as many different scenarios as possible. This is accomplished by bringing innovative ways to teach our students and expose them to scenarios not typically available.

We learned that administrators need to understand the detailed planning needed for both technology and simulation laboratory infrastructure early on. This message needs to start before initial plans are made, not after. Simulation laboratories are not well understood by those outside of healthcare (particularly the complicated networking plans) and cannot be stressed enough times to the installers. LEED® and WELL™ certified buildings are additional challenges that create need for compromise. Lighting and sound accomodations can be made that meet those requirements. In conculsion, while there are many details to be considered, a knowledgeable team of technnology professionals can help to add to the cutting-edge-facility in such a way that everyone is satisfied with state-of-the-art technologies that are innovative and useful.

Acknowledgements

The authors wish to thank DF Chase, Inc. Construction for their willingness to working with my team and accept our countless change orders on our first ever construction project. The Tristar Digital connections, LLC team without your expertise this project would not have gone as smoothly as it did. As well as all the Vanderbilt University faculty and staff that provided input into the project. While faculty and staff as of this writing occupy offices within the building none of the other spaces are yet open, they will be open in January 2019.

References

[1] M. Honey, Undergraduate student nurses' use of information and communication technology in their education, In A.K. Rotegard *et al*. (eds*.), Proceedings of Nursing Informatics*, **250** (2018), 37-40. doi:10.3233/978-1-61499-872-3-37.

[2] SimCapture, (2018). https://blinemedical.com/pdf/simcaprutre.pdf, (accessed November 24, 2018).

[3] L. Shuhong, H. Harrison, Using technologies in nursing research education: A mixed methods case study, *CIN* **66** (2018), 293-304.

[4] AM-200 AirMedia® Presentation System 200 (2018) https://www.crestron.com/Products/Workspace-Solutions/Wireless-Presentation-Solutions/Crestron-AirMedia-Presentation-Systems/AM-200 (accessed November 24, 2018).

[5] B. Mackay, J. Anderson, T. Harding, Mobile technology in clinical education, Nurse Educ Pract **22** (2017), 1-6, doi:doi.org/10.1016/j.nepr.2016.11.001.

[6] K. Coco, IWBI™ announces alignments between the WELL building standard™ and LEED®, (April 21, 2017), https://www.wellcertified.com/en/articles/iwbi-announces-alignments-between-well-building-standard-and-leed (accessed November 25, 2018).

[7] DigitalMedia NVX Series, (2018), https://www.crestron.com/Products/Featured-Solutions/DigitalMedia-NVX-Series (accessed November 24, 2018).

[8] T. Schilhab, M. Stevenson, P. Bentson, Contrasting screen-time and green-time:a case for using smart technology and nature to optimize learning processesFront Psychol **9:773** (2018), doi:10.3389/fpsyg.2018.00773.

[9] C. Mather, E. Cummings, F. Gale, Mobile learning in nursing: tales from the profession, In E. Cummings et al, (eds.), *Connecting the System to Enhance the Practitioner and Consumer Experience in Healthcare*. **252**, (2018), 112-117, doi:10.3233/978-1-61499-890-7-112.

Address for correspondence

Ryan McNew, Vanderbilt University School of Nursing, 461 21st Ave. S., Nashville, TN 37240-1104, USA; E-mail: Ryan.McNew@Vanderbilt.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190434

User-Centered Value Specifications for Technologies Supporting Chronic Low-Back Pain Management

Mark Merolli[a], Charlotte J. Marshall[a], Adrian Pranata[a], Jeni Paay[b], Leon Sterling[b]

[a] *School of Health Sciences, Swinburne University of Technology, Melbourne, Victoria, Australia*
[b] *Centre for Design Innovation, Swinburne University of Technology, Melbourne, Victoria, Australia*

Abstract

Low-back pain (LBP) is a leading cause of disability globally. It is complex and multifactorial, with a miriad of factors interwining and interacting to burden healthcare and individuals. Self-management support is central as part of best-practice to improve outcomes. In recent years, informatics has increasingly been considered to support care; however, due to its complex nature, several factors need to be unpacked in order to consider how technologies might support LBP. The present study utilised semi-structured interviews involving N=20 participants (n=10 practicing clinicains and n=10 indidivuals living with chronic LBP (cLBP)) to collect user-centered perceptions and considerations for key factors central to technology succeeding in supporting cLBP. Six themes were identified: tracking, alterts, user-experience, communication, feedback, and content. Findings lay groundwork for future research aimed at developing technologies that can encourage shared-decision making in supporting cLBP management in a particpatory health paradigm.

Keywords:

Low Back Pain; Informatics; Self-Management

Introduction

Low-back pain (LBP) continues to be listed as the leading cause of disability according to the Global Burden of Disease study by the World Health Organization (WHO) [1,2]. This can be observed via years lived with disability, ballooning costs, opioid analgesic abuse, increasing service demand, poorer quality of life, and inability to work [3,4,5]. Fear-avoidance of movement and poor adherence to self-management are major challenges at the heart of clinical patient management and individuals' self-management of chronic LBP (cLBP). This condition continues to plague health professionals and individuals within a participatory health and shared-decision making paradigm to achieve good outcomes, due to its complex and multifactorial etiology, with significant physical and psychosocial effects underpinning its presentation [4,6,7,8].

With greater self-management support for those living with cLBP, there is an increased likelihood that [9,10]: return to work outcomes will improve; overprescribing of opioids will decrease; unnecessary referrals to public chronic pain services from primary care will be reduced, thus unburdening the public healthcare system; better treatment options and health outcomes can be provided to individuals in rural and remote communities with difficulty accessing services; private health insurance claims may decrease, allowing opportunity for premiums to be reassessed.

Informatics technologies have been considered in its management. User-centered and participatory health enabling mobile health technologies have previously been reported to have a positive impact on motivation, behavior change, adherence to interventions, and pain outcomes in this context [11,12,13,14].

Clinicians, researchers, and informaticians continue to consider digital self-management support technologies for a variety of presentations of LBP; for example, this includes wearable monitoring device for movement and posture detection, and motivation and adherence tracking applications. These have the potential to transform current practices by improving health literacy, thus promoting greater self-management and improved outcomes [15,16,17,18].

However, before further developments in the digital monitoring and support space for cLBP can be achieved, more research is prudent to ascertain key evidence-based, user-centered considerations, in order to gather value specifications and user requirements for technology from both clinicians who treat cLBP and individuals living with cLBP. The present study forms part of a larger project, whose ultimate aim is to develop digital technology to support outcomes for LBP. The present study reports on data collected from individuals regarding perceptions concerning what underpins safe, effective and empowering mobile and/or digital monitoring technologies for this cohort. In other words, what will be required for participatory health enabling technologies for them to have a significant effect on cLBP management?

Methods

The present study recruited clinical health care professionals (HCP) with expertise in managing cLBP and individuals living with cLBP to participate in exploratory telephone-based semi-structured interviews (SSIs) to discuss the complex needs of living with and/or managing cLBP, as well as perceptions around the use and utility of informatics technologies to support care.

The University's Human Research Ethics Committee has approved this project (ID: 2018/135).

Recruitment

Participants were recruited through various avenues, inlcuding: word of mouth, study recruitment posters in clinical settings, and the study investigators' clinical networks spanning public primary health care settings, tertiary multidisciplinary pain services, and private primary care clinics

Once individuals registered interest in participating in the study, they were directed to an online form that included the study plain language statement and screening questions, which then directed them to the informed consent form. The questionnaire screened particpants against study inclusion criteria (which, for eligible participants from the cohort living with cLBP, included completing the Oswetry Disability Index to confirm a self-reported diagnosis of at least moderate LBP), and collected baseline demographics for descriptive statistical purposes (e.g. age range, gender, employment status, clinical speciality, years living with cLBP, etc.).

Data Collection

Data was collected and recorded using SSIs. Interviews took on average 22.7 minutes to complete.

The broader research project is methodologically sound and underpinned in informatics methodologies central to the robust and successful design and development of technologies for digital health interventions. This includes underpinning the research in the rigorous and academically validated roadmap for developing technology in health, the "Centre for eHealth Research Roadmap (CeHRES)", which has been used in a sample of chronic pain patients [19,20,21] http://www.ehealthresearchcenter.org/wiki/index.php/Main_Page . The present study is part of the 'value-specification' stage of CeHRES.

Furthermore, Greenhalgh et al. [22] recently published a robust review proposing a novel informatics model to ensure that future interventions in digital health do not fail the nonadoption, abandonment, scale-up, spread, and sustainability framework (NASSS). The framework consists of seven inter-related domains that should be considered when developing technology for health to support success.

Using NASSS as the methodological underpinning for data collection, the present research is the first (to the investigators' knowledge) to appropriate NASSS into a qualitative data collection instrument to conduct interviews as part of the value specification per CeHRES [20,22]. Hence, this study's data collection tool is a unique offering in itself to the informatics knowledgement management community as well.

Data Analysis

Once conducted, interviews were transcribed verbatim. Inductive thematic content analysis (TCA) was employed to identify themes latent within the data [23]. This was because the primary goal of the present study was to explore and examine user-centered consideration for technology that supportted management of LBP. Thematic analysis aims to identify themes in a set of qualitative data in an attempt to give meaning to the common voices of collective participants [23]. As described, basic descriptive statistics were also collected to quantify simple closed questions, such as demographics and condition-related data.

Three investigators (MM, CM, and AP) analysed the first interviews to be transcribed to create a preliminary coding schema and MM then used this framework to analyse the data.

Results

Recruitment and Participants

The present study successfully recruited and interviewed a total of N=20 participants. This included an even cohort of n=10 clinical health professionals with expertise in managing cLBP, and n=10 individuals living with cLBP. A full prospective co-hort of N=27 were originally screened; however, after applying inclusion/exclusion criteria, five clinical health professionals were excluded due to lack of follow-up after pre-registering. Two prospective LBP patients were excluded for the reasons of lack of follow-up post registration of interest, and non-chronic LBP as per the Oswestry.

Participant demographics of those included can be seen in Table 1. As can be observed, health professionals came from a mixed background, were predominately male, between the ages of 30-39, with a wide range of clinical years of experience (3-25 years post-graduation), and skewed towards practicing in the hospital setting. Of participating patients living with cLBP, age range was also skewed towards 30-39, with an average Oswestry dis-ability score of 39.1 (range = 31.1-51.1).

Table 1 - Participant Demographics

	HCP (n=10)	Individuals with cLBP (n=10)
Gender		
Male	8	5
Female	2	5
Age Range		
30-39	8	5
40-59	2	3
60-69	-	2
Level of Education		
High School or Less	-	4
College/University Completed	2	3
Post-Graduate Degree Completed	8	3

Thematic Content Analysis

A total of seven hours and 35 mintues of interview data was coded. Following the first round of coding, a total of n=52 codes were identified. This was broken down into n=31 individual codes from the health professional interview data, and n=21 codes from the cLBP cohort. This preliminary coding schema was conferred by the investigators and after a second round of coding, refined to group like codes into a resultant categorisation of n=6 themes common across both cohorts pertaining to digital technology for supporting cLBP. Of the n=6 resultant themes, these were broken down into sub-categories: N=19 identified by the clinical health professional cohort, and n=12 identified by people living with cLBP. Themes and sub-categories can be observed in Table 2.

Tracking

Participants from both cohorts idenitified the utility of any technological solutions to be able to track several metrics. For example, these may inlcude: activity tracking, other physiological metrics (i.e. inflammation, heart rate, etc.), posture, sleep, regimen adherence, and pain levels.

"I'd sort of think of something like that, where it feeds back and says oh, you're in this posture or you're in that posture, those sorts of things" (PRT05)

Alerts

Similarly, participants identified that a useful feature of digital technology to support cLBP would be to include alert/reminder features. This might include reminders to move or complete prescribed exercises, or further provide physical prompts and/or

motivating prompts to breath, move, and reinforce good behaviours. Of note, as opposed to the health professional cohort, patients suggested haptic prompts, such as vibrations or prods, which might reinforce positive behaviours or postures.

"..almost like a, you know when the Apple watches, they give you a little buzz if you've been sitting down for too long, that kind of stuff I would be really, I'd find that really useful" (PRT04)

User-experience

User-experience of prospective digital solutions was also featured in the identified themes. This referred to both the platform of any physical platform (which was overwhelmingly suggested to be app and smartphone based), as well as reference to more aesthetic and subjective features. Feedback included suggestions for any technology to be small, portable, wireless, visually appealing, insightful, lightweight, and durable.

Communication

Clinicians were more vocal around their desire for digital solutions that support LBP management to include robust communication features. Whilst individuals living with LBP also indicated that SMS or text messaging features would be useful, clinical health professionals were more direct in their suggestions, recommending secure messaging features, email capabilities, and even social networking features for patients to connect with one another.

Feedback

Feedback was another theme commonly idenitfied. Quite similar to the 'alerts' theme, at the heart of providing feedback, participants indicated that they believed any digital technology designed for this context would be valuable and useful if it were intuitive to positively reinforce positive behaviours. This might include providing insights and reinforcement around good posture, regimen adherence, and gains/improvements.

"I can imagine if you had a wearable device that was, like, you've been standing for X number of minutes and we know that your tolerance is four minutes and you've been standing for three and a half, you need to go and sit down.." (PRT05)

Content

Finally, but perhaps the most strongly represented theme, several codes described key content or, functionality that the technology should include. For example, the most obvious inclusion according to both health professionals and individuals with cLBP, was the provision of education resources that educated individuals about cLBP, its causes, progress, and management. Furthermore, health professionals indicated certain complimentary features, such as, the inclusion of educative (or demonstrative) videos, the ability to prescribe and view exercises from within an app, gamification features to enhance motivation and/or adherence, mental health components (i.e. coping strategies, pacing and mindfulness training), as well as one suggestion to be able to collect pre-screening patient data before they arrive for an appointment. Patients on the other hand, also indicated that teleconsultation features, such as video-based consultation ability with their practitioner might be desirable.

"If there was some application component around what chronicity does to pain, and how that changes how pain is perceived in your brain..that woud be beneficial" (HCP02)

Table 2 - Thematic Analysis of Interviews

Theme	Sub Categories
Tracking	Activity
	Physiological Signals
	Posture
	Sleep
	Regimen Adherence/ Compliance
	Mood
	Pain
	Patient-Reported Outcome Measures
Alerts	Reminders
	Prompts
User-Experience	Platform
	Aesthetics
	Usability
Communication	
Feedback	
Content	Educational Resources
	Videos
	Exercise Prescription
	Gamification
	Mental Health Support
	Screening

Discussion

The data collected from the present study provides prelminary insights into the user-centered needs for informatics technologies that have the potential to support cLBP management. Findings indicate that both health professionals who manage cLBP, as well as individuals living with it, consider a range of factors when envisioning where and how informatics technologies might support management of the condition**, such as** being able to track progress, communicate, provide or receive feedback and reminders, and source educational content. Similar themes have been reported in a previous study [24]. The range of diverse themes identified (n=6) pertaining to techology to support cLBP further highlights the complexity of managing a condition like cLBP, with its mulifactorial nature [4,7,10].

Whilst both cohorts of participants (clinicians and individuals living with cLBP) provided data on the **six** themes, there was a slight difference in their individual perceptions or motivations underpinning these. For example, language pertaining to patient perceptions towards technologies surrounded their desire for technology to support active participation in self-management. Descriptive langauge used by participants pertaining to the sub-categories presented in Table 2 was quite 'active' in that it promoted active engagement in self-management (e.g. move, enter, log, reinforce, motivate, remind me). This supports literature in the participatory health domain, indicating that patients living with complex chronic conditions wish to be empowered shared-decision makers in their rehabilitative journey [14,16]. **Likewise, this** was also evident to some extent in comments from clinicians in describing their desire for technology to include robust communication features to enhance patient-practitioner communication, which has previously been reported to be beneficial [25]. Furthermore, patient comments commonly referenced their desire for technology that supports LBP management to focus on providing them motivation, feedback**,** and reinforcement to perform their exercises and rehabiltiation regimens. This was also broached by one clinician, who suggested gamification may be useful. These comments align to literature promoting best practice for LBP management, indicating that one of the primary factors in supporting the course of cLBP is to enhance

motivation, decrease fear avoidance, and thus improve adherence in order to generate positive outcomes [11,12].

Conversely, descriptive language regarding the same **six** themes as described by health professionals, depicts language that is skewed towards data presentation and key indices of disease-specific progress (e.g. physical activity, measure, monitor, angles, habits, questionnaires, insight, and pre-screening), which has also been previously reported [24]. Whilst this is not suggestive of a disconnect, it does suggest that clinicians and patients do have differing needs and perceptions regarding the utilitiy of informatics in supporting outcomes. Hence, it is recommended that these subtle nuances are taken into consideration when considering technology to support care.

Study Limitations

As seen in the demographics of study participants, the insights generated in this study represent opinions from a range of clinical health professionals. **Eighty percent** of data represents a cohort aged between 30-39, which suggests that data is not to be generalised to a broader age range. This is similarly cautioned regarding the sample size, as well as themes generated not being an exhaustive list**; h**owever, several methodological approaches were included in study design to support and control for these biases. For example, a) the study's clinical cohort represented clincians working in a variety of settings, with a range of years of experience, working in several speciality areas (i.e. pain management, general medicine, physiothera**p**y, occupational therapy, psychology, etc). b) individuals living with cLBP were screened for inclusion using the widely used and validated Oswestry Disability Index, in order to be more confident participants lived with a moderate degree of LBP [26]. c) The entire scoping of this research project **was** guided and underpinned by the validated and published informatics roadmap for digital health development, the CeHRES Roadmap [20]. Finally, d) To the investigators' knowledge, this is one of the first studies to utilise, adapt**,** and appropriate the recently published rigorous informatics methodology published by Greenhalgh et al**.,** NASSS [22]. The authors believe that by considering the 7 domains of NASSS, the relevant insights obtained from the data are in depth and well considered.

Conclusions

This work adds to the informatics community in several ways. A) It provides unique insights into person-centered considerations for developing technology to support LBP, b) it offers a novel appropriation of the well-regarded NASSS framework, and c) suggests that multiple facets of informatics can come together under a single model to conduct research that has the potential to improve healthcare: i.e. user-centered design, evidence-based practice, patient-reported outcomes, and informatics research methods.

The present study provides preliminary evidence of what clinicians and patients perceive to be central considerations for developing digital technology to support LBP. Whilst the study cautions against wider generalizability outside of the present conditions, its findings are underpinned by well-regarded and validated informatics methodologies and offers a novel approach to considering technologies to improve patient-reported outcomes in a participatory health paradigm.

Future research is planned and will progress to a larger project aimed at designing and prototyping technology in this domain.

Acknowledgements

The authors duly acknowledge participants who provided data, and Kieser Australia for supporting recruitment of particpants.

References

[1] R. Buchbinder, F.M. Blyth, L.M. March, P. Brooks, A.D. Woolf, and D.G. Hoy, Placing the global burden of low back pain in context, Best Pract Res Clin Rheumatol 27 (2013), 575-589.

[2] T. Vos, C. Allen, M. Arora, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015; A systematic analysis for the Global Burden of Disease Study 2015, Lancet 388 (2016), 1545-1602.

[3] M.J. Cousins, Unrelieved pain: A major health care priority, Med J Aust 196 (2012), 373-374.

[4] J. Hartvigsen, M.J. Hancock, A. Kongsted, Q. Louw, M.L. Ferreira, S. Genevay, D. Hoy, J. Karppinen, G. Pransky, J. Sieper, R.J. Smeets, and M. Underwood, What low back pain is and why we need to pay attention, Lancet 391 (2018), 2356-2367.

[5] M.N. Hogg, S. Gibson, A. Helou, J. DeGabriele, and M.J. Farrell, Waiting in pain: A systematic investigation into the provision of persistent pain services in Australia, Med J Aust 196 (2012), 386-390.

[6] B.W. Koes, M. van Tulder, C.-W.C. Lin, L.G. Macedo, J. McAuley, and C. Maher, An updated overview of clinical guidelines for the management of non-specific low back pain in primary care, Eur Spine J 19 (2010), 2075-2094.

[7] M.M. Mittinty, D.S. Brennan, C.L. Randall, D.W. McNeil, M.N. Mittinty, and L. Jamieson, Influence of fear of pain and coping strategies on health-related quality of life and patient-anticipated outcomes in patients with chronic pain: Cross-sectional study protocol, JMIR Res Protoc 6 (2017), e176.

[8] A. Ramond, C. Bouton, I. Richard, Y. Roquelaure, C. Baufreton, E. Legrand, and J.F. Huez, Psychosocial risk factors for chronic low back pain in primary care--a systematic review, Fam Pract 28 (2011), 12-21.

9] F.M. Blyth, L.M. March, A.J.M. Brnabic, L.R. Jorm, M. Williamson, and M.J. Cousins, Chronic pain in Australia: A prevalence study, Pain 89 (2001), 127-134.

[10] N.E. Foster, J.R. Anema, D. Cherkin, R. Chou, S.P. Cohen, D.P. Gross, P.H. Ferreira, J.M. Fritz, B.W. Koes, W. Peul, J.A. Turner, and C.G. Maher, Prevention and treatment of low back pain: Evidence, challenges, and promising directions, Lancet 391 (2018), 2368-2383.

[11] A.B. Irvine, H. Russell, M. Manocchia, D.E. Mino, T. Cox Glassen, R. Morgan, J.M. Gau, A.J. Birney, and D.V. Ary, Mobile-Web app to self-manage low back pain: randomized controlled trial, J Med Internet Res 17 (2015)

[12] Q.A. Rahman, T. Janmohamed, M. Pirbaglou, P. Ritvo, J.M. Heffernan, H. Clarke, and J. Katz, Patterns of user engagement with the mobile app, manage my pain: Results of a data mining investigation, JMIR Mhealth Uhealth 5 (2017), e96.

[13] H. Slater, J.E. Jordan, J. Chua, R. Schütze, J.D. Wark, and A.M. Briggs, Young people's experiences of persistent musculoskeletal pain, needs, gaps and perceptions about the role of digital technologies to support their co-care: A qualitative study, BMJ Open 6 (2016), e014007.

[14] L. Yardley, L. Morrison, K. Bradbury, and I. Muller, The person-based approach to intervention development: Application to digital health-related behavior change interventions, J Med Internet Res 17 (2015), e30.

[15] J. Barlow, C. Wright, J. Sheasby, A. Turner, and J. Hainsworth, Self-management approaches for people with chronic conditions: A review, Patient Educ Couns 48 (2002), 177-187.
[16] A.J. Lentferink, H.K. Oldenhuis, M. de Groot, L. Polstra, H. Velthuijsen, and J.E. van Gemert-Pijnen, Key components in eHealth interventions combining self-tracking and persuasive eCoaching to promote a healthier lifestyle: A scoping review, J Med Internet Res 19 (2017), e277.
[17] V.C. Oliveira, P.H. Ferreira, C.G. Maher, R.Z. Pinto, K.M. Refshauge, and M.L. Ferreira, Effectiveness of self-management of low back pain: Systematic review with meta-analysis, Arthritis Care Res (Hoboken) 64 (2012), 1739-1748.
[18] S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, A review of wearable sensors and systems with application in rehabilitation, J Neuroeng Rehabil 9 (2012), 21.
[19] M. Fledderus, K.M. Schreurs, E.T. Bohlmeijer, and M.M. Vollenbroek-Hutten, Development and pilot evaluation of an online relapse-prevention program based on acceptance and commitment therapy for chronic pain patients, JMIR Human Factors 2 (2015), e1.
[20] J.E. van Gemert-Pijnen, N. Nijland, M. van Limburg, H.C. Ossebaard, S.M. Kelders, G. Eysenbach, and E.R. Seydel, A holistic framework to improve the uptake and impact of eHealth technologies, J Med Internet Res 13 (2011), e111.
[21] L. Van Velsen, J. Wentzel, and J.E. Van Gemert-Pijnen, Designing eHealth that matters via a multidisciplinary requirements development approach, JMIR Res Protoc 2 (2013), e21.
[22] T. Greenhalgh, J. Wherton, C. Papoutsi, J. Lynch, G. Hughes, C. A'Court, S. Hinder, N. Fahy, R. Procter, and S. Shaw, Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies, J Med Internet Res 19 (2017), e367.
[23] R. Anderson, Thematic Content Analysis (TCA): Descriptive Presentation of Qualitative Data, in, Institute of Transpersonal Psychology, Palo Alto, CA, 2007, p. 4.
[24] M. Takemoto, B. Lewars, S. Hurst, K. Crist, C. Nebeker, H. Madanat, J. Nichols, D.E. Rosenberg, and J. Kerr, Participants' perceptions on the use of wearable devices to reduce sitting time: Qualitative analysis, JMIR Mhealth Uhealth 6 (2018), e73.
[25] F. Babatunde, J. MacDermid, and N.J.B.H.S.R. MacIntyre, Characteristics of therapeutic alliance in musculoskeletal physiotherapy and occupational therapy practice: A scoping review of the literature, BMC Health Serv Res 17 (2017), 375.
[26] R. Chou and P. Shekelle, Will this patient develop persistent disabling low back pain?, JAMA 303 (2010), 1295-1302.

Address for correspondence

Dr. Mark Merolli. School of Health Sciences, Swinburne University of Technology, Melbourne, Australia
Email: mmerolli@swin.edu.au

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190435

Understanding Perceptions and Attitudes in Breast Cancer Discussions on Twitter

François Modave[a*], Yunpeng Zhao[b*], Janice Krieger[c], Zhe He[d], Yi Guo[b], Jinhai Huo[e], Mattia Prosperi[f], Jiang Bian[b]

[a] *Department of Medicine, Center for Health Outcomes and Informatics Research Loyola University Chicago, Maywood, IL, USA,*
[b] *Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, FL, USA,*
[c] *Department of Advertising, University of Florida, Gainesville, FL, USA,*
[d] *School of Information, Florida State University, Tallahassee, Florida, USA.*
[e] *Departiment of Health Services Research, Management and Policy, University of Florida, Gainesville, Florida, USA*
[f] *Departiment of Epidemiology, University of Florida, Gainesville, Florida, USA*
[*] *Equal contribution, co-first*

Abstract

Among American women, the rate of breast cancer is only second to lung cancer. An estimated 12.4% women will develop breast cancer over the course of their lifetime. The widespread use of social media across the socio-economic spectrum offers unparalleled ways to facilitate information sharing, in particular as it pertains to health. Social media is also used by many healthcare stakeholders, ranging from government agencies to healthcare industry, to disseminate health information and to engage patients. The purpose of this study is to investigate people's perceptions and attitudes related to breast cancer, especially those that are related to physical activities, on Twitter. To achieve this, we first identified and collected tweets related to breast cancer; and then used topic modeling and sentiment analysis techniques to understand discussion themes and quantify Twitter users' perceptions and emotions with respect tobreast cancer to answer 5 research questions.

Keywords:

Analysis, Breast Cancer, Social Media

Introduction

A report from the National Cancer Institute (NCI) indicates that one in eight women will develop breast cancer during the course of her lifetime [1]. An estimated 266,120 new cases of invasive breast cancers, and 63,960 non invasive breast cancers will be diagnosed in 2018 in the U.S. Among American women, breast cancer remains the second most diagnosed cancer, just behind lung cancer [2]. Nevertheless, a recent study shows that physically active women have a lower risk of breast cancer than inactive women [3]. Furthermore, for breast cancer surivivors, physical activities (PAs) have benefits on their mental health, physical conditions, and movement, which ultimatedly improve patients' quality of life [4].

Access to care, access to adequate health information, and health literacy largely remain to be significant issues especially in disenfranchised populations and minorities [5], despite improvements following the implementation of the Affordable Care Act (ACA). However, the widespread availability and uptake of Internet technologies across the socio-economic spectrum has the potential to facilitate health information sharing. Specifically, nearly 9 in 10 Americans have access to high-speed Internet and 7 in 10 use at least one social media platform [6]. Not only are private citizens widely using social media resources, but the various healthcare stakeholders, e.g. industry, governmental agencies, healthcare professionals, have been increasingly using these online platforms to disseminate health information, engage patients, and recruit for clinical trials. Social media platforms provide unique sources of essentially, an endless data stream, voluntarily shared by their users. These user-generated data provide unique insights into public health; and if mined appropriately, these data are invaluable for understanding various social and health issues.

Twitter is a particularly relevant and effective data source to understand how users' perceptions and attitudes towards health-related issues change over time. In our previous work, we used Lynch syndrome as a case study to show that Twitter can be used effectively to explore discussion topics, and how promotional information can impact laypeople's discussions [7]. In this paper, we describe our Twitter analysis pipeline, as it pertains to users' general perceptions and attitudes towards breast cancer and more specifically whether and how PAs were discussed in these breast cancer-related tweets. We specifically addressed the following five research questions (RQs):

- RQ1: How do people's attitudes (i.e., emotions) towards breast cancer change over time?
- RQ2: How do people's attitudes towards breast cancer differ across geospatial regions?
- RQ3: Can we identify latent topics/themes and topic trends in breast cancer-related tweets?
- RQ4: How does promotional information impact laypeople discussion themes over time?
- RQ5: How physical activities were discussed in laypeople's breast cancer-related tweets?

Methods

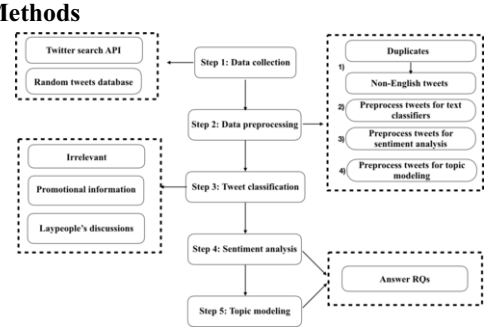


Figure 1. The Workflow of Our Twitter Analysis Pipeline

Our approach started with collecting tweets that are relevant to breast cancer discussions. We then classified these tweets into three groups (i.e., irrelevant, promotional information, and laypeople's discussions), assessed laypeople's attitudes (i.e., emotional states) using sentiment analysis, and explored latent themes using a topic modeling approach on both promotional and laypeople's tweets. Finally, based on the sentiment analysis and topic modeling results, we addressed the 5 RQs. Figure 1 illustrates our analysis workflow in 5 steps.

Step 1: Data collection. The data used in this study were from two different sources: 1) we collected breast cancer tweets from May 27, 2018 to October 13, 2018 (139 days) using a Twitter crawler based on a set of keywords related to breast cancer (e.g., "*breast cancer*" and "*#BreastCancerFighter*"). The keywords were generated through a snowball sampling process, where we started with seed keywords (e.g., "*#BreastScreening*" and "*#lumpectomy*"), then searched on Twitter with these keywords to retrieve a sample of tweets, evaluated the relevance of each tweet, and identified new relevant keywords. We did this process iterately unitl no new keywords were found; and 2) we used the keywords developed above to identify related tweets on a database of public random tweets, which we collected using the Twitter steaming application programming interface (API) from January 1, 2013 to December 30, 2017.

Step 2: Data Preprocessing. We preprocessed the collected data to eliminate tweets that were 1) duplicates across the two sources, or 2) not in English.

To develop tweet classification models (see Step 3), we preprocessed the tweets following the steps used by Glove [8]: 1) replaced hyperlinks (e.g., "http://t.co/xQgeMny5") with "<url>", 2) replaced mentions (e.g., @Channel9") with "<user>", 3) replaced hashtags (e.g.,"#breastcancer") with "<hastag>", and 4) all emojis were replaced with "<emojies>".

For sentiment analysis, we preprocessed the data with the following steps: 1) removed hyperlinks, 2) removed mentions, 3) converted hashtags into original English words (e.g., converted "#breastcancer" to "breastcancer"), 3) removed all emojies, and 4) geocoded each tweets with a geocoding tool we developed previously [9]. For topic modeling, we lemmatized each word and removed stop words (e.g., "this" and "is").

Step 3: Tweet Classification. Even though a tweet contains keywords related to breast cancer, the tweet may not be relevant to the breast cancer discussion (e.g., "*I am going to write #fiction about a 40+ year old mom with #oneboob who finds love and life's meanings who wants to read*"). Thus, we developed a two-step process with two classification models to categorize the massive amount of tweets into 3 groups (i.e., irrelevant, promotional, and laypeople's discussions).

We first annotated 1,774 tweets randomly selected from the overall dataset to create a training set. We then experimented with two different deep learning algorithms: convolutional neural networks (CNN) and long short-term memory (LSTM). We implemented both the CNN- and LSTM-based models in Keras on top of the Tensorflow framework.

A common strategy for building deep learning sentence classifiers is to use word embeddings to transform raw texts into vectors of real numbers as features. Thus, we initialized the embedding layer with a pretrained Twitter word embeddings (i.e., 100 dimension) from GloVe. The same feature matrics were used in the two-step classification process: one that categorized tweets into relevant vs. irrelevant, and another one that further categorized the relevant tweets into promotional information vs. laypeople's discussions. Models whith the best performance were adopted to classify the rest of the tweets.

Step 4: Sentiment analysis. The Linguistic Inquiry and Word Count (LIWC) is a text analysis tool, which can assess indivdiuals' attitudes through counting the percentage of emotional words used in a given text. LIWC has been used widely and its validity and reliability were validated [10]. We used LIWC on all laypeople's discussions to assess their attitudes/emotions in 5 aspects (i.e., positive emotion, negative emotion, anxiety, anger, and sadness).

Step 5: Topic modeling. Topic modeling is a statistical, unsupervised approach that can discover abstract themes in a collection of documents. We used the Biterm algorithm to find the main topics in all relevant tweets (i.e., combined both promotional information and laypeople's discussions). Different from conventional topic modeling approaches (e.g., latent Dirichlet allocation) that are based on word-document co-occurrences, Biterm learns topics by modeling word-word co-occurrences patterns, which performs better on short texts [11]. Although topic modeling is a unsupervised method, the number of topics is a parameter that needs to be determined *a priori*. Based on our previous work [7], to capture as many topics as possible, we set the number of topics as 100, visualized the topics in wordclouds, and then manually evaluated each topic's quality and merged topics with similar themes.

To answer our RQs, we also need to know the topic of each tweet. The Biterm model can infer the topic of a given tweet and return a list of topics and associated topic probabilities. We extracted the topic with the highest probability for each tweet.

Results

Data Collection

Our data came from two different sources as shown in Table 1. First, we collected 1,672,178 tweets using 32 breast cancer-related keywords and the Twitter search API from May 27, 2018 to October 13, 2018. After filtering out duplicates and non-English tweets, 1,467,783 tweets were left. Second, we used the same list of keywords to identify relevant tweets from a database of random public tweets we collected from January 1, 2013 to December 30, 2017. We found 257,045 tweets from this database, within which 167,205 tweets were written in English. Due to the different mechanisms behind the two Twitter APIs (i.e., streaming vs. search), the volume of the tweets from the two data sources were significantly different. For Twitter search API, users can retrieve almost all public tweets related to the provided keywords within 10 to 14 days (i.e., the exact time range is not published by Twitter); while the Twitter streaming API returns a random sample (i.e., roughly 1% to 20% varying acros the years) of all public tweets at the time and covers a wide range of topics. The number of tweets related to breast cancer in the random sample database was expected to be low. We plot the trends of the tweet volumes for the two sources separately as shown in Figure 2.

Table 1. Descriptive Statistics of the Two Data Sources

Data source	Data time range	# of tweets before pre-processing	# of English tweets	# of geotagged tweets
Twitter search API	05/27/2018-10/13/2018	1,672,178	1,467,783	428,041
Random public tweets	01/01/2013-12/30/2017	257,045	167,205	61,273

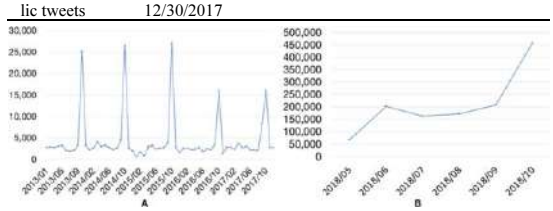


Figure 2. Breast Cancer Tweet Volume Distributions Across Time (A: random public tweets; and B: Twitter search API).

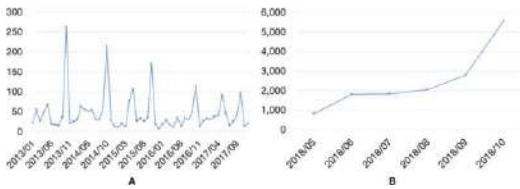


Figure 3. The Volume Trends of Tweets Related to Physical Activity in Laypeople's Breast Cancer Discussions (A: random public tweets; B: Twitter search API)

After integrating and eliminating duplicates and non-English tweets from the two sources, there were 1,634,988 unique tweets, in which 489,314 tweets can be geotagged to a US state.

To identify tweets related to PA from breast cancer tweets, a list of PA keywords (n=133) were developed through a snowball sampling process. Figure 3 shows the volume trends of tweets related to PA in laypeople's breast cancer discussions.

Text Classification

We explored two deep learning classifiers to category the tweets. Both CNN and LSTM have been wildly used in text classifications and achieved state-of-the-art performance. Table 2 shows the performance of the different classifiers and tasks. We used 80% of the annotated data for training and the performance was measured on the remaining 20% as an independent test data. As shown in Table 2, the CNN models outperformed the LSTM models in both tasks (i.e., 1) relevant vs. irrelevant; and 2) promotional information vs. laypeople's discussions). Thus, we adopted the CNN models as the final classifier.

Table 2. A Comparison of Classifier Performance

Classifier	Precision	Recall	F-score
Task 1: Relevant vs. Irrelevant			
CNN	0.886	0.851	0.865
LSTM	0.847	0.797	0.814
Task 2: Promotional information vs. Laypeople's discussions			
CNN	0.943	0.937	0.941
LSTM	0.914	0.898	0.903

The CNN models identified 1,466,292 relevant tweets (out of 1,634,988 breast cancer related tweets). Out of the 1,466,292 relevant tweets, 961,110 are tweets with promotional information and 505,182 tweets are laypeople's discussions.

Sentiment analysis

To answer RQ1 (i.e., "*How do people's emotions towards breast cancer change over time?*"), we assessed 5 emotion aspects (i.e., positive emotion, negative emotion, anxiety, anger, and sadness) of laypeople's breast cancer discussion tweets using LIWC. We visualized the changes of their emotion scores across time as shown in Figure 4. The Y-axis is the emotion scores generated by the LIWC. Even though there were fluctuations, the overall trends of negative emotion, anxiety, anger, and sadness have been decreasing since 2013 with significant Mann-Kendall test scores ($P_{trend} < .05$) as shown in Table 3.

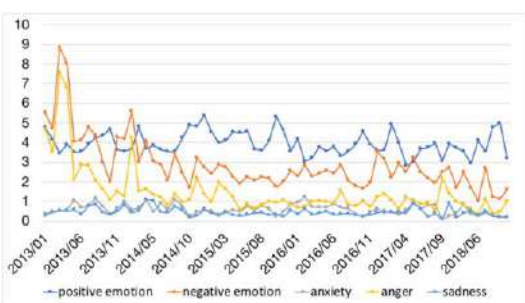


Figure 4. Laypeople's Emotion Changes by Time

Table 3. Mann-Kendall Tests of Laypeople's Emotion Changes

Emotion	P-value	Test-score	Trend
positive emotion	0.17	-1.37	not significant
negative emotion	<0.01	-5.45	decreasing
anxiety	0.01	-2.53	decreasing
anger	<0.01	-5.44	decreasing
sadness	0.01	-2.59	decreasing

We also compared the emotion scores of laypeople's breast cancer discussions across states, as heatmaps in Figure 5. The warmer the color, the higher the emotion score. People in Mississippi had the highest negative emotion and anger when they discussed breast cancer on Twitter. Delaware had the highest positive emotion. Hawaii had the highest sadness. Washington D.C. had the highest anxiety.

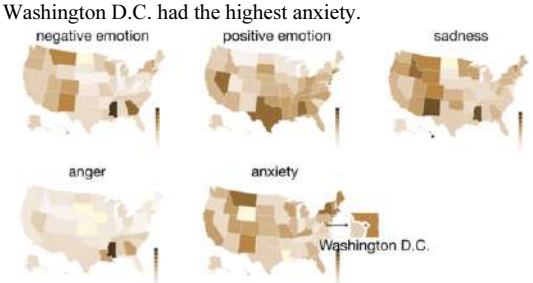


Figure 5. Comparsion of Laypeople's Emotions Towards Breast Cancer Issues Across States

We also analyzed laypeople's emotion changes over time by state. California (n=30,404) and Florida (n=17,983) had the highest number of breast cancer-related laypeople discussions, which gave us sufficient sample sizes to detect the trends. The trends of laypeople's emotion changes for these two states were decreasing as shown in Figure 6 and Table 4.

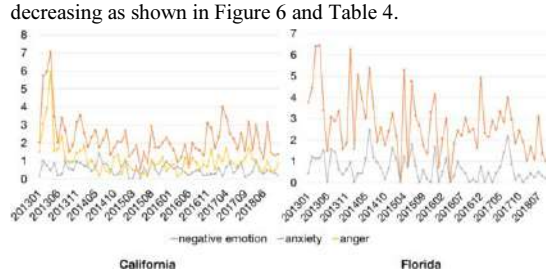


Figure 6. Laypeople's Emotion Changes by Time.

Table 4. Mann-Kendall Test o Laypeople's Discussions from California and Florida

State	Emotion	P-value	Test-score	Trend
California	negative emotion	0.01	-2.47	decreasing
	anxiety	0.02	-2.30	decreasing
	anger	0.03	-2.16	decreasing
Florida	negative emotion	<0.01	-3.09	decreasing
	anxiety	0.01	-2.60	decreasing

Topic modeling

To answer RQ 2-4, we used Biterm to discover the latent topics in our data. Based on experience from our previous work [7], we first set the number of topics to 100 to extract as many topics as possible. We then manually reviewed the 100 topics and a sample of associated tweets to assess topic quality and merge topics with similar themes. We summarized the 100 topics into 12 topics. The results are visualized as wordclouds in Figure 7.

The topic distributions of laypeople's discussions and promotional information are shown in Figure 8. Laypeople discussed more on the topics of "*family*", "*friend*", "*diagnosis*", "*food*", and "*treatment*"; while they talked less about "*risk*" and "*mortality*" compared with topics in promotional information.

Figure 7. 12 Topics Summarized from the 100 Topics Learned with a Biterm Topic Model.

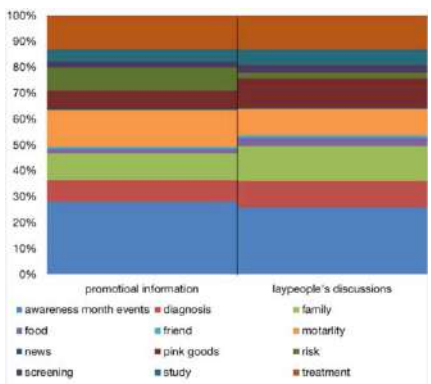


Figure 8. Topic Distributions by Tweet Type (promotional information vs. laypeople's discussions)

To exam the relations between promotional information and laypeople's discussions by topic, we used the random tweet dataset (spanning from 2013 to 2017) and calculated the Pearson correlation coefficient between the two (promotional vs laypeople) based on their monthly tweet volumes for each topic. As shown in Table 5, laypeople's discussions had strong correlations with promotional information on "*awareness month events*", "*risk*", and "*treatment*" and moderate correlations on "*diagnosis*", "*family*", "*friend*", "*news*", "*pink goods*", "*screening*", and "*study*". The monthly trends from 2013 to 2017 for these topics are shown in Figure 9.

Table 5. Correlations Between Promotional Information and Laypeople's Discussions Based on Tweet Volumes by Topic.

Topic	Correlation coefficient	P-value
awareness month events	0.711	<0.001
diagnosis	0.635	<0.001
family	0.682	<0.001
friend	0.526	<0.001
news	0.642	<0.001
pink goods	0.557	<0.001
risk	0.767	<0.001
screening	0.690	<0.001
study	0.683	<0.001
treatment	0.788	<0.001

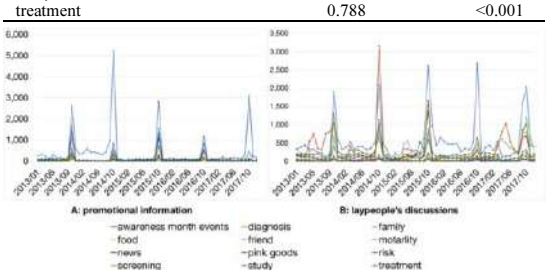


Figure 9. Selected Topic Trends from 2013 to 2017 with Significant Correlations Between Promotional Information and Laypeople's Discussions.

To answer RQ5, we first extracted all laypeople's breast cancer discussions that were aslo related to physical activites (PAs, i.e., discussions that contain PA-related keywords). We then extracted 100 topics using the Biterm method on these PA-related dicssuions. Similar topics were merged into one; and two themes emerged: give support (i.e., give support to breast cancer awareness through sporting events) and reduce risk (i.e., raise awareness that PAs can reduce breast cancer risks). The wordclouds of the two themes are shown in Figure 10; and a few example tweets for each theme are shown in Table 5.

Figure 10. Two Main Themes in Laypeople Tweets Related to Both Breast Cancer and Physical Activities

Table 5. Example of Physical Activity Related Tweets by Theme

Themes	Example tweets
A: give support to breast cancer awareness through sporting events	• *"I don't understand why it happened to me, but hope because I got cancer I can help bring about change." #BreastCancer survivor Shannon O'Fallon supports research with #Obliteride's 5K walk & urges others to help #curecancerfaster."* • *"On May 14th of this year I lost my older sister Ashley to breast cancer. On August 11th my family will walk in her honor"*
B: physical activities can reduce the risk of breast cancer	• *"Important to think about HOW we #exercise, #breastcancer incidence and relapse risk can be reduced by physical activity https://t.co/ajfew2wY5D"* • *"@RepDavidYoung I ran 91 miles in June. One of the reasons I run is bc it reduces my risk of recurrence of my breast cancer. O"*

Discussion and Conclusions

The goals of our study were to examine breast cancer-related discussions on Twitter and to understand people's percetions and attitudes towards breast cancer through their Twitter posts. We were also interested in assessing how promotional information impacts laypeople's discussions on Twitter. Thus, we used well-established text analysis approaches (i.e., sentiment analysis and topic modeling) on breast cancer-related tweets to answer our five RQs.

As shown in our results (Figure 4), laypeople's attitudes towards breast cancer changed from time to time. The overall trends of negative affects (i.e., negative emotion, anger, anxiety, and sadness) have been decreasing since 2013. There have been some controversial issues being discussed on public news outlets related to breast cancer prevention and treatments, which might lead people to think negatively. Taking the discussions on mammography as an example, mammography is a common way to screen for breast cancer as an early detection method. However, many people on Twitter questioned the effectiveness of mammography and raised concerns that it might bring overdiagnoses and overtreatments to patients. Nevertheless, as the negative attitudes are decreasing, it might indicate that stakeholders such as health organizations and agencies are doing a better job educating the public.

We also found that emotions in laypeople's breast cancer discussions on Twitter differ across the states. As shown in Figure 5, people in Mississippi had significantly higher negative emotion and anger when they discussed breast cancer-related issues on Twitter. Neverthless, further investigations are needed to find the reasons behind these state differences. For example, people's attitudes towards breast cancer maybe a reflection of breast cancer screening rates. Mississippi is one of the states with the lowest breast cancer screening rate possibly due to barriers for people to access screening programs. These barriers may lead to negative emotions

expressed in tweets (e.g., "*To cure a patient's disease at the cost of financial ruin falls short of our duty as physicians to serve*").

Laypeople's discussions were correlated with promotional information on a number of topics as shown in Table 5. These strong correlations might, from another perspective, indicate that breast cancer-related promotion strategies to raise public awareness have been rather successful in the past few years. Further, as observed in Figure 9, both promotional information and laypeople's discussions surged in October (i.e., breast cancer awareness month) every year, which suggested that promotional events in media (including social media) are effective ways to gain participants and raise public's awareness. Such raises in awareness would ultimately lead to improved health outcomes. For example, in early 1987, when the American Cancer Society started to focused on raising breast cancer awareness and before breast cancer became an official National Health Observances (NHO) event, only 26% of women in the U.S. had undergone a mammogram in the previous 12 months; while by October of the same year, the proportion had raised to 38% [12].

Over half of the women diagnosed with breast cancer gained weight during treatment with multiple reasons [13]. Chemotherapy often leads to fluid retention, reduced PA levels (due to pain and fatigue), decreased metabolism, and food cravings (that can reduce nausea). Nevertheless, breast cancer survivors should engage in a weight management program focusing on dietary intake and PA even during treatment, if manageable. There is clear evidence that weight management including PAs have a positive impact on mental health during and after cancer treatment [14]. As shown in Figure 10, we obtained two general themes from laypeople's breast cancer discussions that are also related to PAs: 1) people like to give support (to breast cancer awareness through sporting events), and 2) people are aware of the benefits of PAs that can reduce breast cancer risks. When we looked into these tweets in more details, we also found tweets from breast cancer survivors (e.g., "*@raceforlife Fully intended to do this (for the 5th year running; pardon the pun!!) but breast cancer diagnosis and treatment have run off with my fitness. Next year though!*"), which provides evidence data that social media might be a good source of information to study behavioral factors such as PAs related to improving surviorships including their quality of life (QoS). For example, we can further understand the barriers to adopting an exercise program among breast cancer survivors in our future work using social media data sources.

In sum, our study demonstrated that 1) social media such as Twitter are invaluable sources to provide insights into public and consumer health; and 2) natural language processing combined with machine learning are effective tools and methods to assess laypeople's attitudes changes in their health-related tweets, discover laypeople's perceptions towards specific health topics, and understand how promotional informaiton has an impact on laypeople's discussions.

Acknowledgements

The study was supported by the University of Florida Health Cancer Center (UFHCC) Research Pilot Grant and in part by NSF Award #1734134 and NIH UL1TR001427. The content is solely the responsibility of the authors and does not necessarily represent the official views of the sponsors.

References

[1] Howlader N, Noone AM, Krapcho M, Neyman N, Aminou R, Waldron W, Altekruse SF, Kosary CL, Ruhl J, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Chen HS, Feuer EJ, and Cronin KA (eds), SEER Cancer Statistics Review, 1975-2009 - Previous Version - SEER Cancer Statistics Review, *SEER*. (n.d.). https://seer.cancer.gov/archive/csr/1975_2009_pops09/index.html.

[2] CDC, Breast Cancer Statistics, *Breast Cancer Stat.* (2018). https://www.cdc.gov/cancer/breast/statistics/index.htm (accessed March 21, 2019).

[3] Y. Wu, D. Zhang, and S. Kang, Physical activity and risk of breast cancer: a meta-analysis of prospective studies, *Breast Cancer Res. Treat.* **137** (2013) 869–882. doi:10.1007/s10549-012-2396-7.

[4] N. Mutrie, A.M. Campbell, F. Whyte, A. McConnachie, C. Emslie, L. Lee, N. Kearney, A. Walker, and D. Ritchie, Benefits of supervised group exercise programme for women being treated for early stage breast cancer: pragmatic randomised controlled trial, *BMJ.* **334** (2007) 517. doi:10.1136/bmj.39094.648553.AE.

[5] H. Levy, and A. Janke, Health Literacy and Access to Care, *J. Health Commun.* **21** (2016) 43–50. doi:10.1080/10810730.2015.1131776.

[6] Paul Hitlin, Internet, social media use and device ownership in U.S. have plateaued after years of growth, Pew Research Center, 2018.

[7] J. Bian, Y. Zhao, R.G. Salloum, Y. Guo, M. Wang, M. Prosperi, H. Zhang, X. Du, L.J. Ramirez-Diaz, Z. He, and Y. Sun, Using Social Media Data to Understand the Impact of Promotional Information on Laypeople's Discussions: A Case Study of Lynch Syndrome, *J. Med. Internet Res.* **19** (2017) e414. doi:10.2196/jmir.9266.

[8] J. Pennington, R. Socher, and C.D. Manning, GloVe: Global Vectors for Word Representation, in: Empir. Methods Nat. Lang. Process. EMNLP, 2014: pp. 1532–1543. http://www.aclweb.org/anthology/D14-1162.

[9] A. Hicks, W.R. Hogan, M. Rutherford, B. Malin, M. Xie, C. Fellbaum, Z. Yin, D. Fabbri, J. Hanna, and J. Bian, Mining Twitter as a First Step toward Assessing the Adequacy of Gender Identification Terms on Intake Forms, *AMIA Annu. Symp. Proc. AMIA Symp.* **2015** (2015) 611–620.

[10] Y.R. Tausczik, and J.W. Pennebaker, The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods, *J. Lang. Soc. Psychol.* **29** (2010) 24–54. doi:10.1177/0261927X09351676.

[11] X. Yan, J. Guo, Y. Lan, and X. Cheng, A Biterm Topic Model for Short Texts, World Wide Web Conference Committee (IW3C2), Rio de Janeiro, Brazil., 2013.

[12] G.D. Jacobsen, and K.H. Jacobsen, Health awareness campaigns and diagnosis rates: Evidence from National Breast Cancer Awareness Month, *J. Health Econ.* **30** (2011) 55–61. doi:10.1016/j.jhealeco.2010.11.005.

[13] Z. Huang, S.E. Hankinson, G.A. Colditz, M.J. Stampfer, D.J. Hunter, J.E. Manson, C.H. Hennekens, B. Rosner, F.E. Speizer, and W.C. Willett, Dual effects of weight and weight gain on breast cancer risk, *JAMA*. **278** (1997) 1407–1411.

[14] J.M. Beasley, M.L. Kwan, W.Y. Chen, E.K. Weltzien, C.H. Kroenke, W. Lu, S.J. Nechuta, L. Cadmus-Bertram, R.E. Patterson, B. Sternfeld, X.-O. Shu, J.P. Pierce, and B.J. Caan, Meeting the physical activity guidelines and survival after breast cancer: findings from the after breast cancer pooling project, *Breast Cancer Res. Treat.* **131** (2012) 637–643. doi:10.1007/s10549-011-1770-1.

Address for correspondence

Jiang Bian; Email: bianjiang@ufl.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190436

Proposal of Relevant Information Visualization for a Universal Viewer in Oncology

Mukai Masami[a], Nakajima Noriaki[a], Nakatsugawa Minoru[b], Shimomura Yuka[b], Shiokawa Yasunari[b], Mihara Naoki[a]

[a] Division of Medical Informatics, National Cancer Center Hospital, Tokyo, JAPAN,
[b] Canon Medical Systems Corporation, Otawara, Tochigi, JAPAN

Abstract

Medical information, such as physicians' descriptions, nursing records, and various examination reports, is stored separately in the subsystems of an electronic medical record (EMR). To provide efficient treatment, health professionals require a universal viewer that collectively organizes and visualizes this distributed information at each clinical phase (i.e., first outpatient visit, treatment, follow-up observation, etc.). This research investigates a method to enable physicians to easily browse relevant information at each clinical phase, and to reflect and verify information in a universal viewer. We analyzed the patterns of picture archiving and communication system (PACS) reference logs and EMR performed orders to identify relevant information. We found that by considering the locations of display terminals in addition to disease types (clinical departments) and clinical phases, more relevant information could be identified, and the contents displayed on a universal viewer were improved accordingly.

Keywords:

Computer-Assisted Therapy, Information Technology, Neoplasms

Introduction

Our institution is a specialized hospital for treatment of malignant tumors with 597 beds, approximately 2,500 outpatients per day, and 10,000 new patients per year. Physicians specialize by organ, such as head and neck, lung, gastrointestinal, urinary, breast, and so on, in addition to the divisions of internal medicine, surgery, psychiatry, and palliative care. Our hospital information system (HIS) consists of an electronic medical record (EMR) system and approximately 60 departmental systems, such as radiation, biochemical testing, physiological testing, medicine, surgery information systems, and so on.

To fully understand a patient's situation, physicians must display a plurality of screens, such as the EMR description, a record of the first visit, biochemical test results, images from a picture archiving and communication system (PACS), and a pathology report. Information must be manually organized and displayed before each instance of medical care for patients. This complicated task takes significant time away from the medical care that should be prioritized.

Therefore, we have been investigating methods to enable physicians to browse the relevant information (i.e., medical records, nursing records, examination reports, etc.) necessary for clinical practice [1-4]. Morgan showed that integrated information visualization can improve radiologists' use of clinical decision support tools [5]. Jorritsma investigated the usability of PACS by analyzing logs based on pattern mining [6]. De-Arteaga and Zeng showed the possibility of improving retrieved information by analyzing image queries [7,8]. Vest showed a difference in access frequency of radiology reports between outpatient and inpatient users [9]. Our previous research identified that clinical departments (disease types) and clinical phases (i.e., first visit, treatment, follow-up observation, etc.) are important parameters distinguishing the relevant information that should be displayed in a universal viewer [10]. This research investigates whether the information on location of display terminals (i.e., inpatient ward and outpatient clinic) from PACS logs can aid in identifying relevant content for physicians.

It was determined by the ethics committee (institutional review board: IRB) of our institution that a formal review by the IRB was not required for this research.

Methods

Data Set

PACS reference logs from April 1-30, 2018, in respiratory medicine, colorectal surgery, and breast surgery departments were analyzed. The PACS reference logs included the operating physicians' information and referred image information. The operating physicians' information included physician ID, reference date and time, physician department, and terminal IP address and location. Referred image information included patient ID, imaging date and time, modality, and region of images. Modalities included computed radiography (CR), computed tomography (CT), endoscopy (ES), mammography (MG), magnetic resonance (MR), nuclear medicine (NM), positron emission tomography (PT), radio fluoroscopy (RF), x-ray angiography (XA), and others (OT). Image regions included the chest, pelvis, abdomen, head, and breast.

Pattern Analysis of Image References

The number of image references were counted for each clinical department, clinical phase, location of image references, imaging modality, and region of images.

Clinical Phase of Image References

The clinical phases of each image reference were determined by the time of the references. Major clinical events (e.g., first visit, hospitalization, surgery, discharge) and their time were extracted from EMR performed order information. The interval between two consecutive clinical events was defined as a clinical phase (e.g., hospitalization – surgery) [10]. The clinical phase of each image reference was determined by comparing

the time of the image reference and the time of the patient's clinical phases. This research focused on the clinical phases of first visit, before hospitalization, during hospitalization, and after discharge.

Location of Image References

The location of each image reference was determined by the IP address of the display terminal. Locations were categorized into three types: outpatient clinic, inpatient ward, and others. This research compared image reference patterns in outpatient clinics and inpatient wards.

Frequency of Image References

To quantify the frequency patterns of image references by physicians, we counted the frequency of references to each image or combination of images referred to at the same time. A sequence of image references for the same patient within 30 minutes by a single physician on a single terminal was defined as *one reference*. The number of image references was then counted based on this definition.

Days Elapsed from Imaging

To investigate the time distribution of the images referred to by physicians, we measured the number of days elapsed between the imaging date and the reference date. Elapsed days were categorized as shown in Table 1. The range for each category was determined so that reference counts for each category were distributed nearly equally, and so that categories would be weekly or monthly. Figure 1 shows the number of image references in each category.

Table 1 - Categorization of elapsed days

Category	Range of elapsed days	Note
0	0	Same date
1	1 – 3	1 to 3 days ago
2	4 – 7	3 days to 1 week ago
3	8 – 14	1 to 2 weeks ago
4	15 – 21	2 to 3 weeks ago
5	22 – 30	3 weeks to 1 month ago
6	31 – 40	1 month to 40 days ago
7	41 – 60	40 days to 2 months ago
8	61 – 150	2 to 5 months ago
9	151 +	More than 5 months ago

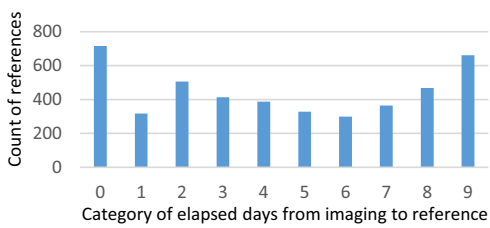


Figure 1 - Image reference counts in each category for days elapsed between imaging and reference

Combination Pattern of Referred Images

To analyze which images were referred to in combination, we analyzed combinations of images in one reference (defined above). The following three combination types were considered.

Type 1: Combination of elapsed days categories
Type 2: Combination of elapsed days categories and modalities
Type 3: Combination of elapsed days categories, modalities, and imaging regions

For example, when a physician referred to an abdomen MRI image from that same day (Category 0) and a chest CT image from more than 5 months ago (Category 9), Type 1 generated the combination (0,9), type 2 generated the combination (0_MRI, 9_CT), and type 3 generated the combination (0_MRI_ABDOMEN, 9_CT_CHEST). If only one image from the same day was referred to, Type 1 would generate (0), for example.

Comparison of Image Reference Patterns

The frequency of combination patterns of referred images was counted and compared among clinical departments, display terminal locations (outpatient clinic, inpatient ward), and clinical phases (first visit, before hospitalization, during hospitalization, after discharge). The relative frequency (ratio), as defined below, was compared.

Relative frequency (ratio) = number of image references for each combination / total number of image references

Analysis Environment

Python 3.7 was used on a computer running Windows 10 64-bit OS with 16 GB of memory.

Display of Relevant Information on a Universal Viewer

The combinations of frequently referred images were considered to be relevant information in each clinical phase and location. We developed a prototype universal viewer [11] that displays the frequent combinations of referred images at each clinical phase and location. The universal viewer displays modality images, reports, laboratory tests, image measurements, clinicians' descriptions, nursing records, and so on, switching the displayed contents to show those that are relevant at each clinical phase (Figure 2). The configuration of displayed content at each clinical phase and location was determined by the above method. The universal viewer runs on a web browser and can flexibly adjust to various sizes/resolutions of displays including multiple displays and tablet displays.

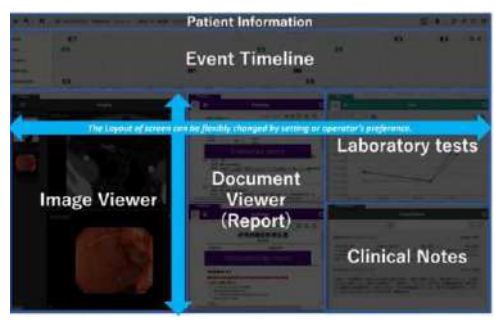


Figure 2 – Screen configuration of a universal viewer

Results

The total number of PACS image reference logs was 24,966. The total number of terminals with log records of usage was 234 (outpatient clinic 19, inpatient ward 121, others 94). Thirty-eight physicians used PACS (14 in respiratory medicine, 11 in colorectal surgery, 13 in breast surgery). The total number of

image references as defined in the Methods section was 3,331 (1,066 in respiratory medicine, 939 in colorectal surgery, 1,316 in breast surgery).

The results of pattern analysis of image references are illustrated in Figures 3-15.

At outpatient terminals, there were many cases of reference to an image from within one week alone, and to combinations of current and several months old images (Figure 3). At inpatient ward terminals, same-day images alone were most referred to (Figure 4).

Analysis of the reference information at outpatient and ward terminals with consideration of clinical phase showed the following tendency. At outpatient terminals upon first visit, there were many cases of references to images from within one week or from several months previous alone (Figure 5). At outpatient terminals after discharge, combinations of same-day images and images from 2 to 5 months ago were referred to (Figure 6). At ward terminals, the number of references was greater before hospitalization than during hospitalization. The reference pattern included images from various periods, particularly images from immediately before hospitalization (Figure 7). At ward terminals during hospitalization, same-day images alone were most referred to (Figure 8).

Adding clinical department as a parameter showed the following tendency. At outpatient terminals in respiratory medicine, same-day images alone or combinations of same-day images and images from 2 to 5 months ago were referred to most, regardless of clinical phase (Figure 9). However, images from a relatively long period were referred to equally at ward terminals (Figure 10). In colorectal surgery, recent images were referred to at outpatient terminals (Figure 11). At ward terminals, references differed before hospitalization (Figure 12) and during hospitalization (Figure 13). During hospitalization, same-day images were referred to. In breast surgery, images (mammography) from one or two weeks previous were primarily referred to at outpatient terminals upon first visit (Figure 14). Same-day images (nuclear medicine) were referred to during hospitalization (Figure 15).

Considering these results, the information display on a universal viewer was configured for each terminal location and clinical phase. Figure 16 shows a display for breast surgery at outpatient terminals upon first visit (left) and a display forward terminal before hospitalization (right). Figure 17 shows the difference in screen configuration between breast surgery (left) and colorectal surgery (right) during treatment.

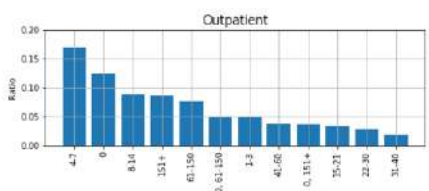


Figure 3 – Image references at outpatient terminals

Figure 4 – Image references at ward terminals

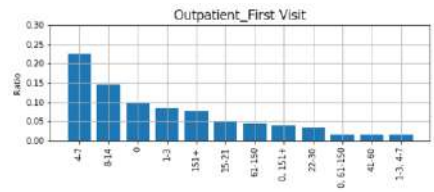


Figure 5 – References at outpatient terminals upon first visit

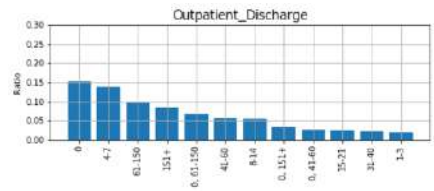


Figure 6 –References at outpatient terminals after discharge

Figure 7 – References at ward terminals before hospitalization

Figure 8 – References at ward terminals during hospitalization

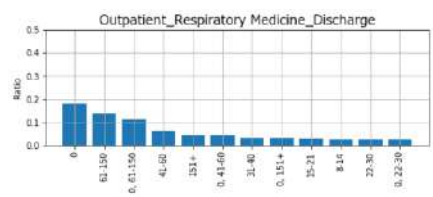


Figure 9 – References at outpatient terminals in respiratory medicine

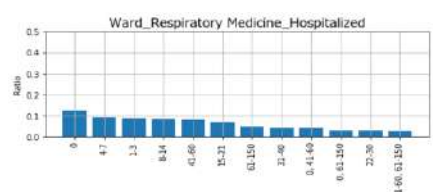


Figure 10 – References at ward terminals in respiratory medicine

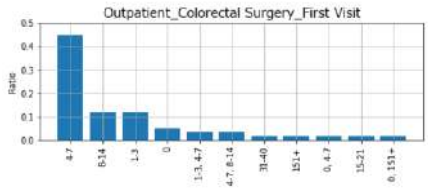


Figure 11 – References at outpatient terminals in colorectal surgery

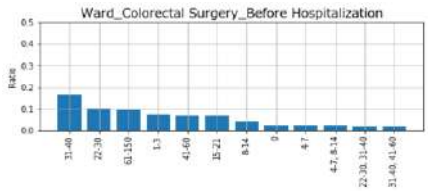


Figure 12 – References at ward terminals in colorectal surgery before hospitalization

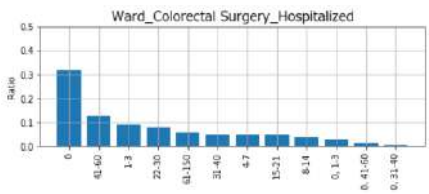


Figure 13 – References at ward terminals in colorectal surgery during hospitalization

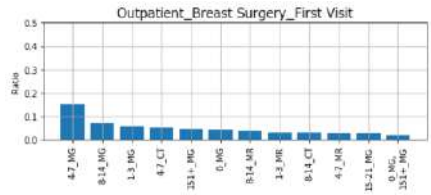


Figure 14 – References at outpatient terminals in breast surgery upon first visit

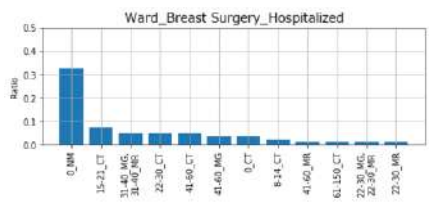


Figure 15 – References at ward terminals in breast surgery during hospitalization

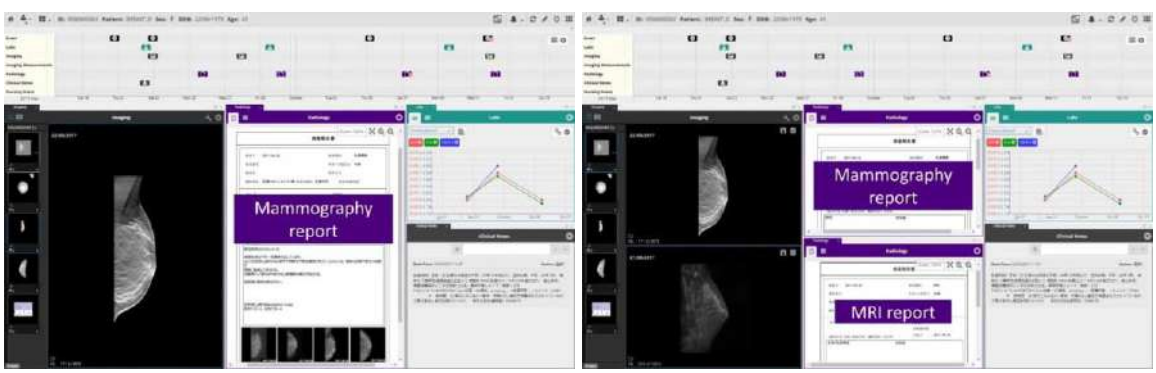


Figure 16 – Example of display configuration for a universal viewer in breast surgery (Left: outpatient clinic upon first visit. Right: inpatient ward before hospitalization)

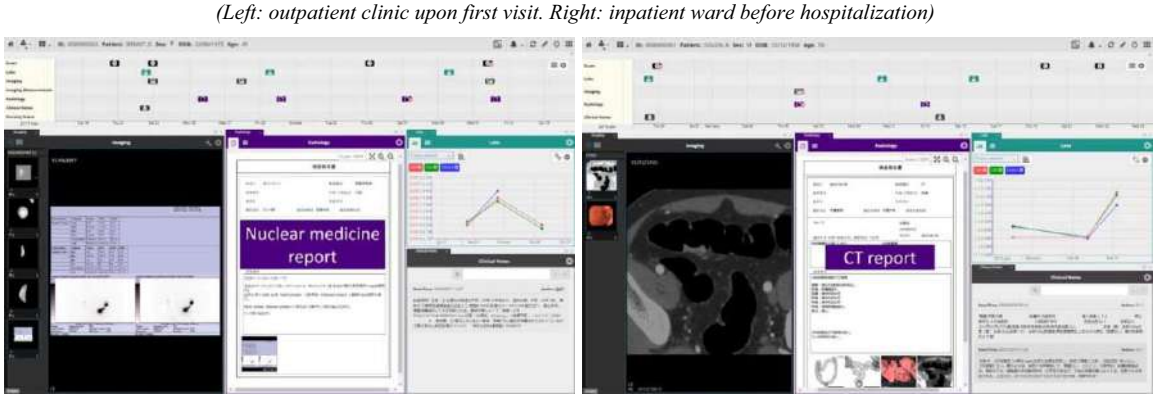


Figure 17 – Example of display configuration for a universal viewer (Left: breast surgery during treatment. Right: colorectal surgery during treatment

Discussion

In previous research, physicians' behavior was analyzed by PACS operation logs or retrieval queries mainly to improve usability of PACS or image retrieval systems [5-9]. However, this research analyzed the PACS operation logs to investigate the method to display relevant information at each clinical phase. Specifically, this research showed that identification of the clinical scene (clinical department, clinical phase and the location of display terminals) and pattern analysis of physicians' image references (imaging date and time, modality, and regions of the images) at the clinical scene allow the universal viewer to display relevant information, which is effective for clinical decision support for diagnosis and treatment.

In our hospital, almost all patients are referrals from other hospitals. Cases in which a physician refers to an image from several months before the first outpatient visit are presumed to be references to the initial examinations at the patient's original hospital to verify the patient's status at the time of the initial diagnosis.

Images from before hospitalization are referred to at inpatient wards for conferencing and consideration of treatment policy. Images are referred to during hospitalization to confirm the treatment situation and conduct necessary immediate examinations. The results of this study's analysis reflect the clinical workflow in practice.

The finding that reference tendencies differ among clinical departments also reflects differences in the requirements of each department. The treatment period is longer in respiratory medicine, and confirmation of status during the course of treatment is important. Colorectal surgery and breast surgery require examinations for recurrence and confirmation of metastasis.

References to mammography images before hospitalization and to nuclear medicine images during hospitalization in breast surgery also reflect clinical practice, considering the burden of mammography on patients and the necessity of metastasis examination.

The limitation of this research is that PACS reference logs are available at the study level, but not the series level. Display of axial images of specific series, sagittal images, and fusion images according to disease condition and reference purpose warrants further investigation

Conclusions

This research showed that the location of display terminals (outpatient clinic, inpatient ward) in combination with clinical phase and disease type (hospital department) can potentially identify relevant contents for display on a universal viewer. The results warrant further investigation for the application of the proposed method to other clinical departments, clinical scenes, and diseases, given that a universal viewer is expected to improve the efficiency of medical treatment for various diseases, not only for specific diseases such as cancer.

Acknowledgements

We would like to thank Editage (www.editage.jp) for English language editing.

References

[1] Mihara N, et.al; Collaborative development of universal viewer in medicine using VNA (Vendor Neutral Archive) technology to accelerate integrated management of clinical data, Monthly report 'Shin-Iryou' (Japanese magazine), 8-13, July, 2018.
[2] Mihara N, et.al; What is the best method of saving image for medicine ~ Report of analyzing for operational PACS log, both short term and long-term storage ~ 2nd report, 73rd annu meeting of Japan Radiol Soc, 2016.
[3] Mihara N, et.al; How should we do to realize saving and viewing image efficiently? Primitive report of analyzing log data of PACS, Annu meeting of Japan Radiol Soc, 2013.
[4] Mihara N, et.al; The prototype of the self-leaning system of the chest X-ray images, Japan J Med Inform, 31:576-79, Nov., 2011
[5] Morgan MB, et.al; Just-in-time radiologist decision support: the importance of PACS-integrated workflow. J Am Coll Radiol, 8(7):497-500, 2011.
[6] Jorritsma W, et.al; Pattern mining of user interaction logs for a post-deployment usability evaluation of a radiology PACS client. Int J Med Inform, 85(1):36-42, 2016.
[7] De-Arteaga M, et.al; Analyzing medical image search behavior: semantics and prediction of query results. J Digit Imaging, 28(5):537-46, 2015.
[8] Zeng QT, et.al; Assisting consumer health information retrieval with query recommendations. J Am Med Inform Assoc, 13(1):80-90, 2006.
[9] Vest JR, et.al; Using a health information exchange system for imaging information: patterns and predictors. AMIA Annu Symp Proc, 1402-11, 2013.
[10] Mukai, et.al; What is required for Universal Viewer - in case of Oncologist - ?, JCMI2018.
[11] Hirashima V; Development of Abierto™ Cockpit that integrates and visualizes clinical information, JIRA technical report, 28(1):28-9, 2018.

Address for Correspondence

MUKAI Masami, e-mail: mmukai@ncc.go.jp

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190437

The Handover from Intensive Care Unit to General Ward: Baseline Performance and Participatory Design of an Electronic Follow-Up Plan

Kija Lin Østergaard[a], Jesper Simonsen[a], Morten Hertzum[b]

[a] Department of People and Technology, Roskilde University, Roskilde, Denmark,
[b] Department of Information Studies, University of Copenhagen, Copenhagen, Denmark

Abstract

The transfer of patients from intensive care unit (ICU) to general ward involves risk to patient health. To mitigate this risk the present study investigates the current use of follow-up plans in the handover from ICU to general ward and proposes a novel design of follow-up plans. On the basis of a record audit we find that follow-up plans exist for only 16% of the audited transfers, that these plans are rarely used, and that 25% of the patients with a plan die within 24 hours of their transfer. In a subsequent series of participatory-design workshops with ICU and ward nurses we devised an electronic follow-up plan that consists of an attend-to list rather than a checklist. The attend-to list specifies the issues of concern but leaves the process of attaining them for the general-ward nurses to decide, thereby acknowledging and utilizing their expertise.

Keywords:
Patient transfer, Patient handoff, Electronic health records.

Introduction

Patients are habitually transferred from one clinical setting to another, such as from ambulance care to emergency department [1], from one ward to another ward within a hospital [2], and from hospital care to primary care [3]. These transfers involve risks of information loss [4], discontinuity of care [3], and adverse events [5]. To mitigate these risks, patient transfers are accompanied by handovers during which information about the patient is transferred to the receiving staff. In the present study we focus on the handover from intensive care unit (ICU) to general ward. This handover warrants attention because the patients arriving from the ICU will be among the weakest patients at the general wards and because the general wards have far fewer resources for monitoring and treating patients than the ICU.

ICUs cater to patients with severe, life-threatening conditions that require constant monitoring, advanced equipment, and specialist medication to sustain normal bodily functions [6]. As a consequence, the mortality rate is high for ICU patients. Table 1 shows the mortality rate for ICU patients in Denmark, the country in which this study is conducted. Not only do many patients die in the ICU, there are also many patients who die shortly after ICU discharge, and this pattern has been stable over the past several years. In the year July 2016 - June 2017 the 4% deaths within 48 hours of ICU discharge correspond to 986 patients. Temporally these deaths coincide closely with the transfer of the patient from ICU to general ward.

At the studied hospital the handover from ICU to general ward was facilitated by a follow-up plan made for the individual patient by the ICU nurses. The follow-up plan covered the initial 24 hours after transfer and consisted of a prioritized list of tasks for the nurses at the general ward, such as that the patient needed supplementary oxygen. For each task it was indicated when the task should be performed and whether there were special issues to consider. The day after the transfer the ICU nurse would visit the general ward to check up on whether the patient was treated according to the follow-up plan. In spite of the follow-up plans 4% of the patients died within 48 hours of their transfer, see Table 2. In addition, patient mortality within 30 days of ICU discharge was substantially above the national average (Table 1 vs. Table 2) and, thereby, called for scrutinizing follow-up practices.

Table 1 – Mortality of ICU Patients at Danish Hospitals

Mortality measure	2012-13 [a]	2016-17 [a]
In-ICU mortality	13%	11% (2958)
Within 48 hours of ICU discharge	4%	4% (986)
Within 30 days of ICU discharge	21%	21% (5966)

[a] The 12-month period July 1 - June 30. Sources: [7], [8].

Table 2 – Mortality of ICU Patients at the Studied Hospital

Mortality measure	2012-13 [a]	2016-17 [a]
In-ICU mortality	15%	14% (91)
Within 48 hours of ICU discharge	4%	4% (23)
Within 30 days of ICU discharge	27%	27% (167)

[a] The12-month period July 1 - June 30. Sources: [7], [8].

In December 2012 the studied hospital installed electronic whiteboards at all wards to support interdepartmental communication and coordination. After the whiteboard had been installed the local clinicians were encouraged to configure it for their needs [9]. The whiteboard provided an opportunity to rethink follow-up procedures and introduce electronic follow-up plans instead of the existing paper-based plans. By using the whiteboard for coordinating patient transfers the follow-up plans would become more visible and interactive.

In this paper we report from our yearlong collaboration with the hospital to investigate existing follow-up practices and design electronic follow-up plans. While the former was achieved through a record audit, the latter involved a process of participatory design. We specifically ask:

- To what extent are follow-up plans made and used when ICU patients are transferred to a general ward?

- What should an electronic follow-up plan look like to support the transfer better than the current paper plan?

Systematic reviews of handover effectiveness recommend the use of electronic tools and of forms that systematize the information to be transferred, but they also report limited effects on patient mortality [4; 10]. In addition, studies of the nursing of former ICU patients show that general wards tend to feel overwhelmed by unrealistic demands [11; 12].

Methods

This study was conducted at a hospital in Region Zealand, one of the five healthcare regions in Denmark. The hospital had 250 beds and about 35,000 annual admissions. Before the study started it was approved by hospital management. The participants individually consented to take part.

Record Audit

To determine the extent to which follow-up plans were made and used we audited the records of all patients admitted to the ICU in 2012. First, hospital records were consulted to extract the patients who met our inclusion criteria, then local ICU records were consulted to obtain these patients' follow-up information. The inclusion criteria mirrored the criteria at the ICU for when follow-up should be performed. Follow-up could be initiated for additional patients at the ICU nurses' discretion, but we restricted our study to the patients who met the general criteria for follow-up. That is, we applied these five *inclusion criteria*:

- Patients >18 years of age, because people under the age of 18 were transferred to other hospitals if they needed intensive-care treatment.
- Mechanically ventilated >24 hours (invasively or non-invasively), because the complications associated with ventilation necessitated close monitoring of the patient after ICU discharge.
- ICU admission >72 hours, because the complications following ICU admission increased with its length.
- Unplanned transfers to a general ward in the evening or night (16:00-07:00), because the general wards had fewer staff resources during evenings and nights and, therefore, reduced capacity for unplanned tasks.
- Transfers to a general ward at the hospital, because follow-up was restricted to the hospital and, thus, did not cover patients transferred to other hospitals.

To avoid skewing the results of the audit we *excluded* patients transferred to the general ward for palliative care, because these patients had a high mortality rate irrespective of whether they were transferred from the ICU or from another ward. On the basis of these inclusion and exclusion criteria a total of 304 of the 946 patients admitted to the ICU in 2012 were included in the audit. The included patients were transferred from the ICU to three general wards: the medical ward for pulmonary diseases, the orthopedic ward, and the surgical ward. For each of the 304 included patients we inspected the records for information about whether a follow-up plan existed, whether it had been read at the general ward, whether ICU staff had made a follow-up visit to the general ward, whether the follow-up plan had been completed, and whether the patient died within 24 hours of the transfer to the general ward.

Participatory-Design Workshops

We applied a participatory-design approach [13] to investigate how an electronic follow-up plan could be designed to support the patient transfers. Following this approach, we aimed at engaging the nurses from the involved departments in mutual learning processes to investigate the realities and challenges of the transfer situation. By iteratively articulating their needs and discussing how these needs might be supported the nurses arrived at a design proposal for a follow-up plan integrated in the electronic whiteboard.

The participatory-design process involved the ICU and the three same general wards as in the record audit: the medical ward for pulmonary diseases, the orthopedic ward, and the surgical ward. To ensure the representation of all nursing staff from these four wards, 24 participatory-design workshops were conducted, each lasting 1-1.5 hours. The workshops were conducted by the first author and spread across day, evening, and night shifts. Each workshop was attended by 1-8 nurses, who were at the same time on call in case the busy clinical environment demanded their attention. Eight workshops were held at the ICU, six at the medical ward, four at the orthopedic ward, and six at the surgical ward. The workshops sought to identify the processes and challenges in the current paper-based ICU follow-up and to discuss ideas for how to address the challenges through an electronic version of the follow-up plan.

On the basis of the 24 workshops a preliminary design proposal was made. This proposal was presented and discussed at two subsequent 3-hour workshops attended by eight representatives, two from each of the four wards. The representatives included four managing nurses, that is head nurse, head nurse assistant, or clinical development nurse. At these two workshops the design proposal was thoroughly discussed and revised into a final design of an electronic follow-up plan [14].

Results

We first report the results of the record audit and then those of the participatory-design workshops.

Baseline Performance

Follow-up plans existed for only 48 (16%) of the 304 eligible patients. That is, follow-up plans were absent for 84% of the patients for whom such plans should have been present. In addition, Table 3 shows that the majority of the plans that did exist were not accompanied by a follow-up visit by an ICU nurse, not read by the staff at the general ward, and not carried through to completion. Because our exclusion criterion eliminated transfers to palliative care the 48 patients with a follow-up plan were expected to survive. However, 12 (25%) of them died within 24 hours of their transfer to the general ward.

Table 3 – Baseline Performance

Category	Number	Percent
Follow-up plan made by ICU	48	100
Follow-up visit performed by ICU	10	21
Follow-up plan read at general ward	20	42
Follow-up plan completed	15	31
Death within 24 hours of transfer	12	25

For all 12 patients who died the follow-up plans prescribed supplementary oxygen and lung therapy. Nevertheless, the cause of death for two of them was hypoxic heart failure (i.e., insufficient oxygen or lack of supplementary oxygen). For four of

the 12 patients the follow-up plans also prescribed physiotherapy in relation to dysphagia (difficulties swallowing; a common complication after invasive mechanical ventilation). However, the cause of death for two of these four patients was aspiration failure (i.e., food or fluids going down the windpipe). These findings suggest that a redesign of the follow-up plans might improve the handover from ICU to general ward and save lives.

Design of Electronic Follow-Up Plans

The participatory-design workshops revealed that follow-up plans were poorly implemented in the transfer of ICU patients to the general wards. The idea of facilitating the transfer with a follow-up plan was initiated by an ICU nurse who studied such plans as part of her continuing education four years prior to this study. The ICU managing nurse appreciated the idea of a written follow-up plan, asked the nurse to make a template, and arranged with the general wards to start using follow-up plans. The follow-up plan was introduced at the involved wards in 2008, but since then no systematic evaluation had been made. The follow-up plan was seen as an extra service provided by the ICU; no resources were specifically allocated to producing the plans.

The nurses, especially the ICU nurses, were astonished when they realized the poor baseline performance indicated in Table 3. The workshops uncovered that the follow-up plans were not integrated in ICU routines: Remarkably few were made and even fewer were followed up by visits at the general wards. The ICU nurses mentioned busyness and lack of management attention as primary causes, along with the experience that many of the follow-up plans they did make were not read at the general wards. At the general wards the situation surrounding the follow-up plans was also characterized by poor integration into work routines, busyness, and lack of management attention. In addition, plans were often displaced or not discovered by general-ward nurses. And the plans were generally considered confusing, overly detailed, and unrealistic. The main source of the perceived lack of realism was that the prescriptions in the plans did not fit the conditions and practices of the general wards, especially their staffing (up to 20 patients per nurse). Therefore, the participating nurses from all wards welcomed the initiative to redesign the follow-up plan and its integration into their work practices.

The workshops investigated two different paths for the redesign of the electronic follow-up plan: the checklist and the attend-to list. The checklist reflected an attempt to turn the paper-based follow-up plan into an electronic substitute with similar features. The initial 24 workshops followed this path. At the two final workshops the checklist design was challenged and revised into the attend-to list.

The paper-based follow-up plan was a checklist that contained evidence-based tasks for the general-ward nurses to perform. As such the paper-based plan followed an approach that originated from aviation and formed the traditional way of handling patient safety issues in healthcare, especially in relation to surgery [15]. A checklist is a cognitive aid that supports memory recall during high-stress situations. It seeks to regulate and standardize processes in order to comply with best practices and reduce errors [16]. A checklist-based follow-up plan functions as an aide-mémoire with a list of specific actions to perform and boxes for ticking off each action when completed.

In designing the electronic follow-up plan the ICU nurses established a base list of all the tasks it could potentially be relevant to include in a specific follow-up plan. This base list was largely adopted from the work leading to the paper-based plan. To create a specific follow-up, plan the ICU nurse would walk through the base list and select the tasks relevant to the patient in question. The proposed design of the electronic follow-up plan included a facility on the electronic whiteboard in the ICU for creating the plan from the base list. When the patient was transferred the follow-up plan would become available on the electronic whiteboard in the general ward. The general-ward nurse responsible for the patient would then follow the plan and indicate the completion of its tasks by signing them off on the whiteboard. This way the administered care would be registered, and it would be made visible how far care had progressed toward completion of the plan. For example, the provision of lung therapy through continuous positive airway pressure (CPAP) would be specified as a task to be performed at 06:00, 10:00, 14:00, 18:00, 22:00 and 02:00 for at least 5 minutes, with a box to tick off the completion of each CPAP instance.

At the two final workshops the general-ward nurses raised concerns about the checklist-based design proposal. They felt that their professional integrity was compromised, and their competences questioned. The rigid checklist format was perceived as the ICU nurses stating in detail how the follow-up process should be performed, as if the general-ward nurses were novices. Furthermore, the checklist prescribed actions to be done according to ICU routines. The general-ward nurses were responsible for performing the tasks on the checklist, but they were not able to do it according to the detailed instructions in the proposed follow-up plan. The prescribed follow-up process did not align with the staffing and routines of the general wards. As an example of misalignment with general-ward routines, the ICU and the general wards calculated the patients' fluid balance at different times during the day. Thus, when a patient was transferred either the ICU or the general ward had to recalculate the patient's fluid balance. The times at which fluid balances were calculated depended on the time at which the nurses in each ward started their shifts, on their workload, and on the need to align their activities with those of other clinicians, such as the physicians' medical rounds. The differences in work contexts and the associated difficulties for the general-ward nurses to comply with the detailed task prescriptions in the follow-up plan were considered main reasons for the non-use of the paper-based plans. The general-ward nurses were compelled to perform only those tasks that were practicable under general-ward conditions and leave the rest of the follow-up plan undone. This problem was not addressed in the checklist-based proposal for an electronic follow-up plan. Rather, it was aggravated by timestamping the general-ward nurses' activities and, potentially, giving the impression of non-compliance and low quality.

To resolve the problems with the checklist-based plans the participants at the two final workshops redesigned the electronic follow-up plan into an attend-to list. This list itemized the issues that required the general-ward nurses' attention. While the list specified the issues of concern, it left it to the general-ward nurses to decide how best to attend to these issues within the constraints of the general ward. That is, the attend-to list described the pursued goals and left the process required to meet these goals for situated specification by local actors, see Table 4. For example, balancing a patient's fluids could be described as maintaining a fluid balance that was positive by 1000 ml. As another example the avoidance of aspiration could be described as regaining the patient's swallowing function or mobilizing mucus, rather than by prescribing when and how to perform CPAP. The workshop participants felt that the attend-to list supported the general-ward nurses' memory and optimized the communication between ICU and general ward, thereby preserving the positive features of the checklist. The ICU nurses were to select the issues of concern from a base list. This approach would make the selection quick for the ICU nurses and

safeguard against accidental omissions. The final design included a base list of 39 issues – seven of which with links to standard clinical guidelines – arranged under five overall tasks.

Table 4 – The Checklist Compared to the Attend-to List

The checklist	The attend-to list
Process-oriented	Goal-oriented
• Prescribes the process of a task but not necessarily its goal	• Describes the goal (issue of concern) without prescribing the process
Controlled processes	Contextual processes
• Processes are standardized to bolster quality and safeguard against errors	• Process decisions are left to the nurse's discretion in the given context
Novice-oriented	Expertise-oriented
• Relies on instruction to tell the general-ward nurse how to care for the patient	• Relies on the general-ward nurse's competence to decide how best to proceed

Discussion

Our results (a) document the risk associated with the handover from ICU to general ward at the studied hospital, (b) reveal a handover reality that is quite different from how it was perceived by, especially, the ICU nurses, and (c) provide a novel design of electronic follow-up plans that consist of an attend-to list rather than a checklist. A follow-up plan existed for only 16% of the audited patient transfers. Furthermore, as much as 25% of the patients with a follow-up plan died within 24 hours of their transfer; in several of these cases the cause of death suggested that the general-ward nurses had remained inattentive to information specified in the follow-up plan. Thus, it is evident that improved handover practices are needed in the ICU, which is responsible for making the follow-up plans, as well as in the general wards, which are responsible for the patients after their transfer.

The present study highlights that handover procedures such as follow-up plans must cater to the differences between the ICU and the general wards. This finding echoes previous studies. For example Kauppi et al. [12] observe that the "gap between how care is structured and practiced in a general ward in comparison with an ICU requires adaptation of care in order to ensure a smooth transition", point to the need for "the best possible preparation and collaboration between the nurses at the ICU and the nurses in the general wards", and suggest to "improve the reporting and documentation prior to the ICU discharge." Similarly, Enger and Andershed [11] emphasize the importance of a good handover report and observe that the general-ward nurses "often found reports to be suboptimal, without a clear caring plan for the patient, or too long, containing too much information about the procedures and medications given in the ICU." The present study shows that failing to heed these insights may have fatal consequences for the transferred patients.

The participatory-design workshops with nurses from the ICU and three general wards constitute an extensive collaborative effort to reflect on the transfer situation and propose ways of improving it. The importance of a collaborative approach has previously been stressed by Enger and Andershed [11], who "encourage an interdisciplinary dialogue" by observing that the general-ward nurses "lacked a greater understanding of the work in the ICU", whereas the "ICU nurses did not understand how much work they had to do on the general ward."

The main result of the participatory-design workshops was the realization that checklist-based follow-up plans were suboptimal, and the design of an alternative based on attend-to lists. A checklist-based follow-up plan imposes a specified process on the general ward. The general-ward nurses perceive this approach as being treated like novices when they struggle to apply the prescriptions from the ICU. As Bosk et al. [17] note they "come to feel that checklists undermine their claims to expertise." The limitations of the checklist-based approach are consistent with how Markus [18] characterizes the knowledge reuse situation of the expertise-seeking novice. The contextual differences between the ICU and the general wards result in a follow-up plan with de-contextualized prescriptions that could easily be re-contextualized back into the ICU but were near impossible for the general-ward nurses to re-contextualize into their work.

Instead of the checklist-based plan the workshop participants propose an electronic follow-up plan based on attend-to lists. This alternative acknowledges the differences between the ICU and the general wards by avoiding detailed process prescription. At the same time the attend-to list enables the ICU nurses to convey their expert knowledge about the patient to the general ward in terms of specified issues of concern. The attend-to list specifies the goals (e.g., maintaining a fluid balance that is positive by 1000 ml) but leaves the process of attaining and sustaining them for the general-ward nurses to decide. Thereby, the attend-to list also acknowledges and utilizes the general-ward nurses' expertise.

The checklist and attend-to list represent two very different strategies for supporting the handover. Checklists subscribe to the same line of thinking as accreditation, a quality-improvement strategy that has dominated the healthcare sector in Denmark for more than a decade [19]. That is, checklists value behavior control [20] and aim to improve quality by standardizing work processes [21]. Conversely, the attend-to list corresponds to outcome control [20] and aims to ensure quality by leaving the decisions about how to conduct work in the hands of trained and skilled individuals [21]. The former is effective when the work can be planned in detail, while the latter is effective when the work is characterized by frequent exceptions and the need for situated adjustments. The attend-to list acknowledges the reality of a "gap between how care is structured and practiced" [12]. This gap limits the possibilities for pre-planning the patient transfer from ICU to general ward. Instead, the proposal to base follow-up plans on attend-to lists leaves the process of deciding how to handle the issues of concern to the discretion – and expertise – of the nurses at the receiving ward.

In interpreting the results of this study, it is important to remember its limitations. First of all, the attend-to list has been devised but its use has not been evaluated in clinical practice. Such evaluation is important future work. The present study provides a baseline against which to compare the results of introducing follow-up plans that consist of attend-to lists. A second limitation is that this study is restricted to one hospital. While its above average patient mortality within 30 days of ICU discharge calls for improving handover practices, it may also indicate differences in the composition of the patient population compared to other parts of Denmark. Such differences may, in turn, influence the work in the ICU and general wards. The handovers at the studied hospital are, however, complicated by issues similar to those discussed in previous studies.

Conclusions

Discharge from the ICU involves risk to patient health. At the studied hospital 25% of the patients for whom the ICU has made a follow-up plan die within 24 hours of their transfer to a

general ward. Part of the reason for these deaths is that the follow-up plans fail to consider the staffing and other practical realities at the general wards and, therefore, mostly remain unread or uncompleted. An alternative design of the follow-up plans was devised in collaboration with ICU and general-ward nurses. This design made use of the electronic whiteboard already in use across the hospital and consisted of an attend-to list rather than a checklist. The attend-to list specifies the issues of concern and leaves it for the general-ward nurses to heed these issues in the way they deem right on the basis of their situated understanding of general-ward work.

Acknowledgements

This study was co-funded by Region Zealand and the studied hospital. In the interest of full disclosure, we declare that prior to the study the first author has worked as an ICU nurse at the studied hospital. Special thanks are due to the workshop participants.

References

[1] C. Owen, L. Hemmings, and T. Brown, Lost in translation: Maximizing handover effectiveness between paramedics and receiving staff in the emergency department, *Emergency Medicine Australasia* **21** (2009), 102-107.

[2] M.-S. Ong and E. Coiera, A systematic review of failures in handoff communication during intrahospital transfers, *Joint Commission Journal on Quality and Patient Safety* **37** (2011), 274-284.

[3] G. Hesselink, L. Schoonhoven, P. Barach, A. Spijker, P. Gademan, C. Kalkman, J. Liefers, M. Vernooij-Dassen, and H. Wollershelm, Improving patient handovers from hospital to primary care: A systematic review, *Annals of Internal Medicine* **157** (2012), 417-428.

[4] D. Flemming and U. Hübner, How to improve change of shift handovers and collaborative grounding and what role does the electronic patient record system play? Results of a systematic literature review, *International Journal of Medical Informatics* **82** (2013), 580-592.

[5] H.B. Andersen, I.M.D. Siemsen, L.F. Petersen, J. Nielsen, and D. Østergaard, Development and validation of a taxonomy of adverse handover events in hospital settings, *Cognition, Technology & Work* **17** (2015), 79-87.

[6] J.C. Marshall, L. Bosco, N.K. Adhikari, B. Conolly, J.V. Diaz, T. Dorman, R.A. Fowler, G. Meyfroidt, S. Nakagawa, P. Pelosi, J.-L. Vincetn, K. Vollman, and J. Zimmerman, What is an intensive care unit? A report of the task force of the World Federation of Societies of Intensive and Critical Care Medicine, *Journal of Critical Care* **37** (2017), 270-276.

[7] Danish Clinical Registries, *Dansk intensiv database: Årsrapport 2013/2014*, Danish Clinical Registries, Frederiksberg, Denmark, 2014.

[8] Danish Clinical Registries, *Dansk intensiv database: Årsrapport 2016/2017*, Danish Clinical Registries, Frederiksberg, Denmark, 2017.

[9] A. Torkilsheyggi and M. Hertzum, Incomplete by design: A study of a design-in-use approach to systems implementation, *Scandinavian Journal of Information Systems* **29** (2017), article 2.

[10] N. van Sluisveld, G. Hesselink, J.G. van der Hoeven, G. Westert, H. Wollersheim, and M. Zegers, Improving clinical handover between intensive care unit and general ward professionals at intensive care unit discharge, *Intensive Care Medicine* **41** (2015), 589-604.

[11] R. Enger and B. Andershed, Nurses' experience of the transfer of ICU patients to general wards: A great responsibility and a huge challenge, *Journal of Clinical Nursing* **27** (2017), e186-e194.

[12] W. Kauppi, M. Proos, and S. Olausson, Ward nurses' experiences of the discharge process between intensive care unit and general ward, *Nursing in Critical Care* **23** (2018), 127-133.

[13] J. Simonsen and T. Robertson, eds., *Routledge international handbook of participatory design*, Routledge, London, 2013.

[14] K.L. Østergaard, J. Simonsen, and H. Karasti, Examining situated design practices: Nurses' transformation towards genuine participation, *Design Studies* **59** (2018), 37-57.

[15] D. Aerden, D. Smets, J. Poelaert, J. Oste, and P. van den Brande, Fighting human error: What surgeons can learn from aviators, *Acta Chirurgica Belgica* **114** (2014), 228-232.

[16] B.M. Hales and P.J. Pronovost, The checklist - A tool for error management and performance improvement, *Journal of Critical Care* **21** (2006), 231-235.

[17] C.S. Bosk, M. Dixon-Woods, C.A. Goeschel, and P.J. Pronovost, Reality check for checklists, *Lancet* **374** (2009), 444-445.

[18] M.L. Markus, Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success, *Journal of Management Information Systems* **18** (2001), 57-93.

[19] J. Simonsen, M. Hertzum, and J.D. Scheuer, Quality development in healthcare: Participation vs. accreditation, *Nordic Journal of Working Life Studies* **8** (2018), 49-69.

[20] K.M. Eisenhardt, Control: Organizational and economic approaches, *Management Science* **31** (1985), 134-149.

[21] H. Mintzberg, Structure in 5's: A synthesis of the research on organization design, *Management Science* **26** (1980), 322-341.

Address for Correspondence

Kija L. Østergaard, Department of People and Technology, Roskilde University, Roskilde, Denmark, email: kijal@ruc.dk.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190438

Designing Tailored Displays for Clinical Practice Feedback: Developing Requirements with User Stories

Veena Panicker [a], Dahee Lee [a], Marisa Wetmore [b], James Rampton [a], Roger Smith [b], Michelle Moniz [b,c], Zach Landis-Lewis [a]

[a] *Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, MI, USA,*
[b] *Department of Obstetrics and Gynecology, Michigan Medicine, Ann Arbor, MI, USA*
[c] *Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, MI, USA*

Abstract

Improving visualizations in clinical quality reports and dashboards may improve the visualization influence on clinical practice. Tailored displays could accommodate individual and situational differences, but these diplays introduce complex requirements across healthcare professionals and teams. We applied user stories, a method for managing complex software requirements, to a user-centered design process for tailored visual displays about postpartum contraception care. We mapped user stories to tailored displays to identify the quantity of displays that were supported by each user story. We developed 9 tailored displays and 11 user stories. Displays varied in their mappings to user stories (mean 5, max 9, min 0), revealing differences in healthcare professionals and teams preferences and information needs. User stories and user-centered design may be useful for healthcare organizations to manage complex requirements of tailored displays in clinical practice feedback.

Keywords:

Feedback, quality improvement, software design.

Introduction

Clinical quality reports and dashboards are widely used to communicate about clinical performance to healthcare professionals and teams [1,2]. Visual displays, such as charts, graphs, and tables, are a central component of dashboards and performance reports because of the ability to reduce the cognitive burden required to understand performance data [3,4]. However, evidence about the use of these reporting tools as feedback interventions to change practice shows a wide range of results [5], suggesting that ideal conditions may exist under which performance feedback is highly influential on clinical practice.

An approach that may achieve ideal conditions is the tailoring of feedback reports where population-level practice feedback could deliver prioritized, actionable information [6]. Tailoring information has been demonstrated to improve cognitive processing of messages and the effects on motivation in the field of health promotion and communication [7]. Tailored feedback could accommodate provider differences such as numeracy and graph literacy, which can vary among healthcare professionals [8,9], and to accommodate changing priorities based on evolving gaps and trends in quality measures. Tailoring feedback requires representations of individual and situational difference characteristics that may change over time across healthcare professionals and teams.

User stories are a method for managing software requirements in complex and dynamic environments from the agile software methodology [10]. A key feature of user stories is that they associate a specific software function (what) with a specific user (who) and a rationale (why). User stories have been applied in an agile approach to clinical decision support [11] and quality dashboard development [12], but to our knowledge have not been applied to the development of feedback reports for the purpose of tailored messaging.

User-centered design is increasingly used in healthcare to develop information resources, including performance feedback reports [13]. We applied user-centered design methods to develop prototype clinical quality feedback reports for healthcare professionals and teams in a department of obstetrics and gynecology at an academic medical center. The focus of this work was to support the implementation of a new practice of offering immediate postpartum long-acting reversible contraception (IPLARC) to patients. The objective of this study was to explore how user-centered design and user stories could support complex requirements management for tailored displays of clinical performance.

Methods

Setting and Participants

This work was done in an academic department of obstetrics and gynecology that provides clinical quality feedback to approximately 40 providers in a regional health system with 8 clinics in the midwestern United States. Healthcare professionals in the department include nurses, midwives, resident physicians, and attending physicians. The department routinely measures the quality of care using EHR and administrative data. Reports are routinely produced in a clinical quality dashboard. A single report is also sent monthly via email about appropriate screening for vaccination. We recruited participants for the user-centered design activities via email. They were not financially compensated. We focused on the measurement of performance for postpartum contraceptive counselling. The study was determined to be not regulated as human subjects research by the University of Michigan Medical School IRB (HUM00140107).

Design Process

We designed tailored displays and user stories in iterative and multi-stage process (Figure 1). The process of tailoring prototype displays involved a qualitative component that served to

generate themes for both the display design process and for user story development. At the conclusion of the design process we mapped tailored displays and user stories to understand the comprehensiveness of the displays in addressing individual and situational difference characteristics described in user stories.

Tailored Prototype Display Development

Prototype design displays for feedback reports in the following steps: 1) iterative low-fidelity prototyping with contextual interviews and usability testing, 2) affinity diagramming, and 3) high-fidelity prototype development.

Iterative Low-Fidelity Prototyping

We conducted 30-minute contextual interviews with healthcare professionals from each provider role in the department (midwife, nurse, resident, attending physician). We audio-recorded interviews and took field notes to identify the existing feedback reports that participants use, the channels through which feedback reports were delivered, and the contexts in which reports were viewed. We also asked open-ended questions about the goals and preferences of healthcare professionals with respect to receiving, understanding and acting on feedback reports.

We collected examples of feedback reports in the OB/GYN department and designs from published examples in the literature to inform the creation of low-fidelity prototypes. We held 4 design meetings to generate a wide range of initial display possibilities in sketches on paper. We prioritized a group of displays from an initial set of sketches for testing and we created initial prototypes for testing with synthetic performance data.

We tested the usability of the prototypes using think-aloud technique, asking each provider to verbalize their thoughts as they viewed each prototype. We also tested the comprehensibility of prototypes by asking participants questions like "what is your clinic's performance this month?". After displays had been tested, we asked participants to compare and express their preferences for the prototypes. We recorded audio and took notes in usability testing sessions.

Affinity Diagramming

We created an affinity diagram to identify themes from interviews and usability testing [14]. We used post-it notes to write down interview quotes, facts, and observations. Similar notes were grouped together and new groups were formed when notes did not fit in an existing cluster. Each cluster of post-it notes addressed one theme or idea. We looked for similarities and differences between different professional groups in terms of preferences to identify requirements for tailoring feedback.

High-Fidelity Prototype Development

After iterating through the display designs, we eliminated displays that proved to be inappropriate, unanimously unpopular, or difficult to comprehend. We created high-fidelity prototypes using Sketch (Hague, NL) design toolkit based on our tested low-fidelity prototype displays.

User Story Development

We wrote user stories based on the themes from affinity diagramming [Table 1]. One member of the study team drafted the stories, and a second team member proposed revisions to create the set of stories for analysis. These user stories followed the three-part format of:

*As a*_____

I want to receive performance feedback _____

So that I can _____.

Mapping User Stories and Prototype Displays

To map user stories with prototype displays, two members of the study team independently selected the prototype displays that appeared to be compatible with each user story. Selections were recorded in a shared spreadsheet. One team member identified differences in selections and then both team members met to resolve differences through discussion.

Results

Prototype Design

Design activities took place over a period of 5 months between June and October in 2018.

Iterative Low-Fidelity Prototyping

We conducted 3 cycles of contextual interviews, low-fidelity prototyping, and usability testing. We interviewed 10 healthcare professionals, including 2 attending physicians, 2 nurses, 4 midwives and 2 residents.

We developed 13 low-fidelity prototype designs for testing.

We conducted 4 usability testing sessions with attending physicians to test the usability of the displays.

As we iteratively developed low-fidelity prototypes, we gained insights that lead us to focus on specific user groups and to refine elements of the low-fidelity prototypes. We learned that the performance measures focusing on IPLARC were not directly actionable for nursing staff, therefore we did not seek to test the usability of prototypes with nurse participants. We learned that resident physician's goals and preferences were focused on immediate feedback about skill

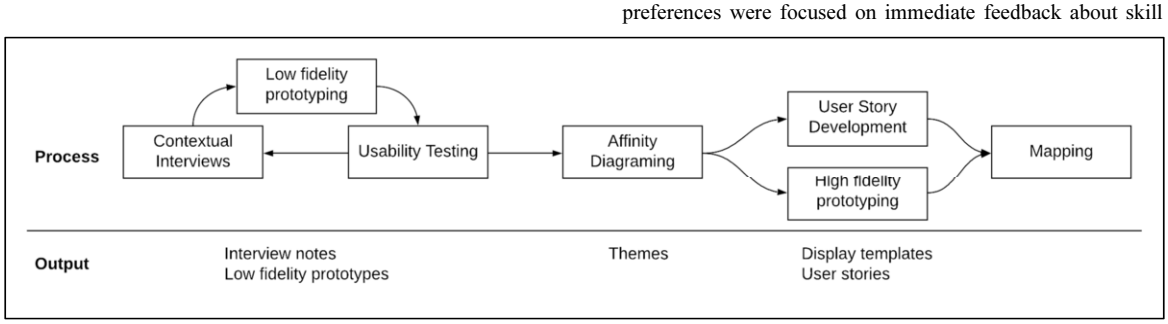


Figure 1– User-Centered Design Process for Tailored Displays of Clinical Performance

Table 1— User Story Development

Quotes	Theme	User Story
"I like seeing c-section rates and the trends" - *Midwife 2* "It's nice to see the trend" - *Midwife 4* "Trend is helpful... what is the reason behind this trend for the clinic?" - *Attending Physician 1*	Providers like seeing trends in the report	**As a** provider, **I want to** receive performance feedback reports that shows trends **So that** I can see how my clinic's performance changes over time.
"I like seeing the comparison of how you're doing in relation to other clinics" - *Attending Physician 4* "I'm competitive, it's nice to see I'm the best...I like comparison" - *Attending Physician 6*	Attending physicians want to see names and performance of other clinics	**As an** attending physician, **I want to** receive performance feedback with data from other clinics **So that** I can compare my clinic's performance to others.

acquisition to the extent that summary performance feedback about individual management of patient and especially for clinic-level performance was perceived to be largely not actionable. We also learned that midwives in the department used a separate performance measurement and feedback process that was distinct from physician-focused quality improvement reporting. Given the available resources for the project we focused our scope on attending physician feedback, who voiced the most support for the utility and actionability of the feedback reports in contextual interviews.

High-Fidelity Prototyping and User Stories Mapping

We identified 11 themes in an affinity diagram. We identified themes about display features and report delivery. We iteratively developed display designs in 3 cycles. We created 9 high fidelity prototype displays for our collection. Each display contains one graph or table with synthetic performance data. We created 11 user stories based on these themes. Each user story corresponded to a single theme from the affinity diagram [Table 2].

The mapping process helped us to identify 2 user stories that were not supported (8 and 9) and two user stories that were each addressed by only a single display (10 and 11). We also identified four user stories(1, 3, 5 and 6) that were supported by all 9 displays. Prototype displays were mapped to an average of 6 user stories (min 5, max 7). User stories were were mapped to an average of 5 prototype displays (min 0, max 9).

Table 2— User Stories

	User Story		
Story ID	**As a...**	**I want to...**	**So that...**
1	provider	receive performance feedback reports regularly	I can check them routinely
2	provider	see a target or goal value in the performance feedback reports	I can track and compare my performance
3	provider	see both percentages and raw data in the performance feedback reports	I can better understand the reports
4	provider	receive performance feedback reports that shows trends	I can see how my performance changes over time
5	attending physician	receive a performance feedback report about counselling, placement and outcomes of IPLARC	I can evaluate my performance holistically
6	attending physician	receive a performance feedback report within the body of an email	I can spend less time opening attachments
7	attending physician	receive performance data from other clinics	I can see how well my clinic is doing compared to others
8	provider	know how my performance data is collected	I can better understand the performance feedback reports
9	resident	receive feedback about my performance immediately	I can improve my skills and techniques
10	attending physician	receive a congratulations message when my clinic performance is consistently high for 4 months	I can recognize our achievements
11	attending physician	receive a "fun" feedback report	I am more likely to look at the report when I receive it

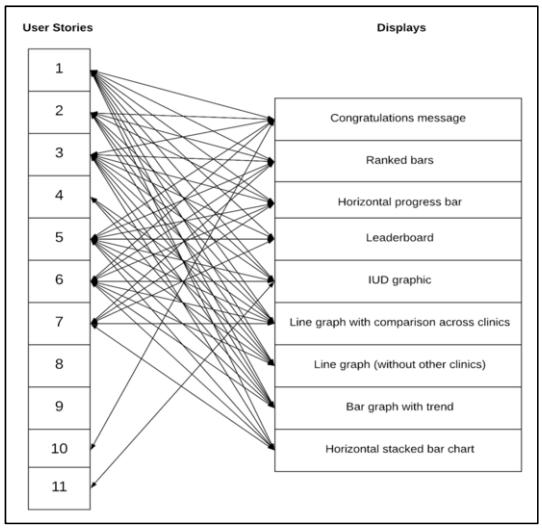


Figure 2– Mapping between User Stories and Display Prototypes. User story numbers correspond with numbers in table 2

Discussion

User stories are a promising approach for representing complex requirements to tailor clinical practice feedback displays to healthcare professionals and teams. Our preliminary experience with a user-centered design method for tailored feedback suggests that user stories may enable improved design, evaluation, and maintenance of tailored feedback interventions. We applied this process to support performance measurement and feedback about IPLARC to healthcare professionals and teams. Incorporating user stories into the design process enabled us to represent contextual factors and preferences of participants that have potential to improve the effectiveness of feedback reports about the quality of clinical practice by supporting the delivery of tailored feedback.

Mapping user stories to displays gave us insight into the ability of a collection of displays to support variable requirements that represent a critical first step toward the delivery of tailored feedback. The mappings served as a preliminary validation of the collection of displays, indicating that each display supported a minimum of 5 user stories. Mapping also revealed the stories that were not supported by any displays. For example, an absence of mappings to Story 9 confirmed what we had learned in contextual interviews about the preferences of residents for receiving performance feedback that was not compatible with the current types of performance measurement that our project supported [Table 2]. Similarly, an absence of mappings to Story 8 revealed a preference that was beyond the scope of this project to address but represented an important next step in the the report design process.

User stories require the story developer to provide a rationale for the user's desire to have a specific function in the information resource. Including a rationale can support the maintenance of tailored reports as user context evolves but also enables feedback report developers to identify stories with rationales that are no longer justified and then to identify the displays that they are supported by for revision or removal from an active collection of tailored displays.

A promising aspect of this approach is the potential to use user stories to develop metadata for tailored displays and to automate the delivery based on user roles and personal profiles. We also anticipate that user stories could be used to improve the user-centered design process if used in a participatory approach [15].

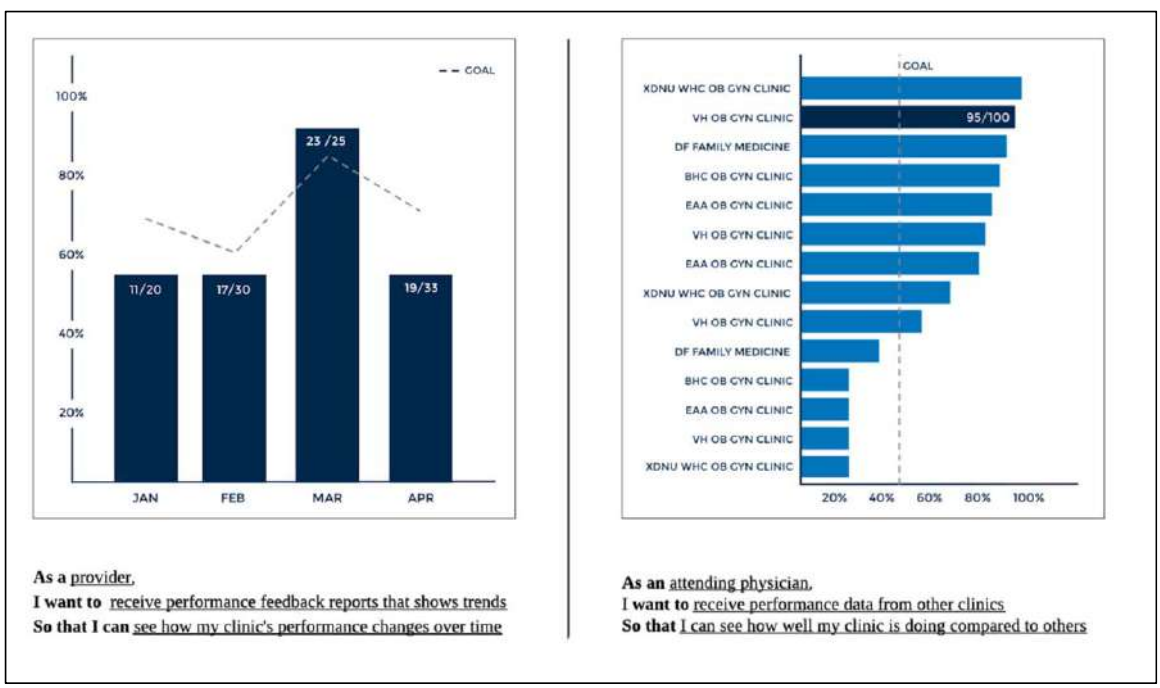


Figure 3– User Story - Display Mapping. Each display is mapped with the user story below it and is incompatible with the other user story.

A limitation we encountered with the user story structure was that it is not appropriate for expressing negative themes. Representation of negative themes is not necessary for general characteristics of reports, but may be necessary to express requirements that differ between users within a context or professional role.

A limitation for this exploratory study is the preliminary nature of the analysis. In some cases, we did not elicit a specific rationale for the report characteristics that participants expressed preferences, leaving a gap for writing the user story rationales. In these cases, we created the rationale based on our interpretation of the general goals and motivations of participants. We plan to conduct follow-up interviews to check our stories and to continue to iteratively improve the requirements for tailored feedback reports. We anticipate that healthcare organizations may apply a similarly iterative approach to refine and adapt user stories.

Conclusions

We have explored the application of a user-centered design methods with user stories to develop tailored displays for clinical performance feedback. We found that these methods enabled the representation of complex requirements across healthcare professional roles. We anticipate that healthcare organizations could use this method to improve the design of visual displays in clinical quality dashboards and reports. We plan to develop these methods to incorporate participatory design approaches and metadata generation to support automated tailoring of clinical performance feedback.

Acknowledgements

We would like to thank Douglas Everhart for his contribution to the visual design of the display templates.

We would also like to thank Cooper Stansbury, Colin Gross and Astrid Fishstrom for their valuable contributions to this study.

Funding for this study was provided by the National Library of Medicine (NLM) through 1K01LM012528-01

References

[1] D. Dowding, R. Randell, P. Gardner et al., Dashboards for improving patient care: Review of the literature, *Int. J. Med. Inf.***84** (2015) 87–100. doi:10.1016/j.ijmedinf.2014.10.001.

[2] B. Brown, P. Balatsoukas, R. Williams et al., Interface design recommendations for computerised clinical audit and feedback: Hybrid usability evidence from a research-led system, *Int. J. Med. Inf.***94** (2016) 191–206. doi:10.1016/j.ijmedinf.2016.07.010.

[3] S. Few, Information Dashboard Design: Displaying Data for At-a-Glance Monitoring, Second Edition, Second edition edition, Analytics Press, Burlingame, CA, 2013.

[4] T. Munzner, Visualization Analysis and Design, CRC Press, 2014.

[5] N. Ivers, G. Jamtvedt, S. Flottorp et al., Audit and feedback: effects on professional practice and healthcare outcomes, *Cochrane Database Syst. Rev. Online*. **6** (2012) CD000259. doi:10.1002/14651858.CD000259.pub3.

[6] Z. Landis-Lewis, J.C. Brehaut, H. Hochheiser et al., Computer-supported feedback message tailoring: theory-informed adaptation of clinical audit and feedback for learning and behavior change, *Implement. Sci.***10** (2015) 12. doi:10.1186/s13012-014-0203-z.

[7] R.P. Hawkins, M. Kreuter, K. Resnicow et al., Understanding tailoring in communicating about health, *Health Educ. Res.***23** (2008) 454–466. doi:10.1093/her/cyn004.

[8] D. Dowding, J.A. Merrill, N. Onorato et al., The impact of home care nurses' numeracy and graph literacy on comprehension of visual display information: implications for dashboard design, *J. Am. Med. Inform. Assoc. JAMIA*. **25** (2018) 175–182. doi:10.1093/jamia/ocx042.

[9] S. Govindan, V. Chopra, and T.J. Iwashyna, Do Clinicians Understand Quality Metric Data? An Evaluation in a Twitter-Derived Sample, *J. Hosp. Med.***12** (2017) 18–22.

[10] M. Cohn, User Stories Applied: For Agile Software Development, 1 edition, Addison-Wesley Professional, Boston, 2004.

[11] M.A. Basit, V. Kannan, and D.L. Willett, Agile Clinical Decision Support Development and Implementation, in: Proc. 2018 ACM Int. Conf. Bioinforma. Comput. Biol. Health Inform., ACM, New York, NY, USA, 2018: pp. 555–555. doi:10.1145/3233547.3233663.

[12] V. Kannan, J. Fish, J. Mutz, A. Carrington, K. Lai, L. Davis, J. Youngblood, M. Rauschuber, K. Flores, E. Sara, D. Bhat, and D. Willett, Rapid Development of Specialty Population Registries and Quality Measures from Electronic Health Record Data: An Agile Framework, *Methods Inf. Med.***56** (2017) e74–e83. doi:10.3414/ME16-02-0031.

[13] H.L. Colquhoun, D. Sattler, C. Chan et al., Applying User-Centered Design to Develop an Audit and Feedback Intervention for the Home Care Sector, *Home Health Care Manag. Pract.***29** (2017) 148–160. doi:10.1177/1084822317700883.

[14] H. Beyer, and K. Holtzblatt, Contextual Design: Defining Customer-Centered Systems, 1 edition, Morgan Kaufmann, San Francisco, Calif, 1997.

[15] K. Kautz, Participatory Design Activities and Agile Software Development, in: J. Pries-Heje, J. Venable, D. Bunker, N.L. Russo, and J.I. DeGross (Eds.), Hum. Benefit Diffus. Inf. Syst. Des. Sci. Res., Springer Berlin Heidelberg, 2010: pp. 303–316.

Address for correspondence

Zach Landis-Lewis, Department of Learning Health Sciences, University of Michigan, 1161J NIB, 300 N. Ingalls St, Ann Arbor, MI, 48109.

zachLL@umich.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190439

User Driven Design: First Step in Involving Healthcare Consumers and Clinicians in Developing a Collaborative Platform to Prevent Cardiovascular Diseases

Sylvia Pelayo[a], Jessica Schiro[a], Pierre-François Gautier[a], Marie-Christine Jaulent[b], Romaric Marcilly[a]

[a] Univ. Lille, INSERM, CHU Lille, CIC-IT/Evalab 1403 - Centre d'Investigation Clinique, EA 2694, F-59000 Lille, France
[b] LIMICS, INSERM UMRS 1142, Université Paris 13, Sorbonne Paris Cité, Paris, France

Abstract

To prevent cardiovascular diseases, eHealth solutions may be used as tools, involving health care consumers in the set-up of their prevention plan, a fundamental condition for improving their long-term adherence to the plan. This paper presents the first step in a web platform design aiming to support the co-elaboration by health care consumers and clinicians of personalized prevention plans. Applying a user driven innovation approach, first, a questionnaire and semi-structured interviews were combined to identify clinicians' needs. Then, three focus group sessions with consumers and clinicians were organized to identify their needs, creating the system workflows, its graphical user interface, and its navigation paths, with the best ideas shaped by paper mockups. An interactive mockup was designed including 30 screens (ex. user dashboards, desk for co-elaborating plan). This user driven approach enabled to design not only the technology and its graphical user interface, but also a prevention plan design process.

Keywords:

Ergonomics; Research Design; Primary Prevention

Introduction

The number of deaths related to preventable risk factors keeps on increasing. High blood pressure is at the top of the list of preventable risks, contributing to more than 10 million deaths worldwide [1]. Tobacco ranks second before obesity and diabetes. When combined, these risk factors are responsible for cardiovascular diseases (CVD), leading to death. Thereby, health prevention actions against those risk factors are essential.

Despite many initiatives, setting up a prevention plan is not easy for health care consumers and clinicians [2]. Prevention is the active and empowered management by a person of his or her health capital. Clinicians mostly initiate prevention actions and require the active involvement of the person. However, from a clinician's perspective, initiating prevention actions is time-consuming while the consultation time is limited, and available tools are unhelpful to support this task [3,4]. As for the health care consumers, they must be highly motivated to be involved in prevention actions [2]. Besides, while healthy people are the main target of prevention actions, prevention is not their priority.

eHealth solutions may be used as tools to modernize prevention practices and get healthcare consumers involved in their prevention plan. Connected devices and web platforms, such as forums and/or social networks, may help health care consumers join health programs (e.g. weight loss or smoking cessation in groups). Nonetheless, so far, the impact of those solutions is limited [5-7]. To be adopted by health care consumers and be used efficiently, those solutions must be well-integrated into clinicians' work environment and must not increase their workload [8]. Behavior change techniques must be embedded in those solutions to help health care consumers get involved in their prevention plan [10]. Finally, the solutions must appropriately support the codesign of a personalized prevention plan by health care consumers and their clinicians [10,11] to increase the likelihood that health care consumers adhere to their prevention plan: finding a consensus between a consumer's preferences, abilities and habits, and his or her clinician's recommendations is essential. [12,13]

This paper presents the PEPS project that aims to design a collaborative clinician-health care consumer web solution for cardiovascular diseases prevention. The project is ambitious in terms of involving end users (i.e. health care consumers, clinicians) into a codesign approach to ensure that the future solution (i) is usable and (ii) encourages and supports the co-elaboration of personalized prevention plans (PPP) by the clinician and the consumer. A user-driven innovation approach was adopted to tap users' explicit and tacit knowledge and develop innovative solutions. The process systematically and actively involved users to get a sound understanding of their actual needs and practices [14:]. With this goal in mind, in addition to researchers in ergonomics and decision support systems and to electronic health record companies, the project board also included the French College of General Medicine (FCGM) and an association of patient associations (Inter-Associative Collective on Health). The FCGM and the association ensured the involvement of primary care clinicians (general practitioners (GPs) - and nurses) and of persons at risk of CVD at all stages of the design and evaluation process. In this paper, we present the first step in the design process, i.e. the application of codesign proven methods to analyze end users' needs and to get mockups and design specifications codesigned by health care consumers and clinicians.

Methods

The methodology proceeded in three steps.

Preliminary Questionnaire and Interviews

An on-line questionnaire was developed to identify current needs of the GPs. It identified: the technology currently used to prevent CVD, appreciated and criticized features, GPs' expectations, and the behavior change techniques they use to prevent CVD. The questionnaire was sent electronically to members of the FCGM.

Semi-structured interviews were performed to get deeper qualitative insights into the results of the questionnaire. They addressed the same questions plus clinicians' CVD prevention practices, their difficulties, their ideas and proposals to improve their practices, and the perceived pros and cons of PPP. Respondents to the questionnaire who were interested were contacted. The FCGM completed the recruitment with some of their members and nurses.

Focus Groups

Three sessions of focus groups were organized successively. They aimed (i) to validate nurses and GP's needs identified through the questionnaire and interviews, (ii) to support the ideation about how the PEPS solution should work with its graphical user interface and its navigation paths, and (iii) to shape the best ideas with paper mockups.

The three focus groups sessions complemented each other: they addressed different parts of the PEPS solution (see Figure 1).

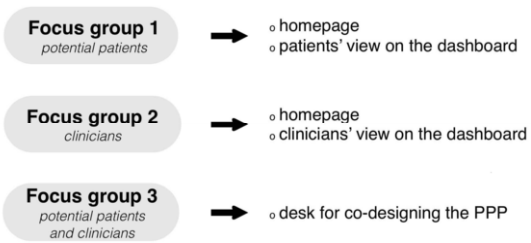


Figure 1– Schematic Representation of the Aims of Each Focus Group Session.

The "studio design" method [15] was the inspiration for the design of the focus groups. This method is a creative problem-solving process to quickly generate ideas and to compare them and achieve a consensus on an agreed-upon solution. It consists in asking participants to sketch mockups in response to a problem. Mockups must be minimal and must enable one to understand the screen's arrangement and content. Several steps of sketching, first individual, then collective, stimulate the creativity and the codesign of a consensual solution that may satisfy the needs of all participants.

Two experts in ergonomics organized and hosted each focus group session.

Focus Group One: Health Care Consumers

This session aimed to get insights regarding the PEPS homepage (accessible without login) and the PEPS prevention dashboard as it appears once logged in with a consumer's account. Participants at risk of CVD (according to their results to the French Federation of Cardiology questionnaire[1]) and healthy participants were recruited.

First, the PEPS project, its aims, and the basics of CVD prevention and of prototyping were introduced. Then, participants were confronted with the following problem: how to engage consumers in CVD prevention? To answer it, participants were divided up into two mixed groups, including both at risk and healthy participants. They had to go through four steps:

1. Brainstorming: they were asked to list techniques and information that may be useful to prevent CVD.
2. Individual prototyping: each participant was asked to sketch how those techniques and information could be displayed and arranged on the PEPS solution.
3. Group prototyping: participants had to explain their own mockup to their group and how their solution answered the initial problem. Then, the group had to combine the best ideas and design elements to create the group's mockup.
4. Inter-groups consensus: then, groups were gathered together. They had to present their own mockup before deciding by consensus on the best ideas and combining them into one new mockup.

At the end of the session, participants had to present their mockup to the experts in ergonomics and to explain how it answered the initial problem.

Focus Group Two: Nurses and GPs

This session aimed to get insights regarding the PEPS homepage and the PEPS prevention dashboard as it appears once logged in with a clinician's account. GPs and nurses were recruited by the FCGM. The problem asked was: how a web solution could help prevent CVD?

This focus group was structured like focus group one except on the first step. After an introduction to the PEPS project, instead of brainstorming, participants had to read the list of needs identified through the questionnaire and interviews and the description of four personas of consumers (see Figure 2). A persona is a fictional character created to represent a person that may use a solution [16]. Then participants were divided in two groups including both nurses and physicians and steps two, three and four occurred like in focus group one. At the end of the session, participants had to present their mockup to the experts in ergonomics.

Focus Group Three: Health Care Consumers, Nurses and GPs

This session aimed to describe precisely the PPP co-elaboration process involving health care consumers and clinicians and to create the screens and the arrangement of information that may support it efficiently. Consumers at risk for CVD, nurses, and GPs who did not take part in the first sessions were recruited. After the introduction to the PEPS project and to the goals of the focus group, participants were split into three mixed groups.

Participants had to answer the following problem: how a website could make easier the co-elaboration of a PPP for CVD? This session proceeded in two steps:

1. Group prototyping: paper cards illustrating the features and the information that may support the PPP elaboration identified through the questionnaire, and the interviews were distributed to both groups. Participants could create new cards if needed. At this step, participants were asked to arrange the cards to follow the PPP design process and to create mockups of the related screens. At the end, the groups were gathered together.
2. Inter-group consensus: each group had to present its process and associated screens' mockups. Together, participants had to decide on the best ideas and to combine them into one new mockup of the solution.

Participants had to present their mockup to the experts in ergonomics and to explain how it answerd the initial problem.

[1] Questionnaire available on jaimemoncoeur.fr

Figure 2 – Instance of Persona Used to Support the Discussion.

Ergonomics Specifications for the Solution

The experts reworked on the paper mockup of the solution with Sketch® to turn it interactive. Then, the computerized mockup was put online with InVision®.To ensure the final mockup was true to the results of the focus groups and met end users' needs, the link to the mockup was sent to 46 former participants/respondents to the three focus groups and to the questionnaire and interviews. This step enabled to get feedback on the designed mockup. Participants could either make comments through the mockup's link or send their comments by email.

Results

Questionnaire and Interviews

A total of 145 GPs replied to the questionnaire; six GPs and three nurses were interviewed. Five main themes arose. First, clinicians need synthesized, unambiguous, updated information, and advice on the prevention of CVD (e.g. under the form of practical memo) along with a synthesized overview of the PPP. The advice must be adapted to the profile of their patient. In addition, clinicians need information about religious and cultural elements they should consider when designing the PPP (e.g. specific diet).

Second, clinicians highlighted a need for interaction through the PEPS website. They need to share the PPP with their patients, and their experiences with colleagues. They need multimedia supports (e.g. video, comics) along with translated supports to communicate with patients who do not understand French fluently. They also need to set the prevention goals and to modify them if needed. Notifications about patients who did not log on for a long time would also be useful.

Third, clinicians insisted on the ergonomics characteristics of the PEPS website: entering and managing data must be easy and quick. The information displayed must always be easy to understand and reusable to save time. Fourth, the PPP should be integrated into their patient electronic record. In this case, the website could be fully interoperable with this record to avoid duplicate data entries. Fifth, clinicians insisted that their patients must be enabled to communicate together (e.g. through social network) to share their experiences and their tricks to achieve prevention goals. Patient must also access an address book of experts.

Focus Groups

Focus Group One: Health Care Consumers

A total of 8 consumers took part to the first focus group: 5 at risk and 3 healthy. Participants expressed that prevention information must be easily accessible through the homepage without being connected nor creating an account. Synthesized medical information on CVD, links to related social network, adapted balanced recipes etc. must be easy to find.

Concerning their account, participants highlighted that it must include "a zone displaying a summary of the statistics on calorie burn and food intake, defined prevention goals, and a way to enter daily those data thanks to a slider, for instance (…) I would need also a calendar to enter the number of cigarettes smoked per day and notes about breathlessness during exercise etc. There would be a kind of side menu displaying my profile, my action plan with its main milestones, my progress, and, if goals were not achieved, a way to change them (…) The GP must also have access to the information".

Finally, health care consumers insisted that the data and the advice must be reliable.

Focus Group Two: Nurses and GPs

Three GPs and four nurses took part in the second focus group.

Clinicians expressed that they should not access the same information as consumers. Once connected, clinicians should see which patients are in a prevention program and, for each patient, charts with statistics to rapidly identify each patient's room for improvement. The website must not be a copy of the patient's electronic record: to avoid double data entry, the website must automatically update data from the patient's electronic record (e.g. automatic retrieval of new lab results). Furthermore, clinicians must be able to access prevention programs and tools and be able to network with colleagues to ask them questions regarding their patients.

As for the health care consumers' accounts, clinicians think that "one should find: the solution must display the PPP with the patient's objective and the level of achievement per week (e.g. under the form of percentage) for each objective (e.g. exercise, level of cholesterol, food intake). The patient must also find

features to enter data such as weight, intakes, amount of exercise, contact details, along with links to recipes and social networks."

Finally, clinicians insisted on the credibility of the information displayed: "sources and scientific references must be displayed".

Focus Group Three: Health Care Consumers, Nurses, and GPs

Two GPs, two nurses, and three consumers took part in the third focus group.

The co-elaboration of the PPP must start with the patients completing a questionnaire about their lifestyle, their medical history, their current risks, and data, such as weight, size etc. Then, each GP will "complete those data with other clinical data gathered during an appointment in such a way as to make a diagnosis". Based on the data entered in the PEPS solution and / or retrieved automatically from the patient electronic record, the solution must propose prevention objectives and associated deadlines. The GP and the healthcare consumer will discuss together the proposed objectives, the ways to achieve them (e.g. exercise, monitoring of intakes) and their implementation into daily life.

To advise consumers and help them make informed decisions, GPs need a toolbox gathering persuasive information (e.g. patient's gain if an objective is achieved). The information must be proposed clearly and be easy to understand. As for the consumers, they must access rapidly a summary of their objectives and actions to take. They should be able to change the objectives if they are too high.

When entering the PPP agreed-upon in the PEPS website, each GP must be able to order tests to check regularly the progress of their patients. Finally, "once the actions to achieve the goals have been chosen with the patients and entered, one must get a summary of the actions decided, and an appointment must be arranged to check how the PPP is followed."

Ergonomics Specifications for the Solution

An interactive mockup of the proposed PEPS solution was designed including 30 screens (e.g. homepage, dashboard for consumers, dashboard for clinicians, desk for the codesign of the PPP, see Figure 3). Navigation paths were represented using flowcharts.

Five former participants provided feedback on the mockup. Overall, respondents appreciated the mockup. Their main concerns were:

- The phrasing of the initial questionnaire: some questions require explanations or should be gender-specific. Scientific references should be added to improve the credibility of the questionnaire.
- Despite its usefulness, the "practical information" screen is too cumbersome. The information density should be reduced.
- Links to other relevant websites should be added.

Discussion

The main objective of the PEPS project was to design a usable tool supporting the co-elaboration of PPP for CVD. The design approach adopted in the project relied on a user driven innovation process to empower end users and give them an active role in the design of the web solution. The "studio design" method involved GPs, nurses, and health care consumers at risk for CVD. As a result, these end users produced mockups of the solution. Even if participants were not used nor trained to graphical user interface techniques, they proved to be able to draw relevant basic mockups, with the information organized within an arrangement. Working on problems they could face during the co-elaboration of PPP helped the participants imagine and design solutions that could help them.

The user driven innovation strategy places the users at the core of the innovation process in a systematic way and facilitates the integration of their knowledge into the increasingly complex design process [17]. Participants' profiles noticeably impacted the design process and its results. Focus group one involved health care consumers, a very heterogeneous population in terms of technology and health literacy, background, needs etc. They proceeded mostly (i) by listing and accumulating different needed information and features and then (ii) by prioritizing those elements to achieve a consensus on the relevant elements the web solution should include. Clinicians who took part in focus group two proceeded differenty. Since they are a quite homogenous group (more than health care consumers even if it represented both GPs and nurses), they expressed similar needs and did not have to reach a consensus on the information and features to display. Consequently, all along the mockups refining process, they questioned in detail how the mockups they produced would actually work *in situ*; they identified blocking elements and found solutions to fix them. Finally, involving both health care consumers and clincians in focus group three enabled them to discuss the PPP design process and to exhange on their respective needs and constraints. The common thinking participants had on the co-elaboration of the PPP allowed them to go deeper than designing the graphical user interfaces of the web solution. Adopting a user driven design process enabled them to design not only a technology but, to a larger extend, a socio-technical system [18].

Comments we get about the online interactive mockup were considered to improve it; an improved version was presented with related specifications to developers. The PEPS web solution is now under development. Once available, its usefulness and its usability will be evaluated with representative end users other than the ones involved in the design process. Then, the solution will be deployed in pilot sites to evaluate its medium and long-term usage. To ensure the PEPS platform is well-accepted and therefore may support the long term prevention of CVD, attention will be paid to acceptability factors by exploring how well the platform is integrated into daily life, how it contributes to transform clinicians' and health care consumers' activities, and to what extent the solution fits end users' personal and social values. [19]

Conclusions

The PEPS project aims to develop a web solution that supports the co-elaboration of the PPP by health care consumers and clinicians. The applied user driven innovation approach enabled to design an interactive mockup of the proposed PEPS solution including 30 screens. This mockup supports the process of co-elaboration of PPP. Ultimately, it must help increase the likelihood that health care consumers adhere in the long run to PPP for cardiovascular diseases.

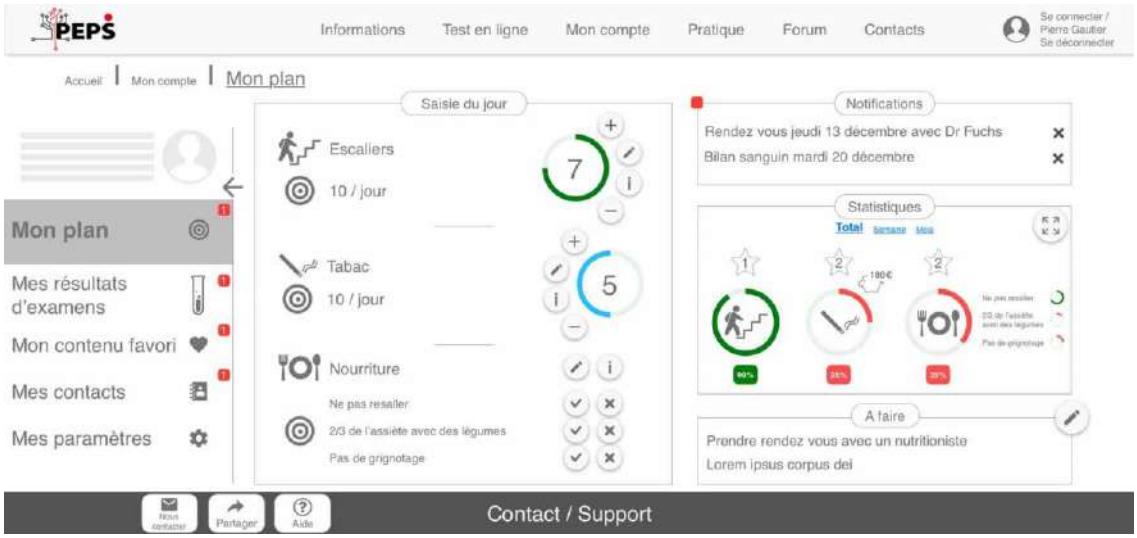


Figure 3– Screens of the Mockup Redesigned by the Experts in Ergonomics - The View the Consumers Have of Their Dashboard

Acknowledgements

This research was funded by the French National Research Agency (ANR), grant number ANR-16-CE19-0018.

References

[1] Rigal L, Saurel-Cubizolles M-J, Falcoff H, Bouyer J, Ringa V. Do social inequalities in cervical cancer screening persist among patients who use primary care? The Paris Prevention in General Practice survey. *Prev Med.* **53** (2011), 199-202.

[2] Klein M, Mogles N, van Wissen A. Intelligent mobile support for therapy adherence and behavior change. *J Biomed Inform.* **51** (2014), 137-151.

[3] Shojania KG, Jennings A, Mayhew A, Ramsay C, Eccles M, Grimshaw J. Effect of point-of-care computer reminders on physician behaviour: A systematic review. *CMAJ Can Med Assoc J J Assoc Medicale Can*. **182** (2010), E216-225.

[4] Jeffery R, Iserman E, Haynes RB, CDSS Systematic Review Team. Can computerized clinical decision support systems improve diabetes management? A systematic review and meta-analysis. *Diabet Med J Br Diabet Assoc.* **30** (2013), 739-745.

[5] Vodopivec-Jamsek V, de Jongh T, Gurol-Urganci I, Atun R, Car J. Mobile phone messaging for preventive health care. *Cochrane Database Syst Rev.* **12** (2012).

[6] Haug S, Schaub MP, Venzin V, Meyer C, John U. Efficacy of a text message-based smoking cessation intervention for young people: A cluster randomized controlled trial. *J Med Internet Res.* **15** (2013), e171.

[7] Shaw RJ, Steinberg DM, Zullig LL, Bosworth HB, Johnson CM, Davis LL. mHealth interventions for weight loss: a guide for achieving treatment fidelity. *J Am Med Inform Assoc JAMIA*. **21** (2014), 959-963.

[8] Damiani G, Pinnarelli L, Colosimo SC, Almiento R, Sicuro L, Galasso R, et al. The effectiveness of computerized clinical guidelines in the process of care: a systematic review. *BMC Health Serv Res*. **10** (2010), 2.

[9] Conroy DE, Yang C-H, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. *Am J Prev Med.* **46** (2014), 649-652.

[10] Redfern J, Usherwood T, Harris MF, Rodgers A, Hayman N, Panaretto K, et al. A randomised controlled trial of a consumer-focused e-health strategy for cardiovascular risk management in primary care: The Consumer Navigation of Electronic Cardiovascular Tools (CONNECT) study protocol. *BMJ Open*. **4** (2014), e004523.

[11] Rupert DJ, Squiers LB, Renaud JM, Whitehead NS, Osborn RJ, Furberg RD, et al. Communicating risk of hereditary breast and ovarian cancer with an interactive decision support tool. *Patient Educ Couns*. **92** (2013), 188-196.

[12] Holbrook A, Pullenayegum E, Thabane L, Troyan S, Foster G, Keshavjee K, et al. Shared electronic vascular risk decision support in primary care: Computerization of Medical Practices for the Enhancement of Therapeutic Effectiveness (COMPETE III) randomized trial. *Arch Intern Med.* **171** (2011), 1736-1744.

[13] Kerns JW, Krist AH, Longo DR, Kuzel AJ, Woolf SH. How patients want to engage with their personal health record: a qualitative study. *BMJ Open*. **3** (2013), e002931.

[14] Pierson J. Users as Innovators. Unpublished presentation in the framework of a COST298 STSM Melbourne workshop. 11 December 2008.

[15] Sullivan BK. The Design Studio Method: Creative Problem Solving with UX Sketching (2015). Routledge.

[16] Cooper A, Reimann R, Cronin D, Noessel C, About Face: The essentials of interaction design. John Wiley & Sons. (2014).

[17] De Moor K, Berte K, De Marez L, Joseph W, Deryckere T, Martens L. User-driven innovation? Challenges of user involvement in future technology analysis. *Science and Public Policy*, **37** (2010), 51–61.

[18] Coiera E. Putting the technical back into socio-technical systems research. *Int J Med Inform.* **76** (2007), S98-103.

[19] Barcenilla J, Bastien JMC. L'acceptabilité des nouvelles technologies : Quelles relations avec l'ergonomie, l'utilisabilité et l'expérience utilisateur ? *Le travail humain* **72** (2009), 311-331.

Address for Correspondence

sylvia.pelayo@univ-lille.fr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190440

Physician Perspectives on Training for an EHR Implementation

Claude J. Pirtle[a], Rollin R. Reeder[b], Christoph U. Lehmann[a], Kim M. Unertl[a], Nancy M. Lorenzi[a]

a Department of Biomedical Informatics, Vanderbilt University, Nashville, Tennessee, USA
b Department of Internal Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA

Abstract

In 2017, a US academic medical center switched to a commerical EHR system using the "specialist training the specialist" model, which combines peer-to-peer training, class-room based training, and web-based training. We conducted semi-structured interviews with physicians at multiple training levels to investigate the impact of this EHR switch and to explore the training experience of physicians and their perception of the training quality pre and post Go-Live. Our team used Grounded Theory methodology to classify the interview information. Themes that emerged from the interviews included stress and anxiety, the desire for more realistic training environments tailored to specialty needs, and concerns about the duration of time between training and implementation. In future implementations, we recommend more data-rich test patients and the demonstration of real-world workflows during training.

Keywords:

Electronic Health Record, Physician

Introduction

Physicians worldwide are concerned about the complexity of today's electronic health record (EHR) systems even though the technology has been available and evolving for more than 50 years. [1] Organizations use multiple techniques to prepare and train clinicians for an impending electronic health record implementation from remote phone training, class-room training, role-based training, web-based training, [2] and the "*specialist training the specialis*t" training model. A 2015 study showed that interns, who had a physician led, skills based EHR curriculum focusing on case-based simulation and skills rather than in-class or even online lectures during on-boarding, performed better on an EHR skills test and rated their training as more useful. [3] A critical question for health care facilities around the world is how to train physicians effectively to use new electronic health record systems.

Our health care system (Vanderbilt University Medical Center (VUMC)) in Nashville, Tennessee, USA selected a replacement system for the existing EHR. This change provided an excellent opportunity for a pre and post implementation study of physicians' perspective of the selected training method.

To judge the success of the implementation, our research effort focused on the perception of physicians regarding their training before and after implementation of the new electronic health record system. VUMC elected to use the "*specialist training the specialis*t" model for implementation. This model combines peer-to-peer training, class-room based training, and web-based training. Due to a paucity of research focused specifically on physician training, we chose to focus the research on the physician's perception of training effectiveness, the effect of new electronic health record system on physicians, and the perceptions of the similarities and differences of the legacy home-grown EHR with the new commercial electronic health record system before and after the implementation.

Methods

In the context of switching to a new electronic health record system, the research goal was to explore physician perception of the training experience and of the training quality pre and post Go-Live. Our sampling goal was to recruit a balanced cohort of residents, fellows, and attendings with the intent to interview 40 participants in each data collection phase. We also collected demographic data for the physicians including medical specialty, gender, age, and training level.

The research process consisted of a written survey and structured interviews with a cross section (senior physician, fellows and residents) of physicians. The pre and post Go-Live surveys/interviews consisted of five topics:

- The physician's expectation for using the new system with patients.
- Use of alternatives to the training to prepare for the new system, and if so, which alternatives were used?
- Anticipations of advantages using the new system with patients.
- Anticipations of disadvantages using the new system with patients.
- Expectations about how the technology might affect the physician on a personal basis.

The actual pre and post Go-Live questions used during the semi-structured interviews are presented in **Figure 1**.

The pre Go-Live interviews occurred in October 2017 (one month prior to the EHR Go-Live) and the post Go-Live interviews occurred in February 2018 and continued into March of 2018 (approximately three to four months post EHR implementation). The selection of physician participants was based on a convenience sampling, and the direct interview process was conducted with the same set of physicians. The data collected from the interviews were classified according to the Grounded Theory methodology.[4]

The Grounded Theory methodology allows respondent answers to be coded by one author and validated by a second (RR and

CP). The codes were then evaluated for central themes. Questions about perception of the training and overall implementation were identified as being amenable for more quantitative analysis.

We focused on physician perception of training pre and post Go-Live and how technology affected the physician personally. We also focused on advantages and disadvantages of the EHR compared to the previously used proprietary system. We also collected the thoughts about training that could have improved their post-implementation work-flow.

Results

During the pre Go-Live phase, we interviewed 37 physicians. This cohort consisted of 21 residents, 3 fellows, and 13 attending physicians from a variety of specialty disciplines. Internal Medicine provided the majority of respondents with 21 physicians. Other specialties included Pulmonary and Critical Care with 5 respondents, Emergency Medicine with 4, Internal Medicine and Pediatrics with 2, Cardiology with 2, Pediatric Emergency Medicine with 2, and Pediatrics with 2 physicians. Our respondents' age range varied, however most of our cohort fell into the 26-30 category with 20 physicians recording their age in this range which accounted for 54 percent of our cohort and a median age range of 26-30. Other categories consisted of 10 physicians in the 31-35 category, 1 in the 36-40 category, 3 in the 41-45 category, 1 in the 56-60 category, 1 in the 61-65 category, and 1 in the 66-70 category.

During the post Go-Live phase, the cohort included 27 physicians: 18 residents, 0 fellows, and 9 attending physicians. The research goal was to interview the same people pre and post implementation. However, 10 physicians were not interviewed in the post Go-Live phase of the study secondary to issues such as schedule conflicts and the inability to speak with us for an extended period of time. Therefore, we were unable to connect with everyone who was originally interviewed. Internal Medicine provided the majority of the cohort with 19 respondents. Other specialties included Cardiology (2), Pediatric Emergency Medicine (2), Pediatrics(1), Pulmonary Critical Care, Internal Medicine-Pediatrics dual appointment, and adult Emergency Medicine all with 1 each.

The respondents' age range also varied in the post Go-Live interviews, however most of our respondents fell into the 26-30 category with 16 physicians recording their age in this range which accounted for 59 percent of our cohort and a median age range of 26-30. Other categories consisted of 7 physicians in the 31-35 category, 1 in the 36-40 category, 1 in the 41-45 category, 1 in the 61-65 category, and 1 in the 66-70 category. (See **Table 1** for a detailed list of respondents).

For the extensive open-ended grounded theory analysis, the codes generated and the number of times they were applied are provided in the ***Figures and Graphs*** section. **Figure 2** shows that the physicians' perception of the training received turned from an overall positive opinion prior to Go-Live to a decidedly negative opinion in the post Go-Live period. The initial opinion began at 54% positive prior to Go-Live, declining to 19% post Go-Live. **Figure 3** shows that the perception of the impact of the EHR technology change turned from a more negative perception before to a more neutral after Go-live with 73% negative opinion declining to 50%.

To illustrate how perceptions of training and EHR experience changed by individual, **Tables 2** and **Tables 3** compare the quantitatively post Go-Live perception to the pre Go-Live response. Perception of training shifted markedly towards the negative for respondents, who had been neutral and positive pre Go Live with 63% and 57% of respondents switching from positive or neutral to native responses. However, the perception of the technology change generally improved with 35% of respondents who had a negative initial response switching to a positive response.

End-User Perceptions of the Technology Change

Based on training experience, 85% of clinicians were positive or neutral prior to Go-Live. We heard comments such as: "That [Epic] is an EMR that I will now be using; Pretty neutral as far as my opinion on that...I'm sure it will be better in some ways and worse in others."

After implementation, there were 65% of respondents who felt negatively regards to the training received. They tended to focus on concerns about better tailoring the training to the actual needs of the clinician. "The training I did with e-Star covered a lot of different Epic functions but not many of them that I use in my clinical practice. It was a broad training but not tailored [to what] I do in my daily work. Currently [I use] stuff [that] I learned in the weeks post, not [in] the training […] beforehand...Having a very broad training applicable to clinicians did not apply very specifically to my work flow in the ICU."

A significant concern related to the need for more practical, on the job training especially immediately after implementation was expressed as: "It seemed like we needed [to do] actual work … to feel out what was necessary. I wish there were more knowledgeable people after Go-Live. Too many people with only half of the right answer."

Desired Training Alternatives

The most common concern regarding training before and after implementation was the lack of verisimilitude of the training. The training provided to all clinicians included a walk-though of a general medicine inpatient, outpatient, and emergency department encounters. Physicians identified gaps in training as compared to their regular activities. One respondent commented, "I do wish they had a class where they would have us admit a patient from step 1 in the playground environment. I felt like 'talking about it and half-way doing it' was not enough."

The need for better personalization of the vendor's system was another major concern. One respondent stated: "It would have been nice to have set up notes, order sets, etc. We did have personalization[,] but I felt like I was left on my own and did not really get the help I needed and had to try and figure things out for myself."

Benefits of New System

Except for two physicians, all respondents described some benefits of the new EHR usually focusing on improved clinical and coding efficiencies. There was little change in perception over time. About half of providers cited increased clinical efficiency as a benefit and the remainder focused on a variety of benefits including improved communication, mobile interface, and billing. However, 15% of clinicians felt there were no benefits to the new system after implementation: " that's a good question...I cannot think of anything...there is literally nothing."

Disadvantages of New System

Prior to implementation, the majority of physicians felt the most pressing disadvantage was the cost associated with changing/transitioning systems: "I think the transition will be rough, […] but in there will be probably some changes to the workflow but overall I think it will be better." After the transition, the physicians' focus was mostly on work-flow

disruption and an increased burden of documentation and the resulting increased time spent interfaced with the EHR. "Too many clicks. Especially in outpatient. When you are trying to write a note, you click on a problem, then assessment, and plan...just let me click on one thing."

Personal Impact of Technology Change

Most clinicians felt they were significantly affected by the technology change, both before and after implementation. Before implementation, anxiety and concern about the transition itself were prominent: “It has caused some anxiety over if it will slow me down in the ER and make it more difficult for me to see patients efficiently.” These nonspecific concerns became more focused on a variety of ways that the system had altered healthcare delivery making processes more and less difficult. Even several months post implementation, many clinicians still expressed stress from the transition: "I finally empathize with all the headlines from 10 years ago with all of the doctors who were 60 years old who thought that it was just too much of a lift and retired."

An especially concerning sentiment was a loss of confidence in the clinician’s ability to deliver reliably high-quality care: "I think in the first couple of months, it was slowing down note writing and chart reviewing although that has gotten better. Ordering is slower and messaging is much more difficult… I don't have confidence that what I am ordering is actually happening."

Discussion

Interviews with physicians before and after implementation were conducted to explore possible opportunities to improve the experience of implementation of a new EHR prior and after the implementation. The physicians interviewed expressed significant concerns in the pre implementation period on how the new EHR would affect productivity and work-flow. While concern is natural given the significant role of the EHR in the modern medical environment, it was notable that there were more concerns prior to implementation than afterwards suggesting that the fear of change was worse than the change itself.

As physician expectations for the new system were low after training and before implementation, an opportunity for improvement may include better communication of the benefits of the new system prior implementation. The importance of robust change management strategies has previously been identified in the implementation of an EHR from paper, but our results indicated its continued importance even when transitioning between EHRs.[5, 6]

Part of the EHR implementation challenge is that user experience varies widely, which can make a workflow solution for one physician disruptive to another. Further, the differing perspectives and priorities of different specialties and provider types challenge any communication effort and make it difficult to train a large array of specialties prior to the implementation. As a result, many work flow solutions had to be discovered by providers once the system was operational.

The nature of the training emerged as a primary concern for our physicians. One limiting factor for the training process included the fact that the system was not completely built while physicians underwent training thus omitting important workflows or features that were finally available when the system went live. This issue created physician frustration with functionality not having been discussed or experienced in training, functionality not working as described in training, or tools not performing appropriately at Go-Live. Further, many users expressed a desire for more realistic simulations of their work-flow during training to allow them expectation setting for Go-Live. This concern was described previously with physician training[7] and emerged in our interviews as one of the most substantial challenges for physicians.

Ideal training would walk physicians accurately through a day in their life in a data-rich practice with complex patients. In our training, the lack of integration of specifically tailored workflows in a specialty clinic or an intensive care unit provided physicians with only a very generalized simulation of the clinical environment lacking verisimilitude to clinical reality and omitting critical details important to efficiently discharge their documentation, ordering, and other duties. While increased training specificity would have further magnified the logistical challenge of training, its lack was found to be frustrating and likely reduced the impact on physician preparedness for clinical workflow optimization and documentation.

The final major themes were the proximity of the training to Go-Live. Many physicians expressed that the long period of time between training and Go-Live resulted in a reduced retention of learned and practiced content by the time implementation took place. Unfortunately, in a large organization, the logistical challenge may be insurmountable to train every end user immediately prior to Go-Live, but physicians suggested an abbreviated refresher course the week prior to Go-Live. While those with a negative opinion of their training prior to Go-Live continued to feel that way after, it is notable that there was a significant shift in perception among physicians, who felt positive pre Go-Live, to negative afterwards.

Conclusions

In conclusion, many physicians felt passionate about the training they received and the EHR technology transition. Training needs to reflect the experience of physicians in their day to day lives and must be in a narrow time window to implementation to be recalled and be impactful. The more practical and tailored to clinical reality the training can be, the more it is appreciated by physicians. Otherwise, as was experienced in this implementation, physicians will experience significant frustration with the quality of training and may develop resentment towards the technology change.

Study Limitations

Our study was limited to the implementation of one large vendor system at a large academic medical center. Because in-person interviews require a substantial amount of time and resources, we had to limit the number of participants in the study and had difficulty scheduling interviews with the same individuals post-implementation. We also were not able to fully represent all specialties in our research sample (e.g. surgical specialties), which may have provided additional perspectives on training. Online surveys may be helpful in broadening the reach of this type of research, although the in-person interviews helped with gaining a deeper understanding of physician experiences and perceptions.

Figures and Tables

Research Questions for Physicians e-Star Implementation – Pre Go-Live

1. After your training and receiving information about e-Star, what is your perception about using e-Star with patients?
2. What other options do you wish you would have experimented with to prepare for the implementation of e-Star?
3. If you had one wish of how e-Star could improve throughput/workflow post implementation, what would it be?
4. What are your expectations of the advantages with using e-Star with patients? What are your expectations of the disadvantages with using e-Star with patients?
5. Tell me a little more about how this implementation and technology change is affecting you.

Research Questions for Physicians e-Star Implementation – Post Go-Live

1. After using e-Star for a little over a month now, what is your perception of your training and preparation for the implantation?
2. Are there any other avenues you wish you would have experimented with to prepare for the e-Star implantation?
3. After using e-Star over the past few weeks (or months), what would you have changed about your training to improve your post-implementation work-flow?
4. What are the advantages or disadvantages you noted when using e-Star with patients over the past few weeks (or months)? How does this compare with StarPanel or is it similar to StarPanel?
5. Tell me a little more about how this implementation and technology change is affecting you.

Figure 1- Research Questions

Table 1- List of All Study Participants by Specialty, Position, Gender and Age Range

Specialty	Position	Gender	Age range
Internal Medicine	Attending	F	26-30
Pulmonary and Critical Care	Attending	M	41-45
Emergency Medicine	Attending	M	56-60
Emergency Medicine	Attending	M	26-30
Pulmonary and Critical Care	Attending	F	41-45
Internal Medicine	Attending	F	26-30
Cardiology	Attending	F	36-40
Cardiology	Attending	M	66-70
Pulmonary and Critical Care	Attending	M	31-35
Peds Emergency Medicine	Attending	F	41-45
Peds Emergency Medicine	Attending	M	61-65
Pediatrics	Attending	F	31-35
Internal Medicine	Attending	M	31-35
Pulmonary and Critical Care	Fellow	F	31-35
Pulmonary and Critical Care	Fellow	M	26-30
Emergency Medicine	Fellow	F	31-35
Internal Medicine	Resident	M	26-30
Internal Medicine	Resident	M	26-30
Internal Medicine	Resident	F	31-35
Internal Medicine	Resident	M	26-30
Internal Medicine	Resident	F	31-35
Internal Medicine	Resident	M	26-30
Internal Medicine	Resident	F	31-35
Internal Medicine	Resident	F	26-30
Emergency Medicine	Resident	M	26-30
Internal Medicine	Resident	M	31-35
IM/Peds	Resident	F	26-30
Internal Medicine	Resident	M	26-30
Internal Medicine	Resident	F	26-30
Internal Medicine	Resident	M	26-30
Internal Medicine	Resident	M	26-30
Internal Medicine	Resident	F	26-30
IM/Peds	Resident	F	26-30
Internal Medicine	Resident	F	31-35
Internal Medicine	Resident	F	26-30
Internal Medicine	Resident	M	26-30
Internal Medicine	Resident	M	26-30

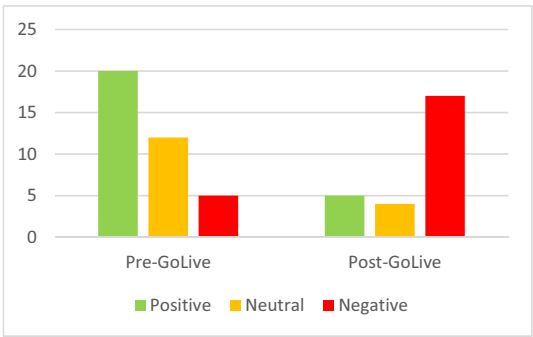


Figure 2- Question 1: Perception of Training

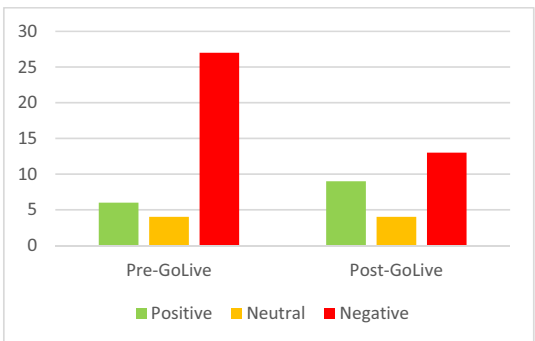


Figure 3- Question 5: Effect of Transition

Table 2 - Quantitative Comparison of the Post Go-Live Perception to the Pre Go-Live Response

		Post Go-Live		
		positive	neutral	negative
Pre Go-Live	positive	4	2	10
	neutral	1	2	4
	negative	0	0	3

Table 3 - Quantitative Comparison of the Post Go-Live Perception to the Pre Go-Live Response

		Post Go-Live		
		positive	neutral	negative
Pre Go-Live	positive	1	1	1
	neutral	1	1	1
	negative	7	2	11

Acknowledgements

Our research team would like to thank all of the study participants for their valuable time during the interview process.

References

[1] Koppel R, Lehmann CU. Implications of an emerging EHR monoculture for hospitals and healthcare systems. J Am Med Inform Assoc. 2015 Mar;22(2):465-71.

[2] M. Tariq Dastagir, et. al. "Advanced Proficiency EHR Training: Effect on Physicians' EHR Efficiency, EHR Satisfaction and Job Satisfaction." *AMIA Annu Symp Proc.* 2012; 2012: 136-143.

[3] Airan-Javia S, Myers J, Bellini L. A Physician Led, Skills-Based Ehr Training Curriculum Can Improve Ehr Skills for Interns [abstract]. *Journal of Hospital Medicine.* 2015; 10 (suppl 2). https://www.shmabstracts.com/abstract/a-physician-led-skills-based-ehr-training-curriculum-can-improve-ehr-skills-for-interns/. Accessed July 29, 2018.

[4] Miller, D.C., & Salkind, N.J. (2002). Handbook of research design & social measurement. Thousand Oaks, CA: Sage.

[5] Boonstra, A; Broekhuis, M. "Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions." https://bmchealthservres.biomedcentral.com/articles/10.1186/1472-6963-10-231. *BMC Health Services Research* 2010 10:231. 6 August 2010.

[6] Lehmann C.U., Unertl K.M., Rioth M.J., Lorenzi N.M. (2016) Change Management for the Successful Adoption of Clinical Information Systems. In: Finnell J., Dixon B. (eds) Clinical Informatics Study Guide. Springer, Cham

[7] Bredfeldt et al.: Training providers: beyond the basics of electronic health records. *BMC Health Services Research 2013*. 13:503.

Address for correspondence

Claude J. Pirtle, MD
Department of Biomedical Informatics
Vanderbilt University
2525 West End Avenue
Suite 1475
Nashville, TN 37203
E-mail: claude.j.pirtle@vumc.org

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190441

The Barriers and Facilitators for Nurse Educators Using Telehealth for Education

Maria Prendergast[a], Michelle Honey[b]

[a] *Nurse Educator, Thames Hospital, Waikato District Health Board, Waikato, New Zealand*
[b] *School of Nursing, University of Auckland, Auckland, New Zealand*

Abstract

Telehealth is available world-wide and in addition to clinical uses, it can also be used to provide education for health professionals, supporting e-Networking. However, within New Zealand the uptake and widespread use of telehealth remains low, and why telehealth is not used more is not understood. This study describes nurse educators use of telehealth for education and identifies barriers and facilitators to increase the uptake of telehealth amongst nurse educators. An on-line survey administered using Survey Monkey had a response rate of 57% (n=19). Equipment that was not user friendly and a lack of initial training were recognised as barriers to their uptake of telehealth. Telehealth training and support, and local champions were identified facilitators to increase the uptake of telehealth. Recommendations include the need for early adopting nurse educators to be recognised and encouraged, to role model good practice in telehealth, and mentor and support others.

Keywords:

Surveys and Questionnaires; Telemedicine; Faulty, Nursing

Introduction

Telehealth and telemedicine are terms often used interchangeably to describe health services provided across distance to both healthcare providers and their clients [1]. Telehealth is available world-wide to support long distance health care and education to both patients' and health professionals [2] but internationally, the uptake and use of telehealth remains disappointing [3]. Within the European Union the expectations of both policy makers and stakeholders fall short [4], whilst in Norway the use of their telehealth system and services is much lower compared to face-to-face options [5]. In Canada, however, telehealth is rapidly increasing in quality, accessibility and popularity [6]. Within New Zealand (NZ), the introduction of ultra-fast broadband is expected to increase the accessibility of telehealth for New Zealanders [7]. To maximise the true potential for telehealth, it is necessary to understand the barriers and facilitators that will lead to an increase in telehealth uptake, not just for service delivery, but also for health professional education.

Background

Internationally, there remains a preference for face-to-face education when compared to providing education via videoconferencing [8]. The term video-conferencing describes real-time, synchronous, two way transmission of digitalised audio and video images between two or more locations [9]. The use of telehealth for education has not developed at the pace and scale expected [10]. The traditional method of gathering as one central group for education does not meet the needs of rural health professionals [11], hence the growth in popularity in the use of telehealth to support their ongoing education. Barriers to the uptake of telehealth include funding, time, infrastructure, skills and preference for the traditional face-to-face approach for education and service [12]. Santos [13] recognises the lack of managerial support, lack of access to and relevance of education sessions, and also workplace culture can be barriers to nurses accessing education via telehealth.

Within NZ, the Ministry of Health has introduced both the NZ Health Strategy and the Health Information Strategy to guide the regional District Health Boards (DHBs) in their introduction and use of telehealth. Despite 17 of the country's 20 DHBs actively using telehealth, a national stocktake in 2014 found the momentum did not reflect the true potential available [14]. This stocktake also recognised the most used telehealth equipment was videoconferencing, which can be used as an adjunct to support health professional education [14]. Yet issues about the use of telehealth, and specifically videoconferencing for educational use, within the DHBs have not been explored. This study will investigate such issues, and form recommendations for future use.

Methods

To explore barriers and facilitators for nurse educators using videoconferencing, a mixed methods approach was selected as it can provide descriptive data including demographics, attitudes and behaviour from a selected group [15]. The participants, employed within one District Health Board in New Zealand, were using videoconferencing as part of their education role. With a population of over 400,000 people, this District Health Board represents the fourth largest region in New Zealand, but over 60% of the people live outside the main city, hence there are a number of small satellite rural hospitals. The nursing staff in these rural hospitals can be disadvanatged by being unable to access continuing nursing education.

For data collection, a questionnaire was created with 13 questions developed, drawing on the available literature. The questions were piloted by three colleagues who did not participate in the survey, and their feedback was used to refine and clarify the questions, finalise the order of questions and also to determine the time needed to complete the survey.

Ethical approval was gained from the University Ethics Committee (Ref. 018105) and the associated DHB (RD: 016114) before the questionnaire was distributed on-line in late 2016, using a commercial survey tool, Survey Monkey.

Survey Monkey was selected for the on-line survey because it was easy to use and familiar to respondents. An email with the link to the survey was sent to all 33 nurse educators who work within one DHB. One reminder email was also sent.

Collation of the responses occurred automatically. Analysis of the numerical data was completed using an Excel spreadsheet, with graphs to present the data. The responses to the open-ended question "Please add any further feedback you would like to share" were thematically analysed based on the process described by Braun and Clarke [16]. The results presented below report the demographics, nurse educators' experience of telehealth, and factors that impact on their engagement with telehealth.

Results

In total, 19 out of 33 nurse educators completed the survey, giving a response rate of 57%.

Demographics

The nurse educators ranged from 31 to over 50 years of age, and the majority (74%) had more than 15 years nursing experience (Table 1).

Table 1- Nurse educator demographics (n = 19)

Demographic	n	%
Age		
< 30 years	0	0
31-40 years	5	26
41-50 years	10	53
>51 years	4	21
Experience as a nurse		
< 5 years	0	0
6-10 years	1	5
11-15 years	4	21
>16 years	14	74

Experience of telehealth

The majority, (90%, n=17) of the nurse educators had not experienced telehealth during their initial nursing education. When asked about their current use of telehealth in their role, nearly two-thirds (63%, n=12) of the nurse educators responded that they were not using telehealth and 31% (n=6) responded that they were using telehealth to some degree; and one respondent (5%) was uncertain. However, nearly 80% (n=13) of the nurse educators felt telehealth was "somewhat important" or "important" for their role (Figure 1).

Of the 19 respondents, 42% (n=8) reported having had "a little" experience with telehealth. Although 31% (n=6) of the nurse educators had either "some" or "quite a lot" of experience, there was still 26% (n=5) of the group who had not experienced telehealth at all. When asked to rate their understanding of telehealth, the majority of respondents (n=13, 68%) reported being "knowledgeable" or "somewhat knowledgeable". There is however, 32% (n=6) of the group who considered themselves "not knowledgeable" at all in the use of telehealth. Furthermore, when asked about the importance of gaining competence in using telehealth, two thirds (67%) believed gaining competence is either "important" or "very important".

Although no respondents felt gaining competence was not important at all, there was still nearly one-third (33%, n=6) of the group who see gaining competence as only "somewhat

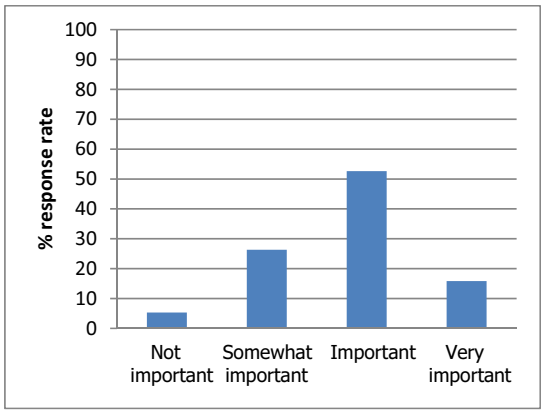


Figure 1- Importance of telehealth for nurse educator role

important."

Respondents were also asked to rate their skill when engaging in telehealth as either novice, competent, proficient or expert, based on the scale of Benner's terms [17]. Approximately two thirds (68%, n=13) of the nurse educators viewed themselves at novice level; 21% (n=4) as competent, and 11% (n=2) rated themselves as proficient. No one considered they are practicing telehealth at an expert level.

Factors for engaging with telehealth

A list of factors that influence staff engagement with telehealth was compiled from the relevant literature, and respondents were asked to identify any factors that applied to them. There was only one statement that all respondents indicated and this was for telehealth equipment to be easy-to-use and reliable (Figure 2). The second most important factor was the need for training and support, with strong leadership and the need to recruit local champions being viewed as third most important. The least important factors identified by the respondents were negative impact on service change and negative impact on staff.

Themes from further comments

Only seven (37%) respondents made further comments in the free text question: "Please add any further feedback you would like to share". To analyse these comments, words with the same or similar meaning were bracketed, then grouped, and from this three themes emerged. These were: engagement in telehealth, the importance of telehealth, and technology.

Engagement in telehealth was the most common theme to emerge in the comments, such as "*Will need to engage in order to be effective*" and, "*I will engage more and more over time*". The importance of telehealth was another theme, with statements such as: "*Telehealth use is very important*". The third theme indicated feelingsabout the importance of telehealth technology, with comments such as, "*vital technology*" and "*equipment and technology is a positive change*".

The positive comments saw telehealth as "*the way of the future*", "*will achieve bang for our educational buck*" and "*likely to improve patient outcomes*". However, the challenges identified by the respondents included the need to "*grow in confidence*" and "*it will take time and support to make it work*".

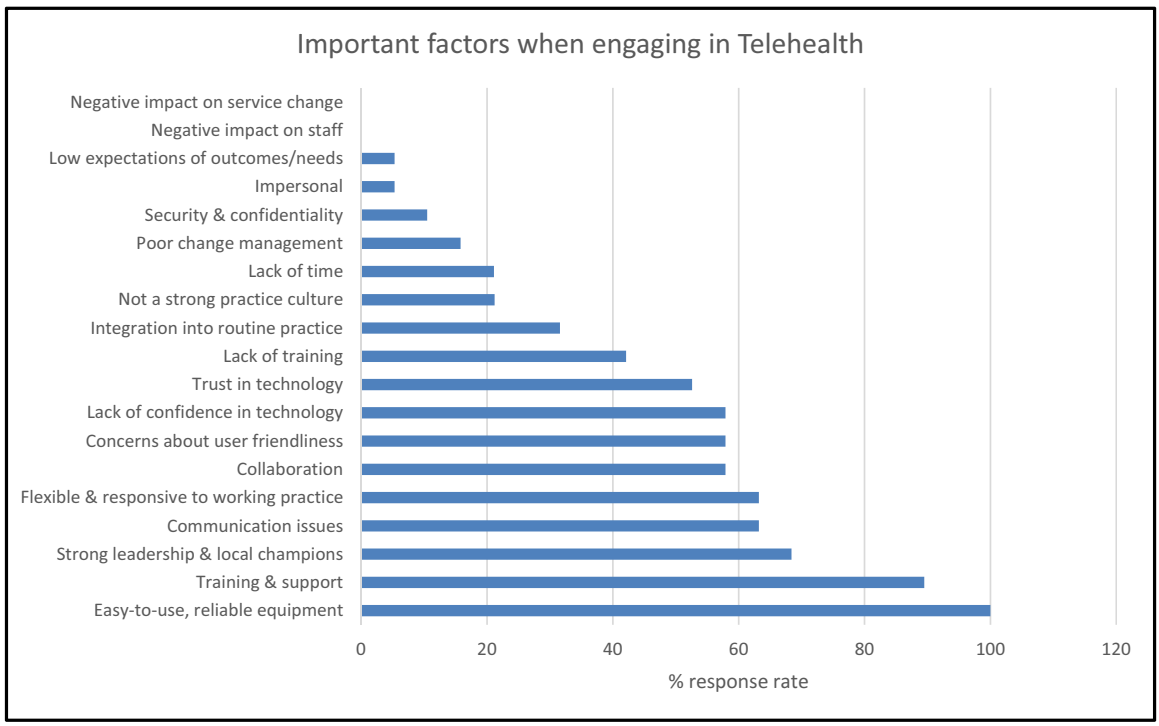


Figure 2: Important factors when engaging in telehealth

Discussion

The results of this survey provide insight into nurse educators' use of telehealth within one DHB. Overall, the results reflect results found in the literature concerning the lack of telehealth uptake, but there is evidence that some of the nurse educators are keen to engage and use telehealth within their role. Telehealth will never remain static, rather it is becoming an increasingly essential element of healthcare [18], so it is important for the nurse educators to become further involved as it provides another format for the provision of education to nurses. Additionally, telehealth has been shown to improve access to educational opportunities for rural and isolated health professionals [11,12].

Using telehealth

Telehealth equipment which is reliable and easy-to-use was recognised as the most important factor in this study, and this is similar to literature that found it to be instrumental in staff acceptance of telehealth [10,19]. Possible issues, echoed in the results of this survey and reflected in the literature, which prevent staff from engaging in telehealth include a lack of adequate equipment [20], and lack of appropriate training [21]. Sligo, Gauld, Roberts, and Villa [22] however, believe there should be a good fit between the technology and the users from the beginning, and the technology should be intuitive so that little staff training is required. Training and support was identified in this study as an important factor for successful implementation of telehealth, both initially and in the longer term. To improve the uptake and sustainability of telehealth use, training should be held in the workplace with staff trained to troubleshoot technical issues [21]. There also needs to be dedicated technical support during the implementation phase [19], as well as later, to consolidate staff knowledge and technological skills [23]. Additionally, providing training and support has been shown to improve staff confidence and engagement in the usage of telehealth [19]. In this study, the nurse educators were not very experienced telehealth users. However, it is essential to acknowledge staff who have more experience as they could be recognised as innovators and possible clinical champions in the future [24].

Leadership and local champions

Successful implementation of telehealth also requires strong leadership and local champions, a finding from this survey and supported by the literature, where champions contribute to the successful implementation [25]. These individuals can be facilitators for the adoption of telehealth [10].

Success of telehealth depends on the competency of the nursing staff [26]. From this survey, nearly two-thirds of the nurse educators felt it was important or very important to gain competence in telehealth, but with a similar proportion rating themselves as a telehealth novice. Based on these findings, further education needs to be planned and implemented. The use of telehealth competencies, including topics such as communication, use of the technology, ethics and coaching skills may be helpful to increase users' skills [27]. Grading the competencies into novice, competent, proficient and expert levels may allow the nurse educators to develop their technical skills in a manner that grows their confidence and competence in a timely manner [28].

Limitations and areas for further research

Although a good response rate for this survey was achieved, there are still nearly half of the nurse educators who did not respond and their perceptions on telehealth remain unknown. The results presented here indicate the experience and perceptions of the nurse educators in one region of New Zealand, therefore repeating this study across a wider area is

warranted. Additionally, further research is needed to explore nurse educators' perceptions as this may indicate strategies that would improve telehealth use for education. A qualitative approach, perhaps using focus group interviews, is recommended which may provide the depth that this study could not.

Conclusions

This study provides insight into the current barriers and facilitators for telehealth use in education from the perspective of nurse educators from one region in New Zealand. Despite low usage and skills with telehealth, this study identifies that most nurse educators are ready to use telehealth to provide health professional education.

The recognition of champions who can act as leaders and assist in the implementation of telehealth is vital to assist in engagement and acceptance by staff. There is also a need to seek early adopters within a health professional group who can role model good practices for telehealth, as well as encourage others to accept telehealth within their daily work. There is a need for educators to have the necessary skills in using telehealth to be able to provide education via this medium.

References

[1] F. Fatehi, R. Wootton, Telemedicine, telehealth or e-health? A bibliometric analysis of the trends in the use of these terms, *J Telemed Telecare* **18,** (2012), 460-464.

[2] A.C. Marttos, F.M. Kuchkarian, P. Abreu-Reis, B.M.T. Pereira, F.S. Collet-Silva, G.P. Fraga, Enhancing trauma education worldwide through telemedicine, *World Journal of Emergency Surgery* **7,** (2012), 1-6.

[3] N.K. Bradford, L.J. Caffery, A.C. Smith, Telehealth services in rural and remote Australia: A systematic review of models of care and factors influencing success and sustainability, *International Electronic Journal of Rural and Remote Health* **16,** (2016), 1-23.

[4] S. Merkel, P. Enste. Barriers to the diffusion of telecare and telehealth in the EU: A literature review. International Conference on Technologies for Active and Assisted Living; 2015.

[5] P. Zanaboni, U. Knarvik, R. Wootton, Adoption of routine telemedicine in Norway: The current practice, *Global Health Action* **7,** (2014), 1-13.

[6] N. Toh, J. Pawlovich, S. Grzybowski, Telehealth and the patient-doctor relationship in rural and remote communities, *Canadian Family Physician* **62,** (2016), 961-963.

[7] Ministry of Business Innovation & Employment, *Broadband and mobile programmes*, 2017.

[8] J. Chipps, P. Brysiewicz, M. Mars, A systematic review of the effectiveness of videoconfence-based tele-education for medical and nursing education, *Worldviews on Evidence-Based Nursing* **9,** (2012), 78-87.

[9] D. Rudel, M. Fisk, R. Roze, Definitions of terms in telehealth, *Informatica Medica Slovenica* **16,** (2011), 28-46.

[10] J. Taylor, E. Coates, L. Brewster, G. Mountain, B. Wessels, M.S. Halwley, Examining the use of telehealth in community nursing: Identifying the factors affecting frontline staff acceptance and telehealth adoption, *Journal of Advanced Nursing* **71,** (2015), 326-337.

[11] P. Lamba, Teleconferencing in medical education: A useful tool, *Australasian Medical Journal* **4,** (2011), 442-447.

[12] J. Moffatt, D. Eley, Barriers to the uptake of telemedicine in Australia - a view from providers, *Rural and Remote Health* **11,** (2011), 1-6.

[13] M.C. Santos, Nurses' barriers to learning, *Journal for Nurses in Staff Development* **28,** (2012), 182-185.

[14] New Zealand Telehealth Forum, *New Zealand Telehealth Stocktake: District Health Boards,* The University of Auckland National Institute for Health Innovation and the Ministry of Health IT Health Board, Wellington, 2014.

[15] P. Moule, M. Goodman, *Nursing research: An introduction.*, Sage Publications., London, England, 2013.

[16] V. Braun, V. Clarke, Using thematic analysis in psychology, *Qualitative Research in Psychology* **3,** (2006), 77-101.

[17] P. Benner, *From novice to expert: Excellence and power in clinical nursing practice*. commemorative ed, Prentice Hall, Upper Saddle River, NJ, 2000.

[18] A.E. Fronczek, N.A. Rouhana, J.M. Mitchin, Enhancing telehealth education in nursing: Applying King's conceptual framework and theory of goal attainment, *Nursing Science Quarterly* **30,** (2017), 209-213.

[19] L. Brewster, G. Mountain, W. Bridgette, C. Kelly, M. Hawley, Factors affecting front line staff acceptance of telehealth technologies: A mixed-method systematic review, *Journal of Advanced Nursing* **70,** (2014), 21–33.

[20] A. Bonney, P. Knight-Billington, J. Mullan, et al., The telehealth skills, training and implementation project: An evaluation protocol, *JMIR Research Protocol* **4,** (2015), e2.

[21] R. Mozer, N.K. Bradford, L.J. Caffery, A.C. Smith, Identifying perceived barriers to videoconferencing by rehabilitation medicine providers, *Journal of Telemedicine and Telecare* **21,** (2015), 479-484.

[22] J. Sligo, R. Gauld, V. Roberts, L. Villa, A literature review for large-scale health information system project planning, implementation and evaluation, *International Journal of Medical Informatics* **97,** (2017), 86-97.

[23] B. Odeh, R. Kayyali, S.N. Gebara, N. Philip, Implementing a telehealth service: Nurses' perceptions and experiences, *British Journal of Nursing* **23,** (2014), 1133-1137.

[24] T. Escobar-Rodríguez, M. Romero-Alonso, The acceptance of information technology innovations in hospitals: Differences between early and late adopters, *Behaviour & Information Technology* **33,** (2014), 1231-1243.

[25] C.M. Shea, C.M. Belden, What is the extent of research on the characteristics, behaviors, and impacts of health information technology champions? A scoping review, *BioMed Central* **16,** (2016), 1-17.

[26] C. Carius, B. Zippel-Schultz, C. Schultz, M. Schultz, T.M. Helms, Developing a holistic competence model for telenursing practice: Perspectives from telenurses and managers of telemedicine service centres, *Journal of International Society for Telemedicine and e-Health* **4,** (2016), 1-17.

[27] C.T.M. van Houwelingen, A.H. Moerman, R.G.A. Ettema, H.S.M. Kort, O. ten Cate, Competencies required for nursing telehealth activities: A Delphi-study, *Nurse Education Today* **39,** (2016), 50-62.

[28] A. Crawford, N. Sunderji, E. Serhal, J. Teshima, Proposed competencies for providing integrated care via tele-psychiatry, *Journal of Technology in Behavioural Science* **2,** (2017), 1-4.

Address for correspondence

Corresponding author Maria Prendergast, email: Maria.prendergast@waikatodhb.health.nz

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190442

Generalizability of Readability Models for Medical Terms

Hanna Pylieva[a], Artem Chernodub[b], Natalia Grabar[c], Thierry Hamon[d,e]

[a] *Ukrainian Catholic University, Faculty of Applied Sciences, Kozelnytska st. 2a, Lviv, Ukraine*
[b] *Grammarly*
[c] *CNRS, Univ. Lille, UMR 8163 - STL - Savoirs Textes Langage, F-59000, Lille, France*
[d] *LIMSI, CNRS, Université Paris-Saclay, F-91405 Orsay, France*
[e] *Université Paris 13, Sorbonne Paris Cité, F-93430, Villetaneuse, France*

Abstract

Detection of difficult for understanding words is a crucial task for ensuring the proper understanding of medical texts such as diagnoses and drug instructions. We propose to combine supervised machine learning algorithms using various features with word embeddings which contain context information of words. Data in French are manually cross-annotated by seven annotators. On the basis of these data, we propose cross-validation scenarios in order to test the generalization ability of models to detect the difficulty of medical words. On data provided by seven annotators, we show that the models are generalizable from one annotator to another.

Keywords:

Natural Language Processing; Terminology; Health Information Systems

Introduction

Specialized areas, such as medical area, convey and use technical words, or terms, which are typically related to knowledge developed within these areas. In the medical area, this specific knowledge often corresponds to fundamental medical notions related to disorders, procedures, treatments, human anatomy, etc. For instance, technical terms like *blepharospasm* (abnormal contraction or twitch of the eyelid), *alexithymia* (inability to identify and describe emotions in the self), *appendicectomy* (surgical removal of the vermiform appendix from intestine), or *lombalgia* (low back pain) are frequently used in the medical area texts. Patients and their relatives usually have some difficulties in the understanding and using of such terms: they show indeed poor health literacy. Some existing studies stressed the difficulty in understanding medical notions and terms by non-expert users, and its impact on a successful healthcare process [1,2]. Yet, it is not uncommon that patients and their relatives must face very technical health documents and information. Examples of this kind are frequent and usually the non-expert users are at loss in such situations: understanding of information on drug intake [3,4], of clinical documents [5], of clinical brochures or informed consents [6], of information provided for patients by different websites [7,8], and communication between patients and medical staff [9,10]. These observations provide the motivation to our work: we address the needs of non-specialized users facing health information and propose to predict the readability of medical words.

In what follows, we first present some related work and introduce the material used as well as the proposed method. We then present and discuss the results. Finally, we conclude with some directions for future work.

Related Work

For studying the readability of medical documents, researchers usually exploit readability measures. Among these measures, it is possible to distinguish classical and computational readability measures [11]. Classical measures usually rely on number of letters and/or of syllables a word contains and on linear regression models [12,13], while computational readability measures may involve vector models and a great variability of features, among which the following have been used for processing the biomedical documents: combination of classical readability formulas with medical terminologies [14]; n-grams of characters [15], stylistic [16] or discursive [17] features, lexicon [18], morphological features [19], combinations of different features [5].

At a more fine-grained level, the readability of words has been addressed much less frequently. In the general language, some research actions are often performed as part of the NLP challenges, such as the SemEval NLP (*www.cs.york.ac.uk/semeval-2012*) challenge held in 2012. This challenge proposed the following task: for a short text and a target word, several possible substitutions satisfying the context have also been proposed. The objective was to rate and to order the substitutions according to their degree of simplicity [20]. The participants applied rule-based and/or machine learning systems. Combinations of various features, designed to detect the simplicity of words, have been used, such as: lexicon from spoken corpus and from Wikipedia, Google n-grams, WordNet [21]; word length, number of syllables, latent semantic analysis, mutual information and word frequency[22]; Wikipedia frequency, word length, n-grams of characters and of words, random indexing and syntactic complexity of documents [23]; n-grams and frequency from Wikipedia, Google n-grams [24]; WordNet and word frequency [25]. The best systems reached up to 0.60 Top-rank and 0.575 Recall. Another work has been done on scholar texts in French written for children with the purpose to differentiate between the texts from various scholar levels and to test various features suitable for that [26]. This system reached up to 0.62 classification accuracy.

In the medical area, we can mention three experiments: manual rating of medical words [27], automatic rating of medical words on the basis of their presence in different vocabularies [28], and exploitation of machine learning approach with various features

[29]. This last experiment achieved up to 0.85 F-measure on individual annotations.

Another issue is to know what are the most suitable data for the analysis of text readability. These data have indeed crucial impact on models created and on their usability. Several approaches have been proposed:

- exploitation of expert judgment, who have an idea on needs of population aimed in the study [30]. The main limitation is that experts may have difficulties to figure out what are the real needs of population;
- exploitation of text books created for population according to their readability levels, such as school books [26]. The main limitation is that such books are usually created by experts using theoretical basis and observations;
- exploitation of crowdsourcing involving large population [30]. The main limitation is that the population involved is uncontrolled and unknown;
- exploitation of eye-tracking methods for a more finegrained analysis of reading difficulties [31,32]. The main limitation is that only short text spans can be used;
- manual annotation by human annotators [33]. In this case, the annotators represent the population, they are part of the controlled population, they can perform more complicated tasks than in case of crowdsourcing, although they are usually less many than in crowdsourcing experiments.

Related to this issue is the question on generalizability of data and of models generated from these data. For instance, it has been observed that data from experts are difficult to generalize over the population [30].

We propose to study the generalizability of the automatic categorization models for a stronger distinction of readability of medical words and distinction of words which may present understanding difficulties to non-experts users. The medical data processed are in French. Seven human annotators participated in creation of the reference data.

Material

The source terms are obtained from Snomed Int [34], which is the most extensive terminology in French, such as available from the ASIP SANTE website (*esante.gouv.fr/services/referentiels/referentiels-d-interoperabilite/snomed-35vf*). Snomed contains 151,104 medical terms organized in eleven axes such as disorders, procedures, chemical products, living organisms, anatomy, social status. For our purpose, we use five axes: disorders, abnormalities, procedures, functions, and anatomy. The assumption is that studying the understanding of these terms is important because they are related to main medical notions and laymen must face then frequently. The 104,649 selected terms are lemmatized and tokenized into words resulting in 29,641 unique words.

The set of 29,641 unique words was annotated by seven French speakers, 25 to 65-year-old, without medical training and without specific medical problems. The annotators are expected to represent the average knowledge of medical words among the population as a whole. The annotators are presented with the list of terms and asked to assign each term to one of the three categories:

- *I can understand the word;*
- *I am not sure about the meaning of the word;*
- *I cannot understand the word.*

The assumption is that terms, which are not understandable by the annotators, are also difficult to understand by patients. The annotators were asked not to use dictionaries during the annotation process. Further to the annotation process, the most frequent category is I cannot understand the word, which gathers between 65 to 70% of terms.

Methods

We propose to tackle the problem through the supervised categorization: the purpose is to categorize terms according to whether they can be understood or not by lay people. The manual annotations provide the reference data. The categorization pipeline is the following: categorization features are computed, they are used for training the classifiers, and the results are evaluated using the cross-validation.

We exploit 11 types of automatically computed features:

- *Syntactic categories*. Syntactic categories and lemmas are computed by *TreeTagger* [35] and then checked by *Flemm* [36]. The syntactic categories are assigned to words within the context of their terms. If a given word receives more than one category, the most frequent one is kept as feature. Among the main categories we find for instance nouns, adjectives, proper names, verbs and abbreviations.
- *Presence of words in reference lexica*. We exploit two reference lexica of the French language: TLFi (*www.atilf.fr/*) and lexique.org (*www.lexique.org/*). TLFi is a dictionary of the French language covering XIX and XX centuries. It contains almost 100,000 entries. lexique.org is a lexicon created for psycholinguistic experiments. It contains over 135,000 entries, among which inflectional forms of verbs, adjectives and nouns. It contains almost 35,000 lemmas.
- *Frequency of words through a non specialized search engine*. For each word, we query the Google search engine in order to know its frequency attested on the web.
- *Frequency of words in the medical terminology*. We also compute the frequency of words in the medical terminology Snomed International.
- *Number and types of semantic categories associated to words*. We exploit the information on the semantic categories of Snomed International.
- *Length of words in number of their characters and syllables*. For each word, we compute the number of its characters and syllables.
- *Number of bases and affixes*. Each lemma is analyzed by the morphological analyzer *Dérif* [37], adapted to the treatment of medical words. It performs the decomposition of lemmas into bases and affixes known in its database and it provides also semantic explanation of the analyzed lexemes. We exploit the morphological decomposition information (number of affixes and bases).

- *Initial and final substrings of the words*. We compute the initial and final substrings of different length, from three to five characters.
- *Number and percentage of consonants, vowels and other characters*. We compute the number and the percentage of consonants, vowels and other characters (i.e., hyphen, apostrophe, comas).
- *Classical readability scores*. We apply two classical readability measures: Flesch [12] and its variant FleschKincaid [38]. Such measures are typically used for evaluating the difficulty level of a text. They exploit surface characteristics of words (number of characters and/or syllables) and normalize these values with specifically designed coefficients.
- *FastText word embeddings* [39] pre-trained on French Wikipedia corpus, which cover up to 56% of the words from our dataset. Tthe embeddings cluster together words that share common contexts and semantics, and can help in generalizing other features over contextually and semantically close words.

The ten first types of features, linguistic and non-linguistic, are called *standard features*, while the embeddings stand for themselves.

The supervised categorization is performed with decision tree (DT) classifier from the scikit-learn library (*scikit-learn.org*).

In the proposed experiments, we learn the model from all the annotations of a given annotator and then test the model on annotations provided by other annotators. In this way, we can measure the ability of the classifier to generalize on all known words, but for unknown annotators. This scenario is realistic to a real-world situation: the reference annotations can be obtained only from a couple of users, presumably representing the overall population, but not from all the possible users. Yet, it is necessary to predict the familiarity of medical words for all the potential users even if they did not participate in the annotations. Hence, the generalizability of models occupies the central position in these experiments.

Results and Discussion

The results obtained are presented in Table 1. The first two columns indicate the annotators. Data provided by each annotator are used for training the classifier (first column). The model generated is then tested on data from all the annotators including the reference annotator (second column). Three sets of such experiments are performed, depending on features exploited: standard features, word embeddings, and combination of all the features available. Each experiment is evaluated with several measures: *P* Precision, *R* Recall, *F* F-measure to evaluate the efficiency in prediction which medical words are understandable or not understandable for a given annotator: the darker background, the better the results.

We can do several observations on these results. Features used show an impact on the results. Thus, standard features usually show better results than embeddings. One explanation is that standard features include 24 individual features covering different aspects of linguistic and non-linguistic description of words, while word embeddings rely only on distribution of words and their similarity. Yet, combination of all the features (standard and embeddings) usually improves overall results, sometimes going to up to 2.9 improvement of F-measure. Our hypothesis is that there exists a robust nonlinear dependency between some subsets of standard features and subword-level components of word embeddings. Testing this hypothesis is the topic of our further research.

Recall values are always higher than Precision values. In each set of experiments, the best results are not obtained when the model of a given annotator is applied to own data. For instance, the *O1* model provides better results when tested on data from annotators *O2, O3* and *A8*. Similarly, the *A7* model shows better results when applied to data from annotators *O1*, *O2*, *O3* and *A8*. This is an important issue because it shows that the models acquired from one annotator can be successfully generalized over other annotators.

Besides, it seems that the annotators form two clusters according to the classification of difficult medical words: one cluster with four annotators (*O1, O2, O3, A8*) and one cluster with three annotators (*A1, A2, A7*). This issue may be related to the health literacy of annotators. This may indicate that the annotation models can be shared by people with similar skills and knowledge. Yet, to confirm this hypothesis, it is necessary to define the level of health literacy of annotators. This task is rather difficult because there is no existing tests created for computing the health literacy level for French-speaking healthy people. Another hypothesis is that some models may be better generalizable than other models. This hypothesis must also be verified with additional experiments.

Another important point is that, while the annotations go forward, the annotators usually show learning progress in decoding the morphological structure of terms and their understanding [40]. This progress is not taken into account in the current models.

Conclusions and Future Work

We proposed to address the detection of medical words which understanding may be difficult for non-specialized users of the medical area. We exploit for this machine learning algorithms, reference data from seven annotators, and several sets of NLP features: standard features (syntactic information, reference lexica, frequency, etc.), distributional features (word embeddings), and their combination. Our results provide several indications. Hence, the combination of all features is the most efficient. Concerning the generalization, we propose to learn model on a given annotator and then to apply it to data obtained from other annotators. This set of experiments indicates that models provide better results when tested on data from other annotators. We consider this to be a positive issue because it is important to be able to generalize annotations provided by a set of users on the whole population. Yet, these results may point out that the users should be apprehended through their health literacy, while currently there is no available tests for measuring it in French-language healthy people.

We have several directions for future work. For instance, we will train our own word embeddings specific to medical data in French, so that they suit better our data. We also plan to implement and test other deep learning/neural networks/NLP methods which use the morphological information of words, such as character-level recurrent neural networks and character embeddings together with 1D convolutions. In addition to the readability of medical words, we will also work on measuring health literacy of French-speaking people.

Acknowledgements

This work was partly funded by the French National Agency for Research (ANR) as part of the *CLEAR* project (*Communication, Literacy, Education, Accessibility, Readability*), ANR-17-CE19-0016-01. Also, we thank again the annotators who participated in preparing the reference data.

Table 1– Portability of models from one user to another

Train annotator	*Test annotator*	*Standard features*			*Embeddings*			*Standard features + embeddings*		
		P	R	F	P	R	F	P	R	F
O1	O1	77.2	82.5	79.7	67.0	72.5	69.3	79.0	82.4	80.2
O1	O2	78.6	81.7	80.1	70.3	74.0	71.2	82.0	84.2	82.8
O1	O3	81.2	85.0	83.0	70.7	75.4	72.6	84.9	87.6	85.9
O1	A1	71.0	74.7	71.2	62.1	63.8	58.8	74.1	75.4	72.2
O1	A2	70.6	78.4	74.0	61.9	68.5	63.3	75.0	80.1	76.2
O1	A7	72.6	77.5	74.2	63.0	66.6	61.9	76.2	78.9	75.8
O1	A8	82.3	84.9	83.5	73.1	76.8	74.5	85.7	87.8	86.6
O2	O1	77.0	82.2	79.1	67.3	72.8	69.6	80.2	83.9	81.1
O2	O2	78.9	82.0	80.0	69.9	73.5	71.3	79.5	81.9	80.3
O2	O3	81.1	85.4	83.0	71.1	75.3	73.0	83.5	86.8	84.7
O2	A1	71.1	72.1	68.2	61.7	64.5	60.2	74.0	75.1	71.5
O2	A2	70.8	77.3	72.7	61.8	68.9	64.2	76.0	79.8	75.5
O2	A7	72.7	75.6	71.8	62.6	67.0	62.8	75.9	78.3	74.9
O2	A8	83.0	86.2	84.4	73.7	77.1	75.3	85.4	88.2	86.7
O3	O1	77.4	82.8	79.7	67.1	72.7	69.4	81.3	84.9	82.4
O3	O2	79.0	82.2	80.2	70.4	74.1	71.6	82.1	84.2	82.8
O3	O3	81.2	85.5	83.2	70.4	74.9	72.3	83.0	85.9	84.2
O3	A1	71.8	73.3	69.5	61.7	64.1	59.6	75.1	75.4	72.1
O3	A2	71.2	78.0	73.5	61.8	68.7	63.9	76.8	80.2	76.3
O3	A7	73.2	76.5	72.9	62.4	66.6	62.2	77.2	78.8	75.8
O3	A8	82.6	85.8	84.1	73.7	77.2	75.2	86.0	88.0	86.9
A1	O1	77.2	82.5	79.8	66.5	67.9	66.6	76.9	79.5	77.6
A1	O2	78.6	81.6	80.1	69.2	69.0	68.5	78.8	79.6	78.9
A1	O3	81.2	84.9	82.9	70.7	69.6	69.2	81.8	82.0	81.0
A1	A1	70.9	74.7	71.3	59.4	64.6	61.8	72.4	75.1	72.9
A1	A2	70.5	78.3	74.0	60.6	66.4	63.2	73.7	78.6	75.0
A1	A7	72.6	77.5	74.2	61.3	66.1	63.6	75.1	79.2	76.5
A1	A8	82.2	84.8	83.5	72.3	70.4	70.4	81.5	81.0	80.5
A2	O1	77.3	82.6	79.8	67.2	72.6	69.6	81.0	82.8	81.8
A2	O2	78.6	81.6	80.1	70.4	74.0	71.9	82.0	82.0	82.0
A2	O3	81.2	84.9	83.0	71.0	75.2	73.0	84.9	85.4	85.1
A2	A1	70.9	74.6	71.2	61.5	64.6	60.4	76.5	76.5	74.7
A2	A2	70.6	78.4	74.0	61.2	68.4	63.7	74.7	77.8	75.6
A2	A7	72.6	77.5	74.2	62.4	67.0	63.0	77.6	78.9	77.3
A2	A8	82.2	84.8	83.4	73.8	77.0	75.3	85.6	85.3	85.4
A7	O1	77.1	82.5	79.7	67.6	73.2	69.9	79.4	81.9	80.3
A7	O2	78.5	81.6	80.0	70.6	74.2	71.8	80.6	81.4	80.9
A7	O3	81.0	84.9	82.9	71.3	75.7	73.3	83.1	83.8	83.0
A7	A1	71.0	74.4	70.9	62.1	64.8	60.3	75.8	78.0	75.7
A7	A2	70.5	78.2	73.8	62.0	69.1	64.3	75.3	79.6	76.5
A7	A7	72.6	77.4	74.0	62.2	67.0	63.1	74.5	77.5	75.3
A7	A8	81.9	84.7	83.3	73.7	77.2	75.3	82.8	82.7	82.4
A8	O1	77.0	82.4	79.6	67.2	72.7	69.6	80.8	84.4	81.7
A8	O2	78.4	81.5	79.8	70.4	74.0	71.7	82.0	84.7	83.0
A8	O3	80.9	84.9	82.8	71.0	75.2	72.9	84.7	87.6	85.6
A8	A1	71.0	74.2	70.7	61.4	64.3	60.0	73.7	75.0	71.5
A8	A2	70.4	78.1	73.7	61.7	68.8	64.1	75.0	80.1	75.9
A8	A7	72.6	77.2	73.7	62.2	66.6	62.5	75.7	78.2	74.9
A8	A8	81.9	84.9	83.4	73.6	77.0	75.1	84.2	86.5	85.2

References

[1] McGray A. Promoting health literacy. *J of Am Med Infor Ass* 2005;12:152–63.

[2] Eysenbach G. Poverty, human development, and the role of ehealth. *J Med Internet Res* 2007;9(4):34–4.

[3] Van der Stichele R. Promises for a measurement breakthrough. In: *Drug regimen compliance. Issues in clinical trials and patient management*, 1999:71-83.

[4] Patel V, Branch T, and Arocha J. Errors in interpreting quantities as procedures : The case of pharmaceutical labels. *Int Journ Med Inform* 2002;65(3):193–211.

[5] Zeng-Treiler Q, Kim H, Goryachev S, et al. Text characteristics of clinical reports and their implications for the readability of personal health records. In: MEDINFO, Brisbane, Australia. 2007:1117–21.

[6] Williams M, Parker R, Baker D, et al. Inadequate functional health literacy among patients at two public hospitals. *JAMA* 1995;274(21):1677–82.

[7] Oregon Practice Center . Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved. Technical report, Agency for healthcare research and quality. Oregon Evidence-based Practice Center, 2008.

[8] Brigo F, Otte M, Igwe S, Tezzon F, and Nardone R. Clearly written, easily comprehended ? The readability of websites providing information on epilepsy. *Epilepsy & Behavior* 2015;44:35–9.

[9] Jucks R and Bromme R. Choice of words in doctor-patient communication: an analysis of health-related Internet sites. *Health Commun* 2007;21(3):267–77.

[10] Tran T, Chekroud H, Thiery P, and Julienne A. Internet et soins : un tiers invisible dans la relation médecin/patient ? *Ethica Clinica* 2009;53:34–43.

[11] François T and Fairon C. Les apports du TAL à la lisibilité du français langue étrangère. *TAL* 2013;54(1):171–202.

[12] Flesch R. A new readability yardstick. *Journ Appl Psychol* 1948;23:221–33.

[13] Gunning R. *The art of clear writing*. McGraw Hill, New York, NY, 1973. [14] Kokkinakis D and Toporowska Gronostaj M. Comparing lay and professional language in cardiovascular disorders corpora. In: Pham T., ed, WSEAS Transactions on BIOLOGY and BIOMEDICINE, 2006:429–37.

[15] Poprat M, Markó K, and Hahn U. A language classifier that automatically divides medical documents for experts and health care consumers. In: MIE. 2006:503–8.

[16] Grabar N, Krivine S, and Jaulent M. Classification of health webpages as expert and non expert with a reduced set of cross-language features. In: AMIA, 2007:284–8.

[17] Goeuriot L, Grabar N, and Daille B. Characterization of scientific and popular science discourse in French, Japanese and Russian. In: LREC, 2008.

[18] Miller T, Leroy G, Chatterjee S, Fan J, and Thoms B. A classifier to evaluate language specificity of medical documents. In: HICSS, 2007:134–40.

[19] Chmielik J and Grabar N. Détection de la spécialisation scientifique et technique des documents biomédicaux grâce aux informations morphologiques. TAL 2011;51(2):151-79.

[20] Specia L, Jauhar S, and Mihalcea R. Semeval-2012 task 1: English lexical simplification. In: *SEM 2012, 2012:347–55.

[21] Sinha R. Unt-simprank: Systems for lexical simplification ranking. In: *SEM 2012, 2012:493–6.

[22] Jauhar S and Specia L. UOW-SHEF: SimpLex – lexical simplicity ranking based on contextual and psycholinguistic features. In: *SEM 2012. 2012:477–81.

[23] Johannsen A, Martínez H, Klerke S, and Søgaard A. Emnlp@cph: Is frequency all there is to simplicity? In: *SEM 2012. 2012:408–12.

[24] Ligozat A, Grouin C, Garcia-Fernandez A, and Bernhard D. Annlor: A naïve notation-system for lexical outputs ranking. In: *SEM 2012, 2012:487–92.

[25] Amoia M and Romanelli M. SB: mmSystem - using decompositional semantics for lexical simplification. In: *SEM 2012. 2012:482–6.

[26] Gala N, François T, and Fairon C. Towards a French lexicon with difficulty measures: NLP helping to bridge the gap between traditional dictionaries and specialized lexicons. In: eLEX-2013, 2013.

[27] Zheng W, Milios E, and Watters C. Filtering for medical news items using a machine learning approach. In: AMIA, 2002:949– 53.

[28] Borst A, Gaudinat A, Boyer C, and Grabar N. Lexically based distinction of readability levels of health documents. In: MIE 2008, 2008. Poster.

[29] Grabar N, Hamon T, and Amiot D. Automatic diagnosis of understanding of medical words. In: EACL PITR Workshop, 2014:11–20.

[30] van Oosten P and Hoste V. Readability annotation: Replacing the expert by the crowd. In: Workshop on Innovative Use of NLP for Building Educational Applications, 2011:120–9.

[31] Yaneva V, Temnikova I, and Mitkov R. Accessible texts for autism: An eye-tracking study. In: ACM , ed, Int ACM SIGACCESS Conf on Comp & Accessibility, 2015:49–57.

[32] Grabar N, Farce E, and Sparrow L. Study of readability of health documents with eye-tracking approaches. In: Workshop on Automatic Text Adaption (ATA), 2018:1– 1.

[33] Grabar N and Hamon T. A large rated lexicon with French medical words. In: LREC. 2016:1–2.

[34] Côté RA, Rothwell DJ, Palotay JL, Beckett RS, and Brochu L. *The Systematised Nomenclature of Human and Veterinary Medicine: SNOMED International.* College of American Pathologists, Northfield, 1993.

[35] Schmid H. Probabilistic part-of-speech tagging using decision trees. In: Int Conf on New Methods in Language Processing, 1994:44–9.

[36] Namer F. FLEMM : un analyseur flexionnel du français à base de règles. Traitement automatique des langues (TAL) 2000;41(2):523–47.

[37] Namer F and Zweigenbaum P. Acquiring meaning for French medical terminology: contribution of morphosemantics. In: AMIA. 2004.

[38] Kincaid J, Fishburne R, Rogers R, and Chissom B. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, Naval Technical Training, U. S. Naval Air Station, 1975.

[39] Bojanowski P, Grave E, Joulin A, and Mikolov T. Enriching word vectors with subword information. *Transactions of the Assoc for Comp Linguistics* 2017;5(1):135–46.

[40] Grabar N and Hamon T. Understanding of unknown medical words. In: BIONLP workshop at RANLP, 2017:1–0.

Address for correspondence

Natalia Grabar
CNRS, Univ. Lille, UMR 8163 STL,
F-59000, Lille, France
e-mail: natalia.grabar@univ-lille.fr
url: http://natalia.grabar.free.fr/

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190443

A Method to Accelerate and Visualize Iterative Clinical Paper Searching

Yiqin Yu[a], Enliang Xu[a], Eryu Xia[a], He Huang[b], Bibo Hao[a], Shilei Zhang[a]

[a] IBM Research China, Beijing, China, [b] Tsinghua University, Beijing, China

Abstract

Clinical paper searching is a major task for clinical researchers to collect authoritative and up-to-date evidences to support their research works and clinical practices. Currently, this task needs huge amount of labor work. Researchers usually spend a lot of time searching on the online repository and iterating many times to get the final paper list. Systematic review is a special case, in which the paper searching process is a critical step. To address this challenge, this paper introduces a method to streamline the iterative paper searching process. It automatically selects the most probably matched papers, and then generates new search strategy. All the intermediate results are visualized based on the paper citation graph. It assembles technologies such as PageRank and Topic-based clustering to accelerate the paper searching tasks. The precision, recall, and execution time of the proposed method are then evaluated by comparing with published systematic reviews.

Keywords:

Paper, Bibliometrics, Literature Based Discovery

Introduction

Clinical paper searching is a major task for clinical researchers to collect authoritative and up-to-date evidences to support their research works and clinical practices. Clinicians/writers of clinical reviews and clinical guidelines also spend a lot of time on clinical paper searching. Starting from a research problem or a research topic, the goal of the searching process is to find clinical papers that mostly depict the corresponding topic. Search strategy is carefully constructed to reveal the research topic, which usually contains at least a set of keywords as search terms, other objects such as fields, in which, part the keywords will be searched in clinical papers, and search logic operators to organize the keywords [1]. Then it is used to do queries via online clinical databases. By further reading and filtering the retrieved papers, clinical researchers can get a comprehensive view about the research topic from the final paper set. A special case is systematic review [2], which aims to collect secondary data from published clinical papers to offer a complete, exhaustive summary for the research question [3].

There are many challenges during the paper searching process from search strategy construction to final paper set confirmation, which include the following:

- In order to include most matched clinical papers, the search strategy is set to relatively broad. That often leads to huge amounts of irrelevant papers being retrieved (precision as low as 0.3% [4]).
- Expert searchers usually use iterative refinement of search queries to improve the precision of paper searching. Nevertheless, this kind of iterative process has not been well automated yet [5].
- There are lack of visualization tools to monitor the searching process, especially when the searchings are iterated for many times, which increase the burden of searchers for capturing the intermediate results.

To use web tools, information retrieval, and machine learning methods to help the clinical paper searching process is a sustained topic since the online literature databases such as PubMed were built up [5,6,7]. AskMEDLINE [8] handles user queries in the form of free-text questions. It can automatically complete the patient, intervention, comparison, outcome (PICO [9]) form by parsing the clinical questions. RefMed [10] retrieves search results based on user queries, learns a ranking function based on the user feedback information, and then applies the ranking function to the retrieved results in the next iteration. Another work named MedlineRanker [11] automatically learns a list of most discriminative words from a set of documents to represent certain topic. Then the learned words are used to score and rank newly published papers on the topic. Chen et. al. [12] used a PageRank based algorithm to assess the relative importance of all publications to find some exceptional papers that are universally familiar to physicists.

This paper introduces a method to streamline the iterative paper searching process. It automatically selects the most probably matched papers, and then generates a new search strategy. All the intermediate results are visualized based on the paper citation graph. It assembles technologies such as PageRank [13] and topic-based clustering [14] to accelerate the paper searching tasks. The contributions of the paper include the following:

- It provides a generalized mechanism to streamline the iterative clinical paper searching process to get most probably matched papers, which is independent of specific research topics.
- A visualization based on paper citation graph is produced to show intermediate results, which can provide evidence and provenance about how the papers are collected and selected during the iterative searching process.
- The precision, recall, and execution time of the proposed method, which assembles technologies such as PageRank [13] and topic-based clustering [14], is evaluated by comparing with published systematic reviews.

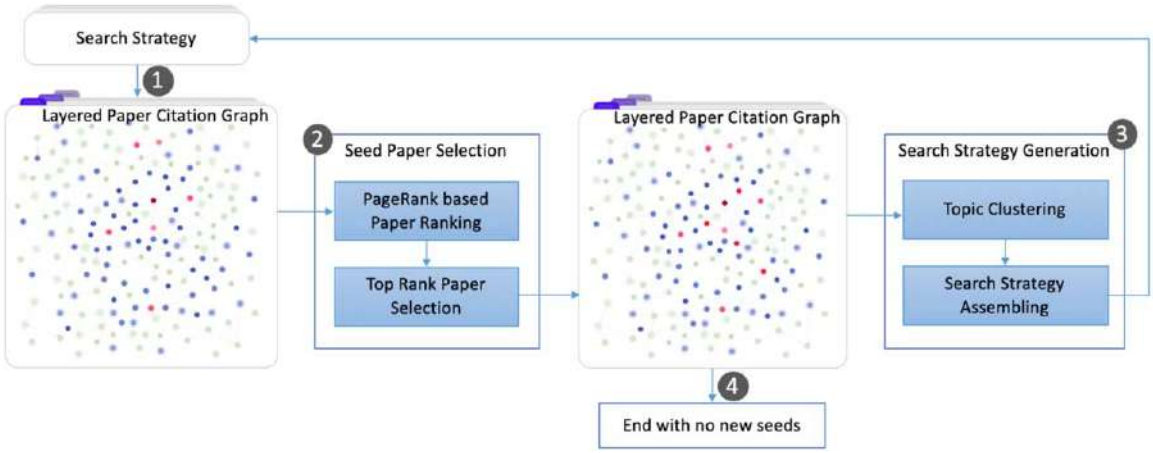


Figure 1– Illustration of the proposed method. (1) When the search strategy is generated, paper querying is performed to retrieve a set of papers, which are then represented into the layered paper citation. (2) Based on the current paper set, papers are ranked by PageRank algorithm, and finally select the top-n papers as seed papers, which are the most probably matched papers. (3) Given the seed papers, a topic-based clustering algorithm is performed to extract key topics and terms from different fields and then assemble them to form a search strategy. (4) The process will be ended until there is no new seed selected.

Methods

The proposed method is illustrated in Figure 1. Starting from a primary search strategy and several papers as seeds, the paper searching process can be iterated automatically until there is no new seed paper selected. A seed paper is a paper identified as a most probably matched paper with the research topic. It is used to generate search strategy for the next iteration, and the overall set of seed papers will be output as the final inclusion set. In spite of the seed paper list, the final results of our method also include a list of search strategies and a list of all retrieved papers. The layered paper citation graph will keep all the intermediate results as provenance. Next we will introduce the main components respectively.

Layered Paper Citation Graph

A Layered Paper Citation Graph (LPCGraph) is a graph representing the intermediate results from different iterations during the searching process (Figure 2). Each node represents a retrieved paper. Edges between the nodes represent the direction of citation relationship. For example, if there is an edge from node A to node B, that means the clinical paper B is cited by paper A. To capture the intermediate results in each iteration, LPCGraph is specifically equipped with the following features, which can provide evidence and provenance about how the papers are collected and selected.

- Layers. Due to the huge number of papers retrieved, it's important to reveal in which iteration a paper is retrieved, and how many times the paper is queried in different iterations. The layer feature in LPCGraph represents the corresponding iteration. Nodes in different layer have different transparency, and nodes with more overlap times will have darker colors.
- Role of nodes. A node in the LPCGraph has multiple roles, which capture different characteristics of a retrieved paper. The layer feature, as mentioned above, represent the iteration in which the paper is retrieved, as well as the overlap times of a paper in different iterations. Another setting is when the corresponding paper is selected as a seed paper, it is represented in red colors. If it is retrieved for multiple times, the red color will be even darker.

For example, in Figure 2(a) and 2(b), node with id *17548730* is a seed paper in the current iteration and it is retrieved multiple times. It is highlighted by deep red color. Its title is *Randomized trial of acupuncture to lower blood pressure*, and the published date is *2007/6/6*.

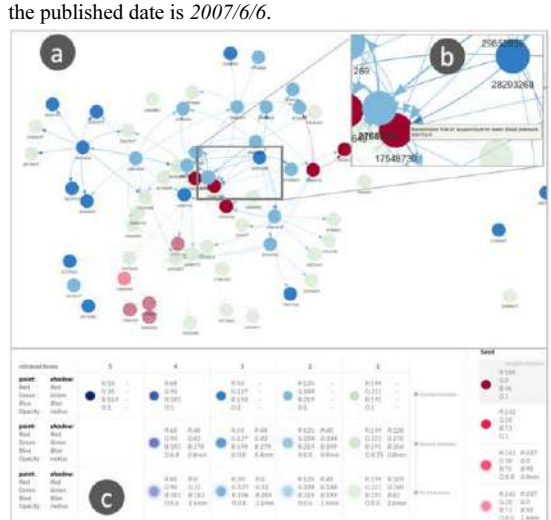


Figure 2– Illustration of the LPCGraph. (a) Overall view of the graph during certain paper searching iteration. (b) Zoom in view of the framed subgraph. (c) Legend of node views for different roles and layers.

Seed Paper Selection

The goal of the paper searching process is to form an inclusion paper set, which could relate back to the research topic. Within our method, the papers are retrieved iteratively, which means that in each iteration, there will be new seed papers selected and included into the set.

Consider the task of seed paper selection in an iteration. Without regard to other information, every paper retrieved by the given search strategy would have the same chance to be selected as a seed paper. If we consider the citation relationships among these papers, the ones with more citations stand out more likely as exceptional because their contents are more influential. Given the hypothesis that the retrieved papers are probably related to the research topic, which is represented by the search strategy, the most cited papers have higher probabilities to match the topic than other less cited ones. The importance level based on citations can be measured by PageRank algorithm [12]. The PageRank algorithm [13] is depicted as following:

$$PR(u) = \frac{1-d}{N} + d\sum_{v\in B(u)}\frac{PR(v)}{L(v)} \quad (1)$$

Where u and v represents a web page respectively. *PR(u)* and *PR(v)* are rank scores of page u and v. d is a dampening factor, which is usually set to 0.85. N is the total number of pages. *B(u)* is the set of pages that point to u. *L(v)* denotes the number of outgoing links of page v.

Search Strategy Generation

If there are new seed papers selected in an iteration, new search strategy should be generated to cover information from these newly included ones and launch the next iteration to further retrieve papers. We apply the idea of topic-based clustering to extract common concepts from different fields of the seed papers and properly assemble them together in the format of search strategy. Topic-based clustering is an unsupervised approach for automatic keyphrase extraction from documents [14]. Instead of extracting words as candidate keyphrases directly, it groups the candidate keyphrases into topics, such that each topic is composed of only those candidate keyphrases that are related to that topic.

The design of search strategy for clinical paper searching follows some guidelines [19]. Checking the common structure of a search strategy in Figure 3, it mainly contains a set of keyphrase candidates as search terms and fields align with the search term that will be searched in clinical papers. Keyphrase candidates which belong to the same topic will be jointly connected by "OR", and "AND" is used to connect bunches of keyphrase candidates belonging to different topics.

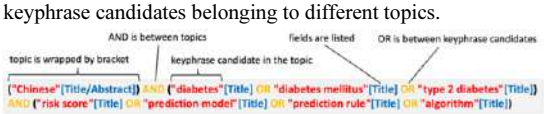


Figure 3– A sample search strategy

In order to generate such search strategy, we apply the idea of graph-based topic ranking for keyphrase extraction [15]. In this work, a graph is built-in, which, vertices are topics represented as clusters of lexically similar keyphrase candidates and all topics are ranked according to their significance. Naturally, the top-n ranked topics and corresponding keyphrases extracted from seed papers could be assembled to form a search strategy, which will represent the main topic of these papers.

$$dist(c_i, c_j) = \sum_{p_i\in pos(c_i)}\sum_{p_j\in pos(c_j)}\frac{1}{|p_i - p_j|} \quad (2)$$

Where $dist(c_i, c_j)$ refers to the reciprocal distances between the offset positions of the candidate keyphrase c_i and c_j in the paper. $pos(c_i)$ represents all the offset positions of the candidate keyphrase c_i.

Results

The selection of benchmark systematic reviews followed some criteria, e.g., whether the full text of the review can be retrieved, whether it reports the search strategy and lists all inclusion papers, and whether its search results can be reproduced on PubMed. Given the research interest of '*treatment of hypertension*' and corresponding search strategy *'treatment of hypertension AND Review[ptyp] AND free full text[sb]'*, the first eligible systematic review [18] was selected from the search results on PubMed. Then the proposed method was applied to mimic the paper searching process in this review for evaluation. The method used PubMed as the literature search service, and did the paper querying via the Entrez search interface [16]. Note that the proposed method is not restricted to specific databases such as PubMed. The method was implemented in Python and the LPCGraph was visualized via visNetwork [17]. The systematic review totally includes 23 papers about RCTs involving acupuncture as a therapeutic intervention for treating hypertension, in which nine papers are retrieved by searching from PubMed. We name the list of nine included papers as *ground_truth_paper_list*.

The experiment started from retrieving papers via the Entrez search interface by using the search strategy from the systematic review (Figure 6(a)). Several papers in the *ground_truth_paper_list* were marked as seed papers manually in the first iteration to guide the clustering of topics. The searching method was run iteratively until there is no new seed selected. All the seed papers were collected together for output, named as *seed_paper_list*. The experiment was executed multiple times with different number of seed papers selected (*number_of_seeds*). All the paper queries were done online via the Entrez search interface to retrieve the newest results. We also use cache mechanism to reduce the time of network connecting. The execution date of the experiment was Nov 23, 2018.

Figure 4 shows the precision and recall of the experiment with different *number_of_seeds*. The calculations of precision and recall are as follows.

$$precision = \frac{|\{x|x\in seed\_paper\_list \& \ x\in ground\_truth\_paper\_list\}|}{|\{x|x\in seed\_paper\_list\}|} \quad (3)$$

$$recall = \frac{|\{x|x\in seed\_paper\_list \& \ x\in ground\_truth\_paper\_list\}|}{|\{x|x\in ground\_truth\_paper\_list\}|} \quad (4)$$

The recall increases when more papers in the *ground_truth_paper_list* are selected as seeds by our method. And the precision decreases because when the *number_of_seeds* becomes larger, more papers are selected into the *seed_paper_list*. It is worth mentioning that when the *number_of_seeds* was set to five, precision and recall together get a balanced value, namely 0.7 and 0.7778, respectively. It indicates that by carefully setting the value of *number_of_seeds*, our method could reach an acceptable and balanced performance, as well as the execution efficiency corresponding to the execution time shown in Figure 5. The results can be further verified by Figure 6. Figure 6(a) is the search strategy copied from the systematic review paper, which contains three topics: acupuncture, hypertension, and randomized trial. Figure 6(b) is the final search strategy automatically generated by our method with *number_of_seeds* as five, which indicates five topics including acupuncture, blood pressure, hypertension, treatment, and group. It is visible that our method kept the original topics, and also expanded the topic of hypertension to blood pressure, which is meaningful.

Figure 2 shows the generated LPCGraph with *number_of_seeds* as five. Totally 135 papers were retrieved and represented in the graph. Ten paper nodes colored with red were selected as seed paper and input into the *seed_paper_list*, among which the five ones with no transparency were automatically selected in the latest iteration

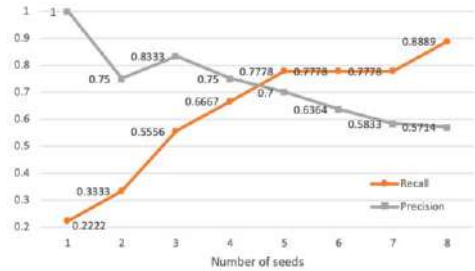


Figure 4– Precision and recall when selecting different number of seeds

by our method. In spite of the LPCGraph, Table 1 also shows the comparison between the *ground_truth_paper_list* and the *seed_paper_list*. There are two papers in *ground_truth_paper_list* but not in *seed_paper_list*. The paper with id *16541853* can be searched in PubMed by id, but can not be retrieved by the original search strategy in the

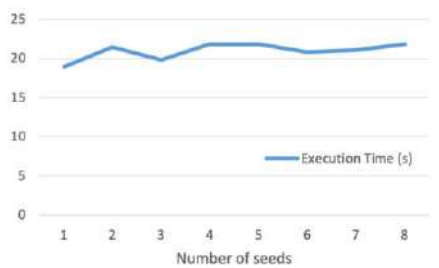


Figure 5– Execution time in seconds when selecting different number of seeds

```
(("acupuncture"[MeSH Terms] OR "acupuncture"[All Fields] OR "acupuncture
therapy"[MeSH Terms] OR ("acupuncture"[All Fields] AND "therapy"[All Fields]) OR
"acupuncture therapy"[All Fields]) OR ("electroacupuncture"[MeSH Terms] OR
"electroacupuncture"[All Fields])) AND (("hypertension"[MeSH Terms] OR
"hypertension"[All Fields]) OR (essential[All Fields] AND ("hypertension"[MeSH
Terms] OR "hypertension"[All Fields]))) AND (("random allocation"[MeSH Terms] OR
("random"[All Fields] AND "allocation"[All Fields]) OR "random allocation"[All
Fields] OR "randomized"[All Fields]) OR ("clinical trials as topic"[MeSH Terms] OR
("clinical"[All Fields] AND "trials"[All Fields] AND "topic"[All Fields]) OR
"clinical trials as topic"[All Fields] OR "trial"[All Fields]))
                                    (a)

("anti-blood pressure"[Title/Abstract] OR "blood pressure"[Title/Abstract] OR "blood
pressure variability"[Title/Abstract] OR "diastolic blood pressure"[Title/Abstract]
OR "diastolic pressure"[Title/Abstract] OR "post-treatment blood pressures"[Title/
Abstract] OR "systolic blood pressure"[Title/Abstract] OR "systolic pressure"[Title/
Abstract] OR "twenty-four hour dynamic blood pressure"[Title/Abstract] OR "untreated
blood pressure"[Title/Abstract]) AND ("arterial hypertension"[Title/Abstract] OR
"diastolic hypertension"[Title/Abstract] OR "essential hypertension"[Title/Abstract]
OR "hypertension"[Title/Abstract] OR "hypertensive group"[Title/Abstract] OR "stop
hypertension"[Title/Abstract] OR "uncomplicated arterial hypertension"[Title/
Abstract]) AND ("active treatment group"[Title/Abstract] OR "days treatment"[Title/
Abstract] OR "pharmacological treatment"[Title/Abstract] OR "treatment"[Title/
Abstract] OR "treatment group"[Title/Abstract] OR "treatment period"[Title/
Abstract]) AND ("acupuncture"[Title/Abstract] OR "acupuncture needles"[Title/
Abstract] OR "saam acupuncture theory"[Title/Abstract] OR "standardized
acupuncture"[Title/Abstract]) AND ("acues group"[Title/Abstract] OR "control
group"[Title/Abstract] OR "ea group"[Title/Abstract] OR "group"[Title/Abstract] OR
"sham-es group"[Title/Abstract])
                                    (b)
```

Figure 6– (a) The search strategy in the first iteration from the systematic review paper. (b) The search strategy in the final iteration generated by the proposed method

systematic review, which leads to an exclusion to our overall paper list. The paper with id *10352369* was retrieved by our method, but due to the lack of citation information, its rank is low and not ever selected as a seed. Another list in Table 1 is the papers in *seed_paper_list* but not in *ground_truth_paper_list*. Considering the content of title and abstract, these papers are highly related to the research topic. But further confirmation need to be done by clinical researchers.

Table 1– Comparison between ground_truth_paper_list and seed_paper_list

Papers in *ground_truth_paper_list* but not in *seed_paper_list*	
16541853	Observation on therapeutic effect of "reducing south and reinforcing north" needling method on hypertension of type of yang-hyperactivity due to yin-deficiency
10352369	Effect of a standardized acupuncture treatment on complains, blood pressure and serum lipids of hypertensive, postmenopausal women. A randomized, controlled clinical study
Papers in *seed_paper_list* but not in *ground_truth_paper_list*	
17359649	Acupuncture, a promising adjunctive therapy for essential hypertension: a double-blind, randomized, controlled trial.
2052631	Effect of acupuncture-point stimulation on diastolic blood pressure in hypertensive subjects: a preliminary study.
16990216	Acupuncture reduces experimental renovascular hypertension through mechanisms involving nitric oxide synthases.

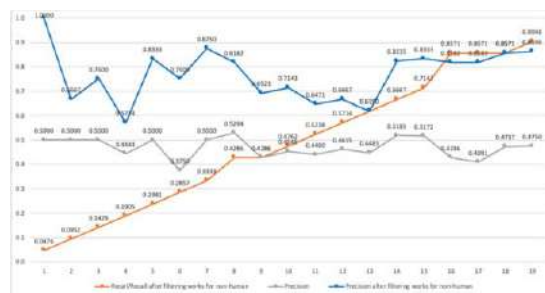


Figure 7– Precision and recall when selecting different number of seeds, and after filtering works for non-humans

Evaluation was performed on another systematic review [20] given the research interest of '*risk of diabetic nephropathy*'. Due to space limitations, we only show the precision and recall in Figure 7. As the *seed_paper_list* contain works for non-human, we also list the adjusted results after filtering the non-human works.

Discussion

The method introduced in this paper provides a generalized mechanism to automate the clinical paper searching task. The implementation of the method is independent of topic specific restrictions. Precision and recall in Figure 4 indicate that by carefully setting the value of *number_of_seeds*, our method could reach an acceptable and balanced performance. Comparing to manual searching, we also demonstrate the concrete execution efficiency of our method via the execution time in Figure 5. In the results of the second experiment, we observed that works for non-human were included as seeds. According to the statement of the review [20], we filtered those works and calculated the adjusted precision and recall in Figure 7. The adjusted results show consistent performance with the first experiment. The reason why the precision before adjustment is relatively low is that the review manually rejects all the works for non-human, but this information is not well

catched by the search strategy. Such kind of 'exclusion' will lead to the use of NOT in the search strategy, which is beyond the scope of the current implementation.

The use of PageRank for ranking and selecting seed papers is intuitive, especially when the original search strategy is not arbitrarily defined by clinical researchers. Given the hypothesis that the retrieved papers are probably related to the research topic, the most cited papers have higher probabilities to match the topic than other less cited ones. When the original search strategy from the review paper is somewhat ambiguous, the seed papers selected by the PageRank algorithm subsequently have more change to deviate from the research topic. As in the second experiment, the original search strategy didn't explicitly depict the exclusion of works for non-human. So the generated LPCGraph have two separated centroids: one relates to works for human and the other relates to works for non-human (mice). As a result, as the PageRank algorithm ranks seed papers from both of them, it leads to lower precision when comparing to the original systematic review. In the future, we will try to address the problem of ambiguous original search strategy, e.g., by comparing seed candidates with the original selected ones.

The topic-based clustering algorithm is applied to generate keyphrases and further cluster them into topics, which are assembled to form a sophisticated search strategy. The main structure of the search strategy follows guidance from Lu [7] and Kitchenham[19]. As discussed above, we will try to include NOT as a connector into the structure of the search strategy in future works.

A main limitation of this method is that it depends on the availability of search interfaces of different online databases. According to our observation during the survey of systematic reviews, most of the reviews search papers from multiple databases such as PubMed, Embase, and Cochrane etc. Lack of access to certain databases will lead to bias of clinical paper selection.

Conclusions

In this paper we proposed an iterative clinical paper searching method to capture the three aspects of the searching process, namely seed paper selection, search strategy generation, and layered paper citation graph visualization. We showed a detailed implementation of the method by leveraging PageRank algorithm, topic-based clustering, and paper citation graph based visualization. We applied the method to mimic the paper searching process in published systematic review, and the results show that our method can be used to accelerate the paper searching process as well as keeping the intermediate results as provenance in a quick and effective way.

References

[1] Lowe HJ, Barnett, GO: Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches, *JAMA* **271 (1994)**, 1103–1108.

[2] Morton S, Berg A, Levit L, Eden J, editors. Finding what works in health care: standards for systematic reviews, *National Academies Press*; 2011 Jun 20.

[3] Systematic review. https://en.wikipedia.org/wiki/Systematic_review (accessed 8 Nov 2018).

[4] Eysenbach G, Tuische J, Diepgen TL, Evaluation of the usefulness of Internet searches to identify unpublished clinical trials for systematic reviews, *Med Inform Internet Med* **26** (2001), 203–218.

[5] Tsafnat G, Glasziou P, Choong MK, Dunn A, Galgani F, Coiera E, Systematic review automation technologies, *Systematic reviews* **3** (2014 Dec), 74.

[6] Jonnalagadda SR, Goyal P, Huffman MD, Automating data extraction in systematic reviews: a systematic review. *Systematic reviews* **4** (2015 Dec), 78.

[7] Lu Z, PubMed and beyond: a survey of web tools for searching biomedical literature, *Database*, 2011 Jan 1.

[8] Fontelo P, Liu F, Ackerman M, ask MEDLINE: a free-text, natural language query tool for MEDLINE/PubMed. *BMC Medical Informatics and Decision Making* **5** (2005 Dec), 5.

[9] Schardt C, Adams MB, Owens T, Keitz S, Fontelo P, Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC medical informatics and decision making* **7** (2007 Dec), 16.

[10] Yu H, Kim T, Oh J, Ko I, Kim S, Han WS, Enabling multi-level relevance feedback on PubMed by integrating rank learning into DBMS, In BMC bioinformatics*, BioMed Central1* **1** (2010 Apr), S6.

[11] Fontaine JF, Barbosa-Silva A, Schaefer M, Huska MR, Muro EM, Andrade-Navarro MA, MedlineRanker: flexible ranking of biomedical literature, *Nucleic acids research* **37** (2009 May), W141-6.

[12] Chen P, Xie H, Maslov S, Redner S, Finding scientific gems with Google's PageRank algorithm, *Journal of Informetrics* **1** (2007 Jan 1), 8-15.

[13] L. Page, S. Brin, R. Motwani, and T. Winograd, The PageRank citation ranking: Bringing order to the web, *Technical report, Stanford Digital Libraries SIDL-WP-1999-0120*, 1999.

[14] Hasan KS, Ng V, Automatic keyphrase extraction: A survey of the state of the art. *In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers) 2014, 1262-1273.

[15] Bougouin A, Boudin F, Daille B. Topicrank, Graph-based topic ranking for keyphrase extraction, *In International Joint Conference on Natural Language Processing* (IJCNLP) 2013 Oct 14, 543-551.

[16] Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Federhen S, Feolo M, Database resources of the national center for biotechnology information, *Nucleic acids research* **40 (**2011 Dec 2), D13-25.

[17] Almende BV, Thieurmel B, Robert T, visNetwork, Network Visualization using 'vis. js' Library, R package version, 2016;1(1).

[18] Zhao XF, Hu HT, Li JS, Shang HC, Zheng HZ, Niu JF, Shi XM, Wang S, Is acupuncture effective for hypertension? A systematic review and meta-analysis, *PloS one* **10** (2015 Jul 24), e0127019.

[19] Kitchenham B, Procedures for performing systematic reviews[J]. Keele, UK, Keele University, **33** (2004), 1-26.

[20] Dellamea BS, Pinto LC, Leitão CB, Santos KG, Canani LH, Endothelial nitric oxide synthase gene polymorphisms and risk of diabetic nephropathy: a systematic review and meta-analysis, *BMC medical genetics* **15** (2014 Dec), 9.

Address for correspondence

Yiqin Yu, MS. IBM Research – China, Building 28, Zhongguancun Software Park, Haidian District, Beijing, China, 100193.

Email: yuyiqin@cn.ibm.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190444

Using an Artificial Intelligence-Based Argument Theory to Generate Automated Patient Education Dialogues for Families of Children with Juvenile Idiopathic Arthritis

Benjamin Rose-Davis[a], William Van Woensel[a], Elizabeth Stringer[b], Samina Abidi[c], Syed Sibte Raza Abidi[a]

[a] *NICHE Research Group, Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia, Canada*
[b] *Division of Pediatric Rheumatology, IWK Health Centre, Halifax, Nova Scotia, Canada*
[c] *Department of Epidemiology and Population Health, Dalhousie University, Halifax, Nova Scotia, Canada*

Abstract

Juvenile Idiopathic Arthritis (JIA) is the most common chronic rheumatic disease of childhood, with outcomes including pain, prolonged dependence on medications, and disability. Parents of children with JIA report being overwhelmed by the volume of information in the patient education materials that are available to them. This paper addresses this educational gap by applying an artificial intelligence method, based on an extended model of argument, to design and implement a dialogue system that allows users get the educational material they need, when they need it. In the developed system, the studied model of argument was leveraged as part of the system's dialogue manager. A qualitative evaluation of the system, using cognitive walkthroughs and semi-structured interviews with JIA domain experts, suggests that these methods show great promise for providing quality information to families of children with JIA when they need it.

Keywords:

Patient Education, Artificial Intelligence, Semantic Web

Introduction

Juvenile Idiopathic Arthritis (JIA) is a chronic rheumatic disease with clear physical and social burdens for those with the condition—and is also known to cause an "emotional rollercoaster" for affected families [1]. Appropriate patient education has been shown to reduce some of these burdens and improve quality of life [2]. Nevertheless, caregivers of children with JIA in Canada have expressed being overwhelmed from the myriad of information sources available to them [3]. Indeed, best practices in patient education call for giving caregivers control over what content they view [4]. Moreover, there seems to exist a strong preference towards education information from trusted sources, such as healthcare providers or the Arthritis Society [3]. Thus, there is a clear need to provide *trusted* patient educational content to (families of) children with JIA, which is not overwhelming and delivers content control and access to the information outside of clinic.

Dialogue systems are automated efforts to imitate a person-to-person communication style to make the interaction with a computer system more intuitive for the user [5]. Unlike traditional verbal and printed patient education mediums, dialogue systems are accessible via computers or mobile phones at the user's convenience. The key feature of dialogue systems is the flexibility of access to the educational content, as the user controls the subject of inquiry. This is unlike paper-based Patient Education Materials (PEM) or static websites, where the information is prepackaged. While dialogue systems have been only sparingly used for patient education, it has been reported that this educational medium results in a significant improvement in knowledge [6] and self-management [7]. Two types of dialogue are relevant to patient education—i.e., information seeking dialogue, where the aim is to provide information to answer a user's question; and inquiry dialogue, where users seek to explore and verify evidence [8].

An AI-based argument theory, based on the Toulmin model of argument, provides a useful model to represent educational content for a dialogue system. Argument theory offers a set of information representational constructs to represent the central elements of any dialogue, such as the dialogue's claim, the evidence backing the authenticity of the claim, the data used to derive the evidence, certain exceptions to the claim and so on. Toulmin's argument model, therefore, supprts the functions of language that are used to justify a claim in terms of a narrative structure that mimics a dialogue [9]. Therefore, there is a case for using argument theory to develop dialogue systems for patient education, as not only are dialogue systems more friendly and intuitive for educational purposes, but they also provide information that is backed up by evidence and can be further investigated for alternative options and/or additional details. Toulmin's model has been used by patient education tools, albeit for question-answering as opposed to a dialogue that manifests a series of follow-up questions [10].

Our objective is to investigate the use of argument theory to model the JIA educational content so it can be delivered via a dialogue system—with the goal of addressing the educational needs of families of children with JIA. Given the requirements of the JIA patient education materials, we leverage the *Extended Model of Argument* (EMA) [11] since it encompasses the diversity of topics and forms of information found in JIA PEM. Using the EMA, we developed an interactive dialogue based JIA patient education system—i.e., the Juvenile idiopathic Arthritis Dialogue-based Education (JADE) system (see Figure 1). In this paper, we present our knowledge management approach to develop JADE [12]. We discuss (**i**) the formulation of an ontology-based EMA knowledge model to formally represent the concepts and relationships underlying (extended) argument theory, resulting in the *JADE Ontology* (JO); (**ii**) the abstraction of relevant education themes from the available PEM, using thematic coding; and (**iii**) computerizing the PEM content, alongside the identified themes, using the JO. JADE has been developed and it contains 32 PEM covering 16 topics, in terms of 931 arguments. We developed semantic web-based reasoning methods to reason over the PEM content, represented using the JO, to formulate an interactive dialogue where the user can ask an initial question, as well as a series of

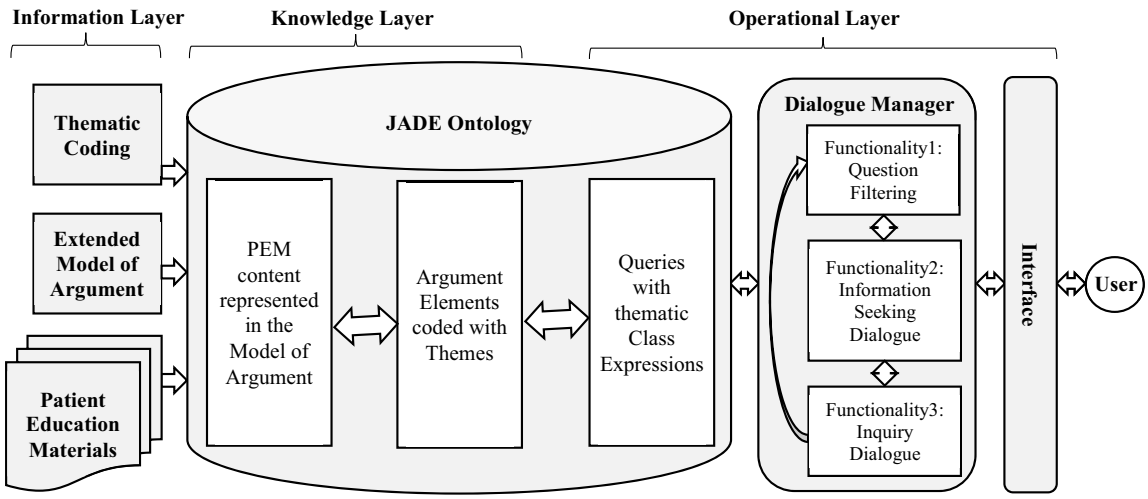


Figure 1– Architecture of the JADE system

follow-up questions. In this way, JADE provides families of children with JIA access to trusted PEM content approved by their healthcare providers—in a way that gives users control over the educational content that they are presented. JADE has been evaluated for content completeness, relevance, and utility.

The novel aspect of our patient education approach is the provision of a dialogue that is dynamically generated (in response to a variety of user queries) by selecting relevant arguments encoded in the JO and chaining a series of arguments using reasoning. According to the literature, one other dialogue system has used the Toulmin model of argument [10], however it is clinician facing and can only respond to a single question with content from a single argument.

JADE Architecture

Figure 1 depicts the 3 layers of the JADE system's design:

1. The information layer contains the EMA, thematic codes, and PEM content described previously [11]. This layer provids the knowledge model which is used to represent the PEM content in the JO.
2. The knowledge layer is comprised of the PEM content, coded to the EMA and thematic coding, represented in an ontology. This layer creates a computerized structure which the operational layer can draw on to provide an interactive dialogue. Thematic coding links individual EMA elements to an initial user question and the relationships between EMA elements, and individual arguments, allow for follow up questions about the content of the initial response.
3. The operational layer is made up of potential user questions encoded in the ontology, the dialogue manager, and the user interface. The dialogue manager receives the user's input through the interface and interacts with the ontology to formulate and deliver responses. It allows the user to find and select questions that are of interest to them, to have an information seeking dialogue by asking and receiving answers to their questions, and to participate in an inquiry dialogue. The latter uses question prompts based on the EMA and the chaining of arguments to help the user navigate through the PEM content.

Methods

The EMA serves as the knowledge model for integrating PEM content into the JADE system and is used to structure the inquiry dialogue it provides. Based on the Toulmin model, which contains 6 constructs: the claim, qualifier, data, warrant, backing, and rebuttal (renamed exception) [9], the EMA's 7 elements represent the parts used to justify a statement [11]:

1. The Claim is the statement being justified.
2. The Qualifier denotes the strength of the claim.
3. The Data represents situations where the claim is true.
4. Warrants, either explicit or implicit, explains how the data relates to the claim.
5. Exceptions are situations where the claim is not true.
6. Elaborations give more information about another element, for example a definition.
7. The Backing is the source of the information.

Thus, the following argument from the pamphlet: 'Using Ice and Heat at Home' (backing) can be coded to this model as follows: 'Some pinking of the skin is normal (exception), however, ice should be removed (claim) if the skin becomes pale and/or pain is felt (data) as this is a sign that skin damage due to cold is beginning (explicit warrant). As well, periodic skin checks are recommended (elaboration to the claim).' This knowledge model was formalized as the JO using Protégé 5.0.0 (Stanford University, 2016). Since the strengths of ontologies include reusability and extendibility, we have based the JO on Vitali and Peroni's Argument Model Ontology [13].

PEM Topic Identification

Thematic coding of the PEM content was used by the JADE system to filter questions, so users can ask about what interests them, and to locate relevant PEM content in the ontology in response to a user's question. Themes were generated inductively from the PEM content using grounded theory [14]. For example, the element 'periodic skin checks are recommended' was coded as *Periodic* and *Skin checks* during open coding. Selective coding merged *Skin checks* with similar codes into the *Monitor* code. *Monitor* and codes such as *Use treatment* were then grouped under the theme *Recommendation* during axial coding.

Knowledge Formalization

The JADE system requires a structured repository of domain specific content, so that it can be accessed by the dialogue manager to formulate responses to user's questions. For this purpose, the PEM content and linked thematic codes were computerized using the JO. The ontology has 4 top-level classes depicted in Figure 2 and described in detail below.

Individual arguments are represented as instances of **the 'ArgumentAsRepresentedByImplicitWarrant' (ARIW) class**. Since two or more arguments could share the same data element, i.e. have the same situation for which a claim is true, such as taking a certain medication, the data element was not unique. Neither was the claim element, as two or more arguments could share the same conclusion, for example, the recommendation to use a treatment. The implicit warrant, usually expressed as 'If Data, then Claim', was unique for all arguments.

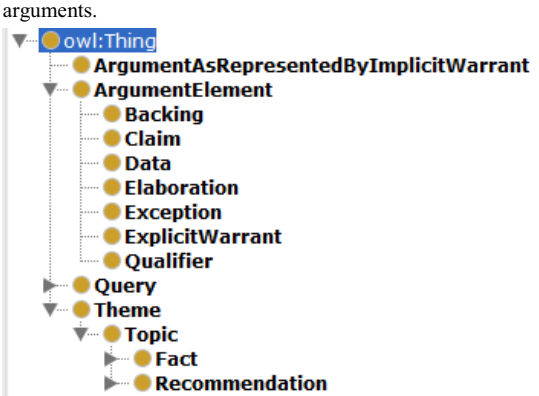


Figure 2– Top-level classes of the JADE ontology

The ArgumentElement class represents the 7 elements of the EMA, which are encoded individually as subclasses. PEM content, coded to the argument elements, was represented as instances of the coded element. For example, the argument: *If using heat and your child's skin turns red* (data)*, although some pinking of the skin is normal* (exception)*, then remove the heat source* (claim) from the pamphlet 'Using Ice and Heat at Home' (backing) was represented in the ontology as five instances: one each of the claim, data, exception, backing, and ARIW classes. Relationships between elements were encoded using object properties as shown in Table 1. Arguments were theoretically chained together when the claim of one argument contained the same content as an element of another argument. For example, the exception in: *There is no cure for JIA* (claim) *although there are medications that can reduce the inflammation and relieve the pain and swelling* (exception), is a claim in another argument: *If you have* JIA (data) *there are medications that can reduce the inflammation and relieve the pain and swelling* (claim). These links needed to be represented in the ontology, so we reused the 'sameAs' object property to link two elements of different arguments that contain the same content.

The Query class represents the questions users of the JADE system require to enter the network of PEM content encoded in the ontology. Individual questions were represented as subclasses of the Query class and were generated inductively from the PEM content. For example, PEM content describing the side effects of methotrexate, a medication used to treat JIA, induced the question: 'What are the side effects of methotrexate?'.

Each Query subclass has a class expression which attempts to translate the English question into a syntax the ontology reasoner can understand. For instance, the question above was expressed as the class expression:

(HasTheme some 'Side effects') and (IsElementOf some (HasElement some (HasTheme value Methotrexate))).

This tells the reasoner to find an argument element instance that has an instance of the 'Side effect' theme and is part of an argument in which an argument element has the theme 'Methotrexate'. The class expressions were written in the format of looking for one topic theme, or class of themes, that describes the subject of the query and one or more themes, or classes of themes, present in the argument describe its context. In the previous example, the topic theme was 'Side effects' and the context theme was 'Methotrexate'.

Table 1– The domain and range of the ontology object properties with inverse properties in brackets

Domain	Object Property	Range
Theme	DefinesThematically (HasTheme)	Argument-Element
Argument-Element	IsElementOf (HasElement)	ARIW
Backing	Backs (HasBacking)	ARIW
Elaboration	Elaborates (HasElaboration)	ExplicitWarrant, Data, Claim, Exception, Elaboration
Exception	Excepts (HasException)	Claim, Data, ExplicitWarrant, Elaboration
Qualifier	Qualifies (HasQualification)	Claim
Explicit-Warrant	Requires (HasRequirement)	Data
Data	Supports (HasSupport)	Claim
Explicit-Warrant	Warrants (HasWarrant)	Claim
Explicit-Warrant	HasAdditionalWarrant	ExplicitWarrant
Claim	sameAs	Argument-Element

The Theme class represents the thematic coding of the PEM. The codes are represented with a hierarchy of subclasses with the leaf codes as instances. For instance, the theme 'Liver damage' is an instance of the class 'Side effects' which is a subclass of the class 'Fact' that represents the context of a user's situation. The object property 'DefinesThematically' linked the instances to the argument elements they were coded to.

To date, 351 arguments extracted from the PEM have been instantiated as argument elements in the ontology. These arguments were prioritized as they were directly relevant or adjacent to topics covered in the scenarios used in the evaluation, which will be described further below. They represented most of the structures i.e. journal articles and pamphlets, topics i.e. treatments and etiology, and formats i.e. lists and images. Thus, this was a representative sample of the arguments extracted from the JIA PEM.

JADE Functionalities

The JADE system was designed with 3 functionalities:

1. A question filtering system offers a way to choose interesting questions that fit their situation.
2. An information seeking dialogue provides quality information from PEM in response to user's questions.

3. An inquiry dialogue allows users to explore the PEM content to determine whether the information provided was trustworthy and relevant to their situation.

Functionality 1: There are currently 163 questions encoded in the ontology. This functionality uses the thematic codes and class expression format to allow users to screen these options for the questions they want to ask. A user selects from a list of themes to find questions looking for: 'Recommendations about Methotrexate and DrinkingAlcohol', yielding 1, namely: 'Can I drink alcohol while taking methotrexate?'.

Functionality 2: By choosing a question, the user engages the information seeking dialogue functionality. This functionality uses the thematic coding represented in the ontology to deliver relevant content from PEM in response to user questions. Each question represents a query subclass and its class expression. The reasoner locates argument elements from the ontology that fit the class expression. In our example, the claim: 'it is best to avoid alcohol' is located because its themes are 'Drinking alcohol' and an instance of the Recommendation class and its data HasTheme 'Methotrexate'.

Question prompts for the inquiry dialogue functionality are also generated here according to the object properties present in the argument of the located element. In our example both HasWarrant and HasBacking properties are present in the located claim's argument. Thus, the full dialogue system response as per our example is:

If taking methotrexate, then it is best to avoid alcohol.

Want to know the reason for this?

Want to know the source of this?

Functionality 3: By choosing a question prompt the user can engage in an inquiry dialogue and explore the PEM content surrounding the answer to their question. The inquiry dialogue functionality leverages the object properties between argument elements, derived from the EMA, to create question prompts the user could use to explore and verify information from the PEM. To continue our example, the user chooses the first prompt: 'Want to know more about this?'. In response, the dialogue system locates the relevant argument element(s) in the range of the object property associated with this prompt, i.e. HasWarrant. This locates the explicit warrant: 'Taking methotrexate and drinking alcohol could harm your liver'. As in functionality 2 question prompts are generated. Although, in this functionality only object properties for the argument element in question, not the entire argument, will be used. The entire response as viewed by the user is:

Taking methotrexate and drinking alcohol could harm your liver.

Want to know the reason for this?

From here the user can use the home button to return to the question filtering of functionality 1, or the back button to return to the information seeking dialogue of functionality 2, or a question prompt to continue with functionality 3.

Evaluation Study

The JADE system was qualitatively evaluated to determine whether its responses were complete, relevant, accurate, and understandable. We aimed to recruit 5-8 healthcare providers from the IWK Pediatric Rheumatology Division for their expert domain knowledge of JIA and familiarity with the PEM content used in the dialogue system. Participants completed a cognitive walkthrough of the dialogue system followed by a semi-structured interview. Cognitive walkthroughs involved the participants verbalizing their thoughts as they interacted with the dialogue system, guided by a scenario and set of tasks [15]. The interview questions were based on the content portion of the O'Grady framework for evaluating interactive applications [16]. Screen capture and audio recordings of the evaluations were analyzed qualitatively using directed content analysis, with the O'Grady framework used as predetermined codes [17]. Approval for this study was given by the IWK Research Ethics Board (approval #1023261).

Results

6 clinicians have participated in the evaluation study: 4 nurses and allied health professionals and 2 pediatric rheumatologists. Two had less than 5 years of experience working in the pediatric rheumatology division, 3 had between 11 and 20 years of experience, and one had more than 20 years of experience. Three claimed to be very comfortable with computers, 1 was moderately comfortable, while two were moderately uncomfortable. Three reported that more than 75% of their clinical encounters involved using PEM, while 3 reported less than 40%.

Eight major themes were identified during analysis. Themes were determined to be major if 40% or more of participants had responses categorized under that code. Saturation was achieved, as evidenced by the fact that no new codes were added after the third evaluation. The major themes are listed below with examples of quotes from participants.

1. Positive responses to the JADE system.

   Participant 6: Having that data…available like this is really exciting!

2. Content was largely accurate outside of some inaccuracies in the PEM.

   Participant 4: [There was] nothing that was incorrect.

   Participant 3: [This PEM is from] 2011, yikes, that'll be from before some of these kids were born.

3. Content was credible but could be improved through better presentation.

   Participant 1: [Having links to the sources] is going to be really helpful.

   Participant 4: For instance, if you have a [hospital] logo… that tends to increase folk's credibility.

4. Content was mostly relevant, with a few exceptions.

   Participant 4: I think it's really relevant to what the patients and families are experiencing.

   Participant 5: Interesting, because I clicked on [a question about] 'ice' and then it starts talking about 'heat' [when I click on a question prompt].

5. Content was mostly complete, with a few exceptions.

   Participant 5: As you continued to 'want to know more about it' I think you did get all the information.

   Participant 3: Is there something about not getting live virus vaccines [while taking methotrexate]? One of your issues is going to be how are all the links set up, because they aren't a single linear link, its lots of things that end up being a complicated Gordian knot.

6. Individual responses were clear and consise, but a broader organization of the content was lacking.

   Participant 2: I think that's simple, to the point, uncomplicated.

   Participant 3: You have things that follow each other that are totally unrelated…with a big long list, people read the first two or three things and then they get bored and miss stuff.

7. Awkward wording occasionally made the information unclear.

Participant 5: 'What does that describe?' ... It's a bit awkward.

8. Accessibility issues to the JADE system.

   Participant 4: This is very much geared towards users that have the ability to be able to navigate it physically. Are there any parameters for folks who have limitations in terms of audio or visual?

Discussion

The results of the evaluation study show that the application of the AI-based argument theory method for delivering patient education dialogue was largely successful. The PEM content was fully integrated and formalized into the JO using the EMA and domain experts found the resulting dialogue content to have utility, outside of a few exceptions. The reasons for the few irrelevant, incomplete, and difficult to understand responses found during the evaluation are discussed here.

The majority of irrelevant or incomplete responses were due to gaps in the existing PEM. The remainder were caused by two issues with how the thematic coding and chaining between arguments were represented in the JO. First there was no representation in the JO of the causation of a side effect by a drug opposed to merely appearing in the same argument as the drug. Second, the sameAs property caused the reasoner to conflate the themes from two chained arguments, which should be separate, leading to irrelevant elements being selected during information seeking dialogue. Future work will address this by using alternate methods of representation.

The 6th theme of the evaluation shows the limits of the EMA, which does not contain constructs to model knowledge beyond the level of individual arguments. Thus, while the EMA and thematic coding have successfully represented a large portion of the knowledge contained in the PEM, another layer of representation is required. The relative importance and overall subject of each argument need to be modelled to organize the elements, relative to each other, in the system's responses. Future work will identify such a model and implement it.

We believe the remaining issues highlighted in the evaluation study, i.e awkward wording, accessibility issues, and credible presentation, can be best addressed in co-design with JIA stakeholders. Future work will therefore seek to engage families of children with JIA and their healthcare providers.

Limitations

The participants recruited for the evaluation study were from a single site and while their familiarity with the PEM used in the system was beneficial to the study it potentially limits the breadth of viewpoints a broader inclusion strategy could have offered. Similarly, the analysis of the results was performed by a single coder which increases the potential for bias.

While the methods used in this work are generalizable to patient education for other conditions, the substantial time and knowledge required, limits their applicability. The coder(s) must have a good working knowledge of the EMA as well as grounded theory for the thematic coding. They must also understand the condition described by the PEM and be able to represent the resulting codes in an ontology. The time needed to do this is an additional barrier to these methods being used.

Conclusions

This work describes a novel method of using a model of argument to create a patient education dialogue system for healthcare users. The evaluation of the JADE system showed that it provided responses that were mostly relevant, understandable, and complete, thus demonstrating its potential to address some of the current gaps in patient education experienced by families of children with JIA.

References

[1] O. Gómez-Ramírez, M. Gibbon, R. Berard, R. Jurencak, J. Green, L. Tucker, N. Shiff, and J. Guzman, A recurring rollercoaster ride: a qualitative study of the emotional experiences of parents of children with juvenile idiopathic arthritis, *Pediatric Rheumatology*. **14** (2016).

[2] H. Cooper, K. Booth, S. Fear, and G. Gill, Chronic disease patient education: lessons from meta-analyses, *Patient Education and Counseling*. **44** (2001) 107–117.

[3] J. Wright, B. Rose-Davis, M. Batthish, T. Cellucci, C.M. Duffy, L. Tucker, A. Huber, B. Lang, D.M. Levy, D. Rumsey, K.W. Duffy, J. Curran, and E. Stringer, Knowledge Translation in Juvenile Idiopathic Arthritis in Canada: A Focus on Parents of Children with JIA [abstract], *Arthritis and Rheumatology*. **69 (suppl 10)** (2017).

[4] A. Syrowatka, D. Krömker, A.N. Meguerditchian, and R. Tamblyn, Features of Computer-Based Decision Aids: Systematic Review, Thematic Synthesis, and Meta-Analyses, *Journal of Medical Internet Research*. **18** (2016).

[5] Y. Wilks, R. Catizone, S. Worgan, and M. Turunen, Some background on dialogue management and conversational speech for dialogue systems, *Computer Speech & Language*. **25** (2011) 128–139.

[6] W.G. Harless, M.A. Zier, M.G. Harless, R.C. Duncan, M.A. Braun, S. Willey, C. Isaacs, and R.D. Warren, Evaluation of a virtual dialogue method for breast cancer patient education, *Patient Education and Counseling*. **76** (2009) 189–195.

[7] S.-A.A. Jin, The effects of incorporating a virtual agent in a computer-aided test designed for stress management education: The mediating role of enjoyment, *Computers in Human Behavior*. **26** (2010) 443–451.

[8] D.N. Walton, Dialog theory for critical argumentation, John Benjamins Pub., Amsterdam; Philadelphia, 2007.

[9] D. Hitchcock, and B. Verheij, eds., Arguing on the Toulmin model: new essays in argument analysis and evaluation, Springer, Dordrecht, 2006.

[10] R.D. Shankar, S.W. Tu, and M.A. Musen, Medical arguments in an automated health care system, in: AAAI Press, Stanford, CA, 2006: pp. 55–63.

[11] B. Rose-Davis, E. Stringer, S. Abidi, and S.S.R. Abidi, Interactive Dialogue-Based Patient Education for Juvenile Idiopathic Arthritis Using Argument Theory, *Studies in Health Technology and Informatics*. (2018) 546–550.

[12] S.S.R. Abidi, Healthcare Knowledge Management: The Art of the Possible, in: D. Riaño (Ed.), Knowledge Management for Health Care Procedures, Springer-Verlag, Berlin, Heidelberg, 2008: pp. 1–20.

[13] F. Vitali, and S. Peroni, The Argument Model Ontology, (2011). http://www.essepuntato.it/2011/02/argumentmodel (accessed January 2018).

[14] M. De Chesnay, ed., Nursing research using grounded theory: qualitative designs and methods in nursing, Springer Publishing Company, LLC, New York, NY, 2015.

[15] C. Wilson, Cognitive Walkthrough, in: User Interface Inspection Methods, Elsevier, 2014: pp. 65–79.

[16] L. O'Grady, H. Witteman, J.L. Bender, S. Urowitz, D. Wiljer, and A.R. Jadad, Measuring the Impact of a Moving Target: Towards a Dynamic Framework for Evaluating Collaborative Adaptive Interactive Technologies, *Journal of Medical Internet Research*. **11** (2009) e20..

[17] H.-F. Hsieh, and S.E. Shannon, Three Approaches to Qualitative Content Analysis, *Qualitative Health Research*. **15** (2005) 1277–1288.

Address for correspondence

Benjamin Rose-Davis, Faculty of Computer Science, 6050 University Avenue, Dalhousie University, Halifax, Nova Scotia, Canada, B3H 1W5; E-mail: ben.rose-davis@dal.ca.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190445

Student Nurse Attitudes and Behaviours when Using Social Network Sites

Roseanne Sadd

Department of Nursing, Toi Ohomai Institute of Technology, Tauranga, New Zealand.

Abstract

The purpose of this study is to describe the effect of education in professional boundaries on the use and management of social media by using quantitative survey methods to ask "What are the social networking behaviours of student nurses following education in professional boundaries?

Findings from this research indicate that student nurses are active SNS users, primarily for personal engagement. Facebook is overwhelmingly the most popular SNS, with snapchat and Instagram also commonly used. While students primarily used SNS for personal reasons, many reported SNS use for educational / professional purposes as well, including to discuss academic related topics. Most students responded that they were aware of privacy settings on SNS, however there is a discrepancy between awareness of privacy settings and the number of students implementing the privacy features.

Keywords:

Privacy, Professionalism

Introduction

Social network sites (SNS) are a form of social media that allow immediate, real-time interaction, utilising Web 2.0 applications to create, share, and interact with, information. In New Zealand, two-thirds of Internet users engage in regular social network use including sites such as Facebook, Twitter, blogs and forums [1]. SNS have become a common form of communication and interaction in our personal and professional lives [2,3]. Nurses too, are increasingly using SNS, both as part of their professional role, and in their private lives [4].

While there is increasing literature regarding student use of SNS, there is a little research that specifically examines students from the health professions despite the risk of breaching the privacy of those the health professionals care for [5]. Few studies involving health professionals (student or otherwise) discuss professionalism when using SNS. Content analysis of the SNS Facebook by Levati, [2] found differences in the use and access of SNS by United Kingdom and Italian nurses. However, both groups had posted personal information and potentially unprofessional information. Hall, Hanna & Huey [6] reported that despite awareness of privacy settings and requirements when using SNS, pharmacy students blurred the lines between personal and professional online behaviours. Similarly, student nurses in South Africa lacked understanding of the risk of posting personal content, many reporting having postings that potentially breached patient privacy [7].

In New Zealand, nurses are educated within a Bachelor of Nursing or similar baccalaureate degree governed by the Nursing Council of New Zealand (NCNZ) under the Health Practitioners Competence Assurance Act 2003. It is the responsibility of educational institutions providing these programmes to ensure student nurses achieve competency to practice, including ethical, legal and professional competency. Although students do not fall directly under the jurisdiction of the nursing council, student nurses must comply with the standards of their educational institution and maintain the same professional conduct as a Registered Nurse. Nursing education needs to keep abreast of changes to develop graduate nurses who are confident and competent to practice within the healthcare environment. Educational institutions have a responsibility to ensure nursing students are aware of the risks of social network use and have strategies to ensure their on-line presence remains professional [8,9]. Currently there is no published research on the use of SNS by nurses or student nurses in the New Zealand context.

In year one of a Bachelor of Nursing programme students complete a course on professional practice with the aim of introducing students to the "professional, ethical and legislative requirements of nursing practice". The question that arises is whether students integrate this learning into their own use of social network; in simple terms, does theory integrate into practice?

The central question asked by this study is "What are the social network site attitudes and behaviours of student nurses following education in professional boundaries".

The scope of this research is the self-reporting attitudes and behaviours of students nurses currently enrolled in a Bachelor of Nursing course. This research is not expected to measure the effectiveness of professional boundary education, but, rather investigate and describe student social network attitudes behaviour to the maintenance of professional boundaries online – have students have put theory into practice when using social network sites?

Methods

This research is a descriptive survey using quantitative methods to provide a snapshot of student nurse attitudes and behaviours regarding the use of SNS, ascertaining frequency and purpose as well as describing the measures taken by student nurses to maintain professionalism.

The objective of this research is to investigate the use and management of social network sites by Bachelor of Nursing students, questioning how students use SNS, how they protect privacy on SNS, and, what their attitudes are to maintaining professionalism when engaging with SNS. Do students recognise the risks involved with social networking? What measures do students take to protect their own information, and

the information of patients they care for, in order to maintain their professional boundaries?

Ethics approval was granted by the Faculty of Business, Information Technology, and Creative Arts, Research Committee of Waiariki Institute of Technology (now Toi Ohomai Institute of Technology).

Sampling

The target population was Bachelor of Nursing students at one tertiary institute who had successfully completed a course in professional boundaries. A face to face information session for potential participants was followed up by an email outlining the research project as their rights as a research participant. Interested participants were emailed a link to the online survey giving assurance of confidentiality, informed consent, and voluntary participation. Consent was gained by participants agreeing (consent) online prior to the survey commencing. Subsequently eligible students were sent an email with an individual participant web-link directing them to the survey. Students were emailed a reminder at one week and before the survey closed using their online learning platform. Once the survey had closed, results were exported from Survey Monkey and prepared for importing into statistical processing software.

Data collection

Data were collected by way of an online self-reporting questionnaire using the web-based format of 'Survey Monkey', (http://www.surveymonkey.com), adapted from that of Hall, Hanna and Huey [6]. The adapted questionnaire was reviewed for face validity by two senior colleagues and content validity by an experienced researcher from another institution. Participants were asked to self-report behaviors and attitudes when engaging with SNS.

Participants were asked to self-report behaviors and attitudes when engaging with SNS. The questionnaire is presented in four sections.

- Use of social networking
- Online profile and privacy
- Professionalism and social networking
- Demographics

The questionnaire asked a series of closed - ended questions answered by yes or no, or, questions to determine responses using a Likert scale ranging from strongly agree to strongly disagree.

Due to the population (BN students within one specific program), no gender or ethnicity data was collected as this information could potentially identify participants. The only demographic data collected was age group <25years or > 25years. This age selection relates to the concept of digital natives (<25 years age) who have grown up with online technology and digital immigrants (>25 years age) who have had to adapt to the technology [10]

Once the survey had closed, results were exported from Survey Monkey and prepared for importing into statistical processing software.

Data analysis

Responses were coded and entered into IBM's Statistical Package for Social Sciences (SPSS) version 24. Data were analyzed using descriptive and inferential statistics providing a description of student nurse behaviours and attitudes when using SNS. Based on the preliminary analyses, it was determined that the research data was primarily categorical and ordinal in nature and non-normally distributed therefore relationships between use, privacy and professionalism were identified using non - parametric tests. An a priori level of less than .05 ($p<.05$) was set as significant. Missing data were not estimated or used in analyses.

Results

228 year one and year two Bachelor of Nursing students were invited to participate in an online survey via an email link to the Survey Monkey website. 102 valid responses were received.

Table 1: Sample population / respondents

Age Group	Total Population	Respondents (valid)	% valid respondents to population
18-25	112	47	42%
26+	116	55	48%
Total	228	102	45%

"How do student nurses use SNS?"

Respondents were asked to select which SNS sites they use. Over 90% reported using Facebook with Snapchat and Instagram being the next most common SNS' reported (Figure 1). 'Other' included sites such as Youtube (*n*=4), Tumblr (*n*=1), and WhatsApp (*n*=1).

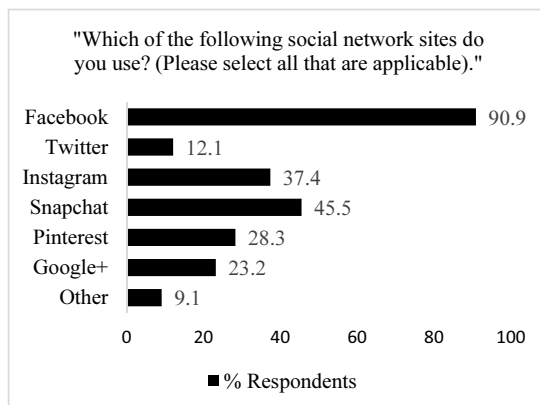


Figure 1: SNS accessed by students

Almost half of the students (49.5%) used three or more SNS while over 70% of students spent between 0.5 – 2 hrs accessing SNS daily. Students access SNS primarily for personal reasons (94.5%), with over half of the respondents reporting that they use SNS for both personal and education purposes (53.5%).

Students reinforced their use of SNS for educational purposes with 65.7% strongly agreeing or agreeing to the question to the questions that they "I *have used social network sites to discuss academic related problems*". However, just under one fifth of students (19.1%) disagreed with this statement. Most students

(92.9%) strongly agreed or agreed that they had "*been made sufficiently aware of the professional behaviour that is expected of me when using social network sites*".

The relationship between number of SNS used and purpose for using SNS showed there was a small positive relationship between the two variables with the higher the number of SNS used, the higher the number of reasons for using SNS, rho = .226, n = 99, p < .05. This relationship was stronger in the 18-25 year age group rho = .311, n = 45, p <.05.

A Mann-Whitney U Test was used to test for difference between the two age groups (18-25, 26+) and responses to the statement "I have used social network sites to discuss academic-related problems". A significant difference was found ranking the 18-25 year age group (Md = 1, n = 45) higher than the 26+ age group (Md = 1, n = 54), U = 965000, z = -2.09, p = .04, r = .21. indicating the 18-25 age group are more likely to use SNS for academic discussion.

"Do students protect privacy when using SNS?"

The first two questions in this section referred to student awareness and use of SNS privacy settings. Although most students were "*aware of the privacy features available on social network sites limiting the amount of information available to the public"*, only three quarters (76.8%) of students reported using these privacy features (Figure 2).

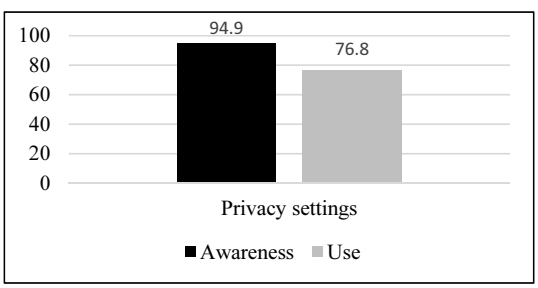


Figure 2: Awareness & use of privacy settings

Over half of the respondents reported that they agreed with the statement "*The profile and activity I present on social network sites websites is an accurate representation of my character*", however, less that one quarter 'strongly agreed' (22.2%), and over a quarter neither agreed or disagreed with the statement When asked about the concern "*that others could post information or photos"* students had a varied response. Just over half agreed (40.4%) or strongly agreed (13.1%) that this concerned them; a further 27.3% neither agreed nor disagreed When asked whether it was "*acceptable to publish whatever I like"* if privacy settings were in place, again the response was varied, 69.7% reporting they 'disagree' or 'Strongly disagree" with the statement, and 30.2% of respondents selecting 'neither agree nor disagree' or 'agree'.

There was also a relationship between the use of SNS privacy settings and the unacceptability of publishing "*whatever I like on social network sites*". The relationship is weak but suggests awareness of maintaining privacy when publishing on SNS, *rho* = .235, $n = 99$, $p < .05$.

There was no significant difference between age groups and awareness of privacy settings. Similarly, there was no significant association between age group and use of privacy settings, χ^2 (1, n=99) – 0, p = 1.0, *phi =0.22*

Do students maintain professional boundaries when using SNS?

The final section of the questionnaire asked questions about students' attitudes to professionalism when using SNS (Table 2). Most students (79.8%), 'agreed' or 'strongly agreed' that they maintain the same professional standards whether using SNS or on clinical placements and 96% of students took responsibility for all information they have posted on SNS including comments.

Table 2: Responsibility for own postings

Strongly agree	67.7%
Agree	28.3%
Neither agree nor disagree	2%
Disagree	2%

A portion of students reported having posted information they would not want either academic staff or an employer to see (Figure 3).

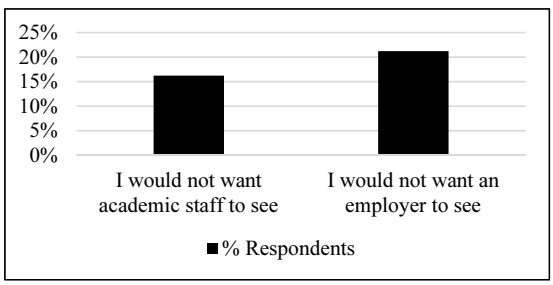


Figure 3: Information posted seen by academic staff / employer

When asked whether it was fair that employers use information from social network sites there was no clear response given, similarly, responses were mixed for the question about whether students were concerned that their professionalism could be judged by their SNS use (Table 3). Just over 50% of respondents reported that viewing of the SNS activities of student nurses by the public could affect their reputation. When qualified or seeking employment 69.7% of respondents said when they would not change SNS behaviours, while 30.3% stated they would change their SNS behaviour when qualified.

Table 3: Judging professionalism by SNS use

Strongly agree	9.1%
Agree	41.4%
Neither agree nor disagree	23.2%
Disagree	23.2%
Strongly disagree	3.0%

The majority of respondents (94%) 'Disagreed' or 'Strongly disagreed' that it was acceptable to comment about patients. While it is encouraging that there was a clear understanding of the importance of respecting patient privacy, it is a concern that this question was not answered with 100% disaffirmation.

Kruskal -Wallis Tests did not reveal any significant difference in either total privacy or profile (composite variable), or total professionalism (composite variable) across the different number of SNS used by students (p<0.5).

Discussion

There are approximately 48000 Registered Nurses in New Zealand [10], but no detail on when, if, or how, they engage with social network sites. Nurses hold a privileged position where patients may share personal information and insights into a patient's life. Nurses are often privy to very personal and sometimes delicate information; therefore have a legal, and ethical, responsibility to protect this information and maintain patient privacy [11]. Inappropriate on-line behaviour by nurses can result in damage to their relationships with patients, colleagues and the profession as a whole. Registered Nurses' use of SNS has the potential to breach professional boundaries, breach privacy and damage trust the public have in the nursing profession.

The aim of this study was to determine if student nurses maintain professional boundaries when engaging on social network sites. While the overall response rate to the online questionnaire was only 45%, the respondent age groups were similar to that of the student nurse population. It is possible that potential participants who did not use SNS most likely did not respond, therefore the analysis of SNS use focused on "how" and "what" SNS do students use.

Findings indicated that students are active SNS users, primarily for personal engagement, in line with other studies [6,12]. Facebook is overwhelmingly the most popular SNS, reflecting general population trends [13]. Student nurses of both age groups in the study, (digital natives, 18-25 year age group, and digital immigrants, 26+ age group) are engaging with SNS, primarily Facebook, with snapchat and Instagram also commonly used.

The similarity in use for the two age groups in the study reflect StrauB and Nentwich's [14] assertion that SNS has become relevant to all ages. While students primarily used SNS for personal reasons, many reported SNS use for educational / professional purposes as well, including to discuss academic related topics. Hall et al. [6] identified a pattern of increasing use of SNS for academic purpose as students entered professional study. It is possible that students who use SNS for education and personal reasons, use multiple SNS sites for different purposes, although the digital natives in the study are more likely to use SNS for academic purposes. Having grown up with online technology, may predispose the digital natives to see SNS as an immediate source of assistance.

Most students responded that they were aware of privacy settings on SNS, however there is a discrepancy between awareness of privacy settings and the number of students implementing the privacy features despite strong recommendations in professional nursing guidelines [9]. This is in contrast to findings by Hall et al., [6], and Kitsis et al. [12] which reported high levels of privacy setting by health profession students.

Students in the study appear to be aware of how they may be judged by their online profile. Despite this, while not evident in this research it is possible that students who did do not use privacy settings are more likely to have inappropriate postings viewed by others, especially as some students reported concern about what others may post about them.

SNS such as snapchat and Instagram (the second and third most popular SNS used by respondents) are mainly used to share images and photos, a format which can inadvertently expose identifiable information by capturing an organisation or emblem in an image, or people in the background of a photo. Students who do not utilise privacy settings may feel their privacy settings are sufficient to protect them, avoid photos or have friends and acquaintances who rarely use post on SNS. Some of the most common posts which portray nurses in potentially unprofessional situations are photos [2].

Professional boundaries need to be maintained not just when dealing with patients or clients. Dual relationships can also develop between a nurse and colleague, employer or other health professionals. Blurring of boundaries results when information gleaned from the professional environment is brought into the personal environment such as discussing the workplace in a social setting. This is exacerbated when in the online SNS setting as information is more able to be inadvertently shared, or, misinterpreted.

The image portrayed by nurses on SNS has an impact on the image of the nursing profession as a whole [16]. Over half of the students in the study reported general concerns with how the nursing profession reputation would be portrayed if student nurse SNS were viewed by public, yet the majority of student nurses in the study took responsibility for content on their SNS and reported maintaining the same professional standards whether on SNS or on a clinical placement suggesting understanding of the importance of professional behaviour. Some students reported they would change their SNS behaviour when becoming qualified (registered), questioning whether students consider themselves as part of the nursing profession while completing their qualification.

Students reported understanding of the importance of maintaining privacy and professional behaviours, yet response were varied when considering how others may view their SNS use, possibly relating to the ethical aspects of others viewing an individual's SNS. Whether it be a potential employer viewing an applicant's SNS or someone judging a nurse's professionalism based on their personal SNS postings, answers could be reflective of whether it is right to use SNS in this way rather than reflective of students postings.

From this research, and other previous research, it is unclear whether students do understand the importance of protecting their personal persona on SNS for today and for the future? Similar findings from the United Kingdom concluded that while nurses are aware of the risks of SNS use, a significant number only partially protected their privacy online [2]. Kitsis et al. [12] acknowledges that medical students can struggle to develop a professional identity separate from their personal persona online, while Hall et al. [6] and Nyangeni et al. [7] reinforce the need for ongoing education in professional behaviour. Findings from this research indicate that student nurses who have completed education in professional boundaries in the first year of their degree are aware of privacy and the need for care when posting on SNS, but it cannot conclude that student do indeed maintain boundaries without further research.

Limitations

Limitations to the research findings result from the small sample size and self-reporting format used for the survey. Non-response bias cannot be eliminated, however the proportion of respondents did reflect the population makeup. A written questionnaire format may have delivered more responses, but could also have pressured potential participants to take part.

Findings could be further explored through a larger study, by interview or by analysis of actual SNS sites in the future. Findings may not offer a specific revelation, but provide a description of student nurse attitudes and behaviours when using SNS within the New Zealand context.

Conclusion

The Nursing Council of New Zealand (NCNZ) have demonstrated the importance of nurses maintaining their professionalism when online through the Code of Conduct and guidelines for social media [11,15]. Despite guidelines from NCNZ, and education in the first year of the Bachelor of Nursing degree, organisational policies are required to set standards of acceptable online behaviour when engaging with SNS. Education must be ongoing throughout the degree, not only in the regard to use of SNS, but in all areas where there is blurring of the personal and professional boundaries of student nurses and the potential to breach professional conduct.

With a respondent's rate of 45% for this research, it is not possible to make conclusions about the wider student nurse population. However, data from the general population and from other studies reporting health professional student use of SNS suggest it is likely that the majority of students are engaging with SNS.

While we know what SNS sites student nurses are using, little is known about the type of content being posted. If students are using SNS more frequently for educational purposes, what type of academic questions are being asked, and where is the information coming from. Content analysis of SNS used by student nurses would validate the self-reported findings and perhaps help explain what nurses consider as acceptable postings on SNS, which would build on this study in an attempt to answer many of the questions raised by the research findings.

Acknowledgements

This research was undertaken in partial fulfillment of Master of Management (Health). The author acknowledges all Bachelor of Nursing students who completed the survey.

References

[1] Statistics New Zealand. *Household use of information technology and communication technology: 2012.* Data file. Statistics New Zealand, 2013, http://www.stats.govt.nz/browse_for_stats/industry_sectors/information_technology_and_communications/HouseholdUseofICT_HOTP2012.aspx

[2] S. Levati, Professional conduct among Registered Nurses in the use of online social networking sites. *Journal of Advanced Nursing*, **70** (2014), 2284-2292, http://doi:10.1111/jan.12377

[3] J. Peck, Social media use in nursing education: responsible integration for meaningful use. *Journal of Nursing Education*, **53** (2014), 164-169, http://doi:10.3928/01484834-20140219-0339

[4] New Zealand Nurses Organisation, NZNO National Student Unit & Nurse Educators in the Tertiary Sector. *Social media and the nursing profession: A guide to online professionalism for nurses and nursing students*, 2012, http://www.nzno.org/resources/nzno_publications

[5] N. Mahapatra, and M.C. Schatz, Social networking among health science university students: examining social network usage, social support, and general wellbeing. *Journal of Human Behavior in the Social Science Environment* **25** (2015), 618-629. http://doi.10.1080/10900359.2015.1011253

[6] M. Hall, G. Huey, and L. Hanna, Use and views on social networking sites of pharmacy students in the United Kingdom. *American Journal of Pharmaceutical Education*, **77,** (2013), Article 9.

[7] T. Nyangeni, S. du Rand, and D. van Rooyen, Perceptions of nursing students regarding responsible use of social media in the Eastern Cape. *Curationis*, **38**, (2015), 9, http://doi:10.4012/curationis.v38i2.1496

[8] T. Schmitt, S.S. Sims-Giddens, and R.G. Booth, Social media use in nursing education. *The Online Journal of issues in Nursing*, **17,** (2012), 15. http://doi:10.3912/OJIN.Vol17No03Man02

[9] J. Green, A. Wylie, and D. Jackson, Social networks for nurse education: Possibilities, perils and pitfalls. *Contemporary Nurse*, **47,** (2014), 180-189.

[10] Nursing Council of New Zealand, *Annual report,* Nursing Council of New Zealand, Wellington, New Zealand, 2014, http://www.nursingcouncil.org.nz/content/download/1399/5240/file/NCNZ-AR2014_WEB.pdf

[11] Nursing Council of New Zealand, *Code of conduct for nurses*, Nursing Council of New Zealand, Wellington, New Zealand, 2012, http://www.nursingcouncil.org.nz/Publications/Standards-and-guidelines-for-nurses

[12] E.A. Kitsis, F.B. Milan, H.W. Cohen, D. Myers, P. Herron, M. McEvoy, . . .M.S. Grayson, Who's misbehaving? perceptions of unprofessional social media use by medical students and faculty. *BioMed Central Medical Journal*, **16**, (2016), 7, http:// doi:10.1186/s12909-016-0572-x

[13] M. Duggan, N.B. Ellison, C. Lampe, A. Lenhart, and M. Madden, *Social media update 2014*. Pew Research Center, 2015, January, http://www.pewinternet.org/2015/01/09/social-media-update-2014/

[14] S. StrauB, and M. Nentwich, Social network sites, privacy and the blurring of boundaries between public and private spaces. *Science and Public Policy*, **40**, (2013),724-732, http://doi:doi:10.1093/scipol/sct072

[15] Nursing Council of New Zealand, *Guidelines: Social media and electronic communication*, Nursing Council of New Zealand, Wellington, New Zealand, 2012. http://www.nursingcouncil.org.nz/Publications/Standards-and-guidelines-for-nurses

[16] Nursing Council of New Zealand, *Guidelines: Professional boundaries*. Nursing Council of New Zealand, Wellington, New Zealand, 2012, http://www.nursingcouncil.org.nz/Publications/Standards-and-guidelines-for-nurses

Address for correspondence

roseanne.sadd@toiohomai.ac.nz

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190446

Multi-Modal Methodology for Adapting Digital Health Tools to New Populations: Adaptation of the Video Information Provider (VIP) for Persons Living with HIV with HIV-Associated Non-AIDS (HANA) Conditions

Rebecca Schnall[a], Jianfang Liu[a], David C. Mohr[b], Suzanne Bakken[a,c], Sabina Hirshfield[d], Karolynn Siegel[e], Samantha Stonbraker[a], Hwayoung Cho[a], Sarah Iribarren[f], Joachim Voss[g]

[a] Columbia University, School of Nursing, NY, NY, USA; [b] Northwestern University, Department of Preventive Medicine, Chicago, IL, USA; [c] Columbia University, Departme4nt of Biomedical Informatics, NY, NY, USA; [d] SUNY Downstate Medical Center, Brooklyn, NY, US; [e] Columbia University, Mailman School of Public Health, Department of Sociomedical Sciences, NY, NY, USA; [f] University of Washington, School of Nursing, Seattle, WA, USA; [g] Case Western Reserve University, Frances Payne Bolton School of Nursing, Cleveland, OH, USA

Abstract

The goal of this study was to illustrate the translation of our extant eHealth intervention (VIP) into an mHealth app for persons living with HIV (PLWH) with HIV-Associated non-AIDS (HANA) conditions, a new clinical problem. We integrated different quantitative and qualitative methodologies from different disciplines to accomplish the task of adapting an eHealth system for a new set of clinical problems. Building off of our past development of the VIP website, we used a multi-modal, iterative user-centered design process to develop the VIP-HANA app. Our process was used to better understand the needs of a national sample of PLWH recruited online. Findings from the usability evaluation demonstrate a potentially useful and easy to use app. Integration of multi-modal methodologies from different fields to accomplish the tasks of adaptation and evaluation of a mobile app is an appealing, rigorous and useful approach.

Keywords:
HIV, Mobile Applications, Symptoms

Introduction

Psychological theories often provide the guiding framework for eHealth interventions, yet the actual development process often relies on developer clairvoyance and the serendipity of the researcher in successfully communicating his/her goals to the software developer. This oftentimes arbitrary process is less than scientific and not ideally suited for efficient, rigorous and reproducible intervention development. The instantiation of behavioral health interventions delivered via technology and the expanded need for translating eHealth tools to a new clinical problem are the impetus for clearly outlining the development process of consumer health informatics tools. The goal of this paper is to illustrate the integration of multi-modal methodologies for adapting an extant intervention for new or growing clinical problems. Our multi-modal iterative development process was comprised of an online web-based survey, expert review, software development, and usability evaluation with experts and end users.

Background

HIV has evolved from an acute to a chronic illness largely due to advances in antiretroviral therapy (ART) [1]. As a result, persons living with HIV (PLWH) are living longer and about 50% of PLWH in the US are currently over 50 years old [2]. As PLWH age,they are developing chronic illnesses and co-morbid conditions at a younger age than individuals who are HIV negative [3, 4]. Fifty to sixty percent of deaths in PLWH occur from HIV associated non-AIDS (HANA) conditions (e.g., cardiovascular disease, liver disease, diabetes, and asthma) and people suffering from these conditions are more likely to be affected by adverse symptoms [5]. An individual's ability to identify and self-manage symptoms of HIV has been shown to improve patient outcomes and quality of life [6-8].. Consequently, a team of researchers at UCSF School of Nursing developed a paper-based symptom management manual with self-care strategies for 21 common HIV/AIDS symptoms. The efficacy of this manual for improving symptoms was established in a 775-person randomized clinical trial (RCT) over three months at 12 sites [9].

To enhance uptake of these strategies by PLWH, we developed and pilot-tested the Video Information Provider (VIP), a web application (app) that delivered HIV-related symptom self-care strategies for PLWH (P30NR010677-03S1). Results from our 3-month pilot study (N=42) overwhelmingly demonstrated the feasibility of the system. Participants reported a decrease in HIV-related symptom frequency and intensity after using the VIP system for 12 weeks [10]. However, the UCSF symptom management manual and VIP were not developed for managing symptoms related to HANA conditions. Little is known about the symptom experience or self-care strategies used by PLWH with HANA conditions. Based on the existing evidence and our previous work, we sought to adapt the VIP system to include tailored self-care strategies to ameliorate HANA condition-related symptoms. Strategies are tailored based on the following characteristics: symptom, sex, race/ethnicity and HANA condition. Following our identification of the most useful self-care strategies for PLWH with HANA conditions, we sought to incorporate these strategies into a user-centered consumer health informatics app (VIP-HANA) to improve health outcomes.

Anonymous Online Survey

Methods

To assess the frequency and severity of HANA symptoms, we distributed an anonymous online survey. Details on the methods and findings for the online survey have been published elsewhere [11]. In short, we posted banner advertisements on CraigsList, POZ.com, BGCLive.com, and Facebook.com to

recruit 769 PLWH living in the US to participate. Those who clicked on the banner ad were directed to a study landing page hosted by Qualtrics software that allowed them to complete a screener to determine if they qualified for study participation. Inclusion criteria included: a) 18 years or older, b) having a HIV diagnosis, c) current US residence, d) a diagnosis of one of the following HANA conditions: arthritis, asthma, cardiovascular disease, chronic obstructive pulmonary disease (COPD), diabetes, liver disease, osteoporosis and renal failure, e) not being pregnant or breastfeeding, f) able to read and understand Spanish or English, and g) having at least one of 28 HANA symptoms. Those who qualified completed the consent form through a clickable link preceding the Qualtrics survey.

The survey assessed the frequency and bothersomeness of 28 symptoms: 1) anxiety, 2) change in appetite, 3) clumsiness, difficulty with balance, 4) difficulty concentrating, 5) constipation, gas, bloating, 6) cough, 7) depression, 8) decreased sex drive, 9) diarrhea, 10) difficulty remembering, 11) lightheadedness/ dizziness, 12) dry eyes, 13) thirst/ dry mouth, 14) difficulty falling asleep, 15) low energy/ fatigue, 16) fever, night sweats or chills, 17) heartburn, 18) problems achieving or maintaining an erection (males only), 19) muscle aches or pain, 20) neuropathy, 21) pain or discomfort during sex, 22) ringing in ears/ noise intolerance, 23) shortness of breath, 24) speech difficulties, 25) difficulty staying asleep, 26) difficulty with urination, 27) nausea/ vomiting, and 28) unplanned weight change. We included open-ended questions for participants to report self-care strategies used to ameliorate each symptom experienced, rate its degree of helpfulness and identify the information source for the strategy.

Results

Survey results yielded a total of 4,036 self-care strategies across the 28 symptoms. These survey findings provided additional information on the usefulness of each strategy as well as the information source for the strategy, which are reported elsewhere [11].

Coding and Analysis of the Self-Care Strategies

Methods

We used multiple steps to rank self-care strategies (very helpful, only a little helpful, etc.) that could then be used to tailor recommendations based on individual characteristics (using branching logic) to be integrated into the app. A qualitative data analytic approach was used to code and organize the 4,036 self-care strategies obtained in the anonymous online surveys. Open-ended responses were categorized into an overall heading. For example, open-ended survey responses included "take a nap," "rest," "rest in the afternoon," and these were all categorized as "take a nap." We then removed the self-care strategies that were rated as only a little helpful or not helpful at all. We also had a group of expert clinicians, HIV researchers, and nurses review each of the strategies to ensure their safety and legality. The purpose of this work was to identify helpful self-care strategies for future use by PLWH with HANA conditions and therefore we did not include strategies which were not helpful, not clinically advised (e.g., do nothing for a depressed patient), or potentially harmful (e.g., using illegal substances or gambling).

After consolidating initial codes of strategies on the basis of conceptual similarities and clinical significance, and after removing unhelpful, illegal and unsafe strategies, we had a final dataset of 728 self-care strategies. We then ranked the self-care strategies based on how frequently study participants rated the self-care strategy as being helpful and completing the review of all of the self-care strategies. Chi-square analysis was performed to test the association between the self-care strategies and each of the following dichotomized variables, which include race/ethnicity (white versus non-white), sex (male versus female), and status (Yes versus No) of each of the nine HANA conditions (Asthma, bronchitis, cardiovascular disease, COPD, diabetes, liver disease, osteoporosis, renal failure, and arthritis). Variables with p-value less than 0.05 were considered to be significantly associated with the frequencies of self-care strategies. If there were more than one single significant variables, then a second-step Chi-square analysis ensued: we first split the data into each category of variable with the lowest p-value from the firsts-step Chi-square, and then further Chi-square analysis was performed between the remaining significant variables and self-care strategies. We then ranked the strategies from the highest to lowest frequency stratified by significant variables identified from each step of Chi-square analysis. We repeated the procedures for each symptom. For example, for the self-care strategies of decreased sex drive, we first identified sex and race as variables significantly associated with this strategy and sex had the lowest p-value. We then split the data into male and female and performed a second-step Chi-square analysis between race and self-care strategies of this symptom by male and female separately. Race was still significant for female participants, but not significant for male participants. Thus, we ranked the self-care strategies of this symptom from the highest to lowest frequency by three separate groups: male, female white, and female non-white. Strategies were further categorized when relevant by sex, race and HANA condition.

Results

Branching logic was developed based on this analysis and provided the knowledge base for the delivery of the self-care strategies to users. Most symptoms had branching logic associated with sex, race and HANA condition. Speech difficulties and pain/ discomfort during sex were the only symptoms with no branching logic. All participants would receive the same ranking and listing of self-care strategies. All other symptoms were tailored by at least one of the three characteristics (sex, race, or HANA condition). Examples of the branching logic are illustrated in Table 1.

Development of the VIP-HANA system

Methods

Once the self-care strategies were finalized, we used Powtoons software to create animated videos (see sample screenshot of video in Figure 1) to illustrate each strategy which would later be integrated into the self-care strategy database. Ruby on Rails, an open source software framework that provides standardized structures for building was used to develop the web-app.

Results

The final VIP-HANA app included the following components: (1) a user profile which collects information at baseline on sex, race, and HANA condition of the user (Figure 2a), (2) reminder system to use the app weekly, (3) weekly symptom assessment (Figure 2c), (4) avatar selection (Figure 2d), (5) an overview report of the recommended strategies (Figure 2e), (6) symptom reports (figure 2f) and (7) a back-end database that provides tailored strategies based on the user profile and their symptom assessment. Videos were hosted on Vimeo software and participants received a video illustrating the self-care strategy every time they used the app and reported a symptom.

Table 1. Sample Self-Care Strategies for Anxiety, Sleep and Sex Drive Symptoms Tailored by HANA Condition, Sex, and Race

Top 3 Ranked Self-Care Strategies for Symptoms of Anxiety tailored by HANA conditions		
HANA Condition	Ranking	Self-Care strategy
Arthritis	1	Try breathing exercises to help with anxiety. An example would be taking a deep inhaled breath into your belly.
	2	Try relaxation or stress-reducing activities, such as meditation, personal "quiet time," or prayer.
	3	Try spending some time alone in a relaxed and calm environment.
No Arthritis	1	Try relaxation or stress-reducing activities, such as meditation, personal "quiet time," or prayer.
	2	Try to prioritize more sleep if possible. A good night's sleep is important.
	3	Try to live an active lifestyle and stay busy. Physical activity has been shown to reduce anxiety.
Top 3 Ranked Self-Care Strategies for Symptoms of Difficulty Staying Asleep tailored by HANA conditions		
HANA Condition	Ranking	Self-Care strategy
Renal Failure	1	Try making time for activities that you personally enjoy.
	2	Don't give up and keep trying different strategies to help you stay asleep.
	3	Some people who have dealt with this have found that taking time to wait while staying calm.
No Renal Failure	1	Try a sleeping aid if you are having trouble falling or staying asleep.
	2	Try to practice good sleep hygiene and relaxation techniques to improve your quality of sleep.
	3	Try making time for activities that you personally enjoy.
Top 3 Ranked Self- Care Strategies for Decreased Sex Drive Tailored by Sex and Race		
Sex/ Race	Ranking	Self-Care strategy
Female/ White	1	Try adding more physical activity to your life. Physical activity can increase blood flow and can help increase your sex drive.
	2	Have an open and honest conversation with your partner about your symptoms and talk about strategies you can use together.
	3	Don't give up and keep trying different strategies that may help you increase your sex drive.
Female/ non-white	1	Some people who have dealt with this have found that taking time to wait or sleeping can help.
	2	Try having sex when you are in different moods. Some people find sex more enjoyable after they've argued, while others find this more distracting.
	3	Try adding more physical activity to your life. Physical activity can increase blood flow.
Male	1	Try using porn to increase your sex drive and increase your sexual desire.
	2	Some people who have dealt with this have found that taking time to wait or sleeping can help.
	3	Try masturbating more often, which can boost your sex drive.

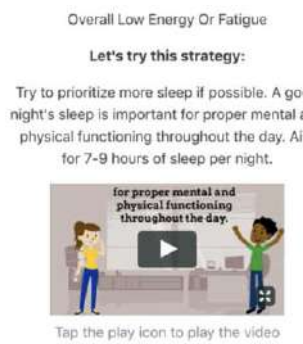


Figure 1. Self-Care Strategy Animated Video

Usability Evaluation of the VIP-HANA system

Methods

The final component of our development process was a laboratory usability evaluation with experts in human computer interaction and PLWH with at least 1 HANA condition. Full details of the usability evaluation are published elsewhere [12]. In brief, participants were provided with a Beta version of the app and asked to use the app to complete a use-case scenario created by the research team. Participants were asked to think aloud as they used the app and Morae screen capture software was used to record the session.

Results

Informatics experts (N=5) and end-users (N=20) rated the VIP-HANA app as a highly usable system. Yet they did make useful recommendations which we incorporated into the system as part of the app development and prior to our trial. End-users had difficulty completing advanced app functions, such as changing their avatar, emailing symptom reports, setting weekly reminders and enlarging videos. Experts and end-users agreed that the app needed to provide more instructions. Experts suggested that this be accomplished through a help button and end-users recommended including in-app instructions with straightforward and simple language. End-users also wanted to see features in big, bright, and bold text with a clear font.

VIP-HANA Trial

Following the implementation of the recommended changes from the usability testing, the software development company updated the app in preparation for a systematic evaluation. Currently, we are conducting a randomized trial with 100 PLWH with HANA conditions living in New York City. The goal of the trial will be to examine the effect of the VIP-HANA app on symptom burden, quality of life and frailty over a 6-month period (NCT03182738).

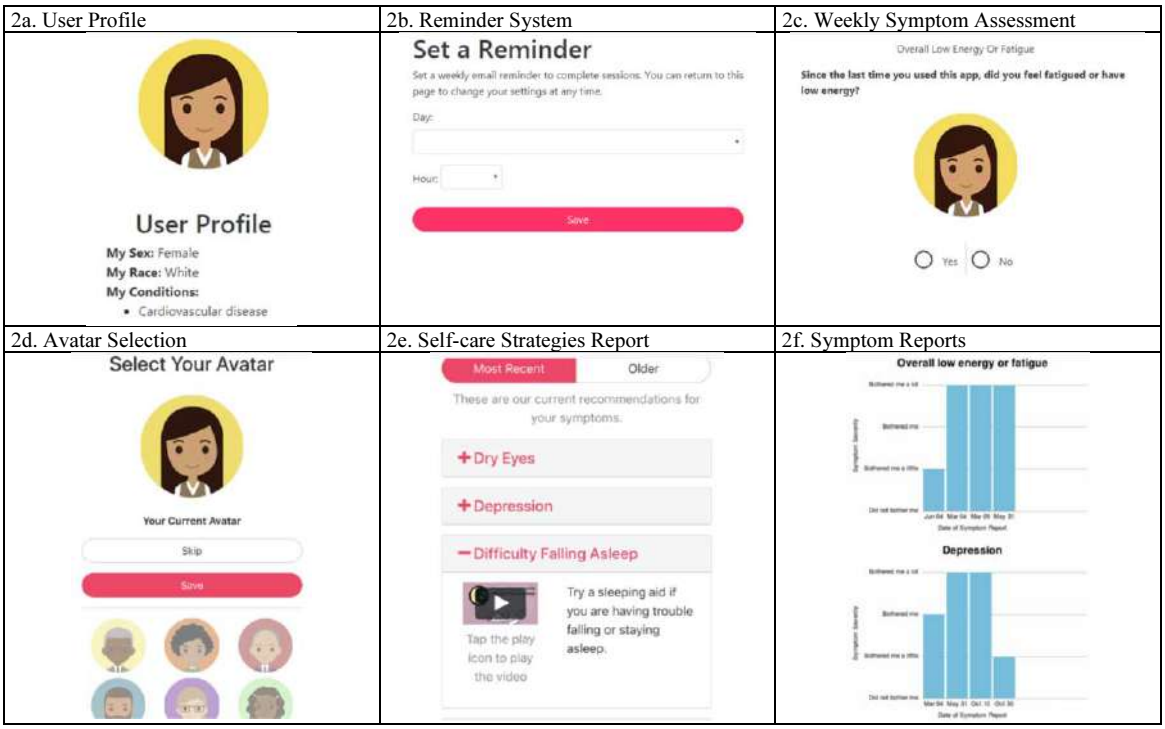


Figure 2. Screenshots of App Component

Discussion

This study demonstrated the ability to successfully translate our extant eHealth intervention for a growing clinical problem, HANA conditions. We incorporated quantitative and qualitative methods to better understand the needs of our target population and then successfully integrated these findings into our VIP eHealth platform, conducted usability testing and then conducted an RCT. Notably, our study population of PLWH with HANA conditions have unmet needs and we had the opportunity to advance our VIP intervention for a new set of users. With the proliferation of eHealth interventions, the aging of the US population and the reduction of healthcare services, there will likely be an increasing need for this translating and adapting eHealth interventions. Our study population is unique in that they are comprised of PLWH who are predominately low-income racial ethnic minorities, suffer from a chronic illness and also an infectious disease. Given the high needs of our study population in terms of social, physical and psychological support and their low socioeconomic status, development of informatics tools that can be used in these persons' everyday settings have the potential to dramatically improve clinical outcomes and quality of life, while also providing savings to our over-burdened healthcare system.

Building on our past approaches to developing mobile technology using end-user feedback and an iterative approach [13-16], this paper presents the integration of innovative methods for the creation of consumer health informatics tools. Notably, our approach included a large national US survey, clinican and expert review of the content, translation of survey data into a database which is the knowledge system of a patient-facing tool and a rigorous usability evaluation. Moreover, our study methods paid close attention to the needs of our study population by developing videos to maintain consumer engagement in the app.

Further, development of consumer health informatics tools can be complicated because of the need to integrate patient preferences, user-centered design principles, the limitations of software tools and the safety and potential efficacy from clinician experts. Given the complex integration of sources, there is a strong need for the development and dissemination of iterative development processes which integrate multiple stakeholders and result in usable and useful health information technology [17]. Finally, this paper illustrates the development of a practical and useful tool with immediate application that enables improvement in population-level quality of life. There is promising potential for integration of these tools to wider popular platforms.

Conclusions

This paper illustrates the development of a consumer health informatics app for ameliorating symptoms in PLWH with HANA conditions. We successfully completed a multi-modal development process incorporating the critical elements of user-centered design to build our app. Findings from the usability evaluation demonstrate a potentially useful and easy to use app. Integration of different quantitative and qualitative methods from different disciplines to accomplish the task of adaptation and evaluation of a mobile app is an appealing, rigorous and useful approach.

Acknowledgements

Research reported in this publication was supported by National Institute of Nursing Research of the National Institutes of Health under award number R01NR015737. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

All authors had full access to all the data and had final responsibility for the decision to submit for publication.

References

[1] S.G. Deeks, S.R. Lewin, and D.V. Havlir, The end of AIDS: HIV infection as a chronic disease, *Lancet* **382** (2013), 1525-1533.

[2] Centers for Disease Control and Prevention, *HIV Among People Aged 50 and Older 2018*, available from: https://www.cdc.gov/hiv/group/age/olderamericans/index.html

[3] A.C. Justice and R.S. Braithwaite, Lessons learned from the first wave of aging with HIV, *AIDS* (London, England) **26 Suppl 1** (2012), S11-S18.

[4] J. Neuhaus, B. Angus, J.D. Kowalska, A. La Rosa, J. Sampson, D. Wentworth D, et al., Risk of all-cause mortality associated with nonfatal AIDS and serious non-AIDS events among adults infected with HIV, *AIDS* (London, England) **24** (2010), 697-706.

[5] K.P. High, M. Brennan-Ing, D.B. Clifford, M.H. Cohen, J. Currier, S.G. Deeks, et al., HIV and aging: state of knowledge and areas of critical need for research. A report to the NIH Office of AIDS Research by the HIV and Aging Working Group, *J Acquir Immune Defic Syndr* (1999) **60 Suppl 1** (2012), S1-S18.

[6] P.K. Nicholas, K.M. Kirksey, I.B. Corless, and J. Kemppainen, Lipodystrophy and quality of life in HIV: symptom management issues, *Appl NursRes* **18** (2005), 55-58.

[7] M.J. Dodd, C. Miaskowski, and S.M. Paul, Symptom clusters and their effect on the functional status of patients with cancer, *Oncol Nurs Forum* **28** (2001), 465-470.

[8] R. Schnall, H. Cho, A. Mangone, A. Pichon, and H. Jia, Mobile health technology for improving symptom management in low income persons living with HIV, *AIDS Behav* **22** (2018), 3373-3383.

[9] D. J. Wantland, W.L. Holzemer, S. Moezzi, S.S. Willard, J. Arudo, K.M. Kirksey, et al., A randomized controlled trial testing the efficacy of an HIV/AIDS symptom management manual. *J Pain Symptom Manage* **36** (2008), 235-246.

[10] R. Schnall, D. Wantland, O. Velez, K. Cato, and H. Jia, Feasibility testing of a web-based symptom self-management system for persons living with HIV, *J Assoc Nurses AIDS Care* **25** (2014, 364-371.

[11] S. Iribarren, K. Siegel, S. Hirshfield, S. Olender, J. Voss, J. Krongold, et al., Self-management strategies for coping with adverse symptoms in persons living with HIV with HIV associated non-AIDS conditions. *AIDS Behav* **22** (2018), 297-307.

[12] S. Stonbraker, H. Cho, G. Hermosi, A. Pichon, and R. Schnall, Usability testing of a mHealth app to support self-management of HIV-associated non-AIDS related symptoms, *Stud Health Technol Inform* ***250*** (2018), 106-110.

[13] R. Schnall, M. Rojas, S. Bakken, W. Brown III, A. Carballo-Dieguez, M. Carry, et al., A user-centered model for designing consumer mobile health (mHealth) applications (apps), *J Biomed* **Inform 60** (2016), 243-251.

[14] R. Schnall, M. Rojas, J. Travers, W. Brown III, and S. Bakken, Use of design science for informing the development of a mobile app for persons living with HIV, *AMIA Annu Symp Proc* **2014** (2014), 1037-1045.

[15] R. Schnall, S. Bakken, M. Rojas, J. Travers, and A. Carballo-Dieguez, mHealth technology as a persuasive tool for treatment, care and management of persons living with HIV, *AIDS Behav*, **19** (2015), 81-89.

[16] H. Cho, P-Y. Yen, D. Dowding, J.A. Merrill, and R. Schnall, A multi-level usability evaluation of mobile health applications: a case study. *J Biomed Inform,* **86** (2018), 79-89.

[17] P-Y. Yen and S. Bakken, Review of health information technology usability study methodologies, *J Am Med Inform Assoc,* **19** (2012), 413-422.

Address for correspondence

Rebecca Schnall PhD, MPH, RN-BC, FAAN
Columbia University School of Nursing
560 West 168th Street
NY, NY 10032
Phone: 1-212-342-6886
Email: rb897@columbia.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190447

Fall Risk Assessment Through a Self-Service Terminal in the Outpatient Setting

Mariana Simón[a], Liliana Giraldo[a], Janine Sommer[a], Giuliana Colussi[a], Julián Perrino[a], Gustavo Staccia[a], Bibiana Schachner[a], Fernando Plazzotta[a], Daniel Luna[a]

[a]*Department of Health Informatics, Hospital Italiano de Buenos Aires, Argentina*

Abstract

Health organizations aim to identify patients with high fall risk in the different attention scenarios in order to provide safety and quality care. In order to address this issue, we designed an assessment tool that surveys fall risk through three questions in self-service terminals in the outpatient setting. Our objectives in this article are to describe the implementation process. A cross sectional design was used for the pilot (between October and November 2018). The issued tickets rate with high fall risk (defined as ⅔ positive responses) was 34.3%. Some adjustments needed to be done because some patients did not have true risk criteria due to self-report. We conclude that this tool will allow quicker identification of patients with true high risk. Effective prevention strategies will be necessary to improve safety after risk fall identification.

Keywords:

Information systems; Risk assessment; Self-management,

Introduction

Health professionals and decision makers in health organizations are aware of the need to provide quality and safe care. Patient safety is usually an institutional policy and is considered part of healthcare quality. Unwanted side effects in healthcare represent a considerable cause of morbidity and mortality in health systems.

The American Geriatrics Society and the British Geriatrics Society have published a clinical practice guideline on detection, evaluation and management of fall risk where it is recommended to annually evaluate all adults over 65 years of fall risk. This test involves asking patients if they have fallen two or more times in the past year, received medical attention for a fall, or experienced unsteadiness or loss of balance when standing or walking. It has been shown that patients who respond positively to any of these questions have an increased risk of falls and should receive an additional evaluation [1].

Patient assessment and identification of characteristics that can increase the probability of falls are fundamental for planning effective prevention strategies [2]. Using a specific tool to identify patients with greater susceptibility to fall can assist in preventing injuries [3]. There are investigations related to falls in different areas [4,5], which have provided insights into fall incidence related to the ambulatory care setting and evaluation of fall risk through validated instruments [6].

Nurses in primary care use the Morse scale for assessing the risk of falling but may not be transferrable to an outpatient setting. For that, evidence-based tools that define the risk of falls in non-hospitalized patients are promising [7].

If the fall rates are not reduced in an immediate future, it is projected that the number of injuries caused by falls will double by 2030 [8]. In our institution fall risk prevention for non-hospitalized patients is part of the 6th objective of the Joint Commission International (JCI), warning that fall risk methods should be appropriate for patient, taking into account risk differences (diversity of populations: pregnant, post-surgical or elderly) [9].To date, we have found no reports related to falls in outpatient settings, or assessments of preventive fall interventions.

The assessment of the Fall Risk Scale is normally performed at the time of the medical consultation or nursing intervention. However, there are no early identification of the risk until admission to the institution. This does not ensure patients receive preventive intervention and are informed about safe practices for patients at fall risk. Thus, the objectives of our paper is to describe the design of an assessment tool for fall risk through self-service terminals in the outpatient setting and to report the results of use in the pilot test.

Methods

Setting

Hospital Italiano de Buenos Aires (HIBA) is a non-profit organization with 165 years of history in Argentina. Its healthcare network includes a university hospital of high complexity that covers health care for outpatient, inpatients, emergencies, critical care, home care, chronic care and medical and surgical specialties. It has its own medical insurance service (health maintenance organization), with more than 160,000 affiliates, and provides health services to 1,500,000 people with other health insurances. Annually, more than 45,000 patients are admitted to HIBA hospitals, and 45,000 surgical procedures and 3,000,000 outpatient visits occur. Since 1998, the HIBA has its own health information system (in-house) that includes the management of clinical and administrative information. Its electronic health record (EHR) is an integrated, modular, problem-oriented and patient-centered system, used in the different clinical scenarios (ambulatory, hospitalization, emergency center and home care) [10,11].

Since 2009, self-service terminals have been implemented into the HIBA, and are currently distributed in all ambulatory care areas. It is a standing self-service touch screen cabinet (Fig 1), with multiple features, developed and manufactured in Argentina. The main features of this tool are: appointment scheduling, ticket printing, as well as data recording. The main function for the outpatient is self-reception when attending a

scheduled appointment. It has a modern and ergonomic design, with great versatility and robustness. The self-service terminals are recommended for high-concurrency spaces, to support smooth to flow of outpatients in our hospital.

Figure 1: Self-service Terminal

Design

In order to assess fall risk fall and identify of patients at risk in the ambulatory area, we decided to incorporate a new functionality in the pre-existing self-service terminal in the ambulatory area at the Hospital. We used a cross-sectional analysis to analyze and report the assessment tool use during the pilot test performed during the period between October 18 and November 7, 2018 at Hospital Italiano de Buenos Aires.

Relative and absolute frequencies included were:

- All patients with potential risk of falls (defined as those requiring priority attention by self-report: pregnant, reduced mobility and / or disability);
- Rate of tickets issued (defined as those patients with high risk: at least 2 out of 3 affirmative responses);
- Number of patients attended and intervened by nursing.

Phase 1: Situation Diagnosis

The 6th version of the JCI Manual recommends having a tool to identify risk of fall in the ambulatory area from the moment patients enter the institution, which allows them to be correctly identified and to establish a secure circuit. The scope is to detect patients at risk of falling, as well as identify risk factors. The educational intervention carried out by the nursing staff must be recorded in the EHR.

Phase 2: Development of the Prototype

This stage is based on the already implemented features design of the self-service terminals (see Figure 2), which includes the incorporation of the fall risk assessment.

Figure 2: Self-service Terminal - Initial Screen Translation: Welcome to the self-service terminal. Which transaction do you wish to perform? Confirm I'm Present (I already have an appointment). Book an appointment. Register (It's my first time at Hospital Italiano). Other tasks.

In the prototype, three questions of assessment are aligned to the left and the possible answers (YES / NO) on the right side of the screen (See Figure 3).

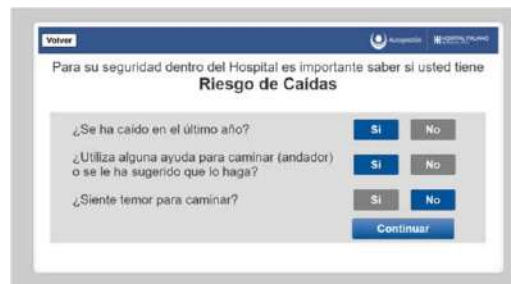


Figure 3: Self-service Terminal - Risk Fall Assessment Screen Translation: For your safety at Hospital Italiano it is important to know your falling risk. Have you fallen the last year? (Yes/No). Do you use any walking assistance, or have you been suggested to do so? (Yes/No). Are you afraid when walking? (Yes/No).

Listed below the questions, is a ' CONTINUE' option that the patient should select and is disabled until the three questions mentioned above are answered. In order to ensure completeness of the scale, all three data fields are mandatory.

For those patients with high risk (defined as two of three affirmative answers), a ticket is printed with the letter C as shown in Figure 4.

Figure 4: Emisión de ticket con Riesgo de caída. Translation: Please take your printed ticket and wait to be attended, thank you.

For the development of the software, we decided to build a web application. Java™ language was used to program all the functionalities of the back-end, with HTML and JavaScript for the front-end. The final set of the application was deployed on web servers and accessed using a Chrome™ browser installed in each self-service terminal, under Windows 7™ or Windows 10™, configured to be used from a touch screen. Eventually the application will also be able to run under Linux™ (Red Hat™, Ubuntu™, Debian™, etc.).

Phase 3: Implementation

The pilot test of the printing ticket feature and process was implemented during the month of October. The care circuit for patients with potential risk fall was established. The administrative staff and nurses were involved in the process. The process begins with the printing of a ticket identify with the letter C (possible patient with risk fall detected), which is generated by answering affirmatively two of the questions, in the self-service terminal. Once the ticket has been issued, the patient waits for his call by the administrative staff, then the patient is identified with a green bracelet in the wrist and is oriented towards a safe place. The nursing staff is informed to locate the patient in a safe place. In addition, the nurse performs Morse scale assessment of falls risk, provides educational materia, and records the intervention in the EHR.

Results

Based on the prototype, the following modifications were made: the size of the letter was increased, and the color was improved and adapted to improve user experience of older adults. The necessary adjustments were made until arriving at the final version, which is currently being used in a pilot test (See Figure 5).

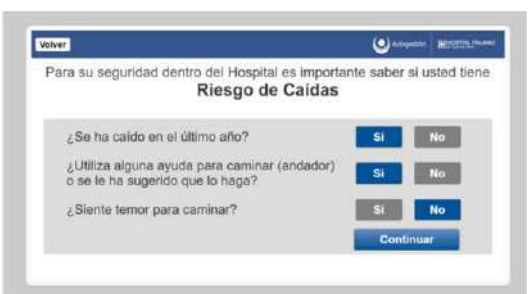


Figure 5: Self-service Terminal - Risk Fall Assessment Screen Adapted from the Original Ddesign.
Translation: For your safety at Hospital Italiano it is important to know your falling risk. Have you fallen the last year? (Yes/No). Do you use any walking assistance, or have you been suggested to do so? (Yes/No). Are you afraid when walking? (Yes/No).

During the study period, there were a total of 1784 patients with potential risk of falls who responded to the assessment, of which the rate of issued tickets was 34.3% (95% CI 32.10-36.56).Figure 6 shows the frequency distribution of patients that carried out the survey and the generation of ticket with fall risk according to the dates. The red line indicates completed surveys that did not generate a ticket, and the blue line indicates the surveys completed that resulted in generation of a ticket.

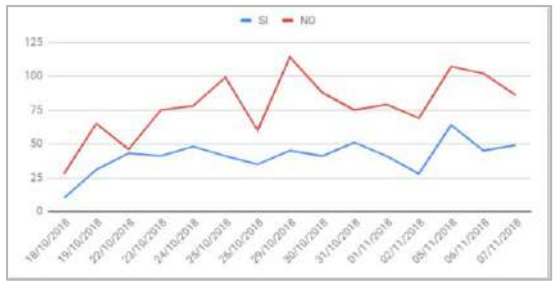


Figure 6: Frequency Distribution of Patients who Completed the Survey that Resulted in Generation of a Fall Risk Ticket

The number of patients who received nursing care each day of the pilot is shown in Figure 7.

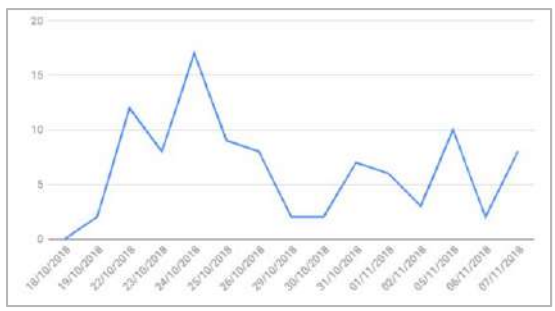


Figure 7: Total Patients Attended by Nursing

An important finding in the validation of patient conditions with potential falling risk was that many of them did not meet a true risk criterion. Table 1 shows the details of this findings.

Table 1: Patients without True Criteria for Falling Risk

Other reasons for consultation (not related to programmed ambulatory appointments)	• Registration application • Request for laboratory printing • Appointment scheduling • Others arrangements
Technical failures in communication	• Patient who wanted priority attention

Discussion

The design of this tool aimed at assessing and detect patients who were at high fall risk within the outpatient setting, in order to plan effective prevention strategies. This process could improve the waiting conditions, adapt specific needs of high fall risk populations, improve patient safety, and potentially reduce the risk of falls within the institution. The rate of tickets issued was 34.3%, higher than expected, due to the logistics and necessary resources required for the implementation of effective prevention strategies, and the coordination of all the actors involved. However, this rate could be overestimated, given that patients reported untrue data, and thus further validation is required. We believe that the step of validation is important, because many patients reported having high risk to avoid the system waiting times, in order to be treated with priority. The process of self-report generates the need to validate the true positive cases of

patients with a high risk of falling, and in some cases the answer of the generated ticket.

The full patient population may also not have been represented, as there were a small number of terminals in which the survey was implemented (not enough for the number of patients), connectivity issues (which sometimes prevented the use of the terminal), and difficulties with the touch screen.

This is our first experience of implementing a technological tool to improve the quality and safety of patients in the outpatient setting. Although, it is a novel tool with great potential and in consistency with JCI priority lines, diffusion, institutional communication and training are needed. For this, it requires more communication and teamwork (security, administrative, nursing, among others).

Furthermore, standard strategies for continuity of care are needed, focused on addressing the diversity of patients with potential risk fall, integrating all the necessary staff of the process with a single objective. At the same time, it highlights the importance of indicators.

As future pending lines of work, we propose: re-design to overcome some of the aforementioned technical difficulties, improving usability and adapting new functionalities to the user needs (incorporation of patient identification through QR or through Personal Health Record).

Conclusions

The design of this tool will allow the rapid identification of patient with true positive high risk of fall in the outpatient setting, and effective prevention strategies will be necessary to guarantee a safe course of action for patients. This could support improving waiting conditions, adapting to the specific needs of the population at high risk, and potentially decreasing the risk of falls within the institution.

Acknowledgements

This study was carried out under the Residency Program in Health Informatics and the Area of Information Technology Oriented to the Community of the HIBA. We want to thank the entire interdisciplinary team (doctors, nurses, designers, developers) for the work you have done.

References

[1] E.A. Phelan, J.E. Mahoney, J.C. Voit, and J.A. Stevens, Assessment and management of fall risk in primary care settings, *Med Clin North Am* **99** (2015), 281–293.

[2] M. de F. Luzia, M.A. de G. Victor, and A. de F. Lucena, Nursing diagnosis risk for falls: Prevalence and clinical profile of hospitalized patients, *Rev Lat Am Enfermagem* **22** (2014), 262–268.

[3] B. O'Rourke, R. Upchurch, J. Bain, L. Tirey, S. Vrchota, and K. Watson, Identification of Fall Risk Factors Through Self-report and Physical Performance, *Topics in Geriatric Rehabilitation*. **34** (2018), 194–199. doi:10.1097/tgr.0000000000000194.

[4] G.P. Forrest, E. Chen, S. Huss, and A. Giesler, A comparison of the Functional Independence Measure and Morse Fall Scale as tools to assess risk of fall on an inpatient rehabilitation, *Rehabil Nurs* **38** (2013), 186–192.

[5] G.F. Del Duca, D.L. Antes, and P.C. Hallal, Quedas e fraturas entre residentes de instituições de longa permanência para idosos, *Rev Bras Epidemio.* **16** (2013), 68–76.

[6] L.Z. Rubenstein, R. Vivrette, J.O. Harker, J.A. Stevens, and B. Josea Kramer, Validating an evidence-based, self-rated fall risk questionnaire (FRQ) for older adults, *Journal of Safety Research*.**42** (2011), 493–499. doi:10.1016/j.jsr.2011.08.006.

[7] [No title], (n.d.). http://nursing.uthscsa.edu/onrs/starmodel/institute/su09/documents/meador.pdf (accessed October 22, 2018).

[8] C. Ni Scanaill, C.N. Scanaill, C. Garattini, B.R. Greene, and M.J. McGrath, Technology innovation enabling falls risk assessment in a community setting, *Ageing Int.* **36** (2010), 217–231.

[9] Website, (n.d.). https://www.jointcommissioninternational.org/6th-edition-in-depth-preventing-falls-in-inpatient-and-outpatient-settings/ (accessed October 22, 2018).

[10] F. Plazzotta, D. Luna, and F. González Bernaldo de Quirós, Health information systems: integrating clinical data in different scenarios and users, *Rev Peru Med Exp Salud Publica*. **32** (2015), 343–351.

[11] D. Luna, F. Plazzotta, C. Otero, F. González Bernaldo de Quirós, A. Baum, and S. Benítez, Incorporación de tecnologías de la información y de las comunicaciones en el Hospital Italiano de Buenos Aires, (2012). http://hdl.handle.net/11362/3959 (accessed November 23, 2018).

Address for correspondence

Mariana Simón. Department of Health Informatics. Hospital Italiano de Buenos Aires, Argentina.

E-mail: mariana.simon@hospitalitaliano.org.ar

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190448

'Hybrid Doctors' Can Fast Track the Evolution of a Sustainable e-Health Ecosystem in Low Resource Contexts: The Sri Lankan Experience

Siribaddana, Pandula[a], Hewapathirana, Roshan[b], Sahay, Sundeep[c], Jayatilleke, Achala[a], Vajira H. W. Dissanayake[b]

[a] *Postgraduate Institute of Medicine, University of Colombo, Sri Lanka*
[b] *Department of Anatomy, Faculty of Medicine, University of Colombo, Sri Lanka*
[c] *Department of Informatics, Faculty of Mathematic and Natural Sciences, University of Oslo, Norway*

Abstract

Although e-health is an area recognized as essential in the rapid development of healthcare systems in low resource contexts, many challenges prevent the emergence of an effective e-health ecosystem. Lack in capacity around health informatics is one of the main challenges. Based on a longitudinal case study gathering data pertaining to a master's program in biomedical informatics in Sri Lanka designed for doctors, in this paper we demonstrate that creating 'hybrid doctors' may be the way forward. We illustrate how hybrid doctors conversant in healthcare and information and communication technology (ICT) are able to facilitate the creation of an e-health ecosystem in a way that it would contribute significantly to the ICT driven healthcare reforms. Through this case study we highlight the importance of multidisciplinarity, participatory design, strategic investments, learning that aligns with developmental needs, networking, gaining legitimacy and re-packaging perspectives on 'health informatics capacity development'.

Keywords:

Health Information Systems; Education, Medical; Medical Informatics,

Introduction

Capacity development around health informatics remains a considerable challenge in low resource settings [1]. In the absence of adequate capacity, many countries lag in incorporating e-health solutions to elevate their healthcare systems and gain better health outcomes [2]. With the recognition that technology could be a main driver in achieving sustainable developmental goals in 2030 [3], identifying ways and means of incorporating technology in healthcare has gained considerable interest.

Developing health informatics capacity among doctors or healthcare professionals has been a focus of attention for a considerable period [4]. The International Medical Informatics Association's (IMIA) initiatives in developing programs to improve health informatics capacity among healthcare professionals have been well received. However, in low resource contexts, capacity development in health informatics remains in the hands of non-governmental organizations or local universities supported by donor agencies [2,5]. Although such programs have contributed to the development of local capacity in health informatics to various degree, there is limited evidence on the impact of these programs in the wider e-health ecosystem of a country.

In 2008, the Postgraduate Institute of Medicine (PGIM), University of Colombo, Sri Lanka in collaboration with the Department of Informatics of the University of Oslo, Norway and the Health Informatics Society of Sri Lanka embarked on an ambitious program in the creation of 'hybrid doctors'. Funded by the NORAD program for Master Studies (NOMA), the program aimed to develop health informatics capacity among doctors in Sri Lanka by developing, implementing and scaling up frugal health information systems without having to depend on imported health informatics expertise. Currently at ten years after implementation, the Master's program in Biomedical Informatics in Sri Lanka at the PGIM has trained more than 150 medical doctors, and the e-health ecosystem in Sri Lanka has been disrupted towards one of the most dynamic in the region. The program has not only generated capacity, but it has also had a high impact on national level health information systems, health informatics policy, innovations, networking, publications and South-South knowledge translation, and expertise sharing. We have been unable to find similar examples of programs in other low/middle income countries that have paved the way towards a standalone specialty in health informatics for doctors.

In this paper, we summarize a decade of work around this capacity development initiative in health informatics in a systematic manner in view of identifying the key factors that contributed to the success of this program. We also discuss how the capacity development effort disrupted the e-health ecosystem in Sri Lanka, a low and middle income country, in a relatively short period of time. In doing so, we address one of the burning questions in health informatics, how to successfully innovate and sustain health information systems in low resource contexts.

Methods

Drawing from data gathered from 2008 to 2018, allowed us to observe the changes taking place within the e-health system that consisted of people, organizations and the environment. In line with Pettigrew's illustration on longitudinal field research [6] and case study research as described by Yin [7], we adopted a longitudinal case study method. The authors of this paper have been involved in this project as principle investigators, project managers, researcher cum educators and as trainees. This has allowed different authors to bring in different perspectives in the interpretation of study data. Using multiple methods, which included semi-structured interviews, focus groups, document analysis, online discussion forums and e-mail communications; we gathered qualitative data pertaining to different yet seemingly overlapping phases of program evolution. Based on our understanding of key developments that took place since the inception of the master's program, these phases were classified as 1) initiation 2) legitimation and institutionalization 3) impact generation and 4) scaling and sustainability phases –

Additional reports have been published on the evaluations of program outcomes as reported by students and the PGIM. Thematic analysis was conducted on the data, where all authors of this paper participated in collectively building themes emerging from the systematically coded data. This approach to data analysis enabled us to overcome the researcher bias that generally plagues information system research [8,9].

Results

In presenting the results, we will provide an overview of each of the phases in general, followed by presentation of the key themes that were developed.

Initiation

The initiation phase consisted of the first two years following signing the memorandum of understanding between the PGIM and the University of Oslo. This phase was highlighted by the creation of a multidisciplinary board comprising of academics and practitioners from various academic disciplines, the Ministry of Health (MoH), professional organizations such as the Health Informatics Society of Sri Lanka (HISSL) and the private sector. The Northern partner, the University of Oslo, provided technical support, but did hold a position on the board. The board had the power to decide how the program should be run in accordance with the university regulations in collaboration with the partners such as the MoH, other universities, professional organizations and private sector enterprises. The key tasks performed by the multidisciplinary board were to develop the curriculum, identifying the training strategy, create learning opportunities for the trainees, negotiate with the stakeholders on the implementation of the program and manage NORAD in the development and implementation of the master's program.

The initiation phase was characterized by a proactive effort towards facilitating multidisciplinarity and participatory design, largely through the creation of a multidisciplinary board. This phase was also characterized by the decisions made on strategic investments in training unit development, trainer training and network building, which contributed to the emergence of a conducive learning environment for teaching and learning. In particular, effective utilization of donor funds to student projects aimed at improving service delivery in various state health institutions . The health institutions in this case benefited in two ways. First, the student projects helped solve issues related to service delivery in a training unit. Second, the fulfilment in establishing a basic infrastructure for these projects facilitated implementation and sustainment of these projects. Given that students were involved in frugal innovations and the scope of the master's projects were limited, the cost incurred for each project was minimal. This enabled the board to support multiple projects across the health care system.

Legitimizing and institutionalizing

The legitimizing and institutionalizing phase of the program represented the first three to four years of the program where the first two batches of graduates became part of the MoH workforce as Medical Officers in Medical Informatics (MOMI). The creation of the post, a unique position for a medical officer within the health sector of a low resource context, was one of the key characteristics identified during this phase. The acceptance of the graduates of the program by the MoH as an integral part of its master-plan for e-health was another critical component during this phase. We identified that this recognition was gained through advocacy and demonstration of competency by the newly graduated doctors and champions of the program. In this phase, roles that were previously filled by experts from other fields (e.g. ICT, engineering) who limited healthcare background transitioned to being within the MoH and were replaced with the MOMIs. In addition, the data collected during this phase illustrated how placement of students cum doctors in various healthcare institutions and programs were solicited by the stakeholders for the master's program. We identified the integration between the academic program and practice as an important theme during this phase.

Other themes recognized during this phase included: South-South collaborations between the program and the regional partners paving way towards sharing of expertise and knowledge regionally. At the same time, recognition of the qualification by the MoH as a means of promotion for the doctors generated renewed interest among potential candidates for the program. This essentially meant that all doctors who qualified this program would be funded by the MoH and would receive full pay after two years. We found this to be a key contributor towards the sustainment of the program beyond the funding period from NOMA.

Impact Generation

The impact generation phase began in from the fourth year and continued to the present. At the beginning of this phase some of the students' projects started attracting the attention of the MoH as proposing potential systems that could be scaled island wide to facilitate national health information flow. This recognition was encouraged by placement of students in strategic locations where there was a need for health information system development or upgrading of existing systems. Additionally, adopting open source software tools made students' projects attractive to the stakeholders.

The phase was also highlighted by the involvement of the graduates of the program in the development and implementation of health information systems, e-health policy and guidelines, consultations with various stakeholders (including donor agencies) and scientific publications presented at various national and international forums. Table 1 highlights some of these impacts.

As highlighted in Table 1, this phase also demonstrated the entrepreneurship potential of the graduates working with various international and private sector organizations on various health informatics projects. The tendency of the private sector and other organizations to seek expertise of the graduates of the program was perceived by us as an indication of further legitimization of the program and development of competencies of the graduates, and facilitation of building an ecosystem across organizational boundaries.

The involvement of the graduates within various aspects of the e-health ecosystem was also facilitated by the HISSL, which became the main representative body of health informaticians of the country. Many of the opportunities gained by the students and graduates to network with the industry, foreign and local experts, as well as donor agencies were facilitated through HISSL.

Table 1 – Impacts of the Biomedical Informatics Master's program (2008 – 2018)

Areas of Impact	Indicators of high Impact
Practical HIS development	National/program wide HISs developed and implemented – more than 12 Institution focused HIS – more than 20
Research papers published	Journal papers – 37; Conference presentations – 88 (up to 2017)
Networking opportunities: International conferences attended	More than 7 regional and international conferences attended (e.g. APAMI, APMEC, e-Health Asia, AEHIN etc.) by students and faculty of the health informatics program.
Networking opportunities: International conferences hosted (organized by HISSL)	e-Health Sri Lanka 2010 and 2014. IFIP 9.4 conference 2014 eHealth Asia 2015 Commonwealth Digital Health Conference 2016, 2017, 2018 APAMI conference 2018 AeHIN conference 2018
Networking and advocacy	Networks established through educational program and supporting activities Linking with organizations such as Health Information Systems Program/DHIS2, OpenMRS, AeHIN, IMIA, Commonwealth Medical Association Active contribution within national bodies such as National eHealth Steering Committee, National Foundation for Open Source Health Software Drafting of National eHealth Policy, National eHealth Standards and Guidelines Pioneering work around Health Identification Number.
Program evolution and sustainability	Eight batches since 2008 Introduction of the MD program - two batches in training (38 students). Gradually increasing demand for the master's program since the introduction of the MD with board certification in Health Informatics Government investment of approximately 40 to 50 million LKR on trainee scholarships since 2008 (up to 2017) Investment set to increase by approximately 100 million LKR each year to support a one-year foreign placement for each MD trainee to gain foreign exposure.
Entrepreneurs hip contributed through capacity development	HISP Sri Lanka was established Multiple e-health companies were engaged with dealing with personal medical records, e-learning, social media, etc. Individual consultations carried out for international development partners such as UNICEF, WHO, USAID, Vital Strategies (VS) and the private sector.

Scaling and Sustainability

This phase was recognized with the renewed interest generated among the potential candidates for the capacity development program following the \approval of the MD program in health informatics. With the creation of this program, those who have successfully completed the master's program would be able to specialize in health informatics similar to any other specialty in medicine. The acceptance of the MD program as a specialist qualification by the MoH further highlighted the ongoing institutionalization process. The return of doctoral level qualified resource personnel in health informatics from their overseas training also demarcated this phase, as it facilitated a path for higher level capacity development directly in Sri Lanka. The creation of a pool of resource personnel from graduated doctors also characterized this phase as an important contributor towards the scaling and sustainability of the program, as it fulfilled the need for trainers with experience in systems development and implementation from the local context.

Discussion

The creation of hybrid doctors who are conversant in healthcare and ICT have enabled Sri Lankan e-health ecosystem to benefit in multiple ways. Firstly, the case study illustrates how government investment in training medical professionals as health informaticians enables them to drive innovations, which are frugal, scalable, sustainable and context sensitive. Secondly, the case study also illustrates the effective use of development funds aimed at capacity development and healthcare system strengthening. In this case, the funds enabled not only training of doctors, but also in creating the learning environments that facilitated innovation, collaboration and sharing of knowledge both within and outside the local context.

However, a capacity development program particularly in low- and middle-income countries cannot exist in isolation without being linked with the development efforts within the healthcare system –[5]. This was clear from the Sri Lankan experience as one of the key factors contributing to the success of the program was its alignment with the national health sector development initiatives. This was achieved through the multidisciplinary multi-stakeholder governing body, the board, and the early exposure of the trainees to problem solving in real life e-health projects. The stakeholders were equal partners in the training similar to the status of the academic institution hosting and running the program and the funding agency providing technical support. The participatory design approach [10, 11] adopted in developing the program ensured that local needs are embedded in the training and that student learning was context oriented, rather than a mere transfer of knowledge from North to South. Networking, therefore, became a key competency for the doctors undergoing training in health informatics, as the knowledge that they were expected to harness was not always transferable but was also tacit in nature [12].

Traditionally, health informatics capacity development in low resource contexts is dependent on donor funding [13,14] and is dominated by academic institutions [15]. These are oth uni-disciplinary and do not necessarily partner with the development efforts in a country. Traditional capacity development programs for health informatics also tend to target people who are competent in ICT [5], which may be partly due to the lack of interest among doctors in becoming health informaticians. From a socio-cultural point of view, a doctor is perceived in many contexts as a person who would treat a person wearing a stethoscope. Even the doctors in these contexts may not pursue their interests as they become confined to the acceptable societal perception of a doctor [16]. In the case

of Sri Lanka, the barriers towards creating hybrid-doctors were broken through multiple means: clear and progressive career paths, guarantee of employment, community of doctors who are also health informaticians, network opportunities with likeminded people both within and outside their own context, and legitimation gained by professionals within their work settings. The perceived change however was not achieved overnight, particularly when it came to change underlying perceptions and beliefs. The case study demonstrated the need to evolve with the changing socio-cultural-political environment for capacity development effort that can be sustained and impactful [17].

As mentioned earlier, one of the key aims of this program was to create an entity who could be the bridge between the domains of ICT and healthcare. In other words, the entity being trained needed to be recognized by the stakeholders of e-health – the MoH, the ICT industry, the development partners, doctors, other healthcare and non-healthcare professionals, academia, etc. Recognition of a doctor as an integral part of the healthcare sector [18] appears to have helped them perform their expected role in bridging the domains of healthcare and ICT. While their role of being doctors enabled them to decode the complexities existing among healthcare professionals in accepting and complying with technology implementations, their expertise in health informatics enabled them to explain the needs of the healthcare system to ICT professionals. During system implementation and scaling up, the doctors had the power to make decisions, negotiate with other professionals, gather and analyze data and intervene in problem resolution. More importantly, they garnered trust among stakeholders of health informatics paving the way for the bridging their dual roles.

While the hybrid-doctors had impact through active participation in the design, development, implementation and scaling up of health information systems, the e-health ecosystem appears to have evolved in several directions. The direct involvement of the doctors in e-health projects within the MoH triggered many more departments and campaigns to identify technology needs and engage in frugal HIS developments. Additionally, the private sector identified the potential of these doctors in serving as idea generators and consultants, which further encouraged private sector involvement in the e-health ecosystem of the country. The development agencies also recognized the potential of hybrid-doctors, which enhanced development support and collaborations. Beyond the local contexts, the e-health ecosystem started to link up with international networks through the connections made by hybrid-doctors, and enabled sharing of knowledge and expertise that extended across different countries. Such networking has long been advocated in the development of health information systems in low resource contexts [19]. The path of evolution however may not always be clear-cut and controllable [20], thus highlighting the need to continuously look at capacity development as part of the evolving e-health ecosystem, rather than as a narrowly focused educational effort. In other words, the creation of hybrid doctors and facilitating the cultivation of an e-health ecosystem reminded us of the Aristotelian quote, 'the whole is greater than the sum of its parts'.

Conclusions

In this paper we illustrated how a capacity development program in health informatics for doctors disrupted the e-health ecosystem in Sri Lanka and achieved high impacts. We attribute the program's success to the multidisciplinary and participatory design approach and integration with ongoing development efforts within the healthcare system . We also highlighted the importance of motivating the students by creating a conducive learning environment, ensuring career pathways and changing the traditional perspective about doctors. Hybrid-doctors was at the core of the emerging e-health ecosystem in Sri Lanka in performing the important role of bridging ICT and healthcare domains.

The nature of the training, which was context sensitive and action oriented, promoted the idea of frugal innovation by linking academia and practice at a very early stage. Not only did the students have to solve real life problems, but they also had the opportunity to contextualize their learning. The learning in this case was not an isolated academic exercise, but a graduated process of integrating health informatics expertise within the healthcare system.

Acknowledgements

We would like to thank the support extended by the NORAD program for Master Studies, which funded the initial phase of this program. We would also like to acknowledge the support extended by the Postgraduate Institute of Medicine of the University of Colombo, the Department of Informatics, Faculty of Mathematics and Natural Sciences of the University of Oslo and the Health Informatics Society of Sri Lanka.

References

[1] J.H. Ledikwe, L. L. Reason, S. M. Burnett, L. Busang, S. Bodika, R, Lebelonyane, S. Ludick, E. Matshediso, S. Mawandia, M. Mmelesi, B. Sento, and B.W. Semo, Establishing a health information workforce: innovation for low-and middle-income countries. *Human resources for health* **11**(1) (2013), 35.

[2] H.C. Kimaro and J.L. Nhampossa, Analyzing the problem of unsustainable health information systems in less-developed economies: case studies from Tanzania and Mozambique. *Information Technology for Development* **11**(3) (2005), 273-298.

[3] United Nations,. Capacity development key to realizing the global goals, (2017). https://www.un.org/development/desa/en/news/capacity/capacity-development-key-for-sdgs.html (accessed November 3, 2018).

[4] J. Mantas, E. Ammenwerth, G. Demiris, A. Hasman, R. Haux, W. Hersh, E. Hovenga, K.C. Lun, H. Marin, F. Martin-Sanchez, and G. Wright, Recommendations of the International Medical Informatics Association (IMIA) on education in biomedical and health informatics. *Methods of Information in Medicine* **49**(02) (2010), 105-120.

[5] S. Sahay, T. Sundararaman, and J. Braa, *Public health informatics: Designing for change-a developing country perspective*. Oxford University Press, London, 2017.

[6] A.M. Pettigrew, Longitudinal field research on change: Theory and practice. *Organization science* **1**(3) (1990), 267-292.

[7] R.K. Yin, Case study research: Design and methods (4th ed.). SAGE Publications, London and Singapore, 2009.

[8] M.D. Myers, and D. Avison, eds, *Qualitative research in information systems: a reader*. SAGE Publications, London, 2002.

[9] S. Sarker, X. Xiao, and T. Beaulieu, Qualitative studies in information systems: A critical review and some guiding principles. *MIS quarterly* **37**(4) (2013), 3-18.

[10] S.K. Puri, E. Byrne, J.L. Nhampossa, and Z. B. Quraishi, Contextuality of participation in IS design: a developing country perspective. In *Proceedings of the eighth*

conference on Participatory design: Artful integration: interweaving media, materials and practices **1 (**2004), 42-52.
[11] E. Byrne, and S. Sahay, Participatory design for social development: A South African case study on community-based health information systems. *Information technology for development* **13**(1) (2007), 71-94.
[12] M. Alavi, and D. E. Leidner, Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS quarterly* **25**(1)(2001), 107-136.
[13] H.C. Kimaro, Strategies for developing human resource capacity to support sustainability of ICT based health information systems: a case study from Tanzania. *The Electronic Journal of Information Systems in Developing Countries* **26**(1)(2006), 1-23.
[14] N. Banks, D. Hulme, and M. Edwards. NGOs, states, and donors revisited: Still too close for comfort?. *World Development* **66**(2015),707-718.
[15] A. Sawyerr,. African universities and the challenge of research capacity development. *Journal of Higher Education in Africa/Revue de l'enseignement supérieur en Afrique* (2004), 213-242.
[16] Fenwick, Tara. Sociomateriality in medical practice and learning: attuning to what matters. *Medical Education* **48**(1) (2014), 44-52.
[17] J. Frenk, L. Chen, Z.A. Bhutta, J. Cohen, N. Crisp, T. Evans, H. Fineberg, P.Garcia, Y. Ke, P.Kelley, B.Kistnasamy, A. Meleis, D. Naylor, A.Pablos-Mendez, S.Reddy, S. Scrimshaw,J. Sepulveda, D.Serwadda, and H. Zurayk, Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *The Lancet* **376**(9756) (2010), 1923-1958.
[18] R.A. Manners, *Professional dominance: The social structure of medical care*. Routledge, 2017.
[19] J. Braa, E. Monteiro, and S. Sahay, Networks of action: sustainable health information systems across developing countries. *MIS quarterly* (2004), 337-362.
[20] T.A. Sanner, T.D. Manda, and P. Nielsen, Grafting: balancing control and cultivation in information infrastructure innovation. *Journal of the Association for Information Systems* **15**(4) (2014), 220.

Address for correspondence

Dr Pandula Siribaddana
Postgraduate Institute of Medicine
University of Colombo
160, Prof. Nandadasa Kodagoda Mawatha,
Colombo 08.
Sri Lanka.
e-mail: Pandula@pgim.cmb.ac.lk

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190449

Perceptions and Preferences About Granular Data Sharing and Privacy of Behavioral Health Patients

Hiral Soni[a], Adela Grando[a], Marcela P. Aliste[a], Anita Murcko[a], Michael Todd[b], Madhumita Mukundan[a], Michael Saks[c], Caroline Horrow[d], Richard Sharp[d], Christy Dye[e], Darwyn Chern[e], Mary Jo Whitfield[f], Mark Callesen[f]

[a] *Department of Biomedical Informatics, Arizona State University, Scottsdale, US*
[b] *College of Nursing and Health Innovation, Arizona State University, Phoenix, US*
[c] *Sandra Day O'Connor College of Law, Arizona State University, Phoenix, US*
[d] *Biomedical Ethics Research Program, Mayo Clinic, Rochester, US*
[e] *Partners in Recovery, Phoenix, US*
[f] *Jewish Family and Children's Services, Phoenix, US*

Abstract

Little is known about data sharing preferences for care and research of behavioral health patients. Eighty-six behavioral health patients (n =37 Latinos; n = 32 with serious mental illness) completed questionnaires, in either English or Spanish, with items assessing their views on privacy and sensitivity of health record information. Most patients (82.5%) considered mental health information as sensitive. In general, there was a direct correspondence between perceived sensitivity of information and willingness to share with all or some providers. A main motivation for sharing data with providers was improving the patient's own care (77.8%). Most participants (96.5%) indicated they would be extremely to somewhat willing to share their data for research with their care facilities and universities. Follow-up patient interviews are being conducted to further elucidate these findings.

Keywords:
Privacy, Electronic Health Records, Surveys

Introduction

Behavioral health problems include substance use disorders, serious psychological distress, suicide, and mental disorders [1]. In the U.S., behavioral health conditions affect over 44 million (18.3%) adults, including 10.4 million (4.2%) adults who have a serious mental illness (SMI) such as major depression [2]. On average, SMI patients have 3.5 times higher emergency room visits, 4 times higher primary care visits, and 5 times higher specialist visits than the general population [3]. It is estimated that 17.8% of the U.S. population is Latino or Hispanic [4]. Of those, 15.6% (over 8.9 million) had a diagnosable mental illness in the past year. The rate of illicit drug use for Hispanic individuals ages 12 and up was 8.9%, while the national average was 10.2% [5].

The growth of health information exchanges (HIEs) and healthcare technologies have stimulated interest in integrated care and data sharing. Behavioral health patients often visit multiple health care providers when receiving behavioral and physical health treatments, and care coordination among those providers could be advantageous [3]. While data sharing among care providers may improve care, it may also impact patient privacy [6,7]. Privacy concerns may lead patients to avoid discussing their problems with providers, delay care and withhold information from providers [8–10]. Additionally, patients with behavioral health conditions frequently experience social stigma, employment and insurance discrimination, legal concerns, and worry regarding disclosing information to others [11,12].

The Office for the National Coordinator for Health Information Technology (ONC) recommends giving patients more granular level control over how and with whom their health record information is shared [13]. Granular control could include giving patients more authority over with which providers, under which circumstances, and for which purposes their data are shared. For example, ability to make decisions over whether or not to share alcohol abuse-related information with a physical health provider. Studies indicate that patients desire more control over their health data for care and research [11,14,15]. A recent study of 394 patients indicated that they were reluctant to share clinical data with for-profit research organizations, and that 32% of them expressed a desire for choices regarding the data sharing category (for example, mental health, substance abuse history, etc.) and data recipient when sharing medical records for care [14]. Understanding the level of granularity that individuals desire is still an open question.

Little is known about data sharing preferences for care and research of behavioral health patients. Grando et al. surveyed 50 English speaking behavioral health patients with no SMI concerning their data sharing preferences [11]. The study showed that behavioral patients may wish granularity over who can access their personal health data for care and research. Understanding patients' data sharing preferences and perceptions may improve education and consent processes that influence their decisions to release or withhold health information.

The purpose of this study was to survey English and Spanish speaking behavioral health patients, including those with SMIs, on their perceptions regarding data sensitivity, willingness to share health data for care and research and related motivations.

Methods

Study Sites

Study site 1 is a community clinic in Arizona providing general mental health (GMH) treatment and social services to adults of all ages. Site 2, also in Arizona, offers case management services to adult patients with SMI.

Survey

We found validated English surveys, such as [16], developed to assess patients' data sharing choices. But there was a lack of bilingual (English and Spanish) data sharing questionnaires validated with behavioral health patients.

Our survey was based on [11]. Demographic information was categorized based on U.S. Census Bureau classifications, except diagnoses, which were adapted per National Institute of Mental Health categorization [17,18]. The sensitive categories used in our instrument to ask questions related to sensitive data were based on those used by the National Committee on Vital and Health Statistics [19]. The resulting survey was translated to Spanish and back-translated to English by native Spanish speakers.

Survey Reliability Testing

Reliabilities of questionnaire items were examined using a test-retest approach with 31 Spanish and English-speaking adult behavioral health patients from study sites 1 and 2. Participants completed the questionnaire, in either English or Spanish, on two occasions, 14-21 days apart. Questionnaire items were revised based on the outcomes of the reliability analyses. The revised questionnaire was used in the current study.

Study Participants

Participants for the study described here were recruited from the same study sites where the reliability study was conducted. Potential participants were identified by study site staff members during routine clinical visits and referred to the recruiters. After the recruiter met with the prospective participant at the facility and explained the study to him/her (in either English or Spanish), the recruiter assessed the participant's decision-making capability (using the UBACC test)[20]. We excluded participants with low consent comprehension (i.e., with UBACC scores < 15). Adult patients (21 years old or older) diagnosed with GMH or SMI who agreed to complete the questionnaire in English or Spanish and were deemed capable of giving informed consent were considered eligible to participate.

Study Design

After initial screening and consenting of eligible participant, the recruiter offered the participant the option of completing the questionnaire either in English or Spanish, and either electronically or on paper. The recruiter was present to help the participant with any questions or technical difficulties. Participants were compensated for their time.

Data Analysis

We used univariate statistics (e.g., frequencies, means, standard deviations, percentages) and plots to summarize the data. Parametric inferential statistical methods were used to analyze perceptions of data sensitivity and willingness to share data among English and Spanish-speaking, Latino and non-Latino participants from GMH and SMI populations.

Results

Demographics

Of the 88 participants recruited, 2 were excluded because of inability to understand and follow the study protocol, as measured by the UBACC test. Table 1 shows the demographics of participants included in the sample. The majority (n = 54; 62.8%) of patients had a GMH condition, while the rest were patients with SMI diagnoses. Most participants (n = 71; 82.5%) opted to have the questionnaire administered in English; the remainder opted for Spanish.

Table 1: Demographic of participants

Participant characteristics (n=86)	*Freq. (%)*
Age (Years)	
21-30	19 (22.1)
31-40	24 (27.9)
41-50	16 (18.6)
51-60	15 (17.4)
61-70	9 (10.5)
>70	2 (2.3)
Unknown	1 (1.2)
Gender	
Male	26 (30.2)
Female	59 (68.6)
Other	1 (1.2)
Race/Ethnicity	
White Alone, Not Hispanic or Latino	34 (39.5)
Black or African American	11 (12.8)
Hispanic or Latino	37 (43.0)
Native American or Alaskan Native	3 (3.5)
Other, Unknown	1 (1.2)
Income	
≤$10000	50 (58.2)
$10001-$20000	23 (26.7)
$20001-$30000	10 (11.6)
>$30001	3 (3.5)
Education	
No Schooling	1 (1.2)
Middle school (grades 6-8)	9 (10.5)
Some high school (no diploma)	14 (16.3)
High school graduate	19 (22.1)
Some college (1-4 years, no degree)	24 (27.9)
Associate degree	14 (16.2)
Bachelor's degree	5 (5.8)
*Diagnoses**	
Anxiety or panic disorder	65
Bipolar Disorder	34
Depression	65
Impulse Control Problems	10
Identity or memory problems	22
Eating disorder	5
Obsessive compulsive disorder	9
Personality disorder	13
Schizophrenia or other psychosis	14
Drug or alcohol addiction	18
Post-traumatic stress disorder or adjustment disorder	36
Chronic pain or somatic disorder	24
Other	1

** As participants may have more than one diagnosis, the percentages are not reported.*

Data Sharing for Care

We asked questions to understand participants' desire for granular data sharing control based on type of information, information recipient and purpose of data usage. We asked participants how likely they were to share their behavioral

health data with different behavioral and non-behavioral providers (Figure 1). Participants were most willing to always or sometimes share their health information with the behavioral providers at the study sites, followed by emergency providers, other non-behavioral providers at the study sites (e.g., primary and specialty care providers, pharmacists), behavioral providers outside the sites, and lastly with other non-behavioral providers outside the study sites. No significant differences in responses were seen in comparisons of English vs. Spanish speakers (χ^2=1.27, p=0.866), Latino vs. non-Latino participants (χ^2=0.78, p=0.941), or GMH vs. SMI patients (χ^2=0.12, p=0.998).

In assessing participants' perceptions about how sensitive different types of health information are, we provided them with eight health information categories: mental health, psychotherapy notes, sexual and reproductive health, domestic violence and abuse information, information on sexually transmitted diseases, drug or substance abuse, alcohol abuse, and genetic data. Most participants considered mental health information the most sensitive, followed by psychotherapy notes (Figure 2). For several categories, the most common single response was '*It does not apply to me'*. Unfortunately, we did not collect with the survey information that could be used to check if the participant did not have certain types of sensitive medical records.

Participants were then asked about the likelihood of sharing sensitive health information with providers outside the study sites (Figure 3). We computed the mean percentage of patients who wanted choices regarding sharing their data with different types of providers. On average, when self-reporting having sensitive information in their medical records (the option *'It does not apply for me'* was not selected), many participants (64.15%) wanted to restrict those records from some or all health care providers.

In general, we observed a direct correspondence between perceived sensitivity of information and willingness to share. The main exception was genetic data. While participants considered genetic data the eighth most sensitive type of information, they ranked it as the third most sharable. However, with the exception of sexually transmitted diseases ($p<.05$, Fisher's exact test), none of the associations between willingness to share a particular category of information and perceived sensitivity of the information).

We asked participants about providers' access to health information when prescribing a new medication. Most participants (78.0%) responded that the providers should have access to all their health data, 12.0% thought that providers should see only
the data to which a patient provides the access, and 10.0% indicated that the providers should see all the health data only when the new medication may have any harmful interactions or effects.

Similarly, we asked patients about emergency providers' access to data in a life-threatening situation. Most of the participants (70.0%), reported that providers should have access to all their data, 19.0% endorsed giving emergency providers access only to data shared by the patients, and 11.0% indicated that providers should have access to all health data only when the emergency may be life threatening.

Participants endorsed sharing their data when it can benefit their own care and treatment (77.8%) or if/when their providers asked them to share their data (61.1%). Large majorities of participants trusted the providers at the study sites overall (87.8%) and trusted them to share only the health data that they consented to share (93.3%).

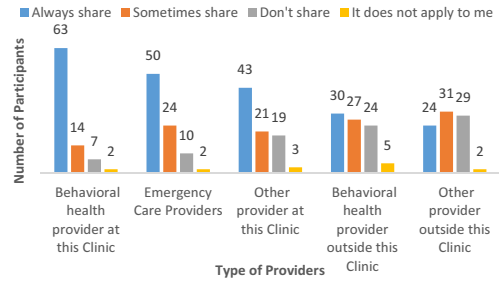


Figure 1: Behavioral health data sharing preferences, based on the type of medical provider

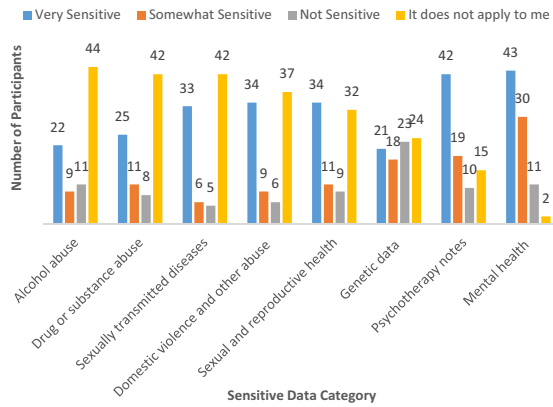


Figure 2: Health categories classification as sensitive information

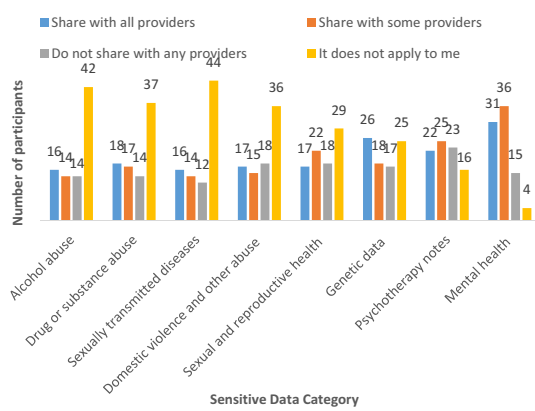


Figure 3: Willingness to share sensitive health data with providers outside study sites

Large majorities also reported that they would be upset if their providers shared their health data without asking them (83.3%) and that they might react by leaving such providers (65.6%). Only 30% of the participants reported worrying about providers knowing that they receive mental health treatment.

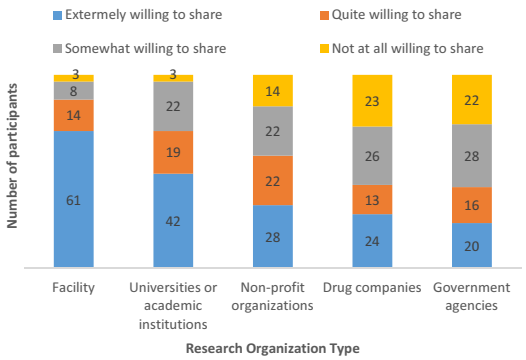


Figure 4: Willingness to share data for research with different types of organizations.

Data sharing for Research

Participants were generally willing to share health information with researchers when their own care (91.1%) or care for others (78.9%) could be improved. About half (51.1%) of participants, indicated they would always share their data for research, while 35.6% indicated that they would share their data for research if they were paid for it. Finally, we asked participants how likely they would be to share their health information with researchers (Figure 4). Participants indicated they would be extremely to somewhat willing to share their data for research purposes with their care facilities (96.5%). Participants appeared less willing to share their health information with drug development companies and government agencies. For each participant, we looked at whether their willingness level varied between different types of organizations. For example; varying willingness to share data with care facility compared to drug companies. Most (78.9%) participants desired control over how they want to share data with different research organizations.

There were no significant differences in data sharing preferences between English vs. Spanish speakers (χ^2=0.29, p=0.990), Latino vs. non-Latino participants (χ^2=0.59, p=0.964), or GMH vs. SMI patients (χ^2=0.25, p=0.993).

Discussion

Consistent with previous studies on behavioral health patients, participants wanted control over how to share sensitive health data with health providers [11]. In general, there was a direct correspondence between perceived sensitivity of information and willingness to share with all or some providers. When we contrast our results to studies from patients without behavioral conditions, it has been reported that patients with and without sensitive information prefer to restrict the sharing of sensitive versus less sensitive EHR information [14,15]. As reported in the literature, most of the participants appeared to be motivated to share health data unconditionally to avoid medical emergencies or drug-drug interactions [15,21,22]. As in previous studies, our participants trusted their providers at the study sites and trust in providers was an important motivation for sharing health information [11,16,23,24]. Additionally, improvement in a patient's own care and treatment was an important motivating factor for sharing health data with providers.

As in previous studies [11,14], patients wanted control over how to share health data with researchers. Consistent with literature, willingness to share data decreases when the recipient is a for-profit research organization and important motivations to share health information for research were benefiting own care or improving care for others [11,25].

The '*It does not apply to me'* response was frequently used when asked to assess the sensitivity of health data and willingness to share sensitive data with providers (34.4% for all types of data, and 29.7% for mental health, psychotherapy notes, drug or substance abuse and alcohol abuse). Participants' lack of understanding of the meaning of certain sensitive data categories, inability to form opinions regarding sensitive categories, or stigma related to disclosing this information could be potential explanations for this response. For some categories, like sexually transmitted disease or substance abuse, it is highly probable that the question did not apply to the participants. For other categories, such as genetic data, the recruiters received frequent requests from participants for clarifications. These results highlight the need for better on-demand education material to address patients' varying data sharing preferences and levels of health literacy.

A limitation of our study is that study participants were sampled from only two outpatient clinics in similar geographic areas with similar social demographics. Additional studies should be conducted on a larger sample of the population to capture more diverse views.

The outcomes of this survey and previous formative studies will guide follow-up card sorting interviews [11,25]. Thirty-six survey participants have given access to their health medical records available through the HIE. In the upcoming interviews, data privacy questions will be asked while study participants have access to a subset of their medical records.

Outcomes from that study will influence the design of an e-consent tool based on the Consent2Share software developed by the Substance Abuse and Mental Health Services Administration (SAMHSA) [26]. The My Data Choices tool will support patient-driven data access based on data sharing interoperability standards. Patient data will be shared in compliance with federal and state confidentiality laws, including protection of confidentiality of substance use disorder (42 CFR Part 2) [27]. On-demand multimedia patient education material will be embedded in the tool to illustrate risks and benefits of cross-organizational data sharing. We aim to pilot test the tool with 270 behavioral health patients in a prospective study.

Conclusions

A better understanding of behavioral health patients' attitudes towards data sharing is needed. The outcomes of this survey indicate consistency between the perspectives on data sharing and privacy of behavioral health patients and other previously surveyed populations of patients with or without sensitive medical records.

In future work, we plan to apply lessons learned from the completed survey to conduct follow-up interviews with a subset of the surveyed patients. The knowledge gained from the interviews will be used for the development of an e-consent tool that will support patient-driven data sharing control and on-demand educational resources to better inform data sharing choices.

Acknowledgements

This research was supported by the National Institute of Mental Health, through My Data Choices, Evaluation of Effective

Consent Strategies for Patients with Behavioral Health Conditions (R01MH108992).

References

[1] Substance Abuse and Mental Health Services Administration, National Behavioral Health Quality Framework, n.d. /data/national-behavioral-health-quality-framework.

[2] NIMH » Mental Illness, (n.d.). https://www.nimh.nih.gov/health/statistics/mental-illness.shtml (accessed October 19, 2018).

[3] SAMHSA-HRSA Center for Integrated Health Solutions. The Current State of Sharing Behavioral Health Information in Health Information Exchanges, (2014). https://www.integration.samhsa.gov/ (accessed June 23, 2018).

[4] SAMHSA, Racial and Ethnic Minority Populations, (2014). https://www.samhsa.gov/specific-populations/racial-ethnic-minority (accessed June 14, 2018).

[5] U.S. Census Bureau QuickFacts: UNITED STATES, (n.d.). https://www.census.gov/quickfacts/fact/table/US/RHI725216#viewtop (accessed June 14, 2018).

[6] M. Abdelhamid, J. Gaia, and G.L. Sanders, Putting the Focus Back on the Patient: How Privacy Concerns Affect Personal Health Information Sharing Intentions, *J Med Internet Res*. **19** (2017). doi:10.2196/jmir.6877.

[7] T. Bodenheimer, Coordinating Care — A Perilous Journey through the Health Care System, *New England Journal of Medicine*. **358** (2008) 1064–1071. doi:10.1056/NEJMhpr0706165.

[8] K. Caine, S. Kohn, C. Lawrence, R. Hanania, E.M. Meslin, and W.M. Tierney, Designing a patient-centered user interface for access decisions about EHR data: implications from patient interviews, *J Gen Intern Med*. **30 Suppl 1** (2015) S7-16. doi:10.1007/s11606-014-3049-9.

[9] C. Campos-Castillo, and D.L. Anthony, The double-edged sword of electronic health records: implications for patient disclosure, *J Am Med Inform Assoc*. **22** (2015) e130-140. doi:10.1136/amiajnl-2014-002804.

[10] P. Sankar, S. Mora, J.F. Merz, and N.L. Jones, Patient Perspectives of Medical Confidentiality, *J Gen Intern Med*. **18** (2003) 659–669. doi:10.1046/j.1525-1497.2003.20823.x.

[11] M.A. Grando, A. Murcko, S. Mahankali, M. Saks, M. Zent, D. Chern, C. Dye, R. Sharp, L. Young, P. Davis, M. Hiestand, and N. Hassanzadeh, A Study to Elicit Behavioral Health Patients' and Providers' Opinions on Health Records Consent, *J Law Med Ethics*. **45** (2017) 238–259. doi:10.1177/1073110517720653.

[12] California Healthcare Foundation. A Delicate Balance: Behavioral Health, Patient Privacy and the Need to Know, (n.d.). https://www.chcf.org/publication/a-delicate-balance-behavioral-health-patient-privacy-and-the-need-to-know/ (accessed June 23, 2018).

[13] Health IT policy committee, privacy and security tiger team, letter to David Blumenthal, Chairman of the Office of National Coordinator for Health IT, 2010. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4419941/#b16-1986238.

[14] H. Kim, E.A. Bell, J. Kim, A. Sitapati, J. Ramsdell, C. Farcas, D. Friedman, S. Feudjio Feupe, and L. Ohno-Machado, iCONCUR: informed consent for clinical data and bio-sample use for research, *Journal of the American Medical Informatics Association*. **24** (2017) 380–387.

[15] K. Caine, and R. Hanania, Patients want granular privacy control over health information in electronic medical records, *J Am Med Inform Assoc*. **20** (2013) 7–15. doi:10.1136/amiajnl-2012-001023.

[16] N. Serenko, and L. Fan, Patients' perceptions of privacy and their outcomes in healthcare, *International Journal of Behavioural and Healthcare Research*. **4** (2013) 101–122. doi:10.1504/IJBHR.2013.057359.

[17] U.C. Bureau, Census.gov, (n.d.). https://www.census.gov/en.html (accessed June 14, 2018).

[18] NIMH » Individual Mental and Behavioral Disorders for U.S., (n.d.). https://www.nimh.nih.gov/health/statistics/disability/individual-mental-and-behavioral-disorders-for-us.shtml (accessed June 14, 2018).

[19] National Committee on Vital and Health Statistics. Recommendations Regarding Sensitive Health Information, 2010. https://www.ncvhs.hhs.gov/wp-content/uploads/2014/05/101110lt.pdf.

[20] D.V. Jeste, B.W. Palmer, P.S. Appelbaum, S. Golshan, D. Glorioso, L.B. Dunn, K. Kim, T. Meeks, and H.C. Kraemer, A new brief instrument for assessing decisional capacity for clinical research, *Arch. Gen. Psychiatry*. **64** (2007) 966–974. doi:10.1001/archpsyc.64.8.966.

[21] V.N. Patel, R.V. Dhopeshwarkar, A. Edwards, Y. Barron, A. Likourezos, L. Burd, D. Olshansky, and R. Kaushal, Low-income, ethnically diverse consumers' perspective on health information exchange and personal health records, *Inform Health Soc Care*. **36** (2011) 233–252. doi:10.3109/17538157.2011.554930.

[22] M. Hiestand, M.A. Grando, A. Murcko, H. Soni, J. Varkey, C. Horrow, M. Saks, R. Sharp, M.J. Whitfield, M. Callesen, C. Dye, and D. Chern, Behavioral Health Provider Perspectives on Health Data Sharing, (2017).

[23] L. Ricciardi, Protecting Sensitive Health Information in the Context of Health Information Technology, Consumer Partnership for eHealth, 2010. http://www.nationalpartnership.org/research-library/health-care/HIT/protecting-sensitive-health.pdf.

[24] P.A. Teixeira, P. Gordon, E. Camhi, and S. Bakken, HIV patients' willingness to share personal health information electronically, *Patient Educ Couns*. **84** (2011) e9-12. doi:10.1016/j.pec.2010.07.013.

[25] E.A. Bell, L. Ohno-Machado, and M.A. Grando, Sharing My Health Data: A Survey of Data Sharing Preferences of Healthy Individuals, *AMIA Annu Symp Proc*. **2014** (2014) 1699–1708.

[26] C2S GitHub Welcome, (n.d.). https://bhits.github.io/consent2share/ (accessed June 22, 2018).

[27] 42 CFR Part 2 - CONFIDENTIALITY OF SUBSTANCE USE DISORDER PATIENT RECORDS, *LII / Legal Information Institute*. (n.d.). https://www.law.cornell.edu/cfr/text/42/part-2 (accessed June 23, 2018).

Address for correspondence

Adela Grando, PhD, agrando@asu.edu

Arizona State University, Department of Biomedical Informatics, 13212 E Shea Boulevard, Scottsdale, AZ 85259, US.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190450

Networking of Young Researchers in the European Area: Relevance, Requirements and Realization Possibilities

Bianca Steiner[a] and Birgit Saalfeld[b]

[a] *Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Braunschweig, Germany*
[b] *Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover, Germany*

Abstract

Networking is a key competence, especially for young researchers in the field of medical informatics. Therefore, it is encouraged in organizations like AMIA. Since, in Europe no such networking possibility is known, concepts and ideas for the implementation shall be established and assessed with regard to their appropriateness. Demands, suggestions and attitudes of the community were collected in an online survey. Based on this, a workshop with international participants was conducted at Medical Informatics Europe 2018 in Gothenburg, Sweden. Following topics were addressed: i) communication channels, ii) activities to be carried out, iii) organizational structures and iv) acquisition of participants. The results show the relevance of such a networking platform. Furthermore, numerous requirements and realization possibilities, but also challenges were identified and assessed during the workshop. Altogether, essential ideas for the implementation of an European Young Researcher Network (EYouRNet) were collected, which can serve as a basis for the realization.

Keywords:

Social Networking; Students; Medical Informatics

Introduction

During the doctoral studies students need not only to exchange with senior professionals, but also within their peer group in order to share their knowledge and experiences. Furthermore, discussing approaches and problems with other young scientists can encourage the tackling of their own topic with new ideas and enthusiasm. Especially in the field of medical informatics, as a highly cross-sectional discipline [1], this interchange is particularly important to learn from one another.

In the international area, various associations provide support for the networking of young scientists through an organizational framework. The "Young European Associated Researchers Network" (YEAR), for example, offers young researchers from all scientific fields the opportunity for training, networking and consulting. However, membership is only possible for young scientists working in an organization that is a member of the European Association of Research and Technology Organisations [2]. Due to this fact, it is rather difficult for individuals to participate in such a network. Apart from this, there is a multitude of other discipline-specific networks, such as the EMES in the area of social enterprise [3]. A special forum for young medical informatics researchers is offered by the American Medical Informatics Association (AMIA) with the "Student Working Group" where (doctoral) students "*[...] can share their educational experiences and viewpoints, as well as information about career and educational opportunities*" [4]. In Germany, once a year a doctoral symposium is held funded by the German Association for Medical Informatics, Biometry and Epidemiology (GMDS). Thereby, the aim is to offer doctoral students from the field of medical informatics the opportunity for networking, and sharing information and experiences [5].

For the interchange of doctoral students within the European area, there is no such networking possibility known. Therefore, as a first step, it is important to clarify to what extent an European network is really needed. In addition, if there is a need, a framework must be created with regard to how such a networking platform should exist, what requirements it should meet and what topics it should deal with.

For this purpose, the idea of the European Young Researcher Network (EYouRNet) as a starting point for a possible realization of such networking platform, was created. Young scientists should have the opportunity to share their experiences, answer questions and get inspired by the work of others. Furthermore, a logo (as seen in Figure 1) was developed beforehand for the recognition value of the EYouRNet-project.

Figure 1 – EyouRNet Logo

Methods

An online survey was used to determine to what extend such a networking platform is considered as relevant by the community members. The survey was distributed via the Council of the European Federation for Medical Informatics (EFMI) and the GMDS mailing list. In addition, suggestions and attitudes of potential future participants and other community members were identified. Based on these results, a workshop was held at the Medical Informatics Europe (MIE) 2018 in Gothenburg, Sweden. The workshop addressed the key issues for the successful implementation of an European networking platform emerged from the survey.

Preliminary Online Survey

A self-developed questionnaire was used for the preliminary online survey. It served to collect suggestions, demands and attitudes of the medical informatics community regarding organizational structures and activities of the proposed EYouRNet. For the implementation of the questionnaire the online survey service "eSurvey Creator" [6] was used in English language to enable the participation of international community members, especially Europeans. To reach the target

group, the survey was distributed via the EFMI-Council and the mailing list of the GMDS.

The self-developed questionnaire contained twelve questions in the form of free text fields, multiple choices and Likert scales. At various points there was the possibility of introducing further ideas and sharing thoughts and comments.

Firstly, demographics and previous experiences with networking possibilities during the doctoral studies were determined. Secondly, the relevance of various topics which could be addressed in the proposed EYouRNet were inquired. Besides the rating of given topics, such as education concepts, lessons learned and career opportunities, the participants could also contribute their suggestions. Finally, the question was asked how a personal contribution can be made.

Definition of Workshop's Thematic Fields

The information provided by the participants was assessed using common descriptive statistical methods. In addition to frequencies and relative frequencies, a content analysis was conducted relating to the free text fields. Here, four thematic fields could be identified, which are decisive for the implementation of the EYouRNet. These were extended by questions stimulating the participants to discuss [7].

1. **Communication channels**
   Which communication channels are necessary and suitable?
2. **Organizational structure**
   Which structures regarding the internal organization are reasonable?
   - Who does what when?
   - Which roles/positions do we need and what are their responsibilities?
   - What may be my own contribution to the EYouRNet?
3. **Activities to be carried out**
   Which activities should be carried out by the EYouRNet?
4. **Acquisition of participants**
   How to gain interested people both participants of EYouRNet and leading people?

Workshop

A 90-minute lasting workshop was organized to take place at MIE 2018 in Gothenburg, Sweden on April 25. The workshop aimed at introducing and improving new networking and collaboration approaches between doctoral students in the European area, especially within the EFMI. The workshop was especially directed to (former) doctoral students of medical informatics and related fields of research. Nevertheless, all conference participants were invited to take part. In the context of the workshop, possible structures, requirements and needs for the networking platform along with potential future participants and other community members should be discussed. Therefore, the workshop consisted of four successive parts: (1) round of introduction (2) presentation of the results of the online survey, (3) introduction of a national concept for networking, and (4) group work and discussion.

(1) Round of Introduction

First of all, the participants were asked to introduce themselves including their name, home country and working context, as well as the information whether or not they already have a doctoral degree. In order to break the ice, the participants were given the opportunity to mark their home country on a world map with a colored adhesive dot.

(2) Presentation of the Results From the Online Survey

To initiate the discussion, the results of the previously conducted online survey were presented. This comprised the results regarding the relevance of the proposed EYouRNet, the experiences of the participants, the topics to be addressed within the networking platform and the future contribution for possible members. Thereby, visualizations like pie charts were used for a better understanding.

(3) Introduction of a National Concept

As third part of the workshop, a short introduction to a national concept for the networking of doctoral students was given. Once a year, German doctoral students of medical informatics have the opportunity to meet up at the "GMDS-Doktorandensymposium". This event is funded by the GMDS and organized by the students themselves at changing locations in Germany. The two to three day event always includes a presentation and discussion of the individual topics of the doctoral theses, an excursion to prospective employers for medical informatics specialists, and a social event [5].

In 2017 the symposium was organized by the authors of this paper at the Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School. That year's excursion led the participants to the accident research division of "Volkswagen AG".

(4) Group Work

The previously mentioned thematic fields (see "Definition of workshop's thematic fields") were presented to the workshop participants and subsequently used for a group discussion. Since the actual number of participants of a workshop at a conference is not known beforehand, two different discussion methods have been prepared.

Whereas for a small group size an open discussion format would have been used, the simple and flexible group discussion method "World café" would have been the choice for a larger amount of participants [7]. When using the second method, the entire group is divided into smaller working groups interchanging independently, for example at different tables. Each of these groups focusses on one subtopic. The discussion points and results of each group (per table) are documented. After a pre-determined time, all but one group member, the table host, switch to a different table and thus to another subtopic. This is the starting point for a second discussion session. Now the table host briefly explains the previous discussion points and results to the new table members. Building on this, the discussion is continued with the new group members. After a predefined number of iterations the results for each subtopic are presented to the entire group. This gives the participants the opportunity to express supplementary thoughts and opinions they had not dealt with before.

Results

Online Survey

The preliminary online survey was available from March 13th to April 16th, 2018. In total 76 scientists participated in the survey. 57 of them completed the whole questionnaire. This results in 19 partial answered questionnaires.

Most participants came from Germany (40), three from the Netherlands, two each from Cameroon, Kenya, Romania and one person each from Austria, Brazil, India, Mexico, Russia, Slovakia, Switzerland and the USA. For 19 participants the home country is unknown, because this question was not included in the beginning of the survey by mistake. The majority of the participants had a Master degree or Diploma (46), followed by a Doctoral degree (15) or a

Habilitation (11). Only two participants had another, unspecified academic degree.

The relevance for the implementation of the networking platform can be derived by the responses to the question of being interested in participating or supporting the EYouRNet. More than 80% of participants are interested in participating or supporting the proposed network (see Figure 2).

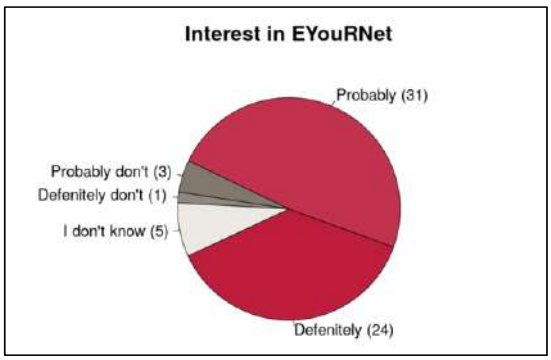


*Figure 2 – Interest of **P**articipating or **S**upporting the EYouRNet*

Participants of the survey, who have already finished their doctoral studies (19), were asked about their network experience as young scientists. Approximately one third of the respondents (7) stated that they had not received any support by networking possibilities. The remaining two-thirds were actively attending working groups (6), made informal contacts at conferences (4) or received support from colleagues (2).

With regard to the question of missed networking opportunities, only about one sixth of the respondents stated that they had not missed anything (3/17). The remaining 14 would have above all liked support in the form of networking events, informal exchanges and advices from experienced scientists, preferably also via an online platform.

The networks contribution to the doctoral student's self-management skills, technical competence and personal development was assessed. The majority of participants stated, that the impact of the networking possibilities is most important for the personal development. The contribution to the self-management skills and the technical competence were equally weighted.

The time to participate as a young researcher in such a network was indicated to be especially inspiring in the beginning of the doctoral studies. Nevertheless, also a later start has been reported as valuable and recommended.

Potential topics for the network to be addressed were also included in the survey. The results show that, apart from training concepts, all proposed topics were ranked as very important. This includes lessons learned during the doctoral studies, discussions on pitfalls in peer groups, training possibilities during the doctoral studies, cooperation opportunities, project support/exchange and career opportunities. In addition, the participants contributed the following suggestions and ideas with regard to further topics:

- Training on writing and presentation skills (2)
- Working groups on subtopics (2)
- Strengthening of social network (1)
- Get in touch with key players in the field of medical informatics and the EFMI-organization (1)
- Site visits around Europe (universities, hospitals) (1)
- Overview of common methods and tools (1)

In the concluding comments, eight participants gave valuable hints and ideas for the realization of the EYouRNet. Whereas, single individuals doubted the feasibility of the network others stated that they were happy to be a part of the realization. A major challenge is to find volunteers (students) to spread the ideas, both to interested students and to supervisors. Supervisor's role is to enable their doctoral students to participate, for example by financing an (extended) conference participation. Furthermore, supervisors play a decisive role in validating discussion outcomes and in the provision of valuable input. This goes hand in hand with the regulation of organizational issues. Here, the workshop verified the fact that face-to-face meetings require a high organizational effort, whereby they are probably most effective because they enable the participants to build direct working relationships (see subsection "Communication channels"). In addition to the face-to-face meetings, online meetings were suggested.

MIE Workshop

The workshop "Networking of PhD Students in the European Area" took place within MIE 2018 in Gothenburg, Sweden on Wednesday 25th April. 16 scientists from the field of medical informatics attended to this workshop. Their home country was mainly Germany, yet also Portugal, the Netherlands, Sweden, Finland and Ethiopia were represented. Among the participants only one had a doctoral degree.

First of all, the results of the online survey and the national concept of "GMDS-Doktorandensymposium" were presented in order to initiate the group work. Due to the amount of participants, the "World-café" format was chosen for the group discussion with two sessions each lasting 15 minutes. Four tables were prepared, each with a thematic field, the corresponding questions, and the ideas as well as the remarks from the online survey.

Communication Channels

Networking is premised on communication with one another. Accordingly, suitable channels must be identified for the communication in the network within the European area. Communication channels proposed by the survey participants were online platforms and real-life meetings. These were already noted on the prepared results document under the heading "Communication channels" (see Figure 3).

Regarding a suitable channel the participants of the workshop uttered the idea of using already existing platforms like LinkedIn, Rocket.Chat, Facebook, Google, Confluence or Zulip (see Figure 3). The use of such platforms offers the advantage of recourse to validated products with low overhead. It is important to consider where and how personal confidential data is stored. The participants explicitly whished no e-mail communication as these is difficult to manage over long periods of time with several participants and offers no possibility for new members to access historical posts.

For real-life meetings they proposed annual meetings, for example in combination with the MIE conference. They thought about a one-day meeting in advance or afterwards the conference to reduce the travel costs and organizational effort in comparison to independent meetings. Nevertheless, additional financial resources may be necessary.

Another mentioned idea was to host regular tutorials. Here, the participants thought about an list where doctoral students can add their topic-related expertise to talk about. Periodically training courses could be organized based on this list by students for students.

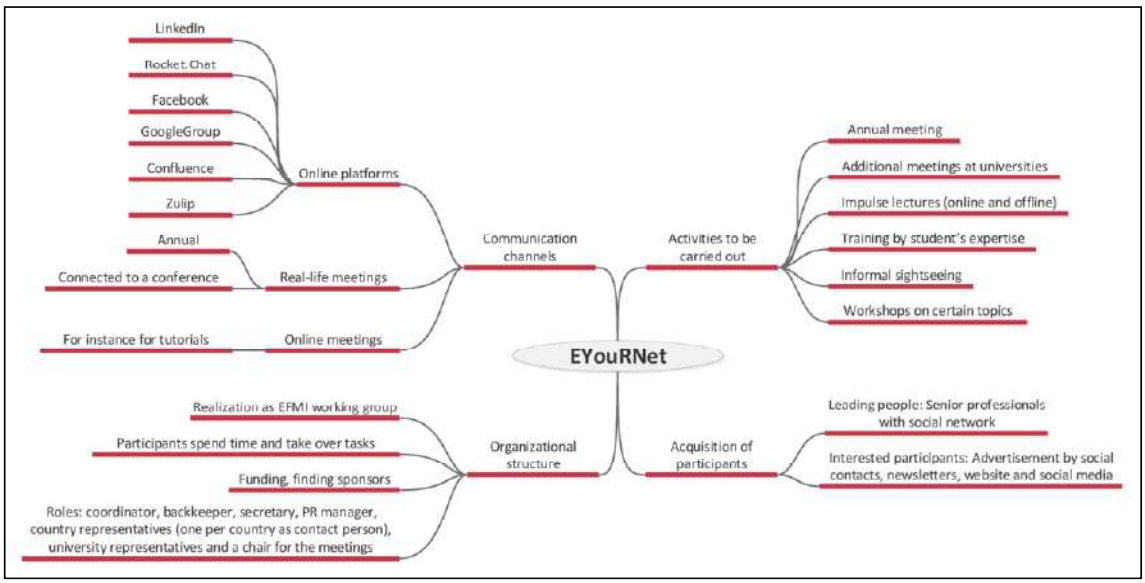


Figure 3 – Results of the World-Cafè – Suggestions and Demands of the articipants

Organizational Structure

In order to increase the chances for the EYouRNet to emerge and continue to exist, the organizational structures should be as complex as necessary, but also as simple as possible. This goes hand in hand with the lowest possible workload for (as many) members as possible. Thus, the obstacle to enter the network is small and the overall time expenditure for active members of the network stays realizable in the everyday work. The establishment of an EFMI working group represents a possibility for the realization within already existing structures.

The question "Which roles/positions do we need and what are their reponsibilities?" was answered with a list containing the following roles: a coordinator, a backkeeper, a secretary, a public relation manager, country representatives (one per country as contact person), university representatives and a chair for the meetings. The responsibilities of university and country representatives are getting in touch with (new) members and make contact to local existing student organizations.

Concerning their own potential contribution to the network the participants proposed time instead of money. They assumed, that every active member could spend up to four hours per month for the network. They also offered the possibility of couch surfing to minimize the traveling costs. Nevertheless, such a network relies on funding. Therefore, sponsors, university funding and student organizations with financial resources need to be identified and asked for their assistance. The following institutions are potential contributors: the EFMI, the GMDS, the International Medical Informatics Association (IMIA), cooperating partners, universities and the "Deutscher Akademischer Austausch Dienst*"* (DAAD), which "*[…] is the world's largest funding organization for the international exchange of students and researchers*" [8].

Activities to be Carried Out

Several activities to be carried out within the EYouRNet were identified by the workshop participants (see Figure 3). But they also perceive the organization of activities as most challenging, due to the fact that someone has to be responsible for the organization. Consequently, this person has a high (unpaid) expenditure. Moreover, the time required to prepare an activity should not be underestimated.

As most important activity annual meetings were suggested in combination with conferences so that the travel costs are minimized in comparision to additional meetings, e.g. at other universities. Organizers have to be aware of the fact, that there are often preconference workshops, which may are in temporal conflict with an EYouRNet meeting. The workshop participants also suggested informal sightseeing at the conference location. A list of sights shall be provided by local students. Formal meetings (also online) on selected topics with invited speakers were also mentioned. In addition, the participants wished for a platform to make calls for papers and other relevant content available to the community with little effort. This also includes self-organized workshops and further training of other doctoral students within the network. A list of the doctoral students' expertise can serve as a basis for such workshops and to find an appropriate topic by vote.

Acquisition of Participants

It is important to gain leading people to keep such a network running. The participants proposed senior professionals as potential leaders. However, it should be born in mind that their workload is often already very high anyway. For this reason, a self-administration of the network by doctoral students should take place.

One other important question is "how to gain the interested people as participants". Therefore, advertisement is needed by social contacts, newsletters, websites and social media. In doing so, ecpecially the personal benefits have to be communicated. For example, invited key players in the field of medical informatics can be attractive.

Discussion

In this paper, the relevance, requirements and realization possibilities for a European networking platform of doctoral students in the field of medical informatics are assessed. Therefore, a self-developed online survey with 76 participants of the community was conducted. Based upon the results of the survey a workshop with 16 participants has been carried out at MIE 2018. Here, four essential thematic fields were dealt with.

The results of the survey show the relevance and demand for the EYouRNet. More than 80% of the participants of the survey

are interested in participating or supporting the proposed network. During the workshop many realization possibilites with varying personal and financial effort were discussed. Besides regular meetings in combination with conferences, existing communications platforms such as LinkedIn, Facebook or Zulip were preferred for the interchange. Overall, the organizational structure should be small to minimize the obstacle to enter the network. In addition to a network coordinator and country representatives many other roles are needed to keep such a network running.

There are, however, some limitations to the described work. The online survey was mainly answered by german people (bias), due to the dissemination via the GMDS mailing list. Furthermore, the dropout rate was 25%. This could have been caused by a lost of interest to fillout the questionnaire. Also the amount and length of questions, as well as the time taken to answer them may have deterred people to complete them. Nevertheless, every participant who has answered the first four questions also completed the whole questionnaire. This confirmes the high level of interest in this topic. Even though not all questionnaires were filled in completely, all answers can be taken into account for the assessment. All participants answered the questions regarding their demographics and the relevance for implementing an European networking platform. The remaining questions were designed independently, thus they can be analyzed separately.

Due to the fact that the majority of the workshop participants came from Germany too, there will be a further bias. This bias could be decreased by additional interviews with scientists from other countries. It should be born in mind that both the online survey and the workshop were conducted in English language to reach international participants. Despite this bias, the results are sufficiently informative, as these are based on statements by members of the target group (young scientists). In-depth considerations of the ideas discussed in the "World-café" during the workshop are necessary in order to select suitable candidates for an initial implementation. Furthermore, maybe the presented results of the survey and the national concept "GMDS-Doktorandensymposium" influenced the participants during the workshop. However, providing general ideas and insights was a conscious decision to initiate the discussions in the rather short time of an workshop. Nonetheless, many ideas and realization possibilities were collected from the workshop participants.

So far, there has been no exchange with other working groups or organizations for the networking of young scientists. A next step could be to exchange with these groups in order to learn from them.

Conclusions

Networking is a key competence, especially for young researchers in the field of medical informatics as a highly cross-sectional discipline. Therefore, it is encouraged in many organizations like AMIA. However, within the European area, there is no such networking possibility known. In this paper the relevance, requirements and realization possibilities for a European networking platform of doctoral students in the field of medical informatics are presented. The relevance as well as first attitudes and suggestions were established by an online survey. In a subsequent workshop within MIE 2018, possible communication channels, organizational structures, activities to be carried out, and acquisition of participants were discussed with potential future participants and other community members.

Altogether, the relvance and the demand for the EYouRNet was shown. Furthermore, the general acceptance, in the surveyed group, to participate and support the network is given. Essential ideas for the implementation of the EYouRNet were collected, which can serve as a basis for the realization.

Acknowledgements

The authors would like to thank all participants of the first EYouRNet workshop at MIE 2018 in Gothenburg, Sweden, for their outstanding cooperation and inspiring discussion. Special thanks is due to Hannes Ulrich (Universität zu Lübeck, Institute of Medical Informatics), who started a first platform realization with Zulip.

Moreover the authors would like to thank Alfred Winter for the support with the implementation and distribution of the online survey and the announcement of the workshop via the mailing list of the GMDS and the EFMI-Council.

References

[1] R. Haux, Medical informatics: Past, present, future, *International Journal of Medical Informatics* **79** (2010), 599-610.

[2] Young European Associated Researchers network [Internet]. Brussels, Belgium: YEAR; c2019 [cited 2019 Mar 25]. About our organization. Available from: http://www.year-network.com/

[3] EMES [Internet]. Liege (Sart-Tilman), Belgium: EMES – International Research Network; c2019 [cited 2019 Mar 25]. Research network for social enterprise. Available from: https://emes.net/

[4] International Medical Informatics Association [Internet]. Geneva, Switzerland: IMIA; c2018 [cited 2018 Nov 8]. IMIA Working Groups & Special Interest Groups. Available from: https://imia-medinfo.org/wp/wgsig/

[5] GMDS [Internet]. Cologne, Germany: GMDS e.V.; c2018 [cited 2018 Nov 8]. GMDS Doktorandensymposium. Available from: https://gmds.de/aktivitaeten/medizinische-informatik/gmds-doktorandensymposien/

[6] enuvo GmbH [Internet]. Zurich, Switzerland: enuvo GmbH; c2018 [cited 2018 Nov 8]. Easily create professional online surveys. Available from: https://www.esurveycreator.com/

[7] S. J. Gill, *The Manager's Pocket Guide to Organizational Learning*, Human Resource Development, Amherst, 2000.

[8] Deutscher Akademischer Austauschdienst [Internet]. Bonn, Germany: Deutscher Akademischer Austauschdienst e.V.; c2018 [cited 2018 Nov 8]. The internationalisation agency. Available from: https://www.daad.de/der-daad/ueber-den-daad/portrait/en/29143-the-internationalisation-agency/

Address for correspondence

Birgit Saalfeld, M.Sc.

Peter L. Reichertz Institute for Medical Informatics

of TU Braunschweig and Hannover Medical School

Carl-Neuberg-Str. 1

30625 D-Hannover, Germany

eMail: Birgit.Saalfeld@plri.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190451

Development of a Mobile Learning System for Nurses' Cultural Competency Training

Sumi Sung[a], Hyeoun-Ae Park[a]

[a] *College of Nursing, Seoul National University, Seoul, Republic of Korea*

Abstract

We developed a mobile learning system that provides cultural competency training courses for nurses at a university hospital in South Korea using the system development life cycle. The contents and functions of the system were identified from the literature review, expert's recommendations, and the users' requirements. An algorithm, database, and user interface were designed. The system was implemented using the Newin Touchclass authoring tool. We evaluated and modified the system based on heuristics evaluation.

Keywords:

Mobile Application, Nursing Education, Culturally Competent Care

Introduction

South Korea is rapidly becoming a multicultural society. As the cultural backgrounds of patients have diversified, the need for culturally competent nursing care has emerged. In order to provide culturally appropriate nursing care to patients with culturally diverse backgrounds, it is important to train nurses on their cultural knowledge, awareness, sensitivity, and skills in nursing practices.

As most nurses work shifts and have intensive workloads, it is difficult for them to attend face to face classes [1]. Meanwhile, mobile learning enables nurses to be accessible from anywhere, can be shared instantaneously with everyone using the same content, brings strong portability, and is cost-effective [2].

The aim of this study is to develop and evaluate a mobile learning system that provides a cultural competency training course for nurses caring for foreign patients working at a university hospital in South Korea.

Methods

We developed a mobile learning system using the four stages of the system development life cycle: analysis, design, implementation, and evaluation.

Analysis

The content and functions of the mobile learning system were developed by literature review, experts' recommendations, the users' requirements through face to face interviews, and validated by experts.

We first reviewed transcultural nursing theories and models to identify a framework for the content of the system. The tools for the pre-post test were also reviewed to evaluate the effect of a mobile learning system.

Four experts, working at a university hospital as medical coordinators, were invited to identify cultural conflict examples and cases based on their experiences when brokering cultural differences between South Korean nurses and foreign patients.

Requirements of the users were identified through face to face interviews with eight nurses at a national university hospital. After Institutional Review Board approval (IRB No. H-1807-109-959), eight nurses were interviewed for about an hour with the aid of semi-structured questions regarding content, method, and time preferences related to the cultural competency training course throughout mobile learning.

The contents and functions of the system were validated based on feedback from six experts working at a university hospital.

Design

Based on the results from the analysis stage, the decision regarding which software to use was made. Choices and decisions regarding content, the configuration of learning modules, and use of multimedia (verbal, visual, or audio) were also made. Several functions available to implement in the software chosen were considered. An algorithm for learning and a database for learning modules was designed. The user interfaces of the pre-post tests and eight learning modules were designed according to the contents and functions from the analysis stage using the Microsoft Office PowerPoint program.

Implementation

We converted the slide presentations to a storyboard of learning modules using the authoring tool of Newin Touchclass on the website.

Evaluation

The heuristics of the mobile learning system on the Touchclass mobile application were evaluated using a mobile heuristic tool by four experts with majors in nursing informatics. If two or more people scored an item as one point or higher, and if any items scored four points, items were considered to have major problems and required revision.

Results

Analysis

We chose the Purnell model for cultural competence [3] as a framework for the content of the system. Two different approaches, cross-cultural approach and categorical approach, were adopted.

The users had a positive attitude toward learning general cultural competent nursing care; however, some of them only wanted to focus on the cultural information of particular cultural groups. The users required HTML-based mobile learning content rather than video content not exceeding 10 minutes to maintain their concentration. The users suggested education related to four particular cultural groups (Chinese, Russian, Mongolian, and Arab from the United Arab Emirates), feedback of cultural experts, sharing the difficulties with other nurses, and guidelines for nursing assessments.

The content of the mobile learning system consisted of eight modules. Based on the users' requests, cultural knowledge of four particular groups and guidelines of nursing assessments and interventions were included.

The functions of the mobile learning system consisted of three mobile learning strategies: guided learning, synchronous sharing, and issue-based learning. These were developed based on the literature review. Based on the users' requests, we decided to develop HTML-based mobile-learning content running less than 10 minutes for each learning module. In total, 17 cultural conflict examples, 10 cases, and three quizzes were included in eight modules. Written text with six tables, three linked YouTube videos, two videos, and 68 images were used.

From the experts' validation, a few expressions which may lead to stereotyping or increased cultural misunderstanding were revised. Common diseases and treatments of four particular groups were added to the biocultural ecology domain in Purnell's 12 cultural domains.

Design

We used the Newin Touchclass software to develop the mobile learning system. Five mobile application functions (registration, login, notices, communication, and alerts) in the Newin Touchclass mobile application development software were chosen.

An algorithm for learning was developed based on requirements of users and module contents developed in the analysis stage. The user's pre-cultural competency level was assessed. After that, the system provided the education, eight modules. When finishing each module, the user had an online discussion on the freeboard. When the user finished the course, their post-cultural competency level was assessed.

The database of the system was drawn in an entity-relation diagram. The data used in the mobile learning system was grouped into four categories: personal information, pre-test, post-test, and learning module. Data dictionaries were created with possible values, data types, and units.

The user interface in this study consists of the 17 interface categories. Login, registration, main page, my course page, noticeboard, freeboard, and pop-up window were originally designed on the Newin Touchclass software. We designed 27 user interfaces on eight learning modules with five videos, 68 images, and six tables (Figure 1). Pre-post tests were placed before and after the course.

Implementation

The operating environment of the app is iOS 8.0 and Android 4.0.3 or higher. The mobile learning modules of the system were developed using the Newin Touchclass authoring tool from September 20, 2018, to November 8, 2018.

Evaluation

In total, 18 suggestions were made by the four experts in the heuristics evaluation. Three suggestions proposed by more than two evaluators were selected, and the system was modified accordingly: "In case of an application error, the contact information is required to be presented", "Unnecessary videos are shown with liked YouTube video", and "The image that indicates going to the next page is confused whether the user should push the screen to the right or to the left."

Conclusions

We developed a mobile learning system for a cultural competency training course. The mobile learning system will be tested and used as part of the on-the-job-training courses for the nursing workforce.

Acknowledgements

This study was financially supported by Seoul National University Hospital.

References

[1] S. Karaman, Nurses' perceptions of online continuing education, *BMC Medical Education* **11** (2011) 86.

[2] Z.L. Berge and L. Muilenburg, editors, Handbook of Mobile Learning, 1st Edition, Routledge, New York, NY, 2013.

[3] L. Purnell, The Purnell model for cultural competence, *Journal of Transcultural Nursing* **13** (2002) 193-196.

Address for Correspondence

Hyeoun-Ae Park, RN, PhD, FAAN, FACMI

Professor

College of Nursing, Seoul National University

103 Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea

Fax +82-2-765-4103

Tel +82-2-740-8827

E-mail hapark@snu.ac.kr

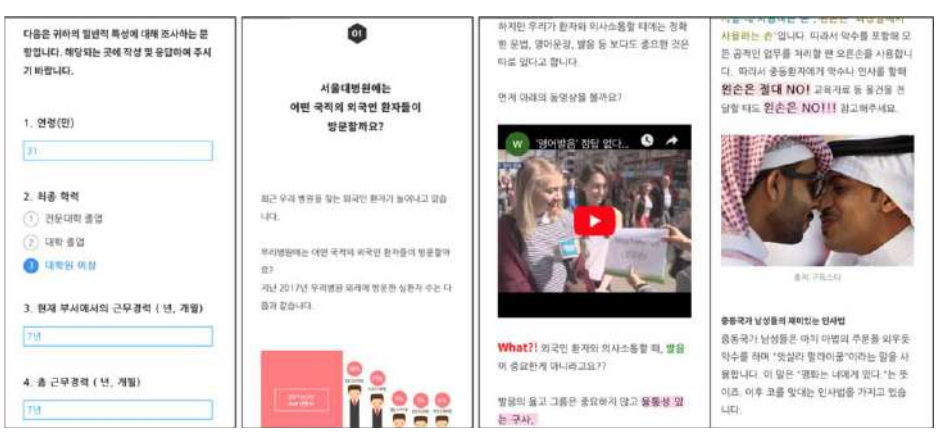


Figure 1 – The screen of the pre-post test and module contents

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190452

Assessment of Traceability Implementation of a Cross-Institutional Secure Data Collection System Based on Distributed Standardized EMR Storage

Katsuya Tanaka[a,b], Ryuichi Yamamoto[c]

[a] *Department of Biomedical Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan*
[b] *IT Integrating and Support Center, National Cancer Center, Tokyo, Japan*
[c] *Medical Information System Development Center, Tokyo, Japan*

Abstract

This paper presents results concerning the traceability of a secure data collection infrastructure based on the Bloom Filter involving standardized electronic medical record (EMR) storage. The objective of this infrastructure is to facilitate rapid secondary use of exported EMR data in cross-patient or cross-institutional analyses based on the Standardized Structured Medical Information eXchange (SS-MIX), Japan's domestic standard for EMR export. Secondary use of EMR data should be based on the principle of patient consent. Therefore, securing the traceability of patient EMR data is important for achieving reliable data collection systems for secondary use. Blockchain is a strong candidate, and we evaluated its performance using several implementations. As shown by the evaluation, it is difficult to realize the method of recording the query history of secondary use on the blockchain. We thus propose a method for recording index information of distributed log data.

Keywords:
Privacy, Data Collection, Electronic Health Records

Introduction

Background

In Japan, the domestic standard for exporting whole electronic medical record (EMR) data to external storage is the Standardized Structured Medical Information eXchange (SS-MIX2) [1], which is based on Health Level-7 (HL7) v2 message files. This system is widely used for backing up data, regional collaboration, disease registries, and other purposes. In this standard, we can represent definite states of 37 clinical events of each patient in a hospital, including admission-discharge history, medication history, and laboratory test results using ADT/RDE/OUL HL7 v2 messages. The directory structure for stored records is based on patient ID, clinical date, and clinical event type. This structure makes it difficult to use exported data for cross-patient analyses, such as epidemiological studies. In May 2017, an amendment to Japan's Act on the Protection of Personal Information [2] designated medical information as important confidential information, requiring strict consent for its provision to a third party. Secondary use of medical information for research is permissible if utilized in accordance with ethical guidelines; however, large-scale data collection and analyses that include secondary use of information by third parties, including commercial companies, is technically difficult because of the requirement to obtain consent from each patient for information disclosure. Therefore, a new law was enacted in May 2018, under which a certified business operator can collect medical information, anonymize the information, and respond to a third party's request for the information for use in analysis. The new law allows secondary usage of patient information on the assumption of consent unless the patient specifically opts out. Patients are provided with an opportunity to opt out. Several forms of data collection are assumed, with large-scale collection using SS-MIX2 being a leading candidate. With SS-MIX2, personal information is collected and transferred to a data center without anonymization. When personal information is collected as raw data, there is a risk of information disclosure. There are also inevitable maintenance costs according to the scale of collection and storage.

To address these issues, we propose an alternative system of collecting and storing EMR data wherein only the necessary items are included, eliminating the need for identifiable patient information to be spread outside the medical institution. The proposed system facilitates EMR data distribution within each medical institution, enabling cross-patient or cross-facility data collection and analysis. Integration and encryption of extracted EMR data are achieved using the Private Set Intersection (PSI) library developed by Miyaji [2]. However, at present, consent management has not been fully discussed [3]. Consent management is a key factor for secondary research use, and several challenges have been reported[4; 5]. Our approach focuses mainly on auditing the secondary usage under opt-out style data collection research using our developed data collection system across participating medical institutions.

Research Objective

This paper investigated how to implement the capability of traceability in the developed system for secure data collection and analysis. Blockchain technology has been recently applied in health care fields, including primary patient care, data aggregation for research purposes, and connecting healthcare providers [6-8]. The system that we are developing has a second purpose: For securing the traceability of EMR data, methods to disclose the logs of secondary usage are needed. In the present situation where patients do not have any common ID, it is difficult for a patient to audit all the secondary usage logs across the distributed hospital storages which he/she visited. By blockchain technology, we expect to provide patients a common search infrastructure with immutable secondary usage logs. Thus, we plan to apply blockchain technology to the aggregation of data extraction log records. This method has several possible implementations, and they must be evaluated assuming operations in real use. The experimental results shown below mainly concern data structure and transaction performance compared to traditional implementation for achieving aggregation of distributed log records of EMR data extraction.

Methods

Overview and Concepts of the Developed System

Figure 1 presents an overview of the system. The key concepts are:

1. Each medical institution has EMR data in SS-MIX2 storage, including billions of HL7 v2 messages.
2. HL7 v2 messages are periodically parsed and stored to relational database management system (RDBMS) tables, maintaining synchronization with the billions of message files in SS-MIX2.
3. Analysis requests from researchers and data collection are managed by the PSI service on the cloud, which communicates with a client agent located at a client terminal and PSI agents located at each medical institution.
4. Target data criteria, such as diseases, age, and gender, must be defined before the PSI executes data collection. The PSI party agent deploys the target dataset in advance from the local RDBMS to memory.
5. Data collection is achieved using PSI software, which is based on Bloom Filter technology for record verification across institutions. The application of Bloom Filter technology is aimed at realizing data matching in which personal information does not leak outside each hospital while the data collection process.
6. The collected dataset can be verified considering the possibility of patient identification using the extracted attributes.
7. Patients can trace the use of their medical records during data collection.
8. If they choose, patients can withdraw consent for secondary use of their data. Consent withdrawal information is assumed to be an input to existing the EMR system and exported SS-MIX2 storage in each hospital.

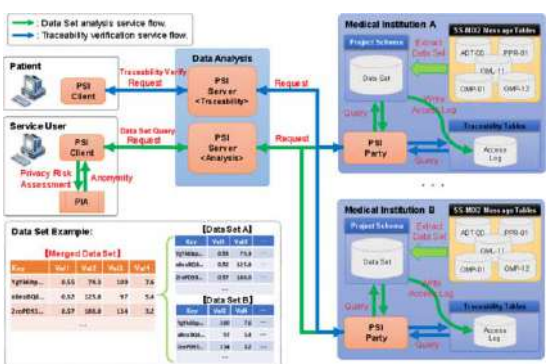


Figure 1. Overview of the Developed System for Secure Data Collection and Analysis (originally published in [3])

Traceability for Patients

EMR storage for the developed secure data collection system is supposed to process queries from clinical researchers using the standard interface implemented by the PostgreSQL database. EMR data are extracted by data extraction requests handled by the PSI service. Thus, the selected records are identifiable based on each query result, and the records represent the disclosure history of EMR data during data collection through use of the developed PSI service. By making the log record of extraction searchable by patients, we suppose that traceability in the secure data collection system will be achieved. However, because storage is supposed to be distributed at each hospital, log records must be aggregated by some secure method to be made auditable.

Log data is assumed to be represented by a combination of the following attributes:

1. Identifier of target patient (Patient Identifier)
2. Storage source (Medical Institution Identifier)
3. Disclosed destination (Extracting User Identifier)
4. Purpose of use
5. Type of extracted EMR data
6. Extraction timestamp

Attribute 1 (patient identifier) is mandatory for patient identification. At present, universal patient identifiers are not available in Japan. We assume that insurance numbers may be desirable for searching log records across medical institutions, because the patient ID at one medical institution is only applicable for searching log records at that medical institution.

Attribute 2 (medical institution identifier) is used to distinguish the institution storing the extracted EMR data.

Attribute 5 (type of extracted EMR data) is represented by HL7 ver.2 message types such as “ADT-00”,”OMP-01”,”OML-11”.

Attributes 3, 4, 5, and 6 are used to distinguish secondary use of target EMR data by patients. By verifying these attributes, patients can determine whether actual secondary usages meet their consent.

Data Structure for Query

A query for EMR storage may extract records of many patients at one time. For disclosing extracted history to patients, the extracted history should be sorted by patient and each history should include the attributes described above.

By focusing on one patient, the extracted history grows as queries hit the target patient EMR record. Moreover, this extracted history is distributed at each EMR storage site across the participating medical institutions.

For achieving desirable response, the aggregation of extracted history should be obtained in a realistic time. This is closely related to the data structure and size of each log record. Future studies should focus on data size of stored log records.

In the performance test, a simple message structure is defined as a JSON (shown in Table 1). Identifiers of patient and institution are represented as hash values. Each log record can be stored separately in the blockchain (separate style) or aggregated in a block by a patient appending records to the corresponding block (appending style). In the former method, we must gather the pieces of the records related to the patient of interest. In the latter method, the block size grows as the system is used. We examined performance differences when the data size of a record to be written is changed.

Experimental Setups

We evaluated the following three approaches to implement traceability function. Of these, two are based on blockchain

Table 1. Sample Data Representing Extraction History.


```
{
        "patientID":"781e5e245d69b566979b86e28d23f2c7",
        "insutitionID":"aabd258c8894b996e8d8561fa868364d",
        "disclosedDestination":"AnalysisUser001",
        "purposeofUse":"DrugDevelopment",
        "typeofRecords":"OMP-01",
        "extractionTime":"2018/11/12 01:23:45"
}
```

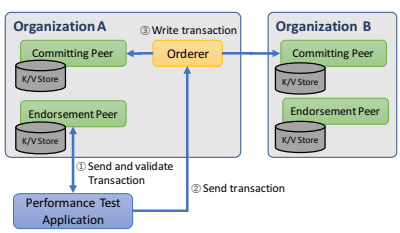

technology. The last approach uses the same method of secure data collection as PSI against log records stored in distributed PostgreSQL databases.

- Hyperledger Fabric[9]
- BigchainDB[10]
- PSI (Bloom Filter)

The experimental settings for each approach are described below. Between Hyperledger Fabric and BigchainDB, key/value store implementation for search use differs from each other.

Hyperledger Fabric

Figure 2 shows the experimental setup using Hyperledger Fabric to store query log records during data collection. Assuming two participating institutions, two nodes were set for the performance test. Native implementation only offers Key-Value storage and is applicable to a separate style. We also evaluated Hyperledger implementation with CouchDB [11], which enables query against the value of the JSON message described above. Thus, we can implement both separate and appending styles.



VM Host Machine		VM * 2 node		Middleware version	
CPU	Intel Xeon Bronze 3106 (16 CPUs @1.70GHz)	CPU	4 CPUs	HyperLedgerFabric	1.3.0
Mem	64GByte	Memory	8GByte	Docker	18.06.1
OS	VMWare ESXi 6.7.0	OS	Ubuntu 18.04.1 LTS	CouchDB	2.1.1

Figure 2. Experimental Settings (Hyperledger Fabric)

BigchainDB

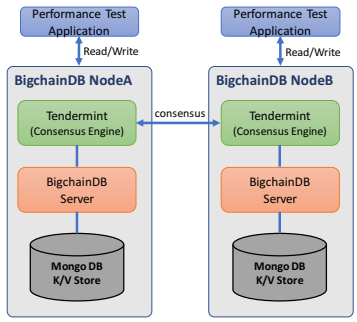
Figure 3 shows the experimental settings for using BigchainDB to store log data. As above, two nodes were prepared for evaluation. MongoDB [12] was selected as the backend database. In this case, both separate and aggregated structures are possible on the same implementation.

Query key candidate is the transaction ID of the stored block or stored JSON value.

PSI (Bloom Filter)

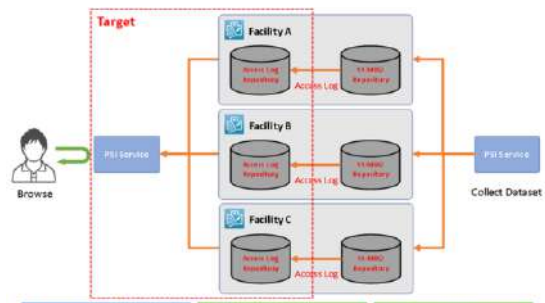
Figure 4 shows the experimental settings in the case of PSI implementation. The log records of data extraction are recorded at the time of extraction. Using the same method of EMR data collection, we can gather the log records against distributed storages under encryption. That is, although the search is performed by specifying the insurance number, the date of birth, the gender by patients, since the matching is performed using the Bloom Filter, these values are not directly disclosed on the infrastructure.

In this test, three nodes were prepared for evaluation, but the performance test measurement was executed on only one node.



VM Host Machine		VM * 2 node		Middleware version	
CPU	Intel Xeon Bronze 3106 (16 CPUs @1.70GHz)	CPU	4 CPUs	Tendermint	0.22.8
Mem	64GByte	Memory	8GByte	BigchainDB	2.0.0b7
OS	VMWare ESXi 6.7.0	OS	Ubuntu 18.04.1 LTS	MongoDB	3.6.3

Figure 3. Experimental Settings (BigchainDB)



VM Host Machine		VM * 3 nodes		Middleware version	
CPU	Intel Xeon E5-2470 (10 CPUs @2.40GHz)	CPU	4 CPUs	Apache Tomcat	8.0
Mem	64GByte	Memory	8GByte	PostgreSQL	9.5
OS	VMWare ESXi 6.7.0	OS	Ubuntu 16.04.1 LTS		

Figure 4. Experimental Settings (PSI)

Results

Performance by Data Size

Figure 5 shows the performance of writing records to blockchain storage by record size for Hyperledger. In the experimental environment, it worked normally for records 7 MB or smaller. As the record size grew, the response became unstable.

Figure 6 shows the same test for BigchainDB. The maximum record size was 0.6 MB, which is much lower than that for Hyperledger. However, the transaction time to commit was larger than that for Hyperledger.

By contrast, the data size for PSI can be as large as allowed by the database system.

Transaction Performance

Figure 7 shows the performance results of writing records to blockchain storage for Hyperledger with/without CouchDB and BigchainDB under one or five thread processing.

In all cases, processing by threads contributed to storage performance, but the throughput did not increase linearly with the number of threads.

Comparing the three implementations, BigchainDB was slightly faster than Hyperledger. Hyperledger with CouchDB had the worst performance; this is likely caused by the cost of indexing within CouchDB. In the best case, 1 million records were written to the blockchain storage in 3-4 hours. This performance is equivalent to writing 10 million records or less in one day.

Comparing these implementations using blockchain technology, the performance of PSI was equivalent to the "insert" performance of the PostgreSQL database used. The necessary time for inserting 1 million records to the database was below 10 min. This performance is about 1000 times faster than the blockchain implementations.

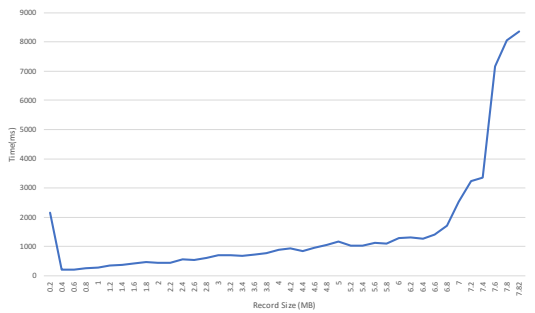


Figure 5. Performance Results by Record Size (Hyperledger)

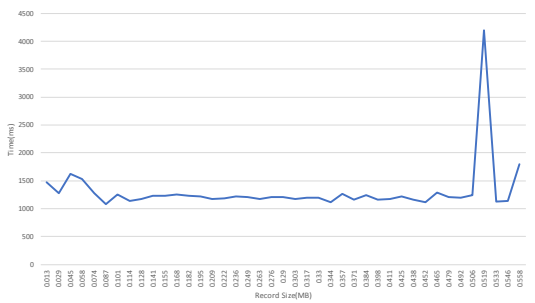


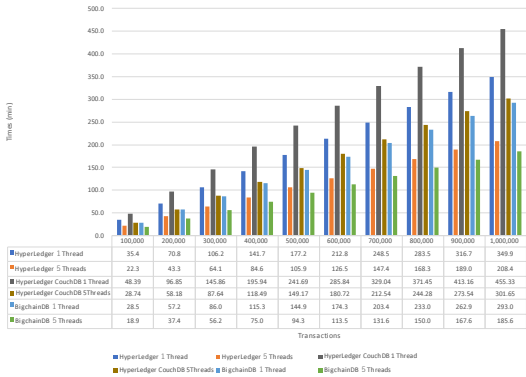
Figure 6. Performance Results by Record Size (BigchainDB)

Query Performance

Figure 7 shows the performance test results of retrieving one record from blockchain storage using four types of implementation. There were no significant differences in the query response times between Hyperledger and BigchainDB. "Hyperledger Key" and "BigchanDB transid" represent the separate style of storage, whereas "Hyperledger Value" and "BigchainDB AssetsText" represent the aggregated style.

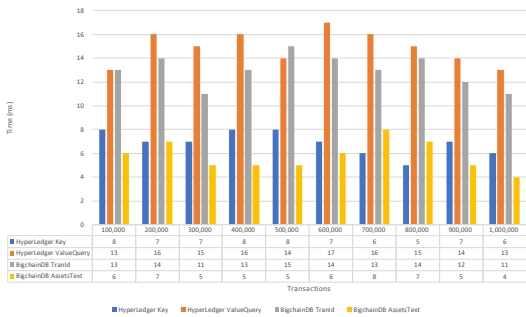
Figure 8 shows the performance test results of query response. Query response is fast enough for actual use in the case of 1 million records in the storage. This result shows hitting 1 record and the response time linearly increases as hit records increase.

On the other hand, PSI implementation needs 1 min or less to aggregate the extracted results across the distributed databases.



	100,000	200,000	300,000	400,000	500,000	600,000	700,000	800,000	900,000	1,000,000
HyperLedger 1 Thread	35.4	70.8	106.2	141.7	177.2	212.8	248.5	283.5	316.7	349.9
HyperLedger 5 Threads	22.3	43.3	64.1	84.6	105.9	126.5	147.4	168.3	189.0	208.4
HyperLedger CouchDB 1 Thread	48.39	96.85	145.86	195.94	241.69	285.84	329.04	371.45	413.16	455.33
HyperLedger CouchDB 5Threads	28.74	58.18	87.64	118.49	149.17	180.72	212.54	244.28	273.54	301.65
BigchainDB 1 Thread	28.5	57.2	86.0	115.3	144.9	174.3	203.4	233.0	262.9	293.0
BigchainDB 5 Threads	18.9	37.4	56.2	75.0	94.3	113.5	131.6	150.0	167.6	185.6

Figure 7. Performance result of writing records



	100,000	200,000	300,000	400,000	500,000	600,000	700,000	800,000	900,000	1,000,000
HyperLedger Key	8	7	7	8	8	7	6	5	7	6
HyperLedger ValueQuery	13	16	15	16	14	17	16	15	14	13
BigchainDB TranId	13	14	11	13	15	14	13	14	12	11
BigchainDB AssetsText	6	7	5	5	5	6	8	7	5	4

Figure 8. Performance Result of Querying Records

Discussion

Based on the initial evaluations, the following recommendations are made.

Transaction Performance

The transaction performance of a blockchain network was quite low for storing massive numbers of log records generated by queries in the developed system. In the case of blockchain, at most 100 transactions per second is best for a node to register to storage. Compared to implementation with PostgreSQL, the total transactions per day will be 1000 times smaller. If we do not implement any aggregation of log records, it will be impossible to process the enormous numbers of log records generated for each EMR item. Some patient-based aggregation of log records should be considered to overcome performance limitation.

Data Size

The results by data size show the upper limit for storing log records to blockchain storage. As writing large records to storage makes the system unstable, writing in the appending style is not suitable due to the long operation time of the system. Considering the transaction performance test results above, the total number of transactions to the blockchain network per day should be limited.

Query Response

As the amount of storage increases, the search function must query all storage in the network. Whole log records thus require some possible indexes for searching by patient. The query performance test results show a good response for searching for a

log record in the blockchain network despite the increase in the number of log records.

Proposed System for Future Implementation

Based on the performance evaluation results, we decided to implement the following policy as the basis for making the search log history visible to patients when using the developed secure data collection system.

- Aggregate log data by patient in each facility
- All log records are stored at each facility
- Record the minimum amount of data, such as the log record identifier key and facility identifiable key, for retrieving index data in the blockchain
- For query log data, use personal identification information, such as insurance number, date of birth, gender, etc.

By following these policies, a patient can search the blockchain and find the storage facility. Moreover, the number of records that must be recorded per period can be reduced to the number of related patients. Figure 9 shows an overview of the proposed log search system. The log records should include:

- Facility identifiable key
- Log record identifiable key
- Digest to audit each log record
- Key to identify each patient; this could be generated by encrypting a patient identifier like insurance number, date of birth, gender, etc.

We plan to develop a log search system with the described structure.

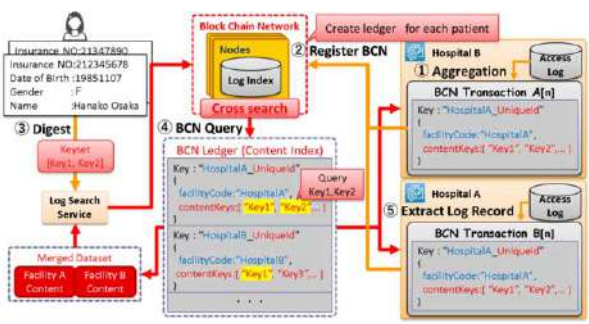


Figure 9. Overview of Proposed Log Search Service Using a Blockchain Network.

Limitations

Because we did not have sufficient time to set up larger records, performance tests were executed for 1 million records or less. As the number of records increases, the test results and system stability may change. Performance tests with more records are required in future work. Similarly, performance should be estimated for larger numbers of nodes.

Conclusions

This paper reports the initial performance results related to traceability for a secure data collection system under development. The desired data structure and system infrastructure were examined. Although blockchain implementation is a strong candidate for establishing an audit infrastructure to verify the use of EMR data for clinical research, there are some challenges for maintaining long-term operation as the amount of data increases. Thus, we proposed a data structure and querying implementation to overcome the implementation performance.

Acknowledgements

This study has been approved by the Research Ethics Committee of Graduate School of Medicine and Faculty of Medicine, The University of Tokyo (Permission number: 11787) and was supported by JST CREST Grant No. JPMJCR1404, Japan.

References

[1] M. Kimura, K. Nakayasu, Y. Ohshima, N. Fujita, N. Nakashima, H. Jozaki, T. Numano, T. Shimizu, M. Shimomura, F. Sasaki, T. Fujiki, T. Nakashima, K. Toyoda, H. Hoshi, T. Sakusabe, Y. Naito, K. Kawaguchi, H. Watanabe, and S. Tani, SS-MIX: a ministry project to promote standardized healthcare information exchange, *Methods Inf Med* **50** (2011), 131-139.

[2] A. Miyaji, K. Nakasho, and S. Nishida, Privacy-Preserving Integration of Medical Data, *J Med Syst* **41** (2017), 37.

[3] K. Tanaka, R. Yamamoto, K. Nakasho, and A. Miyaji, Development of a Secure Cross-Institutional Data Collection System Based on Distributed Standardized EMR Storage, *Stud Health Technol Inform* **255** (2018), 35-39.

[4] T. Ploug and S. Holm, Eliciting meta consent for future secondary research use of health data using a smartphone application - a proof of concept study in the Danish population, *BMC Med Ethics* **18** (2017), 51.

[5] H. Kondylakis, L. Koumakis, S. Hanold, I. Nwankwo, N. Forgo, K. Marias, M. Tsiknakis, and N. Graf, Donor's support tool: Enabling informed secondary use of patient's biomaterial and personal data, *Int J Med Inform* **97** (2017), 282-292.

[6] A. Dubovitskaya, Z. Xu, S. Ryu, M. Schumacher, and F. Wang, Secure and Trustable Electronic Medical Records Sharing using Blockchain, *AMIA Annu Symp Proc* **2017** (2017), 650-659.

[7] M.N. Kamel Boulos, J.T. Wilson, and K.A. Clauson, Geospatial blockchain: promises, challenges, and scenarios in health and healthcare, *Int J Health Geogr* **17** (2018), 25.

[8] J.M. Roman-Belmonte, H. De la Corte-Rodriguez, and E. C. Rodriguez-Merchan, How blockchain technology can change medicine, *Postgraduate Medicine* **130** (2018), 420-427.

[9] *Hyperledger Fabric – Hyperledger*. Available: https://www.hyperledger.org/projects/fabric

[10] *BigchainDB • • The blockchain database.* Available: https://www.bigchaindb.com/

[11] *Apache CouchDB*. Available: http://couchdb.apache.org/

[12] *Open Source Document Database*. Available: https://www.mongodb.com/index

Address for correspondence

Katsuya Tanaka, National Cancer Center
5-1-1 Tsukiji, Chuo-ku, Tokyo, 104-0045, Japan
katstana@ncc.go.jp

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190453

Understanding Patient Attitudes Toward Multifocal Intraocular Lenses in Online Medical Forums Through Sentiment Analysis

Sophia Y Wang[a], Tina Hernandez-Boussard[b], Robert T Chang[a], Suzann Pershing[a,c]

[a] *Byers Eye Institute, Stanford University, Palo Alto CA, USA,*
[b] *Biomedical Informatics Research Center, Stanford University, Palo Alto CA, USA,*
[c] *Veterans Affairs Palo Alto Health Care System, Palo Alto CA, USA*

Abstract

Multifocal intraocular lens implants (IOLs) are a premium option for cataract surgery which patients may purchase to achieve improved spectacle-independence for near vision but may have trade-offs with visual quality. We demonstrate the use of sentiment analysis to evaluate multifocal lenses discussed on MedHelp, a leading online health forum. A search for "multifocal IOL" was performed on MedHelp.org on November 1, 2016, yielding relevant patient posts. Sentiment analysis was performed using IBM's Watson, which extracted 30,066 unique keywords and their associated sentiment scores from 7495 posts written by 1474 unique patient users. Keywords associated with monovision, monofocal, and toric lenses had positive mean sentiment, significantly higher than for keywords associated with multifocals, which had negative mean sentiment ($p<0.001$, ANOVA). Many keywords represented complaints and were associated with negative sentiment, including glare, halo, and ghosting. Sentiment analysis can provide insights into patient perspectives towards multifocal lenses by interpreting online patient posts.

Keywords: Natural Language Processing, Multifocal Intraocular Lenses, Social Networking

Introduction

Cataract surgery is the most commonly performed surgery in the United States, with nearly 4 million surgeries performed per year [1]. Cataract surgery restores clarity of vision by replacing the natural, cloudy lens of the eye with a clear intraocular lens (IOL) implant. Standard monofocal IOLs are designed with a single dioptric power enabling sharp vision at a pre-determined target distance. Most patients desire a distance target, resulting in a need to wear reading glasses to correct presbyopia and achieve clear vision for near- or intermediate-distance tasks. Interest in sharp spectacle-free vision at a range of target distances has paved the way for the development of premium presbyopia- and astigmatism-correcting advanced technology IOLs. Some of these lenses have been designed to mimic accommodation (Crystalens), or provide clear vision at multiple predetermined focal points simultaneously via concentric rings of multiple dioptric powers (multifocal IOLs), or, more recently, offer extended depth of focus across a more continuous range of distances (Tecnis Symfony). These lenses have been promoted as a method of achieving spectacle independence. However, patients must bear the cost of this premium IOL option out-of-pocket as insurance companies do not reimburse for the extra cost, which averages over two thousand dollars per eye in the US [2].

While multifocal and extended depth of focus IOLs (hereafter referred to collectively as multifocal IOLs) may provide a higher chance of acceptably clear spectacle-free vision, they have been reported to be associated with glare, halos, and reduced contrast sensitivity due to their complex optical design, particularly in the earliest generation models [3]. Such side effects can be potentially disabling and sometimes result in patients pursuing a second, riskier, surgery to exchange the implanted intraocular lens [4]. Achievement of improved visual acuity is not always correlated with patient satisfaction with cataract surgery, which rather may relate more to meeting patient expectations surrounding visual functioning [6]. Thus, understanding the expectations and outcomes that matter most to patients is imperative. Given the potentially high patient costs of multifocal lenses, both financial and otherwise, thorough assessment of patient-reported outcomes are important in guiding patients and physicians in their choice of lens implant and may become an important additional endpoint for clinical trials of future generations of IOLs.

MedHelp.org is one of the largest online medical communities where patients may engage in health-related discussions, with over 10.8 million discussions since its inception in 1994. From 2007 to 2014, there was an active eye care "expert" moderated sub-forum operated in collaboration with the American Academy of Ophthalmology (AAO) where members of the community could post questions to be answered by volunteer ophthalmologists. A separate eye care "community" sub-forum is also ongoing. Together, they provide a rich source of unstructured, free-text data on the patient experience with eye health and disease [7].

Natural language processing (NLP) is a rapidly advancing field within biomedical informatics that includes sentiment analysis, a discipline concerned with identifying the sentiment or position of a text towards a particular topic, classified on a scale indicating positive, negative or neutral sentiment. NLP further allows the extraction of structured concepts from free text, such as clinical concepts. With the concurrent increase in computing power and explosion in sources of unstructured text data, NLP and sentiment analysis techniques have been applied to many problems including computing consumer satisfaction metrics, monitoring social media, and forecasting financial markets, but their use is relatively nascent in healthcare. Emerging literature has suggested that patient online social media posts can be well-correlated to more formal measures of patient satisfaction, such as the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey [8]. IBM Watson, one of the most widely-recognized artificial intelligence services, has many NLP capabilities trained to extract concepts, entities, keywords, and sentiment from English language texts and has been used in healthcare to develop decision-support aids in fields as diverse as oncology [9] and radiology [10].

In this study, we sought to apply IBM Watson's NLP and sentiment analysis algorithms towards unstructured text data in MedHelp.org forum posts related to cataract surgery lens options, particularly multifocal IOLs. The goal was to identify the most popular keywords and clinical terms around multifocal IOLs used in patient discussions and their associated sentiment (positive or negative), as a generalized measure of patient satisfaction for each keyword category.

Methods

Data Source and Study Population

MedHelp.org includes an eye care "expert" forum operated by ophthalmologists (2007-2014) and an ongoing eye care "community" forum where patients discuss eye-related issues. Forums contain multiple discussion threads related to a particular topic originated by the initial user post, and each thread consists of multiple posts that are replies made by various individual users to the discussion of that topic.

A keyword search for "multifocal IOL" was performed on November 1, 2016 in the MedHelp.org eye health expert and community sub-forums, yielding the relevant discussion threads in these forums since website inception in 1994. We obtained the full text of the discussion threads, and individual posts of resulting threads were parsed for user, timestamp, and full text. Each user's profile was also analyzed for self-reported age, sex, and location information, if available. Physician profile pages on the site identified physician users, and their posts were excluded from sentiment analysis.

Natural Language Processing and Sentiment Analysis

Sentiment analysis was performed on the full text of each patient post using IBM's Watson proprietary artificial intelligence capabilities via the AlchemyLanguage application program interface (API) [11]. The public model of AlchemyLanguage is trained on English language newspapers and websites to extract concepts, entities, keywords, and sentiment. Each post was given an overall positive or negative sentiment score ranging on a continuous scale from -1 to +1, with the strength of sentiment represented by the magnitude of its numerical value. Neutral sentiments (neither positive nor negative) were assigned a score of 0. AlchemyLanguage also extracted user-generated keywords and phrases from each post and provided an associated sentiment score for each keyword or phrase for that post. Natural language processing was performed to group related keywords by converting to lowercase, removing punctuation, and stopwords (a, and, the, etc.), tokenizing (separating into distinct words, using the Treebank tokenizer [12]), and lemmatizing (transforming each word into its root word, such as making plurals singular, or transforming different forms of the same verb into a root verb, using the WordNet lemmatizer [13]). Keywords and phrases were further processed by string matching to group by clinical concept. Similar keywords and phrases were grouped by considering "IOL" and "lens" to be interchangeable. When IOL manufacturer/brand name were extracted together with the lens model as a keyword, these keywords were aggregated under the name of the lens model. Keywords mentioning "Tecnis Symfony" lenses were grouped into "symfony" whereas mentions of "Tecnis" without "Symfony" were grouped under "tecnis." If key phrases included "Crystalens" and "HD" then they were grouped under "crystalens hd" but if they mentioned "Crystalens" without "HD" they were grouped under "crystalens." All key phrases mentioning "restor", "mplus", "rezoom" were aggregated into groups, regardless of what additional accompanying words were included in the key phrase such as manufacturer (e.g., "Alcon ReSTOR" was grouped under "restor" and "Lentis Mplus" was grouped under "mplus"). If key phrases included both "multifocal" and "toric" they were grouped under "multifocal" but if only "toric" was mentioned then they were grouped under "toric." Key words and phrases associated with contact lenses were disambiguated from IOL lenses. After aggregation of common related keywords, the 250 most commonly identified keywords were examined in greater detail, from which were identified keywords representing multifocal, monofocal, and toric lenses as well as monovision and words that could be related to side effects, complications, or complaints to analyze in greater detail. A flow diagram summarizing the process of online post retrieval, sentiment analysis, and text processing is depicted in Figure 1.

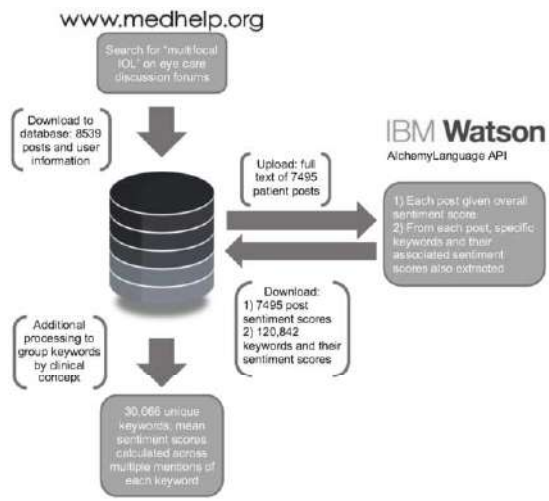


Figure 1 — Flow Diagram for Retrieval of Posts

Statistical Analysis

Mean sentiment scores and their standard deviations were calculated for each keyword across all resulting patient posts. One-way ANOVA was performed to compare mean sentiment scores between different groups of IOLs, which included monofocal, monovision (a strategy whereby each eye is implanted with a monofocal lens but with one eye targeted for clear distance vision and the other eye is targeted for near vision), toric (astigmatism-correcting), and multifocal. For the purposes of defining groups for ANOVA comparisons, the monovision group aggregated keywords "monovision", "minimonovision", "blended vision", and the multifocal group aggregated keywords "multifocal", "premium lens", "restor", "crystalens", "crystalens hd", "symfony", "rezoom", "tecnis", "mplus". Post-hoc Bonferroni comparisons were made to compare sentiment scores between groups.

We performed natural language processing using Python 2.6 (Python Software Foundation, Wilmington, Delaware) and the NLTK package version 3.2.1 (NLTK Project, Philadelphia, PA) [14]. Sentiment analysis was performed using IBM Watson via the AlchemyLanguage application program interface (API) (IBM, Armonk, NY) [11]. Statistical analysis was performed using Stata version 12 (Stata Corp, College Station, TX).

Results

Study Population and Search Results

A search for "multifocal IOL" on MedHelp.org identified 981 threads containing 8539 posts ranging from June 16, 1999 to October 31, 2016. Contributing to these posts were 1488 unique users, of whom 14 were physicians who contributed 1044 posts and 1474 were patients who contributed 7495 posts. Of the patient users, 547 (37.1%) self-identified as male, 536 (36.4%) self-identified as female, and 391 (26.5%) did not specify gender in their profile information. A total of 400 (27.1%) patient users specified an age in their profile information, which ranged from 19 to 87 years, with a mean age of 57.9 years (standard deviation 14.3).

Sentiment Analysis

Overall sentiment scores for patient posts ranged from -0.95 to 0.99, with a mean sentiment score of 0.07 (standard deviation 0.45) across all posts. A total of 120,842 specific keywords and phrases were extracted from all patient posts. Related keywords were grouped by syntax (plurals and singulars, conjugations of the same verb) and clinical concept (for example, "multifocal IOL" grouped with "multifocal lens"), yielding 30,066 unique keywords. Among the top 250 most commonly-occurring keywords were those associated with multifocal lenses and their individual models, standard monofocal lenses or monovision, and keywords that may be associated with complications or complaints (Figure 2).

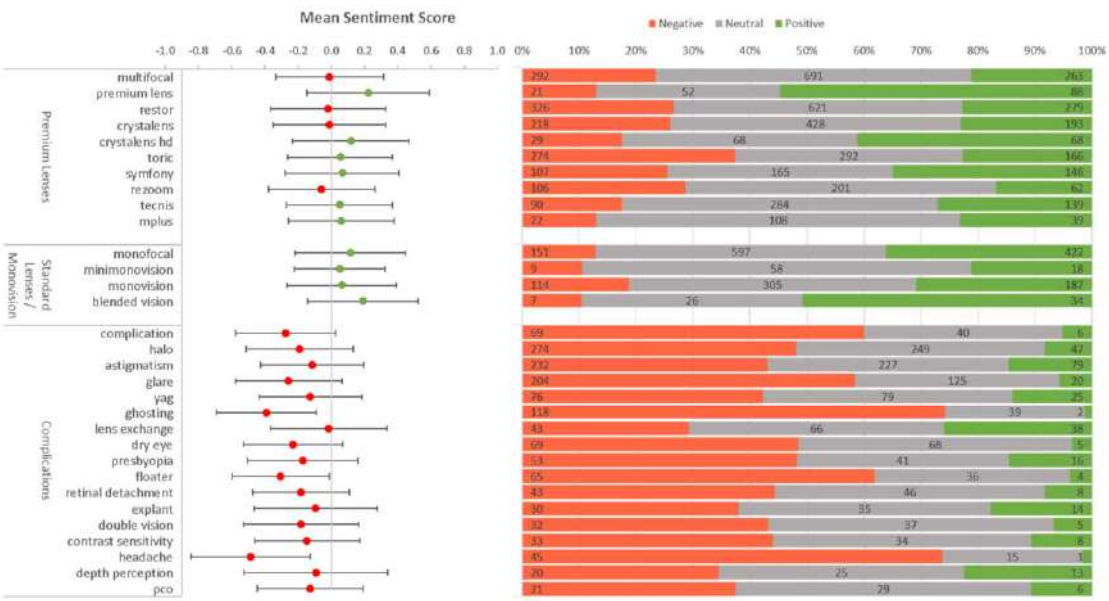


Figure 2 — Sentiment Towards Premium and Standard Cataract Surgery IOL Options and Cataract Surgery Side Effects

(Left) Mean sentiment is shown with standard deviation error bars for keywords related to premium lenses, standard monofocal lenses or monovision, and a variety of words related to cataract surgery complications, complaints, or side effects. Negative and positive mean sentiment are shown in red and green respectively. (Right) Proportion and number of positive, negative, and neutral sentiment instances for each keyword are shown in horizontal bar graph.

Sentiment analyses are displayed in Figure 2. Keywords associated with multifocal lenses (not including specific models) had slightly negative mean sentiment (N=1246, -0.010). Individual models of multifocal or premium lenses were commonly identified as keywords, with those most widely used in the US associated with a slightly negative mean sentiment (ReSTOR N=1226, -0.020; ReZoom N=369, -0.059; Crystalens N=839, -0.012), while newer lenses had more positive sentiment (Crystalens HD N=165, 0.117; Tecnis Symfony N=418, 0.065; Tecnis (non-Symfony) N=513, 0.048). Keywords associated with monofocal lenses had positive mean sentiment (N=1170, 0.11), as did keywords associated with monovision, a strategy whereby each eye is implanted with a monofocal lens but with one eye targeted for clear distance vision and the other eye targeted for near vision (monovision, N=606, 0.06; minimonovision N=85, 0.05; blended vision N=67, 0.19)

A wide range of keywords represented types of side effects or complaints by the patients and were associated with negative mean sentiment. The most common complaints included glare (N=349, -0.257), halo (N=570, -0.190), astigmatism (N=538, -0.115). Other complaints associated with very negative sentiment included ghosting (N=159, -0.392), floater (N=105, -0.305), headache (N=61, -0.485).

There was significant variation in sentiment score between four different IOL implantation strategies: monofocal, monovision, toric, and multifocal (including specific models of multifocal lenses) in ANOVA analysis (F-score [3, 7590]= 35.59, $p<0.001$) (Table 1). Post-hoc Bonferonni multiple comparison tests indicated that when compared to multifocal lenses, monofocal lenses ($p<0.001$), monovision ($p<0.001$), and toric ($p=0.024$) lenses all had significantly higher positive mean sentiment.

Table 1 – ANOVA for Sentiment Scores

Source of variation	df	Sums of Squares	Mean Square	F-Score	p-value
Between Groups*	3	11.98	3.995	35.59	<0.001
Within Groups	7590	851.94	0.1122		

*Monovision group aggregated keywords "monovision", "minimonovision", "blended vision". Multifocal group aggregated keywords "multifocal", "premium lens", "restor", "crystalens", "crystalens hd", "symfony", "rezoom", "tecnis", "mplus."

Discussion

Over 4 million cataract surgeries are performed annually in the US, with premium multifocal intraocular lens implants increasingly used to provide clear vision at multiple focal points. However, potential side effects of these new lenses are a concern, and information on patients' subjective experiences is not readily available. Mining eye-related discussion forums on the MedHelp online health forum, we found patient sentiment towards monofocal lenses was positive, while the sentiment towards multifocal lenses was mixed, but overall slightly negative, especially related to older models such as ReZoom. Furthermore, this approach also identified that newer multifocal lenses, such as Tecnis Symfony, were associated with more positive sentiment. Many side effects associated with multifocal lenses were mentioned in the forums, including glare and halos, as well as ghosting, double vision, and headaches – each associated with negative sentiment. The ability to identify patient concerns with emerging technology is essential to help guide both patients' and clinicians' treatment decisions.

Previous studies have not conclusively demonstrated a benefit in patients' general satisfaction with implantation of multifocal lenses compared to standard monofocal lenses, as overall patient satisfaction following cataract surgery is typically high in both groups [3]. However, studies suggest multifocal lenses are associated with greater reports of postoperative glare, halos, and reduced contrast sensitivity [3]. We found that sentiment towards multifocal lenses was not as positive overall as sentiment towards traditional monofocal lenses. However, when individual models of multifocal lenses were considered, newer lenses such as Symfony and Crystalens HD were associated with more positive sentiments than older lenses. The earliest available multifocal lenses such as ReZoom, Crystalens, and ReStor have been the most extensively studied, and concern for postoperative glare and halos with multifocal lenses is most reflective of these initial studies [3, 5]. As multifocal IOL design has advanced, a few studies have reported that newer IOL models may have fewer visual side effects [15–17] – consistent with our findings that sentiment towards the newer Symfony lens is more positive.

Previous studies have used a variety of questionnaires to measure postoperative satisfaction or visual function, either general visual functioning indices or more often scales developed specifically for a given study, designed to elicit reports of halos or glare [3]. In our analysis of unprompted patient online forum posts, halos and glare were among the most frequently expressed concerns, but we identified many other common concerns with associated negative sentiment including astigmatism, ghosting, dry eye, lens exchange, and double vision. Interestingly, lens exchange was associated with only slightly negative mean sentiment, nearly neutral, suggesting appropriate patient selection and perhaps good visual outcomes following lens exchange surgery. By contrast, mentions of headache, though not exceedingly common, were associated with very strongly negative sentiment.

This study has several limitations. We recognize that our analysis is limited in that it does not account for patient selection or expectations, relying on online forums, which are by nature anonymous, with limited data on users. Participation in online forums may vary by patient demographics, such that represented opinions may be skewed towards younger or more tech-savvy patients. Furthermore, there may also be considerable bias in online discussions, in that patients who are content with surgery may be less likely to post online than highly dissatisfied patients. Posts were limited to 2016 and prior period; however, it represents the dissemination of the key implants. Typographical or auto-correct errors, such as ReSTOR being typed and extracted as restore, were ignored. Users were not always specific regarding which model of lens they were referring to, so multiple models of similarly branded lenses (e.g. ReSTOR and ReSTOR low-add versions) were grouped for sentiment analysis. In addition, sentiment analysis using different algorithms may vary. Although we have utilized IBM Watson as one of the oldest and most recognizable artificial intelligence initiatives, limitations still exist in this approach. We have applied Watson's robust but proprietary general capabilities to the highly specific field of ophthalmology. Future work can improve upon this approach by training algorithms to recognize ophthalmology-specific entities and to distinguish preoperative questions from postoperative concerns. Future work to develop a more ophthalmology-specific model can address these issues.

However, despite these limitations, this study presents a novel application of natural language processing and sentiment analysis techniques to a non-traditional data source—online forum posts—in order to identify insights related to cataract surgery. This approach offers a relatively rapid and low-cost way of identifying many opinions from large groups of people on a particular topic, compared to traditional focus groups, which are labor-intensive and costly to assemble. Using this approach, the overall perception of emerging technology may be quickly surveyed to form an overall community sentiment or to identify unanticipated problems. Automated sentiment analysis may also be useful to supplement the depth and nuance of opinions obtained in focus groups with a broader sweep of opinions from larger populations, or to perform automated sentiment analysis on transcripts of focus group discussions to ensure important concepts are not overlooked. Alternatively, this approach may be utilized to analyze open-ended patient-reported outcomes collected in settings where the population is controlled, such as within the context of a clinical trial, where clinical outcomes are being measured simultaneously. Furthermore, although traditional questionnaires with Likert-scale ratings for specific outcomes allow standardized data analysis, other patient concerns may not be well captured. Using natural language processing to analyze unstructured, open-ended data has the advantage of revealing insights derived from patients' own language, and patient concerns that fall outside the confines of any particular questionnaire. These insights may be used to develop or refine patient-centered outcome measures to be more inclusive of diverse patient concerns, and to better target patient counseling efforts to address the most prevalent or highly impactful patient concerns. Despite the likely presence of some bias in online discussions towards dissatisfied patients, it is still worthwhile to investigate the sentiments of a vocal but potentially unhappy minority as concerns or side effects most discussed online may also be the most impactful when they occur in clinical practice or be the most of interest to preoperative patients who may have come across these concerns while researching their options on the internet. Patients and

surgeons may find these results important in tailoring the choice of lens for the individual patients according to their tolerance of potential side effects.

Conclusions

There is a critical need for evaluating patient-centered outcomes and detecting patients' concerns regarding emerging technology in ophthalmology where rapid surgical innovation with new technologies is transforming care. Internet health forums provide a robust platform for individuals to discuss real-time health concerns and may serve as a resource to identify patient concerns associated with emerging technologies. We demonstrate the use of natural language processing as a powerful tool to gain insight into large amounts of unstructured text data provided by the patient, in this case to understand patient perspectives towards cataract surgery options. We found that sentiment towards monofocal lenses is positive, and sentiment towards multifocal lenses overall slightly negative, though attitudes towards newer multifocal lenses may be more positive. Patients were concerned about both common and some uncommon side effects. Understanding cataract surgery from the patient perspective can be used to improve pre- and postoperative counseling to better address patient-centered concerns and to develop measures of patient-centered outcomes in the future. This study serves as an example of utilizing cutting-edge technology to understand healthcare attitudes and outcomes from the patient perspective and can be applied to many different areas of healthcare and sources of text.

References

[1] Schein OD, Cassard SD, Tielsch JM, Gower EW (2012) Cataract surgery among Medicare beneficiaries. Ophthalmic Epidemiol 19, 257–264.

[2] Segre L (2017) How much does cataract surgery cost?, http://www.allaboutvision.com/conditions/cataract-surgery-cost.htm, Last updated September 5, 2017, Accessed October 24, 2017.

[3] Wang SY, Stem MS, Oren G, Shtein R, Lichter PR (2017) Patient-centered and visual quality outcomes of premium cataract surgery: a systematic review. Eur J Ophthalmol 27, 387–401.

[4] Kamiya K, Hayashi K, Shimizu K, Negishi K, Sato M, Bissen-Miyajima H (2014) Multifocal intraocular lens explantation: a case series of 50 eyes. Am J Ophthalmol 158, 215-220.e1.

[5] Silva SR de, Evans JR, Kirthi V, Ziaei M, Leyland M (2016) Multifocal versus monofocal intraocular lenses after cataract extraction. Cochrane Database Syst Rev 12, CD003169.

[6] Pager CK (2004) Expectations and outcomes in cataract surgery: a prospective test of 2 models of satisfaction. Arch Ophthalmol 122, 1788–1792.

[7] MedHelp (1994-2017), www.medhelp.org.

[8] Huppertz JW, Otto P (2017) Predicting HCAHPS scores from hospitals' social media pages: A sentiment analysis. Health Care Manage Rev.

[9] Patel NM, Michelini VV, Snell JM, Balu S, Hoyle AP, Parker JS, Hayward MC, Eberhard DA, Salazar AH, McNeillie P, Xu J, Huettner CS, Koyama T, Utro F, Rhrissorrakrai K, Norel R, Bilal E, Royyuru A, Parida L, Earp HS, Grilley Olson JE, Hayes DN, Harvey SJ, Sharpless NE, Kim WY (2017) Enhancing next generation sequencing-guided cancer care through cognitive computing. The Oncologist 22, 1-7.

[10] Trivedi H, Mesterhazy J, Laguna B, Vu T, Sohn JH (2017) Automatic determination of the need for intravenous contrast in musculoskeletal MRI examinations using IBM Watson's natural language processing algorithm. J Digit Imaging.

[11] About AlchemyLanguage (2017), https://console.bluemix.net/docs/services/alchemy-language/index.html#index, Last updated July 19, 2017, Accessed October 26, 2017.

[12] MacIntyre R (1995) Treebank tokenization, ftp://ftp.cis.upenn.edu/pub/treebank/public_html/tokenization.html.

[13] Princeton University (2010) "About Wordnet", http://wordnet.princeton.edu, Accessed October 24, 2017.

[14] Bird S, Loper E, Klein E (2009) *Natural Language Processing with Python,* O'Reilly Media Inc.

[15] Kim JS, Jung JW, Lee JM, Seo KY, Kim EK, Kim T-I (2015) Clinical outcomes following implantation of diffractive multifocal intraocular lenses with varying add powers. Am J Ophthalmol 160, 702-9.e1.

[16] Pedrotti E, Bruni E, Bonacci E, Badalamenti R, Mastropasqua R, Marchini G (2016) Comparative analysis of the clinical outcomes with a monofocal and an extended range of vision intraocular lens. J Refract Surg 32, 436–442.

[17] Cochener B (2016) Clinical outcomes of a new extended range of vision intraocular lens: International Multicenter Concerto Study. J Cataract Refract Surg 42, 1268–1275.

Address for correspondence
Sophia Wang, MD, 2452 Watson Ct, Palo Alto, CA 94303
Email: sywang@stanford.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190454

Demand Analysis and Function Design of Health Decision Support System in China

Tao Dai[a,b], Liqin Xie [a], Yan Wang[a], Yanli Wan[a] and Hongpu Hu [a]

[a] *Institute of Medical Information, Chinese Academy of Medical Sciences/Peking Union Medical College, Beijing 100020, China,*
[b] *Corresponding author e-mail: dai.tao@imicams.ac.cn*

Abstract

In order to improve the level of health decision-making, based on health information resources and decision support function types, this study summarized five core functions of health decision support system: information support, monitoring and early warning, analysis and evaluation, trend prediction, comprehensive optimization; And from the perspective of business functions, business processes and business activities of business domains, the demand of Health Decision Support System is refined in to six parts, such as public health, medical care, drug management, medical insurance, comprehensive management, grass-roots health. On this basis, the overall design of the system is carried out.

Keywords:

Software Design, Clinical Decision Support, Expert Systems,

Introduction

Health Decision Support System (HDSS) is an intelligent human-machine interactive information system which is oriented to semi-structured and unstructured decision in the field of medical and health care [1]. It uses the relevant theories and technologies of decision support systems to support decision-making activities of medical personnel [2]. HDSS can assist health managers in scientifically formulating health policies, rationally deploying health resources. It can help doctors to make accurate diagnoses, select effective treatment options. And it can provide assistance to public health, medical insurance, and other related workers [3]. Constructing a health decision support system and building a health informationization development model oriented to information utilization and service decision-making which have important practical significance for improving scientific management level provide practical tools for health decision-making.

This study aims to improve the level of health decision. Based on health information resources, according to the field division of health services, we analyze the needs of support systems in the areas of public health, medical care, drug management, medical insurance, comprehensive management, and grass-roots health. According to the different decision needs, the health decision system that implements the core functions of the above five aspects, such as information support, monitoring and early warning, analysis and evaluation, trend prediction, comprehensive optimization, is designed.

International Experiences

Typical applications of HDSS

Some of developed countries have already widely applied information technology to the medical and health care, formulated strategic plans to promote the use and exchange of health information, and developed many practical decision support systems for medical services, hospital management and public health management, which have achieved remarkable results [4]. Typical applications of HDSS are as follows. First, developed medical expert system for clinical medical practice, such as MYCIN in the USA, Quick Medical Reference and the Map in the NHS in UK [5]. Second, Developed hospital management decision support system to improve hospital management, such as Distributed Hospital Computer Program in the USA [6,7]. Third, developed intelligent drug prescription system to promote rational drug use in clinic, such as Computerized Physician Order Entry and Drug Therapy Screening System in the U.S.A, as well as Eprescribe in Australian. Forth, developed public health management decision support system to enhance disease prevention and control and emergency response capacity of public health emergencies, such as Panorama in Canada and National Electronic Disease Surveillance System in the USA [8].

Enlightenment to China

Compared with some of developed countries, the application of HDSS is not yet widespread in China. It is better to choose the areas where business needs are urgent and decision rules are relatively clear, and then develop HDSS for a specific topic or field. Therefore, the following will start with the needs of the main stakeholders for HDSS. On this basis, we try to preliminarily build the logical architecture of HDSS.

Demand Analysis

The main task of the demand analysis of the health decision support system is to solve the problem of "what to do" in the health decision support system which is the foundation for the design of health decision support system. It mainly includes user analysis, demand type analysis and demand description of the health decision system.

User Analysis

The users of the health decision support system have expanded from the health management department to the fields outside the health management department, from the fields of treatment and prevention and control to the fields of education and commerce, from the biological and physiological fields to the psychological field, and from technical activities to social activities. It forms into a comprehensive service. Users can be divided into the following categories: health administrative staff, medical staff, medical researchers, and the public, teaching staff, students and pharmaceutical marketers.

Health Administrative Staff

The issues they care about are decision, organization, lead, coordination, and control in management. The characteristics of the information needs are systematic and comprehensive.

The information, which is also required universality and accuracy, provides basis to make decision for policy-related, predictive and strategic work.

Medical Staff

Medical staff work in prevention, clinical frontline. Their mainly needs information service in daily work. They are the main target of health information which is mainly for medical staff to provide support in the diagnosis, treatment and rehabilitation of diseases.

Medical Researchers

The main work of medical researchers is to explore the special development policy in various fields of medicine, as well as to carry out knowledge innovation and expand knowledge system. Researchers always want to know the latest developments in related topics at home and abroad. And then they can adjust their research to improve the novelty of their own scientific research. Scientific research work is inherited and continuous, and there are inherent connections between the subjects, so their demand for information is continuous. Moreover, researchers prefer information types such as books, journals, and academic conference papers.

The Public

The public includes patients and their family members, as well as the general population. Their information needs are generally directed at the cause, treatment, prognosis and health of a disease. Health information services to patients, the person who is in a special situation or who is concerned about his or her health. It provides them health information about their illness or health status.

Teaching Staff and Students

Their work is mainly to teach students basic medical knowledge, and some also undertake some scientific research work. Students refer to people at all levels of medical colleges. The information sources they contact are mainly networks and libraries. Their information needs are relatively simple compared to the above-mentioned users, but because students have the characteristics of broad interest of young people, the information they need, in addition to the relevant knowledge of the major they are studying, also covers a wide range of subjects.

Pharmaceutical Marketers

Pharmaceutical marketers mainly refer to business operators who are engaged in market expansion, product sales, customer support, etc. of medical equipment and medicines in market economic activities. Marketers must not only grasp the various information of the products, customer information, but also the situation of partners and competitors, as well as the policy and economic situation of the place of sale. They often access information through informal channels when information services are unable to provide timely, novel, accurate, and reliable information.

Demand Types

In the perspective of the type of decision support function, the five core functions of the health decision support system are summarized.

Information Support

Information support functions are divided into two categories: external information support and internal information support. External information support is mainly to provide health administrative staff with information outside the field needed to assist decision-making. Internal information services provide decision makers with an overall level and dynamics of change of health development, such as health development status, infrastructure conditions, etc.

Monitoring and Early Warning

Through the extraction, integration, monitoring and analysis of relevant indicators in the fields of public health services, medical services, medical insurance, etc., we can timely grasp the operational dynamics of the business and existing problems, and identify the crux of the problem. For example, by analysing the rules of relation between drugs and diseases, forming a prescription model, implementing dynamic monitoring of clinical drugs, drug disputes will reduce, medical quality will improve. Through real-time monitoring of diseases, health hazards, and public health emergencies, we can keep abreast of the development of various diseases and public health emergencies to establish a sound mechanism composed of monitoring, assessment, early warning, and response.

Analysis and Evaluation

Applying appropriate evaluation methods and scientific and reasonable evaluation index system to conduct objective and scientific evaluation and analysis of health industry dynamic monitoring data, defines the actual operational effects of health-related policies, provides support for the formulation of new decisions. For example, through the use of scientific and reasonable evaluation index system, comprehensive analysis and evaluation of the operation of the fund, compensation for expenses, the benefit of the insured personnel, the utilization of medical services, and the flow of medical treatment support managers to develop more scientific and rational medical insurance related programs and systems.

Trend Prediction

Prediction is one of the important means for the health industry to make macro-decisions. It determines strategic policies and means by selecting the influencing factors that have a significant contribution to the forecasting target. It uses highly efficient tools to regularly sample health data and import it into pre-defined intelligent predictive models to predict future developments to support health service management at the regional, provincial or national level. Prediction function can be divided into three cases: predicting future operational trends based on current operating conditions; adjusting some parameter values of the operating system according to the policy, and using the model to predict the future operating status; taking corrective actions and predicting the effectiveness of it when deviations are found in the operation. For example, by analysing the frequency of use of a certain drug in a certain period of time, predicting the demand for a period of time in the future, and optimizing the allocation of drug stocks, the drug supply system can operate benign and efficient.

Comprehensive Optimization

The goal of macro decision is overall optimization, not the local optimum of a certain aspect. Therefore, through comprehensive consideration and analysis of various factors, the health decision support system explores the potential relationships among various links and coordinates them to achieve overall optimization and comprehensive optimization. For example, when dealing with public health emergencies, it is necessary to integrate dynamic monitoring information, emergency resource information, hazard information, various existing typical cases and related plans, as well as relevant knowledge and optimization models in the expert database. Then optimal contingency plans such as personnel evacuation, blood call, emergency resource optimization configuration, etc. can be generate dynamically.

Demand Description

According to the division of users and business areas of the health decision support system, decision-making needs are mainly concentrated in the fields of public health, medical care,

drug management, medical security, comprehensive management, and primary health. Meanwhile, based on the business functions, business processes and business activities of the business domain, the demands for the health decision support system are refined as follows(Table 1):

Table 1– Health Decision Support System Demands Description

Business Demands Analysis	Business Module	Business Refinement
Public Health	Prevention and Control	Infectious Disease Surveillance
		Chronic Disease Surveillance
		Occupational Health Management
		Health Hazard monitoring and Control
	Health Supervision	Health Administrative Licensing and Registration
		Health Supervision and Punishment
	Woman and Child Health Care	Woman Health Care
		Pregnant Woman and Puerpera Health Care
		Child Health Care
		Family Planning
	Command of Emergency Public Health Emergency	Emergency Monitoring
		Emergency Resource Management
		Hazard Information Acquisition and Management
		Emergency Plan Generation and Management
Medical Care	Clinical Medical Service	Disease Management
		Patient
		Surgery
		Treatment
		Clinical Outcome
	Medical Management Service	Medical Service Volume
		Medical Service Quality
		Medical Security
		Medical Cost Control
		Drug Service
Medical Insurance	Basic Medical Insurance of Urban Workers and Residents	Insured Situation
		Medical Cost Control and Prediction
		Medical Insurance Fund Analysis and Supervision
	New Rural Cooperative Medical Scheme	Participation
		Disease Incidence
		Fund Preparation and Availability
		Fund Allocation and Use
		Participant Benefit
		Participation In Medical Expenses and Medical Expenses Control
		Medical Treatment Analysis
		Cost Compensation
		Participation in the Disease Burden
		Urban and Rural Medical Assistance Beneficiaries
	Urban and Rural Medical Assistance	Disease Incidence
		Urban and Rural Medical Assistance Expenditure
Drug Management	Drug Plan	
	Drug Purchase	
	Drug Use	
	Drug Administration	
Comprehensive Management	Health Plans, Guidelines and Policies	Realization
		Classification Statistics
		Impact
	Health Resource Management	Health Institution
		Health Human Resources
		Health Facilities
		Health Expenditure
		Comprehensive Management of Various Types of Health Resources
	Health Supervision	Public Health Supervision
		Medical Service Supervision
		Medical Security Supervision
		Drug Supervision
Grass-Roots Health	Medical Management	
	Resident Health Management	
	Prevention and Health Care	

Public Health

The purpose of public health is to organize social efforts to improve environmental sanitation, to prevent and control epidemics of infectious diseases and other diseases, to cultivate good health habits and civilized lifestyles, to provide medical services, to prevent diseases, and then to promote people's health. Public health institutions defined by the new medical reform include health administrative agencies at all levels, disease control agencies, health supervision agencies, maternal and child health institutions, chronic disease prevention agencies, and public health research institutions. The public health business is more extensive, mainly including: prevention and control, health supervision, maternal and child health care, emergency public health emergency command, etc., so it is divided into several corresponding topics for analysis [9].

Medical Service

Medical services include services for diagnosis, treatment, epidemic prevention, delivery, family planning, and related services such as providing medicines, medical equipment, ward accommodation, and meals. The institutions providing medical services are called medical institutions, including hospitals, health centers, community health service centers (stations), nursing homes, outpatient departments, village clinics, maternal and child health centers, and specialized disease prevention and treatment institutions. Demand analysis is divided into two topics: clinical medical services and medical management services.

Medical Insurance

China's basic medical security system consists of urban workers' basic medical insurance, urban residents' basic medical insurance, new rural cooperative medical scheme, and urban and rural medical assistance systems. The core business of medical insurance is insurance object management,

insurance fee payment management, reimbursement management, and settlement management with medical service utilization units.

The reform of the medical insurance system involves both government responsibility and the interests of the insured units, insured personnel and medical institutions. The medical insurance decision-making is to evaluate the problems in the process of income, to distribution and use of the medical insurance fund in real time, to analyze and master the operation dynamics of the medical insurance fund in time, find out the crux of the problem, and to make scientific decision-making and policy adjustment.

Drug Management

Drug management mainly focuses on the supervision and management of drug uses, including centralized bidding and procurement of medical products of public medical institutions, supervising and managing the use of pharmaceutical preparations and drugs, guiding medical institutions to rationally use clinical drugs, organizing the implementation of national drug policies and national essential drug systems, organizing the handling of adverse drug events in medical institutions, and formulating recommended use lists for drugs, etc.

Drug management mainly includes three major processes, which is procurement, use, and drug administration. It involves many aspects such as drug plan management, drug quality management, drug information management, clinical rational drug management, prescription drug query management, etc. [10]. Due to the variety of drugs, new products, multiple purchase channels, and many complicated and complicated management links, there are often backlogs and shortages of drugs, which makes it difficult for managers to make more effective decisions. Therefore, the development of drug management decision support system is an urgent need for drug management. Drug management will be divided into four aspects: drug planning, drug procurement, drug use, and drug administration.

Comprehensive Management

The main body of comprehensive management is the National Health and Family Planning Commission and the provincial health and family planning administrative departments. The main task of health management decision-making is the corresponding health management decision services carried out by the internal organs of the medical and disease, disease prevention and control center and grassroots health management set up by the health and family planning departments at all levels.

The most basic information query demands of the comprehensive manager mainly include : completing the inquiry of the business indicators and information of each medical institution and various departments within the organization; timely getting the implementation of various health plans, guidelines and policies; timely understanding of the planning and allocation of health resources [11]; timely understanding and getting public health supervision information; timely getting major epidemic information; real-time getting business report data of various medical institutions, keeping abreast of medical quality information and the progress of major business. These provide first-hand information and scientific basis for leadership decision-making. In view of the main responsibilities of the health management department, the specific decision-making and related information demands will be explained in three aspects: guiding health planning guidelines and policies, coordinating health resources and health business supervision.

Grass-Roots Health

The grassroots health service institutions generally refer to the county, township and village level medical institutions, including county-level people's hospitals, community health service centers, township health centers, and village health centers (rooms). It not only provides basic medical services to the residents of the agency's service radiation area, but also undertakes basic public health services such as preventive health care and health education promotion. Its diagnosis and treatment subjects, number of beds, department settings, staffing, infrastructure construction and equipment are adapted to its functional positioning 12]. With the deepening of the reform of the medical and health system, grassroots health has placed greater expectations from synergistic prevention, diagnosis and rehabilitation to a sound grading diagnosis and treatment mechanism. The support of the grassroots health information system has also become an important factor in promoting the development of primary health care services. The functional demands of the grassroots health information system have risen from simple input and query to a broader level to meet the decision-making demands of different levels. According to the business function, the grassroots health decision demands analysis project is divided into three aspects: diagnosis and treatment management, resident health management, prevention and health care.

Overall System Design

The task of demand analysis is to solve the “what” problem of the health decision support system, while the task of system design mainly solves the problem of “how to do”. The design idea is, through making full use of existing decision-making resources such as models, methods, data, and knowledge and combining health decision theory with the experience and subjective wishes of health decision makers, to assist health decision makers with creative thinking, logical reasoning and judgment, help health decision makers clarify health decision-making goals and identify problems, provide various options to health decision makers, evaluate and optimize various options by analyzing, comparing and judging through human-computer interaction functions, provide effective support for health decision makers to make the right health decisions [13].

The overall logical architecture of the health decision support system [14] consists of three levels (Figure 1): 1) Data layer define and maintain the integrity and security of health data, and respond to logical layer requests, access data. This layer reads, extracts, cleans, converts, and summarizes data from hospitals, CDC, health supervision, maternal and child health care, blood centers, communities, medical insurance institutions, and health administrative departments at all levels, etc. through the data import program, and completes the data loading of the data warehouse subtopic. At the same time, the models, methods and expert knowledge of the manual decision-making process are obtained through the interaction with experts in the field of health. And they are classified, organized and managed by method library, model library and knowledge base. And the data, model, method and knowledge resources are called through the corresponding management system; 2) The application layer acts as a bridge between the presentation layer and the data layer. Its role is to respond to the decision-making requests of health decision-makers, to perform decision analysis tasks and to capture corresponding data from the data layer, to analye the corresponding DSS analysis tools such as online analytical processing, data mining, etc. and to get the decision information obtained to the presentation layer. According to different health decision-making needs, the application layer implements functions such as information

support, monitoring and early warning, analysis and evaluation, trend prediction and comprehensive optimization. 3) The presentation layer provides a visual interface for health decision makers, that is, decision-making users can input decision-making demands or obtain decision information through the presentation layer.

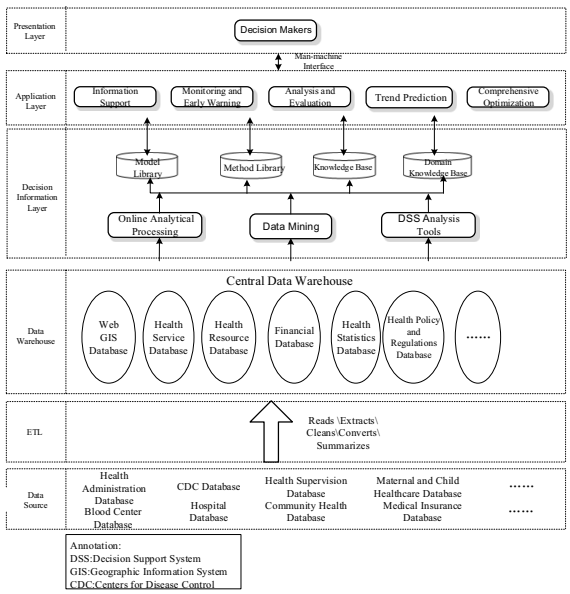


Figure 1–Overall Logical Architecture of HDSS

Acknowledgements

This work was financially supported by National Natural Science Foundation of China (71273280), National Social Science Fund (17BGL184).

References

1. T. Poirier and R. Giudici, Drug interaction microcomputer software evaluation: Drug Therapy Screening System (DTSS), J Hospital Pharmacy **25** (1990), 738-741, 744
2. Dai Tao. Medical insurance new rural cooperative medical system database. Chinese Journal of Health Policy, 2016, **9**(12):1-5.
3. Ma Hao. Health Management decision support system model building research as example of serious illnesses prediction model building in new rural cooperative medical scheme. Peking Union Medical College, 2015.
4. Dai Tao. International Experiences of Health Decision Support System Development. Chinese Journal of Evidence-Based Medicine, 2012, **12** (03):247-250.
5. The Map in the NHS. Http://www.mapofmedicine.com/mapinnhs/mapinnhs/. Department of Health and Ageing. Consultancy inElectronic Prescribing and Dispensing of Medicines (ePrescribing).2008, 1.
6. Graber SE, Seneker JA, Stahl AA, et al. Development of a replicated database of DHCP data for evaluation of drug use. J Am Med Inform Assoc, 1996, **3** (2): 149-156.
7. Lloyd SS. Automated information systems provide health information management support to veterans' health care. J AHIMA, 1992, **63** (6): 63-67.
8. Panorama application overview. Pan-Canadian Public Health Communicable Disease Surveillance and Management Project. 2010-12-25. Http://www-03.ibm.com/indusries/ca/en/healthcare/files/panorama_application_overview_fi nal.pdf.
9. Wang Chuangxin, Zhao Yun, Li Zhen. Design and application of the network report system of neonatal vaccination. China Digital Medicine, 2011, **06** (3):41-43.
10. Ge Menghua, Chen Guohua, Yu lizhang. Drug purchase and supply mode innovation based on the information technology. Chinese Hospitals, 2013, **17** (12):79-80.
11. The white paper on the solution of intelligent decision support platform for regional health management [EB/OL]. http://www.minesage.com/Solution/HealthCare/ [2011-3-24].
12. Zhang Yue, Gao Xing, Chen Quan, Hu Hongpu. Consumer requirement and problem analysis of grass-roots health information system. China Medical Herald, 2016, **13** (21):90-94.
13. Huang Longjun, Dai Liping, etc. Research and implementation of interactive decision supporting system for social insurance. Computer Applications and Software, 2009, 9.
14. Wang Chunrong, Li Daoping. The requirements research of health management decision support system based on regional imformation platform. China Digital Medicine, 2010, **05** (7):42-44.

Address for correspondence

Institute of Medical Information, Chinese Academy of Medical Sciences/Peking Union Medical College, Beijing 100020, China.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190455

Analysis and Measurement of China's Population Health Informatization Development Strategy

Hongpu Hu [a], Liqin Xie [a], Qingkun Chen[a], Xing Gao[a], Yuqi Pi[a] and Tao Dai[a,b]

[a] *Institute of Medical Information, Chinese Academy of Medical Sciences/Peking Union Medical College, Beijing 100020, China,*
[b] *Corresponding author e-mail: xie.liqin@imicams.ac.cn*

Abstract

This paper analyzes the development strategy of population health informatization in China, and summarizes the measurement direction and evaluation elements of population health informatization. ***Methods:*** *Literature and field investigation, expert consultation and PEST analysis were used to determine the development level measurement and evaluation framework.* ***Results:*** *Based on the PEST analysis framework, the development level measurement and evaluation factors were defined, and the evaluation framework was established, which included system construction, IT application, information financing, information personnel, information policy and management, and information application effect. The information from hospitals and grass-roots medical and health institutions was also provided. From the perspective of the level of development, the framework of informatization evaluation is further refined.*

Keywords:

Health Information Management, Population Health, China

Introduction

Population health informatization is a general designation of the process that the organizations in health and family planning system use modern network and computer technology to collect, organize, store, use and provide services for population health information, further which rationally organizes and controls the information activities and factors in the field of health and family planning in order to realize the rational allocation of information and related resources so as to meet the information service and management needs of the health and family planning industry[1]. At present, the construction of population health informatization in China has been greatly developed, and will play a more important role in the development of health and family planning in the future [2]. The planning, input, policy and output of population health informatization should be supported by appropriate evaluation theory and practice [3, 4]. The evaluation theory of population health informatization in China is not feasible, and there is only a few evaluation system applied to practicing and guiding informatization construction [5, 6]. Especially with the government as the main body of evaluation, there are few studies on the evaluation of the development level of regional population health informatization [7]. This study combines the actual and future situation of population health informatization development, comprehensively applies multiple scientific research methods, analyzes the development strategy of population health information in China, and summarizes the measurement direction and evaluation factors of population health informatization to provide reference for the measurement of population health information development level in China and different provinces.

Methods

Research object

This paper determined the weight of the indicator system by the analytic hierarchy process. This study focuses on the development level of population health informatization in the province and below, including the basis of population health information, information resources, technology application, talents, financing and policies, planning, standards, etc. The above are the core elements of the regional population health information construction and the key links involved in the implementation process, as well as the application effects.

Research methods

Literature research

Through systematicly reviewing Wanfang, Tsinghua Tongfang, PubMed, OVID and other domestic and foreign full-text databases, relevant institutions' websites and other network resources, we collect domestic and foreign informationization evaluation theories, indicators and methods, including regional informationization, enterprise informationization, and education informationization, health information, e-health, information systems, etc. we collect and sort out foreign population health assessment frameworks and indicator systems, providing reference for China's population health informationization evaluation framework.

Site investigation

This study conduct on-the-spot investigations on some provinces with more developed population health informatization like Beijing, Jiangsu, Zhejiang and Shanghai. Meanwhile, this paper researches on the development status of informatization, review relevant data, and conduct interviews with insiders. The interviewees mainly include the head of the health administrative department, the person in charge of the information center, and the medical staff at the grassroots level.

Expert advice

In this study, 16 experts were selected to determine the comprehensive evaluation index of population health informationization construction level through letter or expert consultation. Expert selection criteria: leaders of health and family planning administrations, experts and scholars in the field of population health information management, and researchers and business personnel in other related fields; leaders with intermediate or intermediate titles or above; work there in professional field for more than 5 years; be interested in this research and willing to cooperate with expert consultation. Main consultation content: guide and consult on

project design and implementation such as value judgment criteria, determination of evaluation framework and selection of evaluation indicators, weight calculation of indicators, and recommendations for research reports, etc.

PEST analysis

PEST analysis is often used to analyze the background of a company, which evaluates the strategic formulation of a company from the four aspects of politics, economics, society, and technology. This method is increasingly applied to various research fields. This study uses PEST analysis to analyze population health information from four environments: policy (P), economy (E), social culture (W) and technology (T). It will help to grasp the development trend of population health informationization, clarify current obstacles and problems, and propose future development strategies.

Delphi expert consultation

After two rounds of Delphi expert consultation, experts were asked to score the importance of primary indicators. In the third round, experts were invited to measure the weight and combination weight of each index according to the steps of analytic hierarchy process. And then comprehensive index method were used to construct and improve the evaluation framework and index system of population health informatization.

Results

Analysis of population health informationization development strategy

PEST analysis of development strategy

As an important support for the development of health cause, population health informationization is also the inevitable application result of information technology in the field of health care. Its own construction and development is affected by many factors, such as politics, economy, culture, society and technology. Analysis of its influencing factors will help to grasp the development trend of population health informationization, clarify current obstacles and problems, and propose future development strategies. This section analyzes population health informationization from the four environments of policy (P), economy (E), social culture (S) and technology (T) by PEST analysis.

(1) Policy environment: In the field of health care, health departments and other ministries and commissions issued policy documents one after another, or stressed the importance of health information construction, or directly encouraged the construction of health information. From the perspective of policies and regulations at the national level, the health sector has guided, encouraged and regulated population health informationization from the aspects of strategic planning, system construction, and system function norms. In addition to the policies promulgated by the health department, the State Council has also promoted the construction and application of information technology in the field of health care from the perspectives of "big data", "internet+" and "information benefiting the people" [8].

(2) Economic environment: In 2008, the total cost of health in China was 1,435.54 billion yuan, three times that of 2000. The total health expenditure in China in 2016 was 4,414.5 billion yuan, nearly three times that of 2008.Since the new century, the growth rate of total health expenditure has been higher than that of GDP growth, and the proportion of the former in the latter has also increased year by year. It is worth mentioning that the proportion of government health expenditure to total health expenditure has increased from less than 10% to more than 30% in 2016, and the proportion of personal health expenditure has declined year after year. These conditions have put forward new requirements for health care industry to expand supply, optimize service, improve quality and improve management level. However, the reality is that the supply of health care services in China is generally inadequate, and the distribution of quality medical resources is uneven, while the training and establishment of medical institutions is relatively slow. In this case, the use of information technology to extend services, improve efficiency, and improve quality is a very reasonable choice. Corresponding to the rapid growth of the medical industry, China's information industry has also grown rapidly.

(3) Social and cultural environment: In 2014, China's elderly population aged 60 and over reached 212 million, accounting for 15.5% of the total population. This size is comparable to the current population of the three major European countries (Germany, France, and the United Kingdom). Scholars predict that the peak population will appear in the 2030s and 2040s, and the aging rate will accelerate significantly in the next three or four decades. Especially until around 2040, the proportion of the elderly over 65 years old in China will exceed 20%, and this proportion will continue to increase by 2050, reaching 20%-24%. The changes of disease spectrum, health demand and dependency ratio brought about by population aging will bring challenges to the future health care service supply. The rigidity of demand for chronic disease management, remote monitoring, health file management, and combination of medical care and maintenance will become more and more strong, and will become an important development direction of population health informatization [9].

(4)Technical environment: Benefiting from the rapid development and application of information technology, population health information technology has also undergone different technical stages. Technologies such as the Internet of Things and cloud computing have stepped onto the stage, and today the society has entered the era of big data. Telemedicine, teleconsultation, cloud hospital, cloud medical are constantly developing to realize convenient and efficient technical exchanges and cooperation in different time and space. The connotation of medical services is very broad, and it needs to rely on high-quality equipment, a wealth of sample banks and doctors to meet the needs of patients from prevention, medical care, health care to rehabilitation, as well as the experience of patients. "Internet + medical" has become an important breakthrough in the development of hospitals. "Wearable mobile medical devices" have had an important impact on people's lives and work [10].

Analysis of development measures and evaluation factors

Through PEST analysis and strategic analysis of the future development of population health information in China, this study considers that the development level of population health informatization involves multiple dimensions such as information technology utilization and its own business development. Considering the multiple dimensions involved in the development level of population health information, the evaluation elements should include seven aspects: information resource construction, China Unicom and information technology application, information financing, informationized talents, information policy and management, and information application effects.

Improve the measurement index system of the population health information development level

Through PEST analysis and evaluation factors of China's population health information development strategy, the evaluation framework of population health information

development level is formulated, and the measurement index system of development level is further improved according to the optimization strategy. Based on the principle of "measurable and comparable", the measurement index system of population health informationization development level is revised and improved through the method of expert consultation. The evaluation index system optimization focus: continue to optimize the adjustment index weight setting; improve the score differentiation of some indicators; appropriately update the indicator content; simplify some indicators calculation formula. Finally, a three-level structure model including seven primary indicators, 16 secondary indicators and 38 tertiary indicators is established, which includes system construction, unicom and information technology application, information financing, information personnel, information policy and management, and information application effect. Defining the evaluation subject: the competent department of population health informationization; the evaluation target: the level of regional population health informationization development, including the software and hardware level, technology application and output from the perspective of information technology, and the coverage and protection of the population from the perspective of population health development level and information resource sharing.

Table 1–Measurement and evaluation system of population health informationization development level.

Primary indicator	Secondary indicators	Three-level indicator
1 Information resource construction	1.1 Construction and application of electronic medical record	1.1.1 Construction and coverage of electronic medical record database in the region
		1.1.2 Medical institution electronic case database usage rate
		1.1.3 Application of regional electronic medical record database in connection with other business systems
	1.2 Construction and use of resident electronic health records	1.2.1 Electronic health record data update rate
		1.2.2 Regional Electronic Health Archive Database Connection Range
		1.2.3 Business scope covered by the regional electronic health record database
		1.2.4 Application of regional electronic health record database in connection with other business systems
	1.3 Construction and use of the information database for the entire population	1.3.1 Whether the entire population information database is fully covered
		1.3.2 Frequency of update of data of the entire population information base
		1.3.3 Application of full population information database in connection with other business systems
2 Infrastructure and information system construction	2.1 Computer network facilities construction	2.1.1 Business scope covered by the health information private network
		2.1.2 Community Health Service Center / Township Health Center Network Bandwidth
	2.2 Information System Construction	2.2.1 Provincial business information system construction rate
		2.2.2 Use and availability of basic business information system functions
		2.2.3 Construction of health care big data application system
	2.3 Information Security	2.3.1 Machine room security level
		2.3.2 Information system information security and other security levels
3 Unicom and information technology applications	3.1 Unicom and data sharing	3.1.1 Repeated entry of information in primary medical institutions
		3.1.2 Communication between business information system and regional information platform
		3.1.3 Sharing health and medical data with related fields
	3.2 Internet + Medical Services	3.2.1 Proportion of open registration and inspection results of medical service institutions
		3.2.2 Proportion of medical institutions that carry out telemedicine through information systems
		3.2.3 Proportion of medical institutions that conduct two-way referrals through information systems
4 information financing	4.1 Investment scale	4.1.1 Ratio of construction funds and operation and maintenance funds to total health expenditure at the same time

	4.2 Informatization financing sustainability	4.2.1 Construction funds and operation and maintenance funds are included in the regular budget
		4.2.2 Social capital participation in information construction
5 informational talents	5.1 Informational Talents assign	5.1.1 Proportion of informationized personnel in the proportion of health technicians
	5.2 Informational Talent Structure	5.2.1 Proportion of persons with master's degree or above and intermediate titles
		5.2.2 The proportion of composite information personnel
6Information Policy and Management	6.1 policy and planning formulating	6.1.1 policy and planning implementation
	6.2 Construction and management of Information Construction Project	6.2.1 Pre-project evaluation
		6.2.2 Post-event supervision of the project
7 information application effect	7.1 National Health Information Platform Application Effect	7.1.1 Integration of information and data resources
		7.1.2 Key business synergy efficiency
		7.1.3 Data statistics and performance appraisal
	7.2 Health and medical big data application effects	7.2.1 Accurate evaluation of health data
		7.2.2 Application of supplementary medical insurance control fees
		7.2.3 Public health decision-making management capabilities

Multidimensional refinement of the evaluation framework of population health informatization development level

Although the evaluation framework of the development level of population health informationization gives a quantitative analysis to the regional population health informationization on the whole, it lacks effective research and analysis in the sub-segments of population health informationization, such as health and family planning administrative agencies, public hospitals, primary health care institutions, disease prevention and control centers [11]. For example, in the course of the investigation, the overall score of a certain area is relatively high, but the informationization of the primary health care institutions is difficult to meet the requirements of the medical reform policy of graded diagnosis and treatment. Therefore, on the basis of the overall evaluation, it is also necessary to refine the evaluation index framework from multiple angles to achieve more refined and standardized evaluation and analysis. From the perspective of the level of informatization development of hospitals and primary health care institutions, the evaluation framework of population health informationization development level is refined to make up for the shortcomings of the overall evaluation framework. First, the evaluation framework of hospital informatization was constructed with the main framework of capital input, organization and management of hospital informatization, construction of hospital informatization infrastructure, information security, construction and application of hospital information system and performance of information construction. It contains 6 first-level indicators, 20 second-level indicators and 63 third-level indicators. Secondly, this paper constructs an information-based evaluation framework for primary medical and health institutions with five dimensions: information resources and infrastructure, information system construction and application, information sharing and technology application, informationization benefits, and information security, including 5 first-level indicators, 15 second-level indicators, and 53 third-level indicators.

Discussion

Comprehensiveness of evaluation perspective and pertinence of content

The construction of population health informatization in China is still the government-led mode, which is affected by the information technology level differences, policy environment support, and the level of security, so it needs to be considered as a whole. In the past, research focused on the application of information systems in a single business field, which is relatively microscopic, especially in the absence of technical considerations [12]. In the selection of indicators, this study strives to get rid of the single perspective of technology and add more macro policy indicators. At the level of evaluation operation, regional population health information evaluation mainly includes three types: "measurement of development level ", "performance evaluation" and "assessment evaluation". The horizontal measurement focuses on understanding the overall situation, and compares in the space and time; the performance evaluation pays attention to the results, and measures the information efficiency, effect and benefit; the assessment helps the evaluation object to find the lack and improve. At present, there is no unified construction mode for population health informationization construction, and the government-led information construction has a relatively strong public product attribute, and it is difficult to see actual benefits in the short term. Therefore, this study defines this evaluation as the basic regional population health informationization level measurement, appropriately highlighting development strategy at the emergence stage, and specifically reflecting the development of the main attributes in the informatization process.

Method selection of index weight assignment

In terms of the weight assignment of the index system, there are Delphi method, analytic hierarchy process, fuzzy analysis method, data envelopment analysis method, gray evaluation method, etc. Each method differs in subjective and objective assignment, qualitative and quantitative expression, and data integrity. In the measure of the development level of population health information, the subjective opinions of experts are indispensable, and some qualitative data are also inevitable. It is more appropriate to use AHP. The analytic hierarchy process describes the problems of complex systems as well-organized sets of hierarchical structural factors, qualitatively judges the elements of each level, and uses mathematical methods to obtain quantitative weight results, which achieves a

combination of quantitative and qualitative, overcoming the arbitrariness of the subjective valuation method, and avoiding the dilemma of the objective assignment method requiring long-term complete statistical data and the arbitrary judgment of throwing away from the empirical judgment. However, there are certain limitations in the actual use process. The comparison index of the two pairs is generally no more than seven, because the more the comparison, the more likely the logic confusion, the worse the consistency of the score. In order to effectively overcome the above problems, the number of indicators in the same root in each level of this study is no more than four, so that the results of expert scoring have a good logic, which ensures the scientificity of the evaluation system.

Practicability and reliability of index evaluation model

The completion of the hierarchy of the indicator system and the determination of the weight of the indicators are only part of the evaluation work. It is necessary to construct a comprehensive evaluation model, which combines the actual data with the index weight, including the conversion of the actual data into the index value and the combination of the index value and the index system. This study has developed detailed scoring rules for each of the three indicators, including indicator interpretation and calculation formulas, mainly in three cases. Firstly, qualitative indicators are converted into numerical values, and hierarchical values are assigned according to the rank. Secondly, indicators which cannot be directly calculated from the survey data are assigned to the corresponding indicators in segments for convenience, and the continuous "rate" value is converted into grade scores, and then the average calculation of the region is carried out. Thirdly, there are many administrative levels involved in the indicators, and the construction quantity, difficulty and importance of each level are not consistent. Each electronic medical record database cannot be treated equally. It must be dealt with at different levels, and the construction rate of each level corresponds to its own weight coefficient. The construction rate of each level corresponds to the respective weight coefficient. As the current population health informationization has no recognized "gold standard", the associated calibration validity cannot be measured. The content validity is generally not quantitatively measured. This research has been evaluated by experts to construct an evaluation index system, and there are detailed indicators calculation methods, which can be considered as good content validity.

Value of the research

The subjects, methods and objects of population health informationization evaluation in the world are diversified, and the goal orientation is obvious. Compared with international researches, China had carried out health informationization evaluation later than developed countries, and the early evaluation work mainly centers on the construction of hospital information system. In recent years, it has also made some progress in community health informationization, regional health informationization and comprehensive evaluation of health informationization. The evaluation of population health informationization in China is mainly based on index screening and weight testing. The evaluation object is single. It is difficult to measure the effectiveness and development level of health informationization comprehensively and scientifically because of the lack of thinking on how to evaluate and serve the government as the main body of information management. This research has filled in the gaps in this field.

Limitations

Different regions have different levels of health informatization in China, especially grass-roots institutions, which has weak foundation of informatization. Therefore, it had a certain impact on data collection from these institutions. It is necessary to further understand the detailed situation through field investigation in order to improve the accuracy of evaluation.

Conclusions

The index system is scientific and reasonable, and can be used for comprehensive evaluation of the development level of population health informatization, which has a certain guiding significance for the development of population health informatization.

Acknowledgements

This work was financially supported by National Natural Science Foundation of China (71273280), National Social Science Fund (17BGL184).

References

[1] Strategy Research on Health Information Development in China/ LI Xin-wei,HU Hong-pu,GUO Min-jiang,et al // China Digital Medicine .2011 6(6): 50 to 52.
[2] SWOT Analysis on the Development of Population and Health Information in China/ LI Yue-feng,HU Jian-ping,MENG Qun // Chinese Journal of Health Informatics and Management. 2016 13(01):45 to 50.
[3] Current Status and Reflection on the Regional Population & Health Information Development in China/ZONG Wen-hong,ZHOU Zhou,LIU Yue-xing,et al// Chinese Journal of Health Informatics and Management.2015 12(02):196 to 201.
[4] Construction of China's health system standard system/MENG Qun//China Health Standard Management.2013 4(9): 24 to 31.
[5] Study on the Evaluation Index System of the Construction of Regional Health Information Platform/ZHANG Tao,LIU Yue-xing,ZONG Wen-hong//Chinese Journal of Health Informatics and Management.2014 11(06):548 to 553.
[6] Research on Health Informatization Assessment System in China/LIN Bin,WANG Ye-ting,ZHANG Xiao-miao,et al//Journal of Medical Informatics.2013 34(02):10 to 15 plus 27.
[7] Research on Performance Evaluation Index System of Regional Health Information Construction/MA Yu-cheng,XING Ji,ZHOU Jie//Chinese Journal of Health Informatics and Management.2012 9(03):57 to 61.
[8] The Analysis of Reginal Health Information Police/WANG Yan-jun,ZHENG Jian-zhong,ZHANG Ai-lian,et al//China Digital Medicine.2016 11(08):47 to 49.
[9] Analysis of relevant policies to promote the development of electronic medical records in China/LI Guang-qian//China Politic Administration.2011 (10):25 to 29.
[10] PEST Analysis on Development of M-health in China/ZHANG Hui-li,JIANG Yong//Journal of Medical Informatics.2012 33(11):6 to 9.
[11] Development of Health Information Evaluation Perspective Abroad and Its Enlightenment to China/ZHANG Hui-hui,LIANG Li-fan,MA Jing-dong,et al // Journal of Medical Informatics.2014 35(08):2 to 6.
[12] Study on the Synthetical Evaluation System and Empirical Research of Hospital Informationization Level in China. [D].Changsha:Central South University,2007.

Address for correspondence

Institute of Medical Information, Chinese Academy of Medical Sciences/Peking Union Medical College, Beijing 100020, China.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190456

Crowdsourcing Public Opinion for Sharing Medical Records for the Advancement of Science

Chunhua Weng[a], Tianyong Hao[a], Carol Friedman[a], John Hurdle[b]

[a]Department of Biomedical Informatics, Columbia University, New York City, NY, USA
[b]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA

Abstract

This study used Amazon Mechanical Turk to crowdsource public opinions about sharing medical records for clinical research. The 1,508 valid respondents comprised 58.7% males, 54% without college degrees, 41.5% students or unemployed, and 84.3% under 40 years old. More than 74% were somewhat willing to share de-identified records. Education level, employment status, and gender were identified as significant predictors of willingness to share one's own or one's family's medical records (partially identifiable, completely identifiable, or de-identified). Thematic analysis applied to respondent comments uncovered barriers to sharing, including the inability to track uses and users of their information, potential harm (such as identity theft or healthcare denial), lack of trust, and worries about information misuse. Our study suggests that implementing reliable medical record de-identification and emphasizing trust development are essential to addressing such concerns. Amazon Mechanical Turk proved cost-effective for collecting public opinions with short surveys.

Keywords:

Crowdsourcing, data collection, privacy

Introduction

Secondary use of electronic health record (EHR) data offers promise in advancing clinical and translational research. Access to electronic patient data promises efficient research at scale and at lower cost. However, such research also faces significant barriers [1], such as controversy around ownership of medical information and difficulties in sharing information among clinical researchers or across institutions. This can result in information fragmentation and lack of transparency. Due to HIPAA regulations, the need for consent to capture patients' willingness to share their medical records has hindered research by informatics investigators [2]. Patient-led sharing of medical records for research is seen positively in several parts of the world, but barriers to patients' willingness to share medical data for research remain common. In a study conducted in Australia, 95% of participants believed that medical research with data sharing was necessary in general [3]. Only 73% of these participants, however, were willing to share their own data for research. Investigators have studied differences between those willing and unwilling to share their medical records data. Buckley et al. found in a cohort that healthy controls were less willing to share their medical data [4], while Shavers et al. found that race played a factor; for example African Americans were found to be less likely to participate in research [5].

Although several studies have been published regarding different aspects of record sharing, little is known about the feasibility of engaging the public to archive the clinical phenome for medical research [6]. To investigate what lies behind the general unwillingness to share medical data for advancing science and to learn how we can increase the public's willingness to share, we conducted a crowdsourced survey on individuals' willingness to share medical data. This is the first study of thematic trends in free-text survey comments relating why individuals might be unwilling to share their own or their family's medical data.

Methods

We used a survey instrument previously published by a panel of natural language processing experts [7]. It included questions on demographics, willingness to share medical records for research (with and without identifiers), and willingness to share specific types of health information (also a first), such as medications, laboratory test results, and chronic illnesses. We asked respondents if they would be unwilling to share their own or family members' clinical information for research purposes, and, if not, why not.

We surveyed the opinions of the public who live in the United States and who speak English using the Amazon Mechanical Turk (AMT) system. AMT is an online crowdsourcing marketplace that recruits human workers to perform tasks for a nominal fee ($0.10 for completing this entire survey). Only one response was allowed from each AMT worker.

For data quality control we inserted two common-sense knowledge questions: "*What is the first month of the year*?" and "*Who is the current president of the United States*?" in the survey to identify valid responses. Any respondent who failed to correctly answer these two questions was removed from the analysis. Pearson's $\chi 2$ tests were used to examine relationships between respondent characteristics and willingness to share medical records. Free-text responses were categorized thematically.

Following widely accepted guidelines for thematic analysis, two independent coders performed the analysis. Each coder first familiarized themselves with the data and then generated initial codes separately. Then the two coders met to compare the codes and reach consensus. Each coder independently searched for themes in the codes that were relevant to the research question of this study: *Is the public willing to share their medical records for research? What are the barriers?* Multinomial regression was used to generate probability matrices to evaluate the predictive probability of gender, age, and education on willingness to share and reasons not to share.

Results

1,774 AMT workers completed the survey, of whom 1,508 answered the common-sense knowledge questions correctly and were included in the analysis. Among these respondents, 58.1% were male, 54.5% had not completed a college degree, 41% were students or unemployed, and 84.2% were younger than 40. This cohort appears relatively young and well educated. The demographics are shown in **Table 1**.

Table 1 – Demographic Characteristics of Study Participants

Variable	Value	n (%)
Gender	Female	623 (41.31)
	Male	885 (58.69)
Age	18-25	668 (44.3)
	26-40	603 (40.0)
	41-55	176 (11.7)
	56 or older	61 (4.0)
Education	Less than high school	10 (0.7)
	High school/GED	159 (10.5)
	Some college	646 (42.8)
	Bachelor's degree or college graduate	547 (36.3)
	Graduate or professional degree	146 (9.7)
Employment	Religious	3 (0.2)
	Nonprofit organization	31 (2.1)
	Government and public administration	58 (3.8)
	Education	89 (5.9)
	Health care and social assistance	89 (5.9)
	Homemaker	107 (7.1)
	Scientific or technical	142 (9.4)
	Not employed	223 (14.8)
	For-profit business	354 (23.5)
	Student	403 (26.7)

Willingness to Share Medical Records

AMT workers were asked to pick an answer that best described how they felt about sharing medical records for research. This included their perception of how willing they felt the general public was to share their medical records for research, how willing they were to donate medical records of deceased family members for research, and how willing they themselves would be to share their own medical data (both with identifying information and without), as shown in **Figure 1**. Over 74% were at least somewhat willing to donate their de-identified records for research. 32.4% thought others "might be willing" to share medical records of deceased family members for research, while only 4.2% stated they thought that others would definitely share their medical records. 35% said they would share their deceased family member's records for research, with 6.8% saying they would not share their family's records. Only 13.9% expressed willingness to share their identifiable medical records for research. 20% replied they would not be willing to share their data with identifying information, whereas only 5.3% expressed unwillingness to share de-identified medical records. 50.3% indicated willingness to share their de-identified medical records for research.

Willingness to Share Specific Health Information

AMT workers were asked to indicate their willingness to share different aspects of their medical records for research (**Figures 2-3**). Respondents checked whether they were "*willing to share*," "*not willing to share*," "*not sure*," or "*not applicable*" with regards to health information such as lab results, medications, diagnostic reports, chronic illness, mental health, cancer, disabilities, and more. Respondents were most willing to share information on their demographics (**Figure 2**), childhood diseases, substance abuse, alcohol and tobacco use, cancer, and surgeries (**Figure 3**).

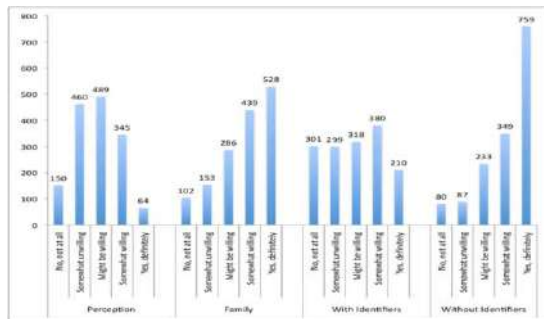


Figure 1 – Willingness to Share Data

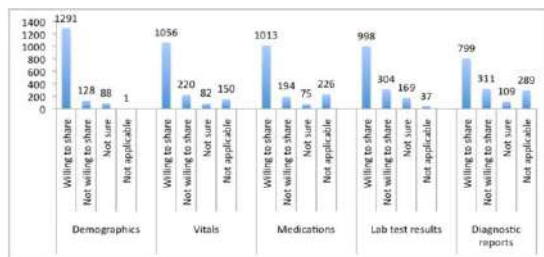


Figure 2 – Willingness to Share Specific EHR Data Types

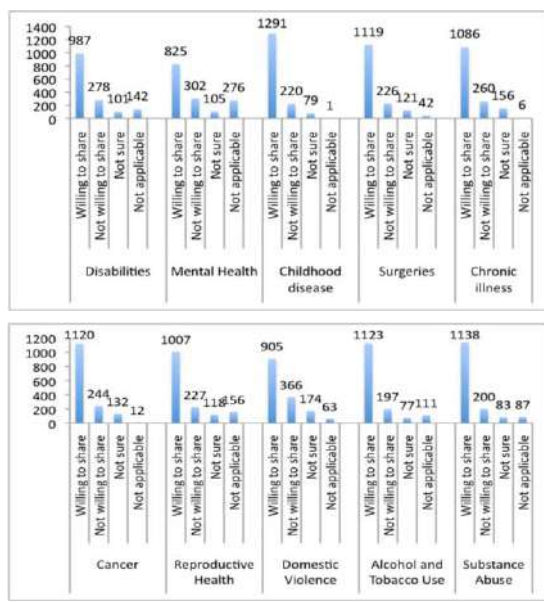


Figure 3 – Willingness to Share Data about Specific Conditions

The five types of health information that the respondents were most unwilling to share were domestic violence, diagnostic reports, lab test results, mental health, and disabilities. Respondents were most unsure about domestic violence, lab test results, chronic illness, cancer, and surgeries. Many respondents found diagnostic reports, mental health, medications, reproductive health, and disabilities to be "not applicable". All descriptive statistics are in **Table 2**.

Table 2 – Descriptive Statistics of Participant Willingness

Willing to Share Records [1=not at all, 5=definitely]	**Mean**	**SD**
What other people think	2.83	1.04
Expired family member	3.77	1.22
Identified	2.98	1.34
De-identified	4.09	1.15
Willing to Share Conditions	**Freq**	**%**
Demographics	1,291	85.61%
Childhood disease	1,291	85.61%
Substance abuse	1,138	75.46%
Alcohol and smoking	1,123	74.47%
Cancer	1,120	74.27%
Surgeries	1,119	74.20%
Chronic illness	1,086	72.02%
Vitals	1,056	70.03%
Medications	1,013	67.18%
Reproductive health	1,007	66.78%
Lab results	998	66.18%
Disabilities	987	65.45%
Domestic violence	905	60.01%
Mental health	825	54.71%
Diagnostic results	799	52.98%

Reasons for Being Unwilling to Share

Respondents were asked to make one to three selections for why they would be unwilling to share their medical records for research, including an option for, "Not applicable, I am willing to share this information." 203 respondents marked two responses, 251 checked three, and 39 marked more than three responses (results in **Figure 4**).

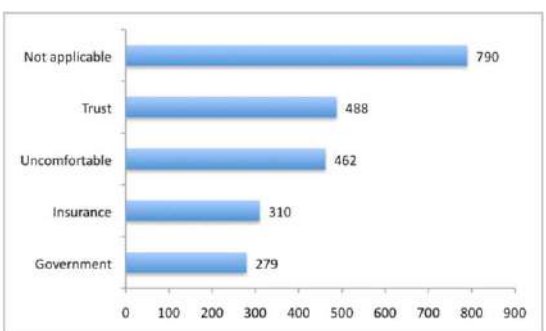


Figure 4 – Concerns Behind Unwillingness to Share

763 (50.6%) AMT workers solely checked, "*Not applicable because I am willing to share my medical information,*" along with 27 who also made other selections. 488 respondents checked "*I don't trust that my information will be kept confidential.*" 463 marked, "*It would make me uncomfortable to share this information.*" 310 respondents checked, "*It may compromise my future health care or insurance,*" and 279 felt, "*I am afraid my information will be used by the government*."

The last question in the survey asked respondents to share any other reasons they would be unwilling to share their medical records for research, with a text box for free responses. Of the 1,508 workers who took the survey, 689 (45.7%) provided a response for the final question in the survey. Only three respondents declared that they were unwilling to share without a reason, while 105 respondents stated they were willing to share. The remaining 581 respondents indicated issues or reasons that caused them to be unwilling to share, or conditions that would make them willing to share their medical information for research. Each response was coded to match one or more of the 13 codes listed in **Table 3**. Respondents had the most concerns with sharing identifying information, privacy, and potential harm caused by sharing. A total of 681 codes for reasons unwilling to share were applied to these responses.

Table 3 – The Thirteen Themes Behind Unwillingness to Share

Theme	**Total**	**% of those unwilling to share**
Risks with identifiable information	155	26.70%
Potential harm	116	20.00%
Risks with privacy	116	20.00%
Unauthorized access to or sharing of information	53	9.10%
Lack of knowledge of research study	49	8.40%
Improper/unauthorized use of information	33	5.70%
Compromised confidentiality	32	5.50%
Uncomfortable sharing medical data/records	29	5.00%
Beliefs on information sharing	28	4.80%
Specific health information	26	4.50%
Medical data handling	23	4.00%
Distrust in government	13	2.20%
Insufficient compensation/no benefit to participant	8	1.40%

Including identifying information generated the most concern from respondents. Respondents expressed worries about the information being traced back to them. One respondent posed the idea of potential harm from unauthorized sharing of medical data, "*I would only be concerned about sharing my personal, identifying information because I'd be concerned it might get shared -- even inadvertently or accidentally*." All respondents expressed some willingness to share their medical data for research if identifiers were removed, and some distinctly stated they would be unwilling to share if identifiers were not removed. Privacy and potential harm tied for being the second most concerning issues.

Of the 690 respondents who provided a response, 116 expressed a desire for privacy. A strong sensitivity to the right to privacy was expressed by several respondents, e.g., that it was "*nobody's business,*" that the information was "*too personal to share*," and that it would be "*an invasion of*

privacy." One stated, "*My medical records are between me and my doctor and I don't believe they are anyone else's business*." Some respondents also felt that their privacy would not be guaranteed if they shared their medical data for research.

Several respondents were concerned that harm may come to them from sharing their medical data. Specific concerns included identity theft, missed job opportunities or loss of employment, and medical care and health insurance discrimination. A total of 112 respondents were unwilling to share medical data because of a perceived risk of harm from sharing their information. One respondent stated, "*I can see the results being used against me for jobs or health insurance*." Another felt that not sharing medical data/records would prevent potential harm, saying he would be unwilling to share to "*Protect identity [and] prevent gossip regarding my health from people who would recognize my name*." Several other respondents likewise were concerned about being judged by their information.

Confidentiality, handling of medical data by the researchers, and the likelihood of information being accessed and shared by people other than researchers was also an issue. Respondents expressed concerns over the researchers' capabilities to keep the data secure and whether their information really would be kept confidential. 33 respondents indicated that they were concerned about confidentiality, and 23 were unwilling to share due to uncertainty in how their medical data would be handled and kept safe. One respondent posted, "*Most importantly, I have no control over what is done with it, and question how securely the data is protected*." 53 respondents believed it was possible that their medical data could be hacked, leaked, or accidentally shared. One stated, "*I would be afraid my personal information would get out and I'm not comfortable with that.*"

3.8% of participants demonstrated concerns about accessibility, applicability, and the necessity of sharing their medical records for research. Some felt that too much information on individuals was already available for others to see ("*I feel that people can find out enough about anyone. Why give out more*?"), or that sharing medical records was not necessary ("*I feel that this information is very personal and doesn't need to be shared for researching*."). Others stated that they could not share because they did not have access to their own records ("*I don't even have access to them*."), or that they were unwilling to share information for specific kinds of research, such as chronic disease, since they did not have a chronic disease or information believed to be pertinent to the research. "*I do not have cancer. There is no domestic violence, nor alcohol, nor substance abuse. If any of this applied to me, I would share that info*."

Sharing specific types of health information was another concern. A variety of respondents indicated that they were unwilling to share some types of health data. A few respondents demonstrated that they would be unwilling to share a given type of information specificially because sharing made them feel embarrassed. One respondent stated, "*I'd be willing to share my medical data/records except for things that are personally embarrassing*."

Respondents also shared that their willingness to share was dependent on specific knowledge of the research study, such as the research purpose and who the researchers and associated academic institutions were. Some participants also stated that they were unwilling to share because they were unable to contribute to the assessments being made by researchers.

32 respondents were unwilling to share because they believed their medical records may be used for research they disapproved of or would be used for marketing purposes or for profit. Some respondents also felt that their records may be used unethically. Less than 4% of respondents were unwilling to share due to distrust in the government or because they felt they would not receive any benefit by sharing. 15 respondents were concerned about government involvement in research and healthcare, feeling the government had too much access and involvement with personal data already. A respondent said, "*NSA already has all this info*," while another stated "*I'm not comfortable trusting the government with my medical information*." For some, being compensated was a significant part of being willing to share medical data for research. One respondent stated, "*If the price is right then I would gladly share my information. That's the only reason. I still wouldn't trust that my information would be kept confidential, but it wouldn't be the motivating factor. I'd want to be compensated.*"

Barriers to sharing medical records included the inability to track the uses and users of their information and a lack of trust in researchers' intent. Explanations for their answers were provided by 960 (54%) respondents. The primary concerns on this point were (1) lack of tracking of the users and uses of shared information; (2) fear of being harmed, including loss of medical care or insurance, identify theft, being discriminated against by future employers, and being selected for targeted advertising; and (3) lack of trust in the data collectors.

Discussion

In this study, Amazon Mechanical Turk proved to be a cost-effective method of collecting public opinions for biomedical research. Using AMT was also more representative of a broader population than conducting surveys among medical center personnel, although respondents did have to be computer literate. Also, the AMT survey took only a couple of days and $20 to collect responses from 1,774 respondents, while it took months for our surveys in medical centers to collect 400 responses.

The majority of participants were willing to share their medical data for research in some manner. We discovered a variety of reasons why patients may be unwilling to share their medical data for research. Some of these reasons are addressable. For example, some respondents stated they would be willing to share if they had more information on the research study, researchers, or the associated academic institution. If respondents were provided with more information and were guaranteed that their data would be protected, kept confidential, and not used by the government, willingness to share increased. Factors that are hard to address but that affected willingness to share were desire for privacy and discomfort sharing specific types of medical data. Reliable de-identification methods for medical records and trust development are critical for addressing these concerns.

Responses to the last, open-ended question of the survey were similar to responses in Weitzman et al.'s focus group study [8]; however, due to the large number of respondents who took the survey and the application of codes to the responses, we were able to determine the strongest and most prevalent concerns with sharing medical data for research. "Identifying information" was the most frequent code. Removing patient identifiers, such as name, address, and social security number, would cause the largest increase in willingness to share. This was confirmed by the increase in respondents indicating they

would be willing to share medical data without identifiers over the number of those willing to share with identifiers. This pattern was also seen in a study comparing attitudes towards sharing of medical data in healthcare settings between U.S. and Japanese populations [9].

Privacy, potential harm, and unauthorized access to or sharing of information were the next most common concerns. These, along with compromised confidentiality, improper or unauthorized use of data and sharing specific types of health information stem from sharing identifying information. This concurs with the findings of a study evaluating the effect of authorization forms used by hospitals on likelihood of consent which found that requesting social security numbers negatively effected the return rate of authorization forms, while distinguishing the hospital name on the forms increased the return rate [10]. The present study disagrees with prior reports of biases in willingness to share specific types of medical data. Respondents were least willing to share information about diagnostic reports, mental health, and domestic violence, while being most willing to share demographics, childhood diseases, and substance abuse.

Comparison to Survey Results from Academic Centers

In a previous study, we conducted a similar survey of 2,140 highly-educated professionals, students, and staff in two academic medical centers [7]. 56% of respondents were "somewhat/definitely willing" to share clinical data with identifiers, while 89% of respondents were "somewhat" (17%) or "definitely willing" (72%) to share without identifiers. Results were consistent across gender, age, and education, but there were some differences by geographical region. Individuals in that study were most reluctant (50-74%) to share mental health, substance abuse, and domestic violence data, but remained fine sharing diagnostic data. Mental health and domestic violence seem to be sensitive areas not likely to be shared in either study cohort. The public in the present study were more willing to share substance abuse information than health professionals.

Limitations

There were several limitations to this study. Presenting most questions with predefined answer options may have prevented more freestyle responses. 54.3% of respondents did not provide an answer to the free-text question of the survey. Results of our study should be applied with care, as respondents were not asked where they currently live nor what their citizenship is. A research group at the University of California-Irvine surveyed MTurk HIT workers to evaluate their demographics and MTurk habits. It was found that 57% claimed to live in the U.S., while 32% were in India [11]. Asking for nationalities of our AMT respondents would have helped us understand how we could overcome barriers in willingness to share medical data in specific regions. Finally, people dealing with serious illnesses often have different conceptions of privacy than those who are not and, therefore, may be more willing to share health information. Our survey did not ask for health status of the respondents. In the future, we would like to study which patients are willing to share and which are not by collecting a richer variety of respondent characteristics.

Conclusions

This study is the first to leverage a crowdsourcing approach to efficiently collect the public's preferences and concerns around sharing medical records for the advancement of science. This study sheds light on the opportunities and challenges when engaging the public on the subject of donating their medical records to support research, as well as on the need to balance privacy and specific patient needs for health information in the age of participatory health, the "measured self," and social media.

Acknowledgments

This work was supported by National Institutes of Health grants R01LM009886, 5UL1TR001873-02, R01LM010981, UL1 TR000040, and R01LM008635. We thank Drs. William Hersh and Justin Fletcher for their assistance and input on the survey. We thank Elhaam Borhaniana for assistance with the analysis.

References

[1] D. Grande, N. Mitra, A. Shah, F. Wan, and D.A. Asch, Public preferences about secondary uses of electronic health information, *JAMA Intern Med* **173** (2013), 1798-1806.

[2] R.B. Ness and Societies of Epidemiology Joint Policy Committee, Influence of the HIPAA Privacy Rule on health research, *JAMA* **298** (2007), 2164-2170.

[3] T. King, L. Brankovic, and P. Gillard, Perspectives of Australian adults about protecting the privacy of their health information in statistical databases, *Int J Med Inform* **81** (2012), 279-289.

[4] B.S. Buckley, A.W. Murphy, and A.E. MacFarlane, Public attitudes to the use in research of personal health information from general practitioners' records: a survey of the Irish general public, *J Med Ethics* **37** (2011), 50-55.

[5] V.L. Shavers, C.F. Lynch, and L.F. Burmeister, Racial differences in factors that influence the willingness to participate in medical research studies, *Ann Epidemiol* **12** (2002), 248-256.

[6] M. Corn, Archiving the phenome: clinical records deserve long-term preservation, *J Am Med Inform Assoc* **16** (2009), 1-6.

[7] C. Weng, C. Friedman, C.A. Rommel, and J.F. Hurdle, A two-site survey of medical center personnel's willingness to share clinical data for research: implications for reproducible health NLP research, *BMC Med Inform Decis Mak* **19** (2019), 70.

[8] E.R. Weitzman, L. Kaci, and K.D. Mandl, Sharing medical data for health research: the early personal health record experience, *J Med Internet Res* **12** (2010), e14.

[9] M. Kimura, J. Nakaya, H. Watanabe, T. Shimizu, and K. Nakayasu, A survey aimed at general citizens of the US and Japan about their attitudes toward electronic medical data handling, *Int J Environ Res Public Health* **11** (2014), 4572-4588.

[10] D. Bolcic-Jankovic, B.R. Clarridge, F.J. Fowler, Jr., and J.S. Weissman, Do characteristics of HIPAA consent forms affect the response rate?, *Med Care* **45** (2007), 100-103.

[11] J. Ross, L. Irani, M.S. Silberman, A. Zaldivar, and B. Tomlinson, Who are the crowdworkers?: shifting demographics in mechanical turk, in: *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, ACM, Atlanta, Georgia, USA, 2010, pp. 2863-2872.

Address for correspondence

Chunhua Weng, PhD, Department of Biomedical Informatics, Columbia University Medical Center, 622 W 168 Street PH-20, New York, NY 10032. Email: chunhua@columbia.edu.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190457

Dialogue Analysis for Clinical Data Query Mediation

Chunhua Weng[a], Amy K. Mir[a,b], David Hanauer[c], James Cimino[a,d]

[a] *Columbia University, New York, NY, USA*
[b] *Harvard College, Cambridge, MA, USA*
[c] *University of Michigan, Ann Arbor, MI, USA*
[d] *University of Alabama, Birmingham, AL, USA*

Abstract

Efficient access to clinical data by investigators is critical for data-driven biomedical research. Mediated query is commonly adopted to facilitate data access for investigators. However, the query mediation process remains a black-box. This study analyzed the email-based dialogues between investigators and clinical data query mediators from three institutions. We identify discussion topics and their frequencies, model task flows, and analyze user needs for query mediation support revealed from the dialogues. While the datasets between different institutions are distinct in some notable respects, we find that together they provide common insights for streamlining data access. From our findings, we conclude an intelligent dialogue-based query support model is feasible to automate human-mediated clinical data access for investigators and stakeholders.

Keywords:

Information storage and retrieval; Interdisciplinary communication; Needs assessment.

Introduction

Big data in healthcare offers immense research potential. In practice, much of the data access is mediated by query analysts who have the adequate technical knowledge, e.g. structured query languate (SQL), to pull information from complex databases. Clinical data access for biomedical researchers is limited not only by patient privacy and data governance but also ineffective communication between clinical researchers and query analysts. In addition to logistic barriers such as lengthy institutional paperwork for approving data requests, there are interdisciplinary communication barriers due to vocabulary differences, conceptualization discrepancies, and knowledge gaps impeding efficient delivery of accurate data sets. With proliferating data, human-mediated data queries are difficult to scale.

The underlying biomedical query mediation (BQM) process is an iterative question-answering process between investigators and query analysts. BQM centers on the investigators' definition of a research statement followed by more precise specifications of the clinical process in question so that query analysts are able to locate the appropriate data elements. In the process of BQM, contextual data constraints often guide revision of the researchers' data queries [1]. The more complex or granular databases are, the more complex BQM may inherently be due to iterative refinements and understanding of the precise information needed from both parties.

Currently, there is minimal literature detailing BQM and much of the underlying communication space exists in a "black-box" in which the actual needs of researchers and the methodology of query analysts' work are not always transparent to each other, which is reflected by the iterative nature of BQM.

Increasing the transparency of the BQM communication space is thus an important first step into understanding how to better streamline data access, as understanding the processes underlying information retrieval can lead to informed redesigns of communication and information flow [2].The biomedical query negotiation can generally be seen as a continuum of automated information retrieval and human-centered communication (e.g. conversations or email exchanges). Depending on the information task, automation can be substituted for in-person communication [3], which is by direct conversation or email, is not always efficient.

In this paper, we investigate email communications during BQM since email exchanges are an excellent candidate for a computational decision-support paradigm in which elements of information flow can be automated. In our analysis of the email communications from three institutions, we attempt to understand how researchers and query analysts negotiate data needs via email and identify knowledge gaps between biomedical researchers and query analysts. We use insights from task and temporal communication patterns, categorization of questions, sender-recipient networks, and the results of predictive text models to better understand email-based BQM in evaluating whether we can feasibly design an automated dialogue-based query support system.

Data and Methods

Email communication for 20 self-reported critical incidents for BQM totaling to 307 emails at an average of 15 emails per case with a standard deviation of 19 were collected from 3 query analysts from three institutions. These cases were exchanges characterized by logistical inefficiencies and communication gaps. Most of the variability in case size comes from Site 3, which also makes up 80% of the email messages. Cases from Sites 1 and 2 average 5 to 6 emails per case with a standard deviation of 2 emails. Cases from Site 3 average to 23 emails per case with a standard deviation of 22. Case 3's variability comes from an outlier case containing 93 emails (Figure 1).

Data De-identification and Structuring

De-identification of sensitive information was done manually by replacing sender names, including titles (e.g. Dr., MD), and associated email addresses with generic codes. In addition to signatures, institutional footers were removed from messages.

Figure 1 - Distribution of cases by number of emails

Telephone numbers, addresses and locations, passwords, and file names of attached documents were deleted from the text. To ensure further data privacy, the delivery context of emails, e.g. 'Sent from my iphone', were removed from all emails. The structure of all emails was standardized to allow easier parsing of the text. Junk delimiters, excluding message separators, such as > and Original Message identifiers were cleaned from the dataset. Non-alphanumeric text was deleted unless it was related to the data or query commands in an email. Any Date delimiters were changed to Sent. Times were truncated from hour-minute-second format to hour-minute format to ensure uniform precision in temporal analysis of emails. Completed recipient information could not be imputed in all emails in cases where copied (cc'd) recipients were not explicitly denoted in abbreviated email headers. Actual email text was demarcated with the MESSAGE tag and additional tags were added to identify task domains and communication parties.

```
--------------------------------
From: S2RID3
Sent: Monday, April 21, 2014 1:41 PM
To: S2RID1
Subject: RE: Request for MRNs given BCN list
General Tags: Resolution
Specific Tags: Resolution Pending
Communication Party: Query Analyst
Target Party: Query Analyst

MESSAGE:
OK, then the estimate is 3 hours, with the
understanding that we may add to it later.
```

Figure 2 - An example de-identified email message

Coding Book for Hierarchical Task Analysis

We defined email tasks, i.e. email events, as what the sender wants to accomplish with the primary recipient. All emails were manually tagged with domains and sub-domains using this coding book (Table 1).

We append an additional identifier Communication Party, an indicator variable of whether a sender is a query analyst retrieving data or biomedical researcher requesting data. Because a sender does not necessarily always communicate with the opposite party, an additional tag Target Party was created to account for researcher-to-researcher and analyst-to-analyst communications in addition to the communications between biomedical researchers and query analysts.

We ran basic email counts across sites, computing min, max median, mode, and standard deviation in the number of emails in a case to get a basic sense of communication volume. We constructed a basic box-plot view of email counts across sites. For each critical incident, we looked at which communication party (researcher, query analyst, or other/unknown) was initiating the thread of the critical incident, as a percentage of the number of cases in each site. We determined the number of emails sent by different communication parties as a percentage of the total number of emails in each site. We additionally determined the number of party-to-party (i.e. researcher to query analyst, query analyst to researcher, researcher to researcher, and query analyst to query analyst) communication flow as a percentage of the total number of emails across sites. We furthermore broke down task domains as well as task sub-domains by researcher and by query analyst across sites.

Table 1 - Domains and Sub-domains for Task Analysis

Domain	Subdomain	Definition
Inquiries		
	Task Request	the recipient is asked to assist with a data request by sending information, modifying the structure of the data, supplying additional information and detail, retrying a solution with new information, or confirming feasibility
	Clarification	the recipient is asked to clarify detail on information sent or confirm and make sure of whatever information was previously provided
	Follow-up	a check-in for updates on progress from recipient
Meetings		
	scheduling	Schedule a meeting
	Meeting confirmed	Confirm the schedule
	Meeting completed	Close the conversation
Resolution		
	Resolution Pending	a sender states that progress will be made or is currently being made. This can include mention of failed attempts at a solution or data request acknowledgements by a data team processing the request
	Resolution Offered	the sender offers at least one part of a larger solution or notes which information is already readily available
Other		
	Notice	a general information email, generated without regard to a particular data request
	Forward	an email in which the purpose of the message is to forward a communication
	Introduction	An email in which a new researcher or query analyst introduces himself or is being introduced into the email thread
	Acknowledgment	an email in which the sole purpose of the message is to acknowledge the resolution offered or the progress being made

The interrogative pronouns *'who', 'what', 'which', the adverbs 'where', 'when'*, and the conjunctions 'if', 'whether' characterize straightforward questions can be automated by a database. The interrogative adverbs 'how', 'why' may be more open-ended and less likely to be candidates for the computational paradigm in communication. Better understanding of these grammatical identifiers helps us better understand the nature of questions in email-based BQM. In a representation based on these

grammatical terms, we categorize our questions as Identification questions involving 'who', 'what', 'where', 'when', 'which' inquiries, Choice questions involving queries of 'whether', 'if', 'or', Quantification questions involving 'how many', 'how much', 'how long', 'how far', 'how often', 'how high', 'how long', and Discussion-oriented inquiries represented by 'why' and other forms of 'how' like 'how do' and 'how would'. We filtered on emails whose task domains include inquiries and whose specific sub-domains exclude stand-alone follow-ups. In the filtering, we were thus able to focus more on questions related to the data request at hand. Our results are represented by researcher and by query analyst.

The bag-of-words model is popular for document classification. Given that emails are labeled with the task domain, task sub-domain, communication party, and target party, we approach task and party identification as a supervised learning problem. Representing email messages as multi-sets of words that only take into account word multiplicity but disregard word order and grammatical structure, word frequencies in the multi-sets are used as a feature in training a multinomial naive Bayes classifier. Multinomial naive Bayes assumes independence between features for a multinomial event model, which are our task and party classifications.

In task identification, we first predicted general domains. Because multiple domains can occur together in an email, we separately ran the model for each general domain and predicted the binary indicator of that particular domain. We then performed a nested prediction, in which we predict specific task sub-domains conditional on the corresponding general domain, extracting only emails for which that domain is the only domain tag. We applied the same algorithm in predicting communication and target parties. For each prediction set, we performed randomized cross-validation in 1000 iterations and noted the range of accuracies with the middle 800 values, i.e. accuracies at the 10th and 90th percentiles.

We developed a simpler classification algorithm premised on the concept of a codebook; that is, a set of key terms and phrases representative of an aspect of the conversation such an email task domain. Unlike bag-of-words, our codebook-based classification gives consideration to word order since it extracts one to four-word phrases in common among at least two messages of a certain category. For domains, we looked first at matches between single-domain messages and then matches of those single-domain messages with multi-domain messages. For sub-domains, we again conditioned on the relevant broader domain first, as in our bag-of-words-based prediction. For domains, sub-domains, and communication parties, we ensured that the codebooks constructed do not overlap between their different classifications so that the set of codes is unique to that particular classification.

The construction of codebooks allows us to understand the language surrounding an email task or communication party. We hence made another probe into the language of biomedical researchers and query analysts using part-of-speech (POS) tagging to determine the most frequent nouns and verbs, and their overlap, among different communication parties. We adapted a script for noun-phrase extraction. Frequent topics are the functional foundation of automated dialogue-based query support and we hence computed the frequencies of our extracted topics across sites.

We examined how tasks follow from one another across emails. We computed a ranked moving average of the number of task sub-domains across emails. We looked at the intra-message and inter-message relationship between tasks, respectively, in computing the most frequents tasks concurring with a specific task and the most frequent tasks preceding or following that specific task. We constructed network visualizations specifically for clarifications and task requests to better understand the events surrounding the iterative question and answer process. For cases without any missing timestamps we visualized the time lengths between conversations in a box-and-whisker plot and analyzed task domains and sub-domain at peaks, i.e lags, in response times. To understand how involved the communication networks are, we plotted a color-coded network visualization of senders, primary recipients, and secondary (cc'd) recipients who are neither senders nor primary recipients. We then generalized the communication interactions to four different network models.

Results

Communication Volume and Communication Networks

In the breakdown of thread origination across sites, we look at which communication party, i.e. researchers or query analysts, are initiating the conversation. While most cases begin with contact by a query analyst with institutions 2 (100%) and 3 (55%), with institution 2 being entirely analyst-initiated, institution 1 is predominantly researcher-initiated (Figure 3a). The researcher to query analyst communication volume among individual email messages is about the same as the query analyst to researcher volume for Sites 1 (40%) and 2 (38%), and is slightly higher with Site 3 (45%) (see Figure 3b).

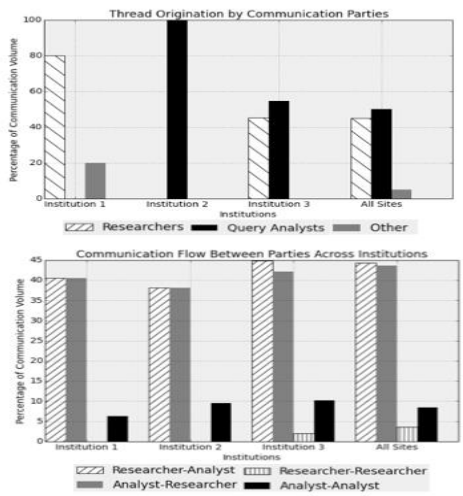


Figure 3 - Breakdown of pairwise communication

Interestingly, exchanges are not simply between opposite parties. Across all three institutions a non-negligible portion of the communication volume occurs between query analysts themselves, and for Site 3, there are also some smaller number of exchanges from researcher to researcher (see Figure 3b).

Communication networks represent the complexity of conversations between senders and primary recipients for an entire critical incident. Two-node interactions are the most common and simplest network type in our data, with nine critical incident cases following this prototype. Linear three-node networks are represented in four cases, where the additional third node can be any party. In linear three-node structures, the two relationships could be bidirectional relationships among the same parties and between different parties. Our cases also include the instance of both bi-

directional and unidirectional interactions among the two pairs of different parties. Triangular three-node structures, represented by four critical incidents, can have one to three bi-directional graph edges and, in our data, roughly the same number of cases involves the majority of the nodes being of either party. Complex networks are the rarest case, represented by only three cases, all within Site 3. The first three network models all exist in exchanges from each of the three institutions.

Task domains are split almost evenly between researchers and query analysts, with insignificant variations between specific domains and sites (see Figure 4a). In the small set of cases from Site 2, there are no emails requesting, confirming, or referencing meetings. Within the inquiry domain, most questions are straightforward identification or choice questions as opposed to quantification questions and more involved, discussion-oriented inquiries (see Figure 4b).

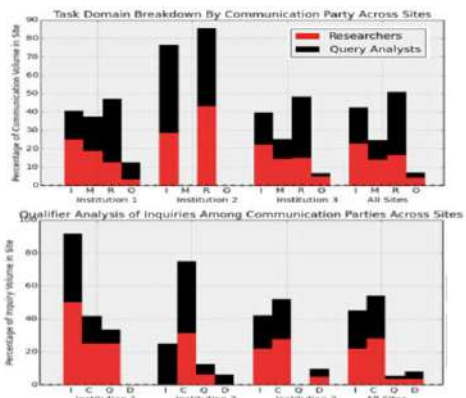


Figure 4 - Breakdown of tasks and classification of inquiries. (a) Task domain breakdown by communication party across all sites. I = Inquires, M = Meetings, R = Resolutions, O = Other. (b) Question breakdown by communication party across all sites. I = Identification, C = Choice Between Options, Q = Quantification, D = Discussion

Predictions of Tasks and Communication Parties

The bag-of-words algorithm with naive bayes assumptions on our multinomial event model for domains and sub-domains has generally good predictive performance. For the middle 80 percent of values within the 10th and 90th percentile of prediction accuracies, we find that single-domain prediction of inquiries, meetings, and resolutions range from 61-74%, 73-84%, and 60-71%, respectively. Because we approached this as binary a prediction for each domain, these results are better than a 50% random prediction.

For the nested prediction in which we condition in the presence of a particular single-domain, we again have better-than-random results. For inquiries, we find that task requests, clarifications, and follow-ups have 80 percent of their accuracies at 50-78%, 50-78%, and 61-83%, respectively. The similar and slightly lower range of prediction accuracies for task requests (i.e. information requests, and clarifications) reflect the similarity in these two sub-domains as they are both inquiries specifically regarding the data at hand, as opposed to questions following up with progress or requesting meetings.

In the conditional predictions with single-domain meetings, we find that meeting scheduling, meeting confirmations, and completed meetings have most of their accuracies at 62-92%, 62-92%, and 100%, respectively. In conditional prediction with single-domain resolutions, we find that pending resolutions and definitively offered resolutions can be mostly predicted within 65-87% and 61-87% accuracies, respectively. Looking at party predictions, we find the party originating the email message can be predicted mostly within 64-77% accuracy for both biomedical researchers and query analysts. The party to which the message is targeted can be accurately predicted within 58-70% and 57-70% accuracy for researchers and query analysts, respectively. The general domains of inquiries, meetings, and resolutions are accurately predicted with 42%, 75%, and 61% accuracy, respectively. Within inquiries, task requests, clarifications, and follow-ups are correctly predicted 20%, 43%, and 85% of the time. Within meetings, meeting scheduling, meeting confirmations, and completed meetings are respectively determined with 54%, 95%, and 96% accuracies. Within resolutions, pending solutions and definitively offered solutions are respectively determined with 33% and 38% accuracy. Communication parties are determined with 50% accuracy.

Topic Extraction

The codebook approach provides a basis for assessing the common language used in dialogue-based query mediation. Looking at the nouns and verbs, the basic elements of topics, we see some overlap among researchers and query analysts and because of the different sizes of word sets, we measure overlap with respect to the smaller word set. Looking specifically at noun topics, we find that between researchers and analysts, there are 43%, 39%, and 58% overlaps across Sites 1, 2, and 3 respectively. Between researchers these overlaps are 15%, 41%, 39%, and between query analysts these are 29%, 34%, 37%. The smaller overlaps within the same parties make sense as each critical incident represents a distinct data request. The more significant overlaps in noun topics between opposite parties is reasonable since a common language is needed to mediate information requests. Looking at the most frequent topics in our subsequent noun-phrase extraction, we see that important common topics across all cases include logistic terms such "MRN," "codes," "data," "warehouse," and "ICD9" as well as more medically-focused concepts such as "adenovirus," "hemoglobin," "ulcer," and "creatinine." Common to all sites are terms such as "lab," "dates," "IRB," "gender," and "diagnosis." Common to Sites 1 and 2 are terms such as "admissions" and "months." Sites 1 and 3 share terms like "control," "age," "medication," "reaction," "demographics," "ICD9," and "MRI." For Sites 2 and 3 we see common terms such as "chronic" and "surgery." These frequent topics serve as the core of dialogue-based query mediation.

Task Flow and Temporal Analysis

In determining the most common tasks preceding, concurring with, and following particular task sub-domains across sites, we found that the email tasks surrounding clarifications and task requests, i.e. information requests, are quite similar across sites. Within the ongoing process of BQM, both task request, i.e. information request, and clarification events tend to be preceded by clarifications, task requests, and definitively offered resolutions, and they both tend to be followed by clarifications and definitively offered solutions. Moreover, these two email events tend to coincide with a clarification or task request alongside an offered resolution.

Whereas meetings will attempt to be scheduled following a clarification, these task events tend to coincide with or precede a task request. Completed meetings also coincide with task requests. Task requests can also commonly occur in response to forwarded messages, whereas clarifications can occur in response to follow-ups suggesting that clarifications intuitively

highlight bottlenecks in the information retrieval process.

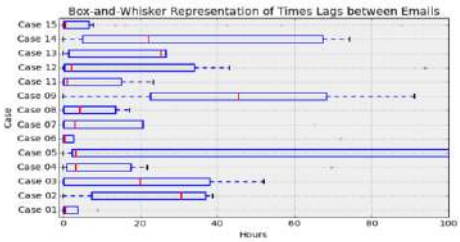


Figure 5 - Distribution of times between emails within a critical incident. Box-plot representations with median and quartiles shown for cases with complete time data. Case 5 has significant variation in email response times with its box-plot extending outside the timescale range used in this comparison chart

There is significant variation in the time durations of our critical incidents (see Figure 5). Analyzing the temporal lags, we find that most frequently preceding time peaks, query analysts have completed a meeting, have offered a resolution, and are currently working on another solution. At the opposite end of the time lag researchers typically contact query analysts with information requests. During the course of exchange, there seems to be no regular pattern among institutions in how the number of distinct task sub-domains change, whether they are generally increasing, decreasing, or peaking midway, emphasizing BQM to be a non-linear process.

Discussion

Communications between query analysts in some of these critical incidents suggests that database querying is not a straightforward process that can always be directly handled by a single mediator. The somewhat smaller proportion of researcher-to-researcher communications suggests that an understanding of the clinical process can also be quite involved. However, these represent a minor proportion of the communication flow and for the most part, BQM is an interaction between opposite parties, mostly having simple communication network structures. That task domains are roughly broken down equally between parties suggests that there is as much of a need and refinement both for the specifications of the clinical process and for the data obtained.

This study complements earlier work in BQM focusing on in-person communication. In-person conversations that comprise non-electronic BQM are an important part of the information retrieval process [4]. Indeed, in 14 out 20 threads that set up meetings, 7 of those threads reference information from that conversation. However, while meetings are important, they are not critically necessary to BQM. Not only are meetings the smallest volume of task domains among our three major task domains, Site 2 does not involve meetings at all and six critical incidents across the entire dataset altogether exclude this task domain. This strongly suggests that BQM, or at least parts of it, can be taken to a computational model.

In fact, there is notable structure and predictability in BQM. On one hand, the volume of distinct events over the course of an exchange does not seem to follow any consistent patterns among our sites. On the other hand, our network analysis of task events surrounding information requests and clarifications confirms the iterative question-and-answer model of BQM discussed in earlier research. Our topic analysis is integral to understanding this iterative nature. Our findings of both logistically-oriented terms such as "codes," and "ICD9," in addition to the medical jargon, can guide the construction of an automated dialogue-based query support, as these codebooks can inform references within automated dialogue-based support that model the actual conversational support found in emails.

Intelligent query formulation relies on the ability to automatically understand the flow of a conversation and we have observed that much of the mediation process can be predicted at a high level. The content of email conversations moreover suggests that BQM is an ideal candidate for the computational design-support paradigm as many of the questions are straightforward inquiries, e.g. questions regarding basic identifying information and questions regarding choices between alternative sets of information. The relevant information for such inquiries can be autonomously accessed by researchers given the appropriate interface support [5; 6].

This study is limited by the small sample of emails from a small number of institutions for 20 critical incidents, thus not providing a complete representation of the entire BQM space. Nonetheless, we offer important initial insights into the dynamics of human-mediated query support, confirming earlier research in BQM as well as providing promising insights into the potential for automating BQM. Our codebook-based algorithm extracted a unique set of codes for a particular domain/sub-domain classification, codes common to two different domains are not used in the determining the classification of a message. This suggests that many domains may occur together, as we see in later analysis.

Conclusions

This study contributes an analysis method for understanding interdisciplinary communicaiton and early insights into the black-box biomedical query process. We conclude that Dialogue-based decision support is needed and feasible.

Acknowledgements

This publication was supported by Grants UL1TR001873 and R01LM009886 from National Institutes of Health of USA.

References

[1] G.W. Hruby, J.J. Cimino, V. Patel, and C. Weng, Toward a cognitive task analysis for biomedical query mediation, *AMIA Jt Summits Transl Sci Proc* **2014** (2014), 218-222.

[2] J.H. Gennari, C. Weng, J. Benedetti, and D.W. McDonald, Asynchronous communication among clinical researchers: a study for systems design, *Int J Med Inform* **74** (2005), 797-807.

[3] E. Coiera, When Conversation Is Better Than Computation, *Journal of the American Medical Informatics Association* **7** (2000), 277-286.

[4] G.W. Hruby, M.R. Boland, J.J. Cimino, J. Gao, A.B. Wilcox, J. Hirschberg, and C. Weng, Characterization of the biomedical query mediation process, *AMIA Jt Summits Transl Sci Proc* **2013** (2013), 89-93.

[5] H. Lieberman and C. Mason, Intelligent agent software for medicine, *Stud Health Technol Inform* **80** (2002), 99-109.

[6] L. Laranjo, A.G. Dunn, H.L. Tong, A.B. Kocaballi, J. Chen, R. Bashir, D. Surian, B. Gallego, F. Magrabi, A.Y.S. Lau, and E. Coiera, Conversational agents in healthcare: a systematic review, *J Am Med Inform Assoc* **25** (2018), 1248-1258.

Address for correspondence

Chunhua Weng, Department of Biomedical Informatics, Columbia University, 622 W 168th Street, PH-20 room 407, New York, NY 10032, USA. E-mail: chunhua@columbia.edu .

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190458

Understanding Patient Information Needs About Their Clinical Laboratory Results: A Study of Social Q&A Site

Zhan Zhang[a], Yu Lu[a], Yubo Kou[b], Danny T.Y. Wu[c], Jina Huh-Yoo[d], Zhe He[b]

[a] *Department of Information Technology, Pace University, New York, NY, USA,*
[b] *School of Information, Florida State University, Tallahassee, Florida, USA,*
[c] *Department of Biomedical Informatics, University of Cincinnati, Cincinnati, OH, USA,*
[d] *Department of Media and Information, Michigan State University, East Lansing, MI, USA*

Abstract

Clinical data, such as laboratory test results, is increasingly being made available to patients through patient portals. However, patients often have difficulties understanding and acting upon the clinical data presented in portals. As such, many turn to online resources to fill their knowledge gaps and obtain actionable advice. In this work, we present a content analysis of the questions posted in a major social Q&A site to characterize lay people's general information needs concerning laboratory test results and to inform the design of patient portals for supporting patients' understanding of clinical data. We identified 15 information needs related to laboratory test results, and clustered them under four themes: understanding the results of lab test, interpreting doctor's diagnosis, learning about lab tests, and consulting the next steps. We draw on our findings to discuss design opportunities for supporting the understanding of laboratory results.

Keywords:

Consumer Health Information, Information Seeking Behaviors, Patient Portals.

Introduction

Advances in personal health record technology, such as online patient portals, empower patients with easy and full access to their clinical data (e.g., laboratory results, radiology reports, and clinician notes) [1]. This access, in line with general interest in patient-centered care, has proven to foster patient engagement, enhance patient-provider communication, and ultimately, improve health outcomes [2]. These potential benefits, along with the financial incentives provided by the U.S. government [3] and the OpenNotes initiative [4], encourage healthcare organizations to increase patients' access to their clinical data via patient portals.

Among the many portal functionalities, access to laboratory test results is an area of high interest to patients; growing evidence suggests that patients are increasingly interested in timely and easy access to laboratory test results [5]. However, patients' current use of test result data is significantly limited due to several reasons [6]. For example, many portals present clinical data to patients in the same way as it is shown to healthcare providers, while patients may not have sufficient health literacy to process and understand the technical nature of the language (e.g., medical jargons) used in the laboratory test reports [7,8]. In addition, patients hope to find useful information, such as actionable knowledge, in online portals, rather than just reviewing the data [9,10]. These findings suggest that while healthcare organizations are increasing patients' access to their clinical data via patient portals, this technology has not adequately met patients' information needs.

Therefore, patients often turn to online resources (e.g., search engines, health forums, and social media) to fill their knowledge gaps. In fact, a recent Pew Research Center study reported that over 70% of adult Internet users in the U.S. searched online for health information [11]. As one of the most popular activities online, health information seeking has been the focus of many studies over the past few decades [12]. However, to date, only a few studies have started looking into patients' online health information seeking behaviors in the context of understanding laboratory results. For example, Reynolds et al. [8] examined the type of supports patients need related to their laboratory data through analyzing questions in an online health forum (medhelp.org). In particular, they found that patients tend to ask questions pertaining to several topics: diagnosis, management/treatment, laboratory report, test, risk, and prognosis. Their study also preliminarily assessed the feasibility of identifying and characterizing the nature of patients' questions related to laboratory results. Building upon their work, we conducted an explorative study to gain further insights into patients' general information needs concerning laboratory test results.

In this paper, we analyzed the questions that users posted on a major social Q&A site, Yahoo! Answers. Among various online forums containing health communities, Yahoo! Answers allows patients to seek information through raising questions and receiving answers from others (e.g., peer patients, health professionals) who are willing to share their knowledge and opinions [13]. Unlike queries submitted to search engines, the questions posted on these platforms are expressed in natural language. These questions also tend to contain more contextual information, such as patients' medical histories and symptoms. Therefore, Yahoo! Answers is a good resource for examining lay people's health information needs [12,13]. As this study is exploratory in nature, we chose to focus on a specific chronic disease, i.e., diabetes, rather than multiple conditions. Diabetes is an ideal condition for us to investigate lay people's information needs regarding laboratory test results. That is, diabetes is recognized as one of the most important public health problems with escalating health concerns [14], requiring long-term management and regular laboratory tests. Addressing the barriers of understanding laboratory test results will benefit a broad population and the society at large. Furthermore, laboratory tests vary across different types of conditions and diseases. Thus, focusing on one condition allowed us to generate comprehensive search terms for data collection.

In this study, we began our inquiry by asking: What are lay people's information needs in making sense of their laboratory test results? We identified 15 information needs related to laboratory test results, and clustered them under four themes: understanding the results of lab test, interpreting doctor's diagnosis, learning about lab tests as a source of information, and consulting the next steps. This study highlights the need to address the gap between patient knowledge and limited contextual information presented on their lab reports.

Methods

Data Collection

Using the application program interface (API) of Yahoo! Answers, we collected a total of 58,422 questions in the diabetes category between 2009 and 2014. The questions were downloaded and loaded to a MySQL database. We then extracted 8,655 posts using keywords suggested by the guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes, such as HbA1c, glucose, and creatinine [15]. The complete search terms and the number of posts retrieved by each term are listed in Table 1. The terms "glucose" OR "blood sugar" yielded the most posts (87.1%). The study was approved by the institutional review boards at Pace University and Florida State University.

Table 1. Search terms and the number of retrieved posts.

Search Terms	Number
"lab" OR "laboratory"	243
"A1c" OR "HbA1c" OR "hemoglobin A1c"	427
"glucose" OR "blood sugar"	7,536
"blood pressure" OR "systolic" OR "diastolic"	338
"creatinine"	111

Data Analysis

We generated a random sample of 1,619 posts of the potentially relevant question posts (8,655 posts containing keywords). Then two researchers independently reviewed posts for relevance. Duplicate or irrelevant posts were discarded. The posts were determined to be irrelevant if they did not contain any laboratory results or questions related to laboratory tests. This screening resulted in 967 posts eligible for further analysis. The relevant posts were then transferred into NVivo, a program for organizing, storing, and manipulating qualitative data. The research team performed content analysis on these relevant posts. The analysis was performed independently by three researchers and consisted of multiple steps (Figure 1).

The first step was to iteratively develop a codebook using the open coding technique. Two coders, C1 and C2, independently analyzed 240 randomly sampled posts until saturation was reached. The initial list of codes was generated and then discussed in a group session to determine which codes to keep, merge, or remove. After the list of codes was set, we created a data dictionary defining each code to standardize the coding process. Our final coding scheme contained a total of 15 codes, which were clustered under four themes: understanding the results of lab test, interpreting doctor's diagnosis, learning about lab tests as a source of information, and consulting the next step (see Table 2).

Next, a third coder (C3) coded 100 randomly sampled posts from the rest of the posts to check for exhaustiveness of the themes. Once confirming that the themes were comprehensive, C1 and C2 independently coded another set of 100 posts to check for inter-rater agreement using Cohen's Kappa coefficient. The resulting kappa value was analyzed using the kappa interpretation scale suggested by Landis and Koch [16]. The coders presented "Almost Perfect" agreement (kappa value of 0.851). The disagreements were mainly due to the interpretive differences attributed to "Confused about doctor's suggestions or diagnosis" and "Seeking confirmation of doctor's diagnosis" codes; all the disagreements were resolved through discussion. After conflict resolution, C1 and C2 coded the rest of the posts to conclude the analysis.

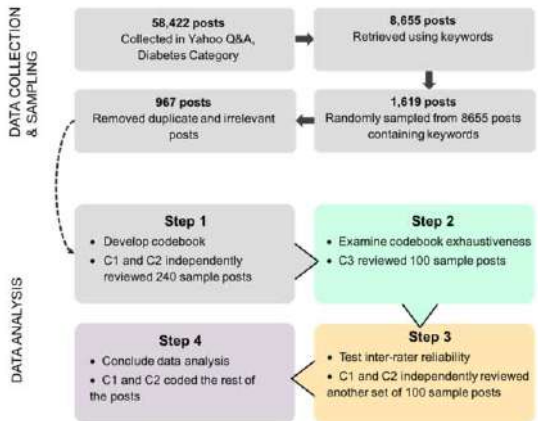


Figure 1– Data Collection, Sampling, and Analysis Process

Results

People come to Yahoo! Q&A to request advice, suggestions, information pertaining to laboratory test results. Their information needs are multi-faceted, manifested in their different but related questions. In this section, we will describe these information needs that people expressed in the questions. We will use representative quotes to illustrate the salient themes.

Understanding the Results of Lab Test

Requests for understanding laboratory test results were by far the most common in this sample (85% of the total posts). That is, posters shared parts of the report content and asked the community to explain their lab test results. As Table 2 shows, posters needed help in understanding different aspects of their lab content, including the meaning of lab value, specific terminology, and the effects and causes of abnormal and/or inconsistent results.

The most common questions in this category were related to understanding the meaning of lab values (74% of the total posts). We found that people had different needs in making sense of their lab results, which may be due to different levels of health literacy, knowledge and experience [17]. For example, a post sought clarification on whether a specific lab value falls into the normal range: *"Is GFR of 73 and creatinine 1.1 normal?"* In other cases, posters often asked the community for diagnosis or opinions, by providing substantial portions of their reports and relevant medical history, medication information, and symptoms. For example, a post sought opinions on what the lab results indicated:

"I am an 18-year-old male. […] Some of the statistics from the report were as follows: high alkaline phosphatase levels, elevated T3, elevated Hbg levels. […] My AbC1 level was 6.1.

What do these elevated levels seem to point to? Can anyone make sense of what might be wrong from my lab results?"

Posters also requested explanation of technical jargon, i.e., terminology. For example, a post asked for the clarification of a specific term: *"Does anyone know what is the meaning of 'Lymph' on blood labs?"* This observation suggests that people have difficulty understanding medical terminologies, even though some patient portals have started implementing consumer-friendly vocabularies [18].

Finally, people wanted to know the effects and/or causes of abnormal lab results. For example, a post asked for advice on the consequences of high creatinine level: *"A recent pathology test states that my creatinine is 6.28. […] What are the effects of such high levels?"* In other cases, people expressed concerns about inconsistent lab results they received from different laboratories or over a period of time, as one post stated: *"My creatinine level increased from 1.0 to 1.1 with a span of 10 days' period. What is the reason?"*

Table 2. Summary of themes. Some posts fell into multiple themes, so percentages add up to more than 100%.

Theme	% (n)
Understanding the results of lab test	
Meaning of lab value	74.3% (418)
Specific terminology	1.7% (9)
The effect of abnormal/inconsistent results	1.1% (6)
The cause of abnormal/inconsistent results	7.3% (41)
Interpreting doctor's diagnosis	
Confused about doctor's suggestions/diagnosis	1.6% (9)
Seeking confirmation of doctor's diagnosis	4.6% (26)
Concerned about doctor's misdiagnosis	0.7% (4)
Learning about lab tests as a source of information	
Inquire information about a specific lab test	34.1% (192)
Ask for lab test recommendations	1.2% (7)
Look for comparison among tests	1.6% (9)
Concerned about lab procedure	2.1% (12)
Consulting the next steps	
Healthcare consultation	27.7% (156)
Treatment options	4.1% (23)
Taking medication	4.1% (23)
Life-style	21.3% (120)

* The percetages are calculated using the number of posts in each category divided by the total number of posts (N=967).

Interpreting Doctor's Diagnosis

Sometimes people posted questions after they discussed the results with their physicians and cited several reasons for doing this. First, people may have doubts about, disagree with, or mistrust their physician's diagnosis, thus seeking a second opinion on their physician's conclusions and/or interpretations (referred to as *seeking confirmation of doctor's diagnosis* in Table 2). For example, in one post the person stated: *"My 4-year-old [child] had all the symptoms and signs of type 1 diabetes so his doctor run test for him. What came back was Glucose, Blood 71, Insulin, Fasting 1.2, Low, C Peptide 0.4 Low. Doctor says there are a few low things, but nothing to worry about. I in my gut don't think that is right. Can someone else help me out?"*

Second, people seemed to be confused about their physician's diagnosis or suggestions as to what to do next and whether or not the treatment is needed. Therefore, they turned to online forums to seek clarification or explanation regarding the information they received from their physician: *"Why do I need to test my creatinine level every three months as being suggested by my doctor?"*

Lastly, a few posters talked about perceived misdiagnosis by their physicians. Often the language used by these patients indicated some level of distress, fear, or other negative emotions. In one post, for example, the poster wrote: *"My wife's doctor, at a prominent San Diego hospital, failed to notice her declining kidney function until she was in end stage kidney failure. […] Is it common for physicians to ignore kidney function and obsess over diabetes labs?"*

These findings reveal a communication gap between health care providers and consumers. Misunderstanding or confusion about doctor's diagnosis may adversely affect patients' access to health information, resulting in poor patient understanding, trust, and satisfaction.

Learning about Lab Tests as a Source of Information

This category concerns questions related to lab test itself. For example, lack of sufficient knowledge about lab tests led people to inquire general information about them (referred to as *inquire information about a specific test* in Table 2), as shown in one post: *"What is creatinine cholesterol?"* In other cases, people asked for some other general information about lab tests, including relationship between lab tests and symptoms (e.g., *why are urea and creatinine levels raised with dehydration?*), how often should one take a specific lab test (e.g., *how often should creatinine and eGRF levels be checked?*), and treatment options (e.g., *my creatinine is 1.6, what is the treatment for it?*).

People also inquired about the diagnostic abilities of a specific test and sought recommendations on which lab test to take (referred to as *ask for lab test recommendations* in Table 2). As this data sample focused on a diabetes online community, the questions therefore were related to lab recommendations for diabetes: *"Which laboratory test is diagnostic for diabetes?"* Similarly, people also sought comparison among different types of test (referred to as *look for comparison among tests* in Table 2): *"Advantages and disadvantages of creatinine clearance test vs. plasma creatinine?"*

Lastly, posters asked questions about the lab procedure. Sometimes, they posted questions while they were waiting for the tests. At this stage, posters asked questions concerning various aspects of the lab procedure, such as what they will go through during the test: *"I am going to the lab to get tested for hypoglycemia (low blood sugar) tomorrow, what exactly will they do?"* Others looked for information as to what they should do or not do to prepare for the upcoming tests: *"This is a lab test for diabetes, blood sugar, cholesterol etc. And I am wondering how long should I fast and can I drink water?"* Similarly, people also posted questions after taking their tests to inquire the turnaround time of their test results: *"How long should it take for a doctor office to call you about lab results?"* These posts tended to exhibit language indicative of distress: *"I had lab work done last Thursday and I am still waiting to hear what my A1C and all else [the doctor] had me*

tested for. Shouldn't they call you with results sooner? What if something is really wrong?"

Consulting the Next Steps

Sometimes people also consulted the community about what they should be doing next. One reason was that people may be waiting for an appointment to discuss the results with their physician, but they wanted to obtain actionable suggestions from the online community first: *"I have lupus [and] my routine blood work shows the ck enzyme at 271 (ref range is 26-192). I have an upcoming doctor appoint. What can I do?"*

They also asked for the community's assistance in assessing the need for a healthcare consultation or further lab test (referred to as *healthcare consultation* in Table 2). For instance, a poster expressed the lack of confidence in the accuracy of lab results and asked for advice as to if it is necessary to re-do the test or take a different test: *"High blood sugar – should I get a second opinion from a different lab? This is too important not to double check with a different lab; last reading was 6.4. This doctor was wrong before about different things."*

Of these posts, people also asked for treatment advice (referred to as *treatment options* in Table 2). For example, one poster wrote: *"A recent pathology test states that my creatinine is 6.28. Does it require dialysis to be done? What can cure this high level?"* In such cases, people also wanted to know what medication and/or whether changing life style (e.g., diet and exercise) could be of any help (referred to as *taking medication* and *life-style* in Table 2), as one post stated: *"My mother aged 45 and has only one kidney. [Her] creatinine level [is] 4.2, Urea [is] 50. What diet she should take and what medicine?"*

Discussion

In the study, we characterized lay people's general information needs related to laboratory test results, such as understanding test results, interpreting doctor's diagnosis, learning about lab tests, and making decisions on the next steps. This study presents an early investigation for our long-term goal of guiding the design of patient portals that can provide more informative and personalized healthcare information. Building upon prior work [8], our study provides a more comprehensive, fine-grained description of lay individual's information needs about their laboratory test results. For example, Reynolds et al. [8] highlighted that patients have confusion about the laboratory report; our study further revealed the aspects of laboratory report that patients had difficulties with, such as the meaning of lab values, medical terminologies, and the causes and effects of abnormal lab results. While this study only examined a subset of questions in an online forum setting, our findings reveal that people need support in interpreting and acting on clinical data, as well as making personalized decisions. Below, we draw on our findings to discuss five design opportunities for supporting the understanding of laboratory results in patient portals.

Providing consumer-friendly and credible information to assist the reading of lab results. Our findings suggest that the design of test results in patient's portal seems to assume that patients have sufficient medical knowledge about their test results. Consequently, patients often did not receive explanatory information or result interpretation in the portal at the time they received the result, and they would search online to make sense of their results. It is therefore crucial to provide more useful information that patients need at the point of viewing their laboratory results in patient portals. For example, patient portals could provide links to consumer-friendly and credible information sources (e.g., entries in MedlinePlus) to help patients better understand the lab results; the portal could also suggest basic healthcare management advices, such as diet and life style.

Accomendating people with different health literacy. People have different levels of health literacy and numeracy as well as potential biases and personal beliefs. For patients who were recently diagnosed, they may not be literate enough to understand the terminology and the results, and thus may ask basic questions such as whether a particular lab value falls into the normal range. In contrast, some patients who have had chronic conditions may have been self-educated on relevant health knowledge (e.g., medical terminology, normal ranges of a test) and therefore need help with more comprehensive questions (e.g., how to interpret the lab results in the context of their medical history). Given such a fact, patient portals need to be designed taking into consideration of people's health literacy differences [19].

Considering the temporality and illness trajectory of patients. We also observed that patients' information needs had a temporal dimension—the nature and extent of the needs may be different at different stages of patients' illness trajectory [12]. For example, right before getting a medical test, patients may want to know how to prepare for the test and what they will go through during the test. Upon receiving their test results, patients may ask for interpretation of what the results mean and what they should be doing next (e.g., make an appointment with their physicians). This observation shed light on portal design with regard to temporal organization of information materials so as to provide relevant health information to patients according to their illness trajectory.

Facilitating shared decision making through personalized and contextualized information along with lab results. An interesting observation is that patients provided contextual information (e.g., medical history, symptom) along with their lab results in order to seek personalized advice and treatment options. This observation suggests that the same lab results may have different indications in different contexts (e.g., family history). In addition, prior work has recognized that personalized healthcare information within a shared-decision making framework leads to better patient engagement, better outcomes, and an increased level of trust between healthcare providers and patients [20]. As such, patient portals should provide more personalized content.

Supporting the sharing of personal stories between patients who are "in the same boat". Sometimes, patients sought health information due to their suspicion about a certain diagnosis made by their physician. This means that patients not only need objective explanations of terms and values in test results, but also other patients' opinions and experiences. Such behavior constitutes reflection upon and distrust in doctors' explanations. It seems that when authoritative explanations lost credibility in certain cases, patients were in urgent need of a second opinion, especially from patients with similar symptoms and conditions. This observation suggests that a social network in patient portals could benefit patients by connecting them with peers (even anonymously) who have similar conditions. This also suggests that patient portals should provide a more streamlined communication channel between healthcare providers and patients in order to resolve any misunderstandings in a timely manner.

Limitation and Future Work

Our study has several limitations. First, our study focused on a single disease, namely diabetes, one type of health information, laboratory test results, and one health forum, Yahoo! Answers. While our findings pertain to the

characteristics of the specific domain, the results may not be generalizable to other types of diseases and types of health information. Our future work will expand to other health conditions (e.g., cancers), other health forums (e.g., eHealth.com, healthboard.com), and include other types of health information (e.g., radiology report, physician notes, discharge summaries) to assess the generalizability of our findings. Second, we did not discuss how the questions were answered on this Q&A site. In our future work, we will synthesize the types of information people gain from the online communities and how these answers were constructed to meet their needs. Lastly, due to social media data availability, we could not collect posters' demographic data, such as level of disease severity, gender, age, and different stages of life/illness trajectory.

Conclusions

This study explored lay people's various information needs related to lab results through analyzing forum posts collected from a social Q&A site. Our results highlighted the need to address the gap between patient knowledge and limited contextual information presented on their lab reports, and provide essential insights into improving the design of patient portals to fully meet patient needs in understanding the lab results. Our findings provide a foundation for our future work, including qualitative studies (e.g., interview with clinicians and patients) and analysis of medical record data to understand how to best provide personalized information and present clinical data in patient portals.

Acknowledgements

We would like to thank Dr. Sanghee Oh for collecting the data from Yahoo! Answers and providing the data for this study.

References

[1] Individuals' Ability to Electronically Access Their Hospital Medical Records, Perform Key Tasks, in https://www.aha.org/guidesreports/2016-07-14-trendwatch-individuals-ability-electronically-access-their-hospital. Accessed November 10, 2018.

[2] S. Peacock, A. Reddy, S.G. Leveille, J. Walker, T.H. Payne, N.V. Oster, and J.G. Elmore, Patient portals and personal health information online: perception, access, and use by US adults. *Journal of the American Medical Informatics Association* 24 (2016), e173-e177.

[3] S.T. Mennemeyer, N. Menachemi, S. Rahurkar, and E.W. Ford, Impact of the HITECH act on physicians' adoption of electronic health records. *Journal of the American Medical Informatics Association* 23 (2016), 375-379.

[4] OpenNotes and AMIA join forces to improve patient access to health records, in https://medcitynews.com/2017/11/opennotes-amia/. Accessed November 10, 2018.

[5] F. Pillemer, R.A. Price, S. Paone, G.D., Martich, S. Albert, L. Haidari, G. Updike, R. Rudin, D. Liu, and A. Mehrotra, Direct release of test results to patients increases patient engagement and utilization of care. *PLoS One* 11 (2016), 1–9.

[6] P.M. Gee, D.A. Paterniti, D. Ward, L.M.S. Miller, E-Patients perceptions of using personal health records for self-management support of chronic illness. *CIN: Computers, Informatics, Nursing* 33 (2015), 229-237.

[7] M.K. Hong, C. Feustel, M. Agnihotri, M. Silverman, S.F. Simoneaux, and L. Wilcox, Supporting families in reviewing and communicating about radiology imaging studies. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 5245-5256. ACM, New York (2017).

[8] T.L. Reynolds, N. Ali, E. McGregor, T. O'Brien, C. Longhurst, A.L. Rosenberg, S.E. Rudkin, and K. Zheng, Understanding Patient Questions about their Medical Records in an Online Health Forum: Opportunity for Patient Portal Design. In *American Medical Informatics Association Annual Symposium Proceedings*, pp. 1468-1477. American Medical Informatics Association (2017).

[9] T.D. Giardina, J. Baldwin, D.T. Nystrom, D.F. Sittig, and H. Si, Patient perceptions of receiving test results via online portals: a mixed-methods study. *Journal of the American Medical Informatics Association* 25 (2017), 440-446.

[10] J.M. Alpert, A.H. Krist, R.A. Aycock, and G.L. Kreps, Applying multiple methods to comprehensively evaluate a patient portal's effectiveness to convey information to patients. *Journal of Medical Internet Research* 18 (2016), e112.

[11] K. Zickuhr, Generations 2010. Pew Internet and American Life Project (2010).

[12] Y. Zhang, Contextualizing consumer health information searching: an analysis of questions in a social Q&A community. In *Proceedings of the 1st ACM International Health Informatics Symposium*, pp. 210-219. ACM, New York (2010).

[13] Z. He, Z. Chen, S. Oh, J. Hou, and J. Bian, Enriching consumer health vocabulary through mining a social Q&A site: A similarity-based approach. *Journal of biomedical informatics* 69 (2017), 75-85.

[14] World Health Organization, Global Report on Diabetes, in http://apps.who.int/iris/bitstream/10665/204871/1/9789241565257_eng.pdf?ua=1. Accessed November 10, 2018.

[15] D.B. Sacks, M. Arnold, G.L. Bakris, et al. Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. *Diabetes Care* 34 (2011): e61-99.

[16] J. Landis and G.G. Koch, The measurement of observer agreement for categorical data. *Biometrics* 33 (1977), 159-174.

[17] U. Sarkar, A.J. Karter, J.Y. Liu, N.E. Adler, R. Nguyen, A. Lopez, and D. Schillinger, The literacy divide: Health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of Northern California (DISTANCE). *Journal of Health Communication* 15 (2010), 183-196.

[18] T. Irizarry, A.D. Dabbs, and C.R. Curran, Patient portals and patient engagement: a state of the science review. *Journal of Medical Internet Research* 17 (2015), e148.

[19] M.S. Park, Z. He, Z.W. Chen, S.H. Oh, and J. Bian, Consumer's use of UMLS concepts on social media: Diabetes-related textual data analysis in blog and social Q&A sites. *JMIR Medical Informatics* 4 (2016), e41.

[20] A.K. Ghalibaf, Z.M. Khorasani, M. Gholian-Aval, H. Tabesh, and M. Tara, Design, implementation, and evaluation of self-describing diabetes medical records: A pilot study. *JMIR Medical Informatics* 5 (2017), e10.

Address for correspondence
Zhe He, PhD. School of Information, Florida State University, 142 Collegiate Loop, Tallahassee, Florida, USA. Email: zhe@fsu.edu. Phone: 001(850)644-5775

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190459

How Do Healthcare Professionals Personalize Their Software? A Pilot Exploration Based on an Electronic Health Records Search Engine

Kai Zheng[a], Yunan Chen[a], Julia Adler-Milstein[b], Andrew L. Rosenberg[c], Danny T.Y. Wu[d], Qiaozhu Mei[e], David A. Hanauer[f]

[a] Department of Informatics, University of California, Irvine, Irvine, California, USA,
[b] Department of Medicine, University of California, San Francisco, San Francisco, California, USA,
[c] Department of Anesthesiology, University of Michigan, Ann Arbor, Michigan, USA,
[d] Department of Biomedical Informatics, University of Cincinnati, Cincinnati, Ohio, USA,
[e] School of Information, University of Michigan, Ann Arbor, Michigan, USA,
[f] Department of Pediatrics, University of Michigan, Ann Arbor, Michigan, USA

Abstract

To improve user experience, many health IT systems provide personalization options allowing end users to tailor the software to their needs and preferences. However, few studies have investigated if healthcare professionals actually make full use of this feature. As an initial step towards understanding end users' software personalization behavior in healthcare, we conducted a pilot study to examine how clinicians, staff, and researchers customized a search engine designed to facilitate information retrieval from electronic health records. The results show that a majority of the end users (82.4%) did not make an effort to modify the system's default settings. Among those who did, they more often changed its 'look-and-feel' than its functionality offerings. We conclude that future research is warranted to study the rationale underlying healthcare professionals' software personalization decisions both to optimize user experience and to avoid building complex and costly personalization options that are unused or underutilized.

Keywords:

Software, Electronic Health Records

Introduction

The lack of usability has been a prominent barrier to the adoption and use of current generation health IT systems [1–4]. Numerous efforts have been attempted to address the issue, e.g., through introducing usability guidelines and incorporating user-centered design requirements into the health IT certification processes [5, 6]. However, no "one-size-fits-all" solutions exist capable of meeting the diverse needs of all healthcare organizations and individual users [7]. Different healthcare organizations have distinct contexts and requirements. At the individual level, users of the same profession and medical specialty may have vastly differing opinions regarding how a health IT system should be designed, reflecting their unique cognitive styles, job routines, prior experiences, and perception of the role of technology.

Through software modularization and local configuration, vendor products can be tailored to meet the needs of specific healthcare organizations. To accommodate variation at the individual user level, one approach is to allow the software to be "personalized" based on the preference of each of the individual users [8]. Personalization strategies have been widely applied in modern software, including operation systems, productivity suites, and games (e.g., Windows' "Control Panel" and Macintosh's "System Preferences") [9]. Modifying a software system's default settings allows end users to not only fine-tune its visual appearance, such as font and color, but also its functions and interaction modalities, such as the expert vs. novice mode.

A significant challenge to software personalization is balancing flexibility and uniformity. While a rigid software design could diminish user experience and result in inefficiencies and dissatisfaction, a fully flexible system that does not work optimally without extensive customization may also cause confusion and fail to deliver 'best' experience uniformly to all users [8]. Further, systems that offer a wider range of personalization options are more difficult to develop, maintain, and support. Different software companies seem to favor distinct design philosophies. For example Apple Inc. values simplicity and provides users as few personalization options as possible, whereas other companies opt for the other extreme.

Health IT personalization design decisions are by and large made by vendors and software developers. While work has been done in other industries [8, 9], no research, to the best of our knowledge, has investigated how healthcare professionals personalize their software. This paucity of knowledge could result in suboptimal software designs that adversely affect user experience. It may also incur unnecessary costs in implementing complex and costly personalization options that may later prove to be of little value. In this paper, we report a pilot exploration that represents an initial step toward addressing this knowledge gap.

The system that we studied, the Electronic Medical Record Search Engine (EMERSE), is a full-text search engine designed to assist clinicians, administrators, and researchers retrieve information from narrative documents stored in electronic health records (EHR) [10]. It provides many personalization options for end users to tweak its look-and-feel, as well as functionality, such as how synonyms should be handled in a search. This study leveraged the audit trail data automatically captured in EMERSE, recording each user's interactions with the system, including modifications made to its default preference settings. In the study, we are primarily interested in answering three empirical questions: (1) whether the users of EMERSE modified the system's default appearance and functions; (2) what types of default settings tended to be altered more often than other settings; and (3) whether certain user characteristics, such as gender and clinical role, may be associated with how end users personalized the software.

Methods

The EMERSE System and the Empirical Study Setting

EMERSE is an application developed at Michigan Medicine [10]. The system was designed to facilitate information retrieval from EHRs to support administrative coding, chart auditing, and patient eligibility screening for clinical trials [11–13]. Its target users include clinicians, medical coding personnel, and clinical and translational researchers.

The system functions similarly to Google™. End users construct and submit their search queries via a web-based interface. Then, the system processes the queries in its information retrieval engine against a large repository of EHR documents, and returns documents ranked by key attributes such as relevance, genre, and time of creation. Many visual aids help users browse returned documents to quickly identify the information of interest. Since its initial deployment in 2005, the system has been routinely used by many clinical, operational, and research teams at Michigan Medicine.

Table 1 (on page 5) lists all personalization options that EMERSE offers. Items 1 through 10 fine-tune the visual appearance of the system (e.g., color and font); Item 11 instructs the system how to highlight matched keywords in monochrome printing; and Items 12 through 15 set parameters governing the system's functional behavior such as how to process user-submitted search queries and how to retrieve documents from the backend EHR database. Some simple personalization options are provided on the screen where actions take place. More complex, system-wide options can be modified through a dedicated "preferences" feature.

Figures 1 through 3 are screenshots of several representative personalization settings in the system. As shown in Figures 1 and 2, most of these settings are straightforward to use and only require one or two mouse clicks. The only exception is the heat-map preference pane (Figure 3), which offers a range of relatively complex parameters that users can tweak to control heat-map visual elements, such as colors and color gradations, for highlighting the relevance of search results.

Figure 1- A sample personalization setting for altering the system's visual appearance

Figure 2 - A sample personalization setting for altering the system's functionality

Empirical Dataset and Data Analysis Methods

The empirical data analyzed in this study were collected from

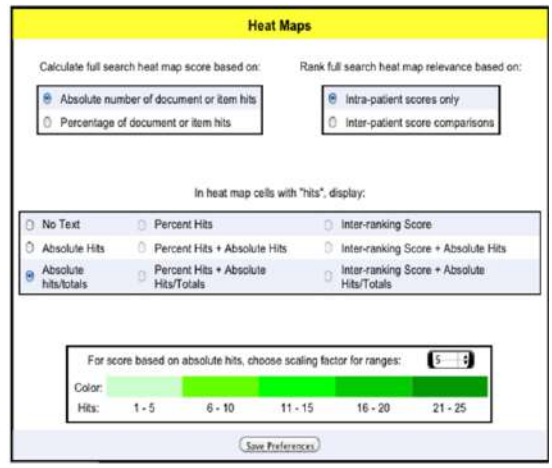


Figure 3 - Heat-map settings: the most complex personalization option provided in EMERSE

the audit trail logs automatically recorded in EMERSE. A total of 558 active users were registered when this study was conducted. Inactive users, who created an account but logged into the system less than five times, were excluded.

A user's usage level of the system might be a factor affecting personalization behavior: the more often a user used the system, the more likely s/he would make an effort to personalize it for better user experience. To test this hypothesis, we used "the number of times a user logged into the system" as the measure to determine a user's level of utilization (hereafter referred to as "system usage"). Other potential measures, such as number of searches performed or number of EHR documents retrieved, were also considered, but were not used as they are heavily task dependent. Number of searches performed, for example, might be artificially inflated as a user could revise and resubmit the same search query multiple times using different combinations of date ranges and patient lists.

Because of the small sample size of this pilot study, we only included two dichotomous variables to characterize the users: gender (female vs. male) and clinical role (clinician vs. non-clinician). Clinicians were those who had active patient care duties at Michigan Medicine based on the employee database provided by the human resources office. Non-clinician users were typically administrators, medical coding staff, IT personnel, and researcher coordinators and research assistants. We included the clinical role variable because frontline clinicians might use the system in distinctive ways compared to other 'back-office' types of users.

To determine whether the three explanatory variables examined in the study, i.e., system usage, gender, and clinical role, may predict an end user's software personalization behavior, we estimated three logistic regression models in which the dependent variable respectively represents (1) whether the user accessed the preferences feature provided in EMERSE (Model 1); (2) whether the user modified any of the system's default settings (Model 2); and (3) if the default settings were modified, whether the user customized the system's functionality or its visual appearance (Model 3). In addition, we also performed a drill-down analysis to examine if heavy users of the system might demonstrate distinct software personalization behavior compared to casual users.

To determine an approximate cutoff between 'experienced' and 'novice' users, we iterated the Model 2 regression by incrementally dropping infrequent users until the association

between system usage and personalization behavior disappeared. Then, we re-estimated each of the models using the data respectively generated by the 'experienced' user group and the 'novice' user group. In all models, we treated system usage as a continuous variable, and gender and clinical role as binary variables.

Results

Table 2 shows the descriptive statistics of the data. The 558 active EMERSE users logged in 157 times on average (median: 42); 378 of them were females (67.7%), and most were non-clinicians (83.3%). Less than one third of the users accessed the preferences feature provided in the system (29.9%); only 98 users (17.6%) actually modified the system's default settings. Among those who did, only 17, or 17.3% of the 98 users who personalized modified the default functionality settings.

Table 2 - Descriptive statistics

Measure	Result
Number of active users	558
Female	378 (67.7%)
Clinician user	93 (16.7 %)
Users who accessed any of the personalization settings	167 (29.9%)
Users who modified default settings	98 (17.6%)
Users who only modified default settings on functionality	17 (17.3% of 98)

Table 3 reports the modification frequency of personalization options. A majority of the frequently altered settings were related to the system's visual appearance. Very few functional settings were ever changed. Heat-map settings, the most complex personalization option provided in EMERSE (Figure 3), were barely touched by users. Nonetheless, "Heat-Map: Color Palette" was the second most frequently modified personalization setting. As detailed in Table 1, this setting is provided on the search results screen that users could access without invoking the heat-map preference pane.

Table 3 - Modification frequency of personalization settings

Type	Personalization setting	Modifying freq.
Visual effect	Highlight Viewed Notes	41
Visual effect	Heat-Map: Color Palette	39
Visual effect	Surrounding Context	19
Print	Highlight in Print	22
Visual effect	Color of Viewed Notes	13
Functional	History of Recent Queries	10
Functional	Spelling Alternatives	5
Visual effect	Line Wrapper	3
Visual effect	Heat-Map: Text	2
Functional	Synonyms: Negation List	2
Functional	Auto-Correct Search Terms	2
Visual effect	Heat-Map: Scaling Factor	1
Functional	Synonyms: Search Query	1
Visual effect	Show All Matches	0
Visual effect	Heat-Map: Score	0
Visual effect	Heat-Map: Relevance	0
Functional	Recommend Synonyms	0

Table 4 shows the regression results of the logistic models. System usage is significantly associated with whether a user would access (Model 1) the personalization options or change (Model 2) the system's default settings ($P < 0.05$ and 0.01, respectively). While gender does not predict a user's actual personalization behavior, females were more likely than males to view the system's personalization options (Model 1; $P < 0.05$). "Clinical role" does not predict a user's personalization behavior; clinicians and non-clinicians behaved similarly across all models. Lastly, whether a user would only modify the system's visual appearance, rather than its functionality, is not associated with any of the explanatory variables (Model 3).

Table 4 - Logistic regression results

Variable	Model 1: Assessing personalization settings		Model 2: Modifying system defaults		Model 3: Modifying visual appearance only	
	Odds Ratio	95% CI	Odds Ratio	95% CI	Odds Ratio	95% CI
System usage	1.00*	[1.00, 1.00]	1.00**	[1.00, 1.00]	1.00	[1.00, 1.00]
Female	1.54*	[1.01, 2.35]	1.22	[0.74, 2.03]	1.42	[0.44, 4.54]
Clinician user	1.04	[0.62, 1.76]	1.05	[0.56, 1.97]	0.54	[0.14, 2.08]

* $P < 0.05$; ** $P < 0.01$.

The drill-down analysis results show that the significant association between system usage and personalization behavior disappeared among those who logged in 48 times or more often (mean: 316; median: 137). Among the 'experienced' users ($N = 259$), 81 were females (31.3%) and 198 were non-clinician users (76.4%). With this information, we re-estimated each of the three models and found that among the 'experienced' users, gender no longer predicts their viewing of the personalization options. The results on the effect of clinical role continue to hold for both usage groups.

Discussion

Software personalization is a common strategy to accommodate heterogeneous needs and preferences by individual end users, especially when customization is flexible, and easy to use and understand [14]. In addition to improving utility and efficiency, personalization is associated with increased sense of control and identity [15]. However, in the human–computer interaction literature, it has been documented that users often do not take advantage of the benefits that personalization offers [8, 16, 17]. For example, Mackay (1991) studied how users personalized a Unix environment and found that most did not want to 'sacrifice' their time to modify the software's default settings [8]. Likewise, Manber et al. (2000) examined how web surfers personalized Yahoo! Applications, and concluded that people simply "take what is given to them and never customize" [16]. The results of this pilot study align with these earlier findings, as less than one third of EMERSEs active users accessed any of its personalization options, and only 17.6% modified the default settings.

This exploratory study is not positioned to provide direct evidence as to why the EMERSE users did not fully utilize personalization. The literature does provide some leads. The time and mental effort it takes to customize a software system is frequently quoted as a key reason for end users' lack of personalization [8]. Further, people fear errors of commission more than errors of omission (omission bias), thus are willing to forgo improvement opportunities in order to avoid making a bad choice [18]. People are also change averse (status quo bias), and tend to trade longer-term benefits (e.g., better future user experience) for short-term discomfort (e.g., coping with

suboptimal default settings) [8, 19]. This law of "power of defaults" has been noted in medicine, and advocated as possible leverage to improve quality of care through its strong influence on the decision-making processes of patients, families, and providers [20].

Although inconclusive and not necessarily generalizable, the findings of this pilot exploration might demonstrate the potential importance of 'defaults' in health IT design and the need to carefully calibrate the default behavior of a health IT system before releasing to end users. For example, as shown in the study, considerable time and effort was spent on building complex personalization settings in EMERSE, such as the heat-map preference pane, which were barely utilized by any of the end users. This suggests that understanding healthcare professionals' personalization rationale is crucial to informing design decisions, such as what personalization options to offer, and what not to, in order to avoid unnecessary software development costs. Further, the fact that the EMERSE users frequently modified the "Heat-Map: Color Palette" setting suggests the users still desired and were willing to personalize, but only through intuitive and readily accessible options that do not require laborious and cognitively challenging steps.

This study has several limitations. First, we only investigated a homegrown EHR search engine designed to support very specific information retrieval tasks. The idiosyncrasies of the system and of its user population may thus limit the generalizability of our study findings. Second, while this pilot exploration quantifies the EMERSE users' personalization behavior, it does not provide direct insights as to why some users chose to personalize the software and more importantly, why some others did not. Future work should therefore use qualitative approaches, such as observations and interviews, to better understand the rationale underlying healthcare professionals' software personalization decisions. Third, due to the small sample size, we were only able to characterize the users using two dichotomous variables: gender and clinical role. Future work may consider incorporating additional individual characteristics, such as age, medical specialty, computer literacy, and clinical experience.

Conclusions

In this paper, we reported a pilot exploration that examined whether and how end users utilized the personalization options in an EHR search engine system. We found that a majority of the users did not make an effort to modify the system's default settings. When they did, most modifications were related to the system's visual appearance rather than its functionality offerings. The lack of end user adoption of the personalization feature might result in lost opportunities for improving the system's performance and user experience. It might also diminish the value of building and implementing complex and costly personalization options in the system. As personalization can be essential to the usability and effectiveness of computerized systems in healthcare, future work is needed to develop a better understanding of healthcare professionals' software personalization behavior in order to inform more useful and more cost effective health IT development efforts.

Acknowledgements

This study was supported in part by the US National Cancer Institute through Grant U24CA204863, and the US National Center for Advancing Translational Sciences through Grant UL1TR001414.

References

[1] M.A. Ellsworth, D. Mikhail, J.C. O'Horo, et al., An appraisal of published usability evaluations of electronic health records via systematic review, *J Am Med Inform Assoc*, **24** (2017), 218–226.

[2] National Research Council, *Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions*, Washington, DC: National Academies Press, 2009.

[3] B. Middleton, M. Bloomrosen, M.A. Dente, et al. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA, *J Am Med Inform Assoc*, **20** (2013), e2–8.

[4] AmericanEHR, *Survey on Physician Use of EHR Systems - Ease of Use Baseline Data*, http://www.americanehr.com/research/reports/use-of-ehr-systems-ease-of-use-baseline-data.aspx, accessed November 17, 2018.

[5] S.Z. Lowry, M.T. Quinn, M. Ramaiah, et al., *Technical Evaluation, Testing, and Validation of the Usability of Electronic Health Records*, NISTIR 7804 National Institute of Standards and Technology, 2012.

[6] The US Office of the National Coordinator for Health Information Technology, *Test Procedure for §170.314(g)(3) Safety-Enhanced Design*, https://www.healthit.gov/sites/default/files/170.314g3safetyenhanceddesign_2014_tp_approved_v1.3_0.pdf, accessed November 18, 2018.

[7] B.T. Karsh, M.B. Weinger, P.A. Abbott, et al., Health information technology: fallacies and sober realities, *J Am Med Inform Assoc*, **17** (2010), 617–623.

[8] W.E. Mackay. Triggers and barriers to customizing software, *Proceedings of the 9th SIGCHI Conference on Human Factors in Computing Systems (CHI '91)*, 1991, 153–160.

[9] R. Oppermann, and H. Simm, Adaptability: user-initiated individualization, In: Oppermann R, ed. *Adaptive User Support*. Erlbaum Associates Inc., Hillsdale, NJ, 1994.

[10] D.A. Hanauer, Q. Mei, J. Law, et al., Supporting information retrieval from electronic health records: A report of University of Michigan's eight-year experience in developing and using the Electronic Medical Record Search Engine (EMERSE), *J Biomed Inform*, **55** (2015), 290–300.

[11] D.A. Hanauer, M.J. Englesbe, J.A. Jr Cowan, et al., Informatics and the American College of Surgeons National Surgical Quality Improvement Program: Automated processes could replace manual record review, *J Am Coll Surg*, **208** (2009), 37–41.

[12] L. Seyfried, D.A. Hanauer, D. Nease, et al., Enhanced identification of eligibility for depression research using an electronic medical record search engine, *Int J Med Inform*, **78** (2009), e13–18.

[13] K. Zheng, Q. Mei, and D.A. Hanauer, Collaborative search in electronic health records, *J Am Med Inform Assoc*, **18** (2011), 282–291.

[14] J. McGrenere, R.M. Baecker, and K.S. Booth, A field evaluation of an adaptable two-interface design for feature-rich software, *ACM Trans Comput Hum Interact*, **14** (2007), Article 3.

[15] S. Marathe, and S.S. Sundar What drives customization? Control or Identity? *Proceedings of the 29th SIGCHI Conference on Human Factors in Computing Systems (CHI '11),* 2011, 781–790.

[16] U. Manber, A. Patel, and J. Robison, Experience with personalization of Yahoo!, *Commun ACM*, **43** (2000), 35–39.
[17] C.A.C. Lampe, E. Johnston, and P. Resnick. Follow the reader: filtering comments on slashdot. *Proceedings of the 25th SIGCHI Conference on Human Factors in Computing Systems (CHI '07),* 2007, 1253–1262.
[18] A. Tversky, and D. Kahneman. The framing of decisions and the psychology of choice, *Science*, **211** (1981), 453–458.
[19] D. Kahneman, J.L. Knetsch, and R.H. Thaler. Anomalies: the endowment effect, loss aversion, and status quo bias, *J Econ Perspect*, **5** (1991), 193–206.
[20] S.D. Halpern, P.A. Ubel, and D.A. Asch. Harnessing the power of default options to improve health care, *N Engl J Med*, **357** (2007), 1340–1344.

Address for correspondence

Kai Zheng, PhD
6095 Donald Bren Hall
Irvine, CA 92697
USA
Phone: +1 (949) 824-692
Email: kai.zheng@uci.edu

Table 1. Personalization settings provided in EMERSE

	Personalization setting	Description	System default
1	Highlight Viewed Notes	A composite option that instructs the system how to render the font color of a returned document that has been previously viewed (i.e., visited links).	Off
2	Color of Viewed Notes		Blue
3	Line Wrapper	A toggle option instructs the system whether to place each matched keyword and surrounding document snippet on its own line for easy human inspection. When turned off, all snippets would be strung together to save screen space.	On
4	Show All Matches	Instructs the system whether to display only the first matched instance of a keyword that may have multiple matches in a returned document.	On
5	Surrounding Context	Sets the character length of the text snippets to be displayed adjacent to the matched keyword in the source document. Longer snippets provide more contexts but are more likely to result in visual clutter.	20
6	Heat-Map: Scaling Factor	A set of parameters that defines the appearance of heat-maps provided to visualize the level of 'relevance' of the returned documents to the search query. On a heat-map, color gradations are proportional to the number of matches found for a particular patient and for a particular document category (e.g., pathology vs. radiology reports); see Figure 1*a* in Seyfried et al. (2009) [12].	5
7	Heat-Map: Relevance		0
8	Heat-Map: Score		0
9	Heat-Map: Text		6
10	Heat-Map: Color Palette	Defines a range of colors that will be used to highlight matched keywords on a heat-map visualization; see Figure 1*a* in Seyfried et al. (2009) [12]. Unlike the other heat-map settings, the Color Palette can be assessed through the search results screen where the heat-map is displayed, without invoking the heat-map preference pane.	Blue, orange, green, purple, red, and black
11	Highlight in Print	As color highlights do not work well in monochrome printing, users may use this option to instruct the system to highlight matched search terms using a combination of bold, italics, or underline font styles when appearing in a printed document.	Off
12	Spelling Alternatives	EMERSE has the ability to detect spelling alternatives, i.e., words contained in a document that may appear to be the alternative or misspelled forms of the search terms that the user entered. This option instructs whether the system should identify spelling alternatives in the source document (when the option is turned on) or ignore them (when the option is turned off).	Off
13	Auto-Correct Search Terms	When this option is on, the system will automatically correct misspelling (e.g., "infectoin") or logic errors (e.g., an incorrect regular expression formula) found in the query.	On
14	History of Recent Queries	Sets the number of recent search queries to be saved in the system for future reuse.	100
15	Recommend Synonyms	EMERSE provides a synonym recommendation function that suggests interchangeable or related terms which the user may consider including in the query (e.g., "renal" ⇔ "kidney", "flu" ⇔ "influenza"). The "Search Query" option instructs the system to expand the keywords included in the search query itself; whereas the "Negation List" option instructs the system to expand the keywords included on the list set to be ignored in the search.	On
	Synonyms: Search Query		On
	Synonyms: Negation List		On

POSTERS

This page intentionally left blank

I. Interpreting Health and Biomedical Data

This page intentionally left blank

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190462

Skin Lesion Detection with Support Vector Machines on iOS Devices

Bianca Schnalzer, Baptiste Alcalde

Institute of eHealth, University of Applied Sciences FH JOANNEUM, Graz, Austria

Abstract

Automated wound detection has become a common issue in health care. A broad variety of image processing algorithms already exist, but they are very power consuming on mobile devices. Meanwhile the use of machine learning algorithms is on the rise and new frameworks have been developed to use these techniques with improved on-device-performance such as Apple Core Machine Learning Interface. In this paper, we evaluate the performance of libSVM for wound detection in practice.

Keywords:
Wounds and Injuries, Machine Learning, Image Processing, Computer-Assisted

Introduction

Skin lesion boundary detection is a fundamental necessity for medical documentation and treatment evaluation. Several well known image processing techniques already exist for identifying edges on images. However, it is a time-/resource consuming task on mobile phones. Trained machine learning models used on Apple Core Machine Learning (ML) focus on on-device-performance and are therefore less power consuming [2]. Once a trained algorithm has been integrated in mobile application, it can easily be used for detection of various features. The goal of this paper is to identify, implement and evaluate suitable and supported frameworks for skin lesion boundary detection

Methods

First, a literature review was conducted to identify suitable models and algorithms for skin lesion boundary detection. Second, the suitable models were analysed in more detail to identify pros and cons. Then, this paper deals with the feature selection and algorithm training as well as testing the usability of training and implementing such models.

The main families of techniques that are described in the literature related to wound classification are image processing algorithms, neural networks and Support Vector Machines. Although image processing methods and algorithms can provide good results, they often need several processing steps and possible server communication and can be a drain. Image processing was used for skin lesion boundary detection for instance in [4]. [6] summarized current work on wound segmentation, monitoring and software tools.

Support Vector Machines (SVM) are suitable to perform edge detection [7][8]. Features were composed of color, texture, region morphology and topology. SVMs appeared to be mature enough for the purposes of our project

Therefore, we focus on SVM techniques and evaluate SVM frameworks for iOS devices in this paper.

Results

Technical Information

Since iOS 11.0, Apple has provided Core Machine Learning Tools for the integration of machine learning models in apps. An advantage of having on-device machine learning algorithms is to enhance security, privacy and improve speed.

To date, many different kind of functionality have been built on the top of the CoreML layer such as Vision (for object or face detection) or Natural Language (for speech identification or speech recognition). The complete list of pre-trained models of Apple Core ML is available under [1]. However, none of these models could answer the needs of this project. Therefore, one of the contributions of this research was to develop a trained model for wound detection for iOS devices.

To use a trained model in a SWIFT project in XCode (developer environment for iOS), the model has to be converted. Conversion was done via Python [11] and coreml-tools [3]. Coreml-tools provides various scripts to model, utilize and convert supported models. CoreML restricts the choice of tools based on the desired model. To use SVM within CoreML, scikit-learn 0.18 and LIBSVM 3.22 can be used. In this paper, LibSVM [5] was identified as the most suitable framework.

Image Processing and Feature Detection

Several pre-segmented and labelled datasets are available for machine learning [10]. Unfortunately, no segmented medical image datasets were available. Therefore, another contribution of this paper was to take images from [13], preprocess and label them. Many features were examined during image processing. Texture and color information were identified as the main image features, as confirmed by [14].

Framework Evaluation

As mentioned before, LibSVM [5] was identified as a suitable framework for the needs of the current research. However, LibSVM only supports numerical data and the file format has to comply with a specific structure, where an indexed set of features is associated to a class. For image processing, MatLab [12] as well as ImageJ [9] were used.

The only supported version of LibSVM at the time of our experiments was 3.22. In this project, LibSVM was used via command line and Python. A main step was to evaluate suitable

training parameters for the given problem. LibSVM supports one-class and multi-class classification as well as regression.

The most important parameters, e.g., kernel-type (-t), degree of kernel function (-d), coefficient, cost function (-c) and gamma of kernel function (-g) as well as epsilon (termination criterion) can be set [5]. An essential criterion in classification result was to keep gamma high, because gamma reduces misclassification. Short classification time was also an advantage of the reduced model.

The image data set contained 16 kinds of wounds. Several patches (image crops) were created to categorize pixels in 3 classes (wound, skin and background).

Initially, default training was conducted with all image patches. Additionally, randomly selected data points were taken to train the classifier. Producing the training set was carried out via a Python script (subset.py), which is available within the LibSVM library. Table 1 gives a brief overview of the training parameters of the five best-performing models.

Table 1 – Overview of training parameters of the five different models.

Model	Parameters
reducedDataSet	-s 0 -t 2 -c 10 -g 1000 -h
reducedDataSet2	-s 1 -t 2 -c 10 -g 1000
woundDetectionReduced5000	-s 1 -t 2 -c 10 -g 1000
woundDetectionReduced50000	-s 1 -t 2 -c 10 -g 100
woundDetectionFULL	-s 0 -t 2 -c 1 -g (1/num of features)

The model with 5000 randomly taken data points had the best performance (76%) during the testing process and was selected for further refinement and integration in an app. The model expects a pixel input in a multi array and provides a pixelwise classification.

Discussion

We clearly see an opportunity of future works in the direction of CoreML tool compatibility. For instance, TensorFlow has a huge acceptance in the developer community and is to-date not included in the list of supported tools, even if Keras can use it as Backend.

Related to this project, feature selection enhancement is still a possible task. Mean colour and dominant colour can possibly be added as additional features. Some researchers postulated integration of texture information to gain better results. Feature dimension reduction, e.g., with PCA (Principal Component Analysis), could possibly reduce training time.

In this work, we used only freely available skin lesion images. As a result, our image dataset was limited. In order to have a performant app, we would need a larger skin lesion image dataset and a professional labelling of these images.

Conclusions

The goal of the current paper was to evaluate machine learning techniques and frameworks for skin lesion boundary detection suitable for deployment on iOS devices, and to develop an App using the selected solution. We identified SVMs as a mature candidate for our project. Then, we developed a labelled image dataset based on [13], as well as a MatLab script for the conversion into the LibSVM data format. Furthermore, we implemented a pre-trained model for skin lesion boundary detection, integrated this model in an App and performed a practical evaluation.

References

[1] Apple Inc., Converting Trained Models to Core ML. https://developer.apple.com/documentation/coreml/converting_trained_models_to_core_ml; 2017. (accessed 08.01.2018)

[2] Apple Inc., Core ML, https://developer.apple.com/documentation/coreml, 2017. (accessed 08.01.2018)

[3] Apple Inc., coreml-tools 0.8., https://pypi.python.org/pypi/coremltools; 2017. (accessed 28.02.2018)

[4] P. Babu, G. Vimalarani, Boundary tracing algorithm for automatic skin lesion detection in macroscopic images. World Engineering & Applied Sciences Journal 2017; 8(1):25-33.

[5] C.C. Chang, C.J. Lin, LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, Vol. 2, No. 3, Article 27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[6] M.F.A. Fauzi, I. Khansa, K. Catignani, G. Gordillo, C.K. Sen, M.N. Gurcan, Computerized segmentation and measurement of chronic wound images. Computers in biology and medicine 2015; 60:74-85.

[7] H. Gómez-Moreno, S. Maldonado-Bascón, F. López-Ferreras, F. Acevedo-Rodríguez, P. Martín-Martín, Edge detection by using the support vector machines. European Conference on Circuit Theory and Design, 2001.

[8] H. Gómez-Moreno, S. Maldonado-Bascón, F. López-Ferreras, P. Gil-Jiménez, A new and improved edge detector using the support vector machines, https://pdfs.semanticscholar.org/1857/0a140e98df1c77f8bc1ce1ffba94e097ff5b.pdf. (accessed 28.02.2018)

[9] W.S. Rasband., ImageJ, U.S. National Institutes of Health, Bethesda, Maryland, USA, https://imagej.nih.gov/ij; 2012.

[10] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. Proc. 8th Int'l Conf. Computer Vision. volume 2; 2001. p. 416-423.

[11] Python Software Foundation, Python. https://www.python.org/; 2017. (accessed 28.02.2018)

[12] The MathWorks, The mathworks. Inc, Natick, MA 1998;5:333.

[13] S. Thomas, Medetec Wound Database: stock pictures of wounds. http://www.medetec.co.uk/files/medetec-image-databases.html; 2018. (accessed 22.02.2018)

[14] H. Wannous, Y. Lucas, S. Treuillet, Combined machine learning with multi-view modeling for robust wound tissue assessment. VISAPP 2010-Fifth International Conference on Computer Vision Theory and Applications. volume 1; 2010. p. 92-104.

Address for correspondence

bianca.schnalzer@fh-joanneum.at

baptiste.alcalde@fh-joanneum.at

FH JOANNEUM

Eckertstrasse 30i, Office111,

8020 Graz, Austria

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190463

An Ontology for Describing Health IT Interventions: Methodological Considerations

Elske Ammenwerth[a], Verena Dornauer[a], Maryam Ghalandari[b], Franziska Jahn[b], Nicolet de Keizer[c], Alfred Winter[b]

[a] *Institute of Medical Informatics, UMIT – University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria*
[b] *Institute of Medical Informatics, Statistics and Epidemiology, University of Leipzig, Germany*
[c] *Department of Medical Informatics, Amsterdam University Medical Center, Netherlands*

Abstract

Health informatics as a young, interdisciplinary discipline lacks a unified terminology in some areas. This is especially true when trying to properly describe health informatics interventions developed and deployed to improve quality and efficiency of patient care. We aim at developing a health IT ontology which allows systematically describing health IT interventions. To achieve this, we combine a deductive and an inductive approach. First results are promising and may later be extended by a folksonomy.
Keywords:
Medical informatics applications, terminology, classification

Introduction

As part of evidence-based health informatics, researchers and practitioners need to be able to describe health IT interventions precisely to allow the publication and aggregation of related evidence. Unfortunately, available terminological systems such as MeSH terms, ACM computing classification systems or IEEE thesauri do not allow a specific description of health IT interventions. As another challenge, terminology in health IT is full of synonyms and homonyms increasing the difficulty to identify a specific intervention. Also, health IT interventions can range from quite small interventions such as a functionality (e.g., an overdosing alert) to more complex interventions (e.g., a full-fledged CPOE system).
The reason for missing comprehensive terminological systems for health IT interventions may be that health informatics is a comparably young and interdisciplinary discipline, and fast technological developments make clear and stable definitions of health IT interventions challenging. We observed this problem when developing and maintaining a health IT evaluation database [1]. We did not find a practical health IT intervention terminological system that we could use for describing and searching for health IT evaluation studies.
Our idea is to build an ontology for health IT interventions. An ontology is a (formal) specification of concepts, relations and functions in a domain. Such an ontology would help us properly describe various types of health IT interventions and their impact by defining concepts and their semantic relationships.
Our research question is therefore: How can we develop an ontology for describing health IT interventions that allows describing health IT interventions in a structured, sufficiently detailed way? In this paper, we describe our methodology consideration for ontology development. The research takes place in the context of the project "HITO: A Health IT Ontology", funded by the Austrian Science Fund FWF (I 3726-N31) and the German DFG fund (WI 1605/11-1).

Methods

We started with the methodology for taxonomy development by Nickerson [2]. A taxonomy aims at structuring and organizing knowledge (i.e., concepts) of a field, often in a hierarchical way [2]. An ontology extends a taxonomy by adding information on semantic relationships between concepts. We thus found the structured approach by Nickerson adequate as a first step.
The method by Nickerson proposes a mix of deductive and inductive steps. Nickerson argues that intuitive, "ad hoc" methods for building a taxonomy, where the researcher just relies on his knowledge, should be avoided. We endorse this and strive for a reproducible taxonomy development.
The inductive approach involves observing empirical cases (here: health IT evaluation studies) that are then analyzed to determine dimensions and characteristics of the taxonomy. The deductive approach derives a taxonomy from theory or conceptualization. Both approaches are used in an iterative manner until predefined ending conditions are met [2].
Before starting inductive or deductive activities, Nickerson proposes to determine the meta-characteristics of the taxonomy that will serve as the basis for defining the purpose of the taxonomy and the expected users.

Results

Following the Nickerson method for taxonomy development, we first define the meta-characteristics, users and intended use of our planned health IT intervention ontology.
We define a health IT intervention as any IT-based intervention that has a potential impact on the process or outcome of patient care. The term "IT-based intervention" can comprise of an IT-based functionality within an application component (such as a new dosing alert within a medication system), an application component with a set of functionalities (such as radiology information systems), or a combination of application components (such as a clinical information system comprising a medication system and a patient portal). AnIT-based intervention can also comprise of a set of physical data processing components (such as tablets). Most often, an IT-based intervention will comprise of both software and hardware components.
The intended users of the health IT ontology are health IT researchers, health IT managers, and health IT users. These

groups are interested in evidence on the impact of health IT interventions on patient care. They are more interested in high-level characteristics of a certain type of health IT intervention than on technical specifications, as these are quite context-dependent and may quickly change over time due to technological developments.

The ontology should help these user groups in the following situations:

- A researcher wants to publish a health IT evaluation study. In the publication, he wants to clearly describe the health IT intervention using the ontology.
- An IT manager searches for available evidence on the impact of a specific type of health IT intervention. The ontology allows him to clearly describe which type of health IT intervention he is looking for.
- A researcher wants to conduct a systematic review on the impact of a given type of health IT intervention. The ontology allows him to clearly describe which type of health IT intervention he is looking for to find all published evaluation studies on this intervention.
- A researcher wants to develop a map for evidence-based health informatics. This map should visualize how much evidence has been published for which types of health IT interventions. The ontology is the basis of this map.

Our ontology thus aims at carefully describing health IT interventions and to distinguish between various types of health IT interventions and their characteristics. We started with a manual review of available health IT taxonomies through a literature search and by analyzing health IT taxonomy reviews such as [3]. We identified e.g. the following taxonomies:

- A taxonomy for rule-based functionality of CPOE systems with four dimensions and up to six characteristics in each dimension by Wang (2002);
- A list of fifteen types of application components by Ammenwerth & de Keizer (2005);
- A taxonomy of CDSS systems with five dimensions and up to six characteristics in each by Berlin (2006);
- A taxonomy of hospital functions by Winter (2011);
- A taxonomy of front-end capabilities of CDSS systems with six dimensions and up to 14 characteristics in each dimension by Wright (2011).

At the moment, we are mapping the identified taxonomies to each other to identify both overlapping and unique dimensions and/or characteristics helping to describe unique attributes of health IT interventions. The result of this deductive approach will be a first proposal of dimensions and characteristics for our ontology.

We also started with the inductive approach and conducted a manual qualitative content analysis of five randomly selected recent evaluation studies using the software MAXQDA. From these studies we were able, in the first iteration, to identify dimensions and related characteristics for describing a health IT intervention:

- User group: Physician, nurse, other clinical user group, patients/relatives, administrative staff, other.
- Organizational unit: Location of intervention; such as surgery, psychiatry, primary care, etc.
- Type of application system: For example, lab information system, nursing documentation system
- Name of software product
- Offered software features: For example, taken from the HL7 EHR-System Functional Model [4].
- Supported enterprise function: For example, hospital administration, quality management, decision making, or patient information.

Before continuing the inductive approach with other types of health IT interventions, we will now combine the results of the deductive literature review with the results of the inductive approach. The results will be a first iteration of an ontology of health IT interventions. Further, inductive steps will analyze evaluation studies of other frequent types of health IT interventions. Results will then be added iteratively to the ontology, until the ending conditions as defined by Nickerson are reached.

Conclusions

The presented methodological considerations should lead to an ontology for health IT interventions that is parsimonious regarding the number of dimensions, robust to clearly differentiate health IT interventions, comprehensive to be able to describe studies, extendible and explanatory, as demanded by Nickerson [2]. We plan to make the contents available via Linked Open Data.

One idea for future work is to use the resulting health IT ontology to infer new knowledge, such as a better definition of various types of application components via a cluster analysis of characteristics.

Another idea for future work is using the vision of a folksonomy [5] which is a user-generated system of classifying and organizing health IT evaluation studies by the use of metadata such as tags. These tags could be used to identify and add missing parts in our ontology.

References

[1] E. Ammenwerth, and N. De Keizer, An inventory of evaluation studies of information technology in health care: Trends in evaluation research 1982-2002, *Methods Inf. Med.* 44 (2005) 44–56.

[2] R.C. Nickerson, U. Varshney, and J. Muntermann, A method for taxonomy development and its application in information systems, *Eur. J. Inf. Syst.* 22 (2012) 1–24. doi:10.1057/ejis.2012.26.

[3] U. Varshney, R.C. Nickerson, and J. Muntermann, Taxonomy Development in Health-IT, *Proc. Ninet. Am. Conf. Inf. Syst. Chicago, Illinois, August 15-17, 2013*. (2013) 1–10.

[4] HL/ International, HL7 EHR-System Functional Model, R2, (2018). http://www.hl7.org/implement/standards/product_brief.cfm?product_id=269 (accessed March 30, 2019).

[5] A. Noruzi, Folksonomies: (Un)Controlled Vocabulary?, *Know Org.* 33 (2006) 199–203.

Address for Correspondence

Elske Ammenwerth, UMIT, Institute of Medical Informatics, Eduard Wallnöfer Zentrum 1, 6060 Hall in Tirol, Austria, elske.ammenwerth@umit.at

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190464

PREDIMED: Clinical Data Warehouse of Grenoble Alpes University Hospital

Svetlana Artemova[ab], Pierre-Ephrem Madiot[ab], Alban Caporossi[ac], PREDIMED group Pascal Mossuz[ae], Alexandre Moreau-Gaudry[abf]

[a] *CHU Grenoble Alpes, F-38000, Grenoble, France,*
[b] *Clinical Investigation Center-Technological Innovation, INSERM CIC1406, CHU Grenoble Alpes, Univ. Grenoble Alpes,*
[c] *TIMC-IMAG, Univ. Grenoble Alpes, CNRS, Grenoble INP, CHU Grenoble Alpes, F-38000 Grenoble, France*
[d] *Grenoble Institut des Neurosciences, INSERM, Univ. Grenoble Alpes, F-38000, Grenoble, France*
[e] *IAB, Univ. Grenoble Alpes, F-38000, Grenoble, France*
[f] *UMS GRICAD, Univ. Grenoble Alpes, F-38000, Grenoble, France*

Abstract

Grenoble Alpes University Hospital (CHUGA) currently deploys a clinical data warehouse PREDIMED to integrate and analyze for research, education and institutional management the data of patients treated at CHUGA. In this poster, we present the methodology used to implement PREDIMED and illustrate its functionality through three first research use cases.

Keywords:
Big data, medical informatics, data warehouse

Introduction

Large quantities of data are generated by healthcare and the volume is increasing every year [1]. This data is diverse and includes electronic health records, -omics, imaging, sensor and textual data, which are complex, heterogeneous, poorly annotated and generally unstructured [2]. The application of big data in healthcare is, nowadays, a fast-growing field [3-4]. Health datasets have been widely reused for research in numerous studies [4]. In particular, Clinical Data Warehouses (CDW) collect and reuse healthcare data for various applications covering all domains of medicine worldwide. For data of such volume, novel powerful technologies are required: non-relational and distributed databases for data storage and handling, artificial intelligence, machine learning and, more specifically, deep learning algorithms for data analysis. A number of reviews have been published focusing on the applications of deep learning in healthcare [2], in particular, in medical imaging [5] and –omics [6]. For diabetic retinopathy [7], skin cancer [8] and breast cancer [9] deep learning algorithms have demonstrated the ability to diagnose pathologies using medical images with at least the same performance as medical experts.

CHUGA is the 12th largest hospital and the first trauma center in France. Having a powerful technical environment, with all patients' data including new data sources (genetics, internet of things, medical imaging, surgery videos), would certainly increase the re-usage of the data produced by CHUGA.

In the following, we describe the methodology used to implement PREDIMED (acronym in French for: Plateforme de Recueil et d'Exploitation des Données bIoMEDicales), a CDW designed to reuse the data of patients treated at CHUGA. We then present the chosen architecture and illustrate the functionality of PREDIMED through use cases. We believe that this work is valuable for any healthcare institution considering building a CDW.

Methods

To implement a CDW, a group was created gathering specialists from transverse departments of CHUGA (Public Health, Imagery, Biology, Information Systems, and Pharmacy) as well as specialized medical experts. The expected work was divided into 9 work packages (WP) that we describe below.

Results

WP1 Possible CDW use cases. A survey was taken on the possible use of a CDW by medical staff and researchers and the results fall into four major categories: *research* (population cohorts and analysis of the corresponding data), *education* (providing real-life cases for teaching), *institutional managemen*t (healthcare organization and optimization) and *healthcare* (patient-specific diagnostics and treatment). It was decided to use PREDIMED for the first three use cases and not to consider the last one for legal and maintenance reasons.

WP2 Choice of the technical solution. State-of-the-art solutions were compared in their functionality, maturity, scalability, security, conformity to the applicable laws, and costs (both material and human resources). Modular open source architectures were chosen for PREDIMED to freely develop new functionalities and algorithms, as well as (we believe) to reduce costs. Powerful and scalable big data technologies were chosen over relational databases as they allow for distributed storage and analysis as well as for text data exploration. Modern machine learning, deep learning and data mining algorithms perform well on these architectures.

WP3 Data model and data import. Graph data model suitable for a general hospital was created to avoid 'data swamp'. The advantages of a graph model are more flexible navigation and easier data interconnection and refinement. The first PREDIMED data sources are: patients, visits, diagnoses, texts, healthcare professionals, laboratory results, structure of medical units, medical imaging, and drug prescriptions. Some of the data was loaded partially for the proof-of-concept.

WP4 Confidentiality and security. PREDIMED contains personal data and thus has to be authorized by the independent supervisory authority in terms of the GDPR (CNIL in France) to switch to the operational phase. This authorization is expected in 2019. Every project using PREDIMED data must also be declared to the supervisory authority. A high-level of security is provided for PREDIMED, e.g. all access to the data will be tracked and daily logs will be analyzed automatically.

WP5 Architecture implementation. We deployed several tools for multiple views on the raw data. Hortonworks Data Platform (HDP) enables agile application deployment, machine and deep learning workloads, security and governance. ArangoDB (a powerful multi-model NoSQL database) and ELK stack allow to navigate through the data model for cohort data selection, analysis and visualization. PREDIMED data is accessible via scripting notebooks (with spark, scala, python, R, SQL, Hive, SolR, etc.), and via developed API features.

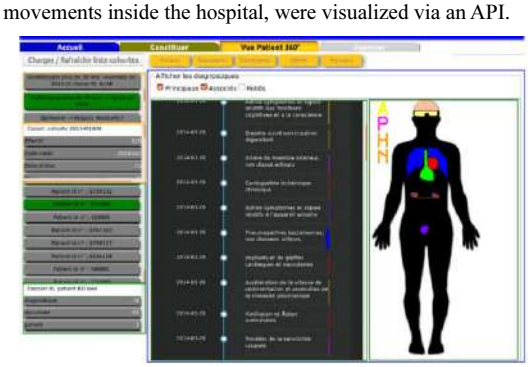
WP6 Governance. To regulate the access to the PREDIMED data, a governance model was developed. Every project requesting PREDIMED data must be approved by: Steering Committee (evaluating the feasibility and estimating the necessary investment), Scientific and Ethics Committees and obtain the legal authorizations if needed.

WP7 Interoperability. A survey on the thesaurus used in CHUGA by medical specialty was taken and has shown that in medical practice local codes are more common than international dictionaries (with rapid mappings), but for medico-administrative activities standardized codes are used.

WP8 Business model. A business model was developed as it was shown in literature that the CDW can generate benefits [10]: each project using PREDIMED data participates in its development in terms of human resources and/or hardware.

WP9 Proof-of-concept through three use cases. A proof-of-concept demonstrator was built and PREDIMED usage was illustrated on 3 research use cases on populations of moderate size (the same code can be used in the future on more data).

In the first use case, we studied the feasibility of real time selecting a population of patients according to diagnoses (diabetes), demographics, and treatment criteria, using the graph query interface that we designed using ArangoDB graph navigation features and visualizing corresponding data using ELK stack. The data of individual patients, such as timelines summarizing their documents, prescriptions, diagnoses and movements inside the hospital, were visualized via an API.



In the second use case, we investigated somatic mutations in 60 onco-hematology patients to determine the genetic signature of the patients' clinico-biological characteristics by linking multiple data sources, including genetics. We characterized patients' populations by mutated gene and combined the mutational status with biological values. We looked for keywords in the patient's documents to find the frequent comorbidities and cancer relapses by mutated gene. Kaplan-Meier survival curves by mutated gene were plotted.

In the third use case, we selected 300 patients with lung cancer by searching for keywords in patients' documents, and linked all data available for these patients with their medical images (raw data and metadata of DICOM files), and plotted the greyscale histograms. The two last cases were coded on the HDP platform via Zeppelin scripts by one engineer.

Conclusions

We have reported the CHUGA's experience of implementing a CDW PREDIMED: we show that we are able to take data out of silos, link this multimodal multisource data and access new data (textual and genetic data, raw data as image pixels), explore, analyze and visualize data, and thereby implement use cases for research, education and hospital management.

PREDIMED aims to be as exhaustive as possible in terms of data available for the CHUGA patients. Thus, important future work includes integration of more internal data sources (service-specific databases), and external databases, such as research, environmental and open data. Another direction of future work concerns the heterogeneous quality of the source data that should be measured and controlled. Interoperability (technical and semantic) is crucial for projects at a national and worldwide level and CHUGA is an active participant of a French national initiative aiming to group local CDW into an interoperable network – a Health Data Hub.

Acknowledgements

The authors thank CHUGA for the support and financial investment, Grenoble Centre of Excellence in Neurodegeneration (GREEN) for financial support, startup Hurence, David Rozier and Cecilia Hughes for the joint effort of many complementary skills and all the members of the PREDIMED group: Brigitte Cohard[a], Marjolaine Bodin[ad], Jean-François Blatier[a], Jérôme Fauconnier[a], Frédéric Olive[a], Daniel Pagonis[ac], Dominique Le Magny[a], Jean-Luc Bosson[ac], Katia Charrière[ab], Ivan Paturel[a], Bruno Lavaire[a], Gabriel Schummer[a], Joseph Eterno[a], Jean-Noël Ravey[a], Ivan Bricault[ac], Gilbert Ferretti[a], Sébastien Chanoine[ac], Pierrick Bedouch[ac], Emmanuel L. Barbier[d], Julien Thevenon[a], Christophe Cancé[f], Christian Lenne[f]

References

[1] I.D. Dinov, Volume and value of big healthcare data, *J Med Stat Inform 4* (2016).
[2] R. Miotto et al., Deep learning for healthcare: review, opportunities and challenges, *Brief Bioinform* (2017).
[3] J. Luo, M. Wu, D. Gopukumar, Y. Zhao, Big data application in biomedical research and health care: a literature review, *Biomed Inf Insights 8* (2016).
[4] W. Wang., E. Krishnan, Big data and clinicians: a review on the state of the science, *JMIR Med Inform* 2 (2014).
[5] G. Litjens et al., A survey on deep learning in medical image analysis, *Med. Image Anal.* (2017) 42, 60–88.
[6] Z. Zhang et al., Deep learning in omics: a survey and guideline, *Briefings in Functional Genomics* (2018).
[7] V. Gulshan et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA 316* (2016), 2402–2410.
[8] A. Esteva et al., Dermatologist-level classification of skin cancer with deep neural networks, *Nature 542* (2017), 115-118.
[9] R. Dezs, et al., Detecting and classifying lesions in mammograms with Deep Learning. *Sci Rep 8* (2018).
[10] R. Nalichowski, D. Keoghm, H.C. Chueh, S.N. Murphy, editors, Calculating the benefits of a research patient Data Repository, *Proc AMIA Annu Fall Symp* (2006).

Address for correspondence
Svetlana Artemova
Email: SArtemova@chu-grenoble.fr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190465

Patient Summary Management with Electronic Health Records: A Descriptive Study in General Practice

Vincent Bello [a], Virginie LaCroix-Hugues [a,b] , Marc-André Guerville [a], David Darmon [a,c]

[a] *Université Côte d'Azur, Department of Education and Research in General Practice, Nice, France*
[b] *Archet 1 Hospital, Department of Public Health, Nice, France*
[c] *Aix Marseille University, INSERM, IRD, SESSTIM Sciences Economiques & Sociales de la Santé & Traitement de l'Information Médicale, France*

Abstract

To describe information content in an automatically generated patient summary worksheet (PS) comparatively to electronic health records (EHRs) for 90 patients. The PS was more focused on the cure than person-centred care. Ergonomic solutions based on the users' needs should enhance shared decision-making and improve the healthcare professional-patient relationship.

Keywords:

Electronic health records, electronc data processing, patient education handout

Introduction

General practitioners (GPs) were assigned with the mission to update a patient summary dataset with electronic health records (EHRs), according to the French interoperability framework of health information systems (CI-SIS) **[1]**. The patient summary worksheet (PS) automatically generated from the patient summary dataset was expected to improve care coordination and personalised care planning **[2]**, especially in multidisciplinary primary care organisations and for telemedicine. Despite incentives, little was known about the patient summary management in primary care.

Objective

To describe information content in the automatically generated PS comparatively to information contained in EHRs.

Methods

A cross-sectional study was conducted in 2017 in three primary care structures in the department of Alpes-Maritimes in the southeast part of France. Nine GPs were recruited from the Department of General Practice, University of Nice. Five GPs used one electronic health record system (system A), and four used another one (system B). After performing tests, 90 EHRs (10 per GP) were selected and compared to the 90 PS. Items were classified as present (present in PS) or absent (in EHRs but absent from PS). Medical history was recorded according to the second edition of International Classification of Primary Care (ICPC-2). Multimorbidity was defined as the presence of two or more chronic conditions [3]. Negative signs and spontaneous recovery from isolated acute medical conditions like seasonal or common viral infections were not compared. Drugs and vaccines were recorded according to the Anatomical Therapeutic Chemical (ATC) classification. Drugs in the PS were compared with the regular treatment reported in EHRs. A logistic regression was performed to identify factors associated with a low rate of absent items. Statistical significance was defined as $p<0.05$. All analyses were performed using R® packages for statistical analysis.

Results

EHRs were selected for 49 females (median age: 71, 29-91 years) and 41 men (median age: 69, 29-91 years). 86% of the patients were multimorbid.

2311 items were reported: 1747 items (mean: 19, 1-52 items) were present in patient summaries while 564 items (mean: 6, 0-38 items) were in EHRs but not in PS (Table 1).

Table 1– Number of items in the patient summary by system compared to information in EHRs

	A		**B**		**A+B**				**Present+Absent**	
	Present	**Absent**	**Present**	**Absent**	**Present**		**Absent**			
	649	**280**	**1098**	**284**	**1747**	**(100%)**	**564**	**(100%)**	**2311**	**(100%)**
Medical history	**401**	**177**	**474**	**204**	**875**	**(50%)**	**381**	**(68%)**	**1256**	**(54%)**
Surgeries	22	32	88	14	110	(6%)	46	(8%)	156	(7%)
Past medical history	166	74	117	55	283	(16%)	129	(23%)	412	(18%)
Current conditions	172	28	219	59	391	(22%)	87	(16%)	478	(21%)
Risk factors	20	33	32	67	52	(3%)	100	(18%)	152	(7%)
Allergies and adverse reactions	21	10	18	9	39	(2%)	19	(3%)	58	(3%)
Drugs	**248**	**48**	**624**	**25**	**872**	**(50%)**	**73**	**(13%)**	**945**	**(41%)**
Vaccines	**0**	**55**	**0**	**55**	**0**		**110**	**(19%)**	**110**	**(5%)**

Table 2 – Factors associated with low absent items

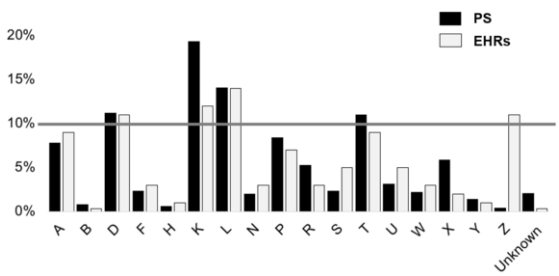
Absent items	Simple logistic regression			Multivariate logistic regression		
	Coeff.	**OR [95% CI]**	***p***	**Coeff.**	**aOR [95% CI]**	***p***
Patient age	0.01	1.01 [0.98-1.04]	0.535			
Male gender	–0.19	0.83 [0.35-1.97]	0.671			
Chronic conditions	0.005	1.004 [0.86-1.17]	0.951			
EHRS – system B	1.47	4.33 [1.77-11.15]	0.002*			
Medical history	–0.04	0.96 [0.90-1.01]	0.108	–0.12	0.88 [0.81-0.96]	0.004*
Present drugs	0.17	1.19 [1.10-1.31]	< 0.001*	0.24	1.28 [1.15-1.46]	<0,001*
Ended or stopped drugs	0.12	1.13 [1.01-1.28]	0.041*			

Coeff: coefficient; aOR: adjusted odds ratio; * significant *p*-value

Fifty-four percent of the drugs in PS did not match with the regular treatment: 46% (401 drugs) were ended or stopped, 8% (72 drugs) were redundant. Only 8% of drugs were in EHRs, but 84% of them had an important therapeutic value and 37% added at least one interaction.

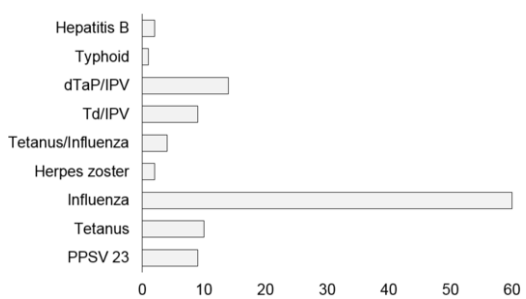
In PS, 78 items were redundancies in medical history, especially 72 redundancies between past medical history and current conditions. The rate of items in PS was higher than 10% for digestive (D), cardiovascular (K), musculoskeletal (L) and metabolic (T) chapters. The chapters with a rate of items higher than 10% in EHRs were almost the same as those in PS, but social problems (Z) were also pointed out in EHRs (Figure 1).

Figure 1– ICPC-2 chapters



All vaccines were in EHRs, but none were in PS (Figure 2). Items in EHRs were retrieved more frequently in medical notes (33%) and in discharge summaries or after-visit summaries from other healthcare providers (32%).

Figure 2– Items for immunisation in EHRs but absent in PS



After univariate analysis and elimination of too strongly correlated variables, two variables remained in the stepwise regression model: the number of items in medical history and the number of drugs present in PS (Table 2).

Conclusion

The PS was more focused on the cure than person-centred care. The rate of absent items would be a poor relevant indicator of the quality of a patient summary. Further studies could be conducted to assess the influence of physical and cognitive ergonomics upon the healthcare professional-patient relationship and the shared decision-making process. However, the secondary use of patient summary datasets to build predictive models in context of data-driven medicine raises major ethical issues.

Acknowledgements

No competing interests. We would like to thank the voluntary GPs for their participation.

References

[1] The Shared Healthcare Information System Agency *(ASIP Santé*), The French health authority, *Volet de synthèse médicale élaboré à partir du modèle métier de synthèse médicale*, 2013, Available at: http://www.has-sante.fr/portail/upload/docs/application/pdf/2013-11/asip_sante_has_synthese_medicale.pdf [accessed on March, 2017].

[2] The European Commission, *Guidelines on the electronic exchange of health data under Cross-Border Directive 2011/24/EU – Patient summary for unscheduled care*, Release 2, November 2016, Available at: https://ec.europa.eu/health/sites/health/files/ehealth/docs/ev_20161121_co10_en.pdf [accessed on March, 2017].

[3] K. Nicholson, A.L. Terry, M. Fortin, T. Williamson, M. Bauer, and A. Thind, Examining the prevalence and patterns of multimorbidity in Canadian primary healthcare: a methodologic protocol using a national electronic medical record database, *J Comorbidity* **5**(1) (2015), 150-61.

Address for correspondence

David Darmon,

e-mail: david.darmon@univ-cotedazur.fr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190466

Regional Professionals Network to Support the Renal Epidemiology and Information Registry in Ile-de-France

Zoubair Cherqaoui [a], Mohamed Ben Said [a], Lucille Mercadal [b,g], Xavier Belenfant [c, g], Eric Gautier [e,g], Evelyne Ducamp [a], Jean François Desassis [f,g], Houssem Tebbakh [a], Paul Landais [c], Jean-Philippe Jais [a,g]

[a] Paris Descartes University, Faculty of Medicine, Biostatistics Unit, APHP - Necker Enfants Malades Hospital, Paris, France,
[b] Urology – Nephrology – Renal Transplantations Department – APHP - La Pitié Salpêtrière Hospital, 75013, Paris, France,
[c] EA2415 Clinical Research Institute, Montpellier University, 34093, Montpellier, France,
[d] Nephrology – Hemodialysis Department- Centre Hospitalier Intercommunal André Grégoire – 93100, Montreuil, France,
[e] Hemodialysis Department - Hôpital Privé de l'Est Parisien – 93600 Aulnay Sous-Bois, France,
[f] Hemodialysis Department - Clinique Médicale et Pédagogique Edouard Rist, 75016, Paris, France,
[g] Association des Néphrologues d'Ile-de-France - ANEDIF, 75015, Paris, France.

Abstract

We present the regional professional network to support the Renal Epidemiology Information Network (REIN) registry in maintaining high quality data production and information analyses in Ile-De-France region. The network is based on a long term partnership between the nephrologists and a regional methodology support unit. It integrates clinical research assistants for data quality control. We also present organizational methods on maintaining the registry and enhancing information analyses and automating analyses reports.

Keywords:

Registries, Data Science, End-Stage Renal Disease, Data Quality, Summary Report.

Introduction

The Renal Epidemiology and Information Network (REIN) registry [1,2,3] is organized at French national and regional levels, around a network of professionals and public health decision makers involved in ESRD health care and concerned by questions raised by kidney replacement therapies. The network contributes to the elaboration and assessment of health care strategies to improve ESRD prevention and patient care. REIN registry is supported by the Agence de la Biomedecine (ABM) [4] and by the Société Francophone de Néphrologie Dialyse et Transplantation (SFNDT) [5]. Data production started in January 2002, and was progressively extended to cover all French metropolitan regions by the end of 2010 and French overseas regions in 2012. REIN registry relies on primary data sources directly collected at bedside in the dialysis units. In contrast, the United States Renal Data System (USRDS) [6] and with the European Renal Association - European Dialysis and Transplant Association (ERA-EDTA) [7] both rely on secondary data sources.

Regional organization level was privileged since the inception of the REIN project. The region represents an appropriate level of proximity for patient data collection, quality control and data analyses. It is adapted for networking between professionals and decision makers at the regional institutions. REIN regional committees involve nephrologists, decision makers, public health insurers, methodologists and patient representatives. A nephrologist is elected as the REIN-regional program coordinator. Partnership between nephrologists and data analysis methodology departments, are highly recommended to guarantee resources, expertise and added value for data quality control, analyses and epidemiology studies. Nephrologists in dialysis care units in Ile-de-France region (about 12 million inhabitants in 2016) [8] have been participating since 2005 in the REIN registry. Data production in Ile-de-France was considered exhaustive and stable since year 2007.

Methods

REIN Regional Network Organization

The "Agence Regionale de Santé Ile-de-France" (ARS-IDF) represents the Public Health Ministry in the region. It plans ESRD related care strategies and manages resource allocations and authorizations for replacement therapy centers. Nephrologists gathered in "Association des Néphrologues d'Ile-de-France (ANEDIF)", to assure the deployment, development and management of the REIN project in the Ile-de-France region [9].

The Biostatistics Unit at Necker University Hospital actively participates in the promotion of the REIN project and developed an application called MSIS (Multi-Source Information System), used from 2002 to 2014, to facilitate REIN' data production and regional deployment [10]. The Biostatistics Unit assists and accompanies ANEDIF members in:

- Supervising the clinical research assistants training and missions.
- Developing methods and applications for assuring data integrity and quality.
- Providing methodology and application support for data analyses.

Clinical research assistants participate in ensuring the respect of the REIN charter [4] and interacting with nephrologists in the care units' environment. They consolidate data and give the start-signal for data analyses. Clinical research assistants facilitate communication and play a major role in maintaining the vivacity of the regional network and registry.

Analyses corpus contents

Regional data to be analyzed are presented in the form of CSV files and require minimalistic data re-arrangement. R statistics scripts are developed for analyses [11]. A data model adapted to statistical analyses is developed as a first step. Study patient populations are then filtered as a function of the analysis purpose: demographic, geographic, public health, disease and specific research topics, etc.

General outline of the statistical analyses subsets are agreed upon over time. They are systematically and longitudinally represented in the annual report. A final collegial proof-review validates the information in the synthesis report and presented to the annual assembly of the ANEDIF.

Report enhancements

Methods to automate the analysis processes from data source reading to producing analytic results in a suitable, easy to read and quality document are developed. R statistics remains the statistical software for data analyses. Related packages for tabular and graphics are used. Knitr package, is tested to automatically edit the results using LaTeX editor and to publish the report in a portable format document (.pdf) [12].

Results

Network status

Nephrologists in 215 care units participated in the producing quality data to the REIN regional registry. Their commitment through the ANEDIF association to the regional network remained active from more than a decade. The role of the clinical research assistant is essential for assuring the completeness of data production and quality control as well as for the regional network animation.

Data production and analyses improved over the years in the region. ANEDIF and the regional supporting Biostatistics Unit at Necker University hospital are the privileged interlocutors of the ARS-IDF for ESRD related decision making and care strategy policies. Nephrologists used the information presented in the annual report for improving their knowledge and as a support in discussions with local supervising authorities.

Report automation

REIN regional annual report in Ile-de-France has been finalized and automated. Its contents and organization have been accepted among professionals involved in the regional REIN network. It took few minutes to edit an updated report. Annual variations were commented as necessary. Results were presented in the form of free text comments, tables and graphics. For example : Flow diagram of ESRD patients in the region represent ESRD patient incidence in Ile-de-France which concerned 1864 new cases of resident patients, 12 non-residents and 95 ESRD patients who received transplantation as inaugural replacement therapy. ESRD prevalence in Ile-de-France 8106 ESRD resident patients in Ile-de-France were treated by dialysis on 31^{first} 2017, 58 resident patients were treated out of the region. Departmental disparities representing standardized ESRD incidence in Ile-de-France, were observed. Prevalence evolution shows a regular growth of an average of 160 patients per year during a 6 years study. The causes of re-hospitalizations in the context of ESRD are dominated by the context of pre- or post-kidney transplantation, vascular diseases, heart diseases, infectious disease and dialysis arterial-venous fistula.

Conclusion

Regional professional network to support REIN registry in maintaining high quality in data production and information analyses are described. The network organization is based on a long term partnership between the medical domain professionals, organized in agreed-upon association, a regional epidemiology and methodology supporting unit. It integrates from the start the clinical research assistants, for supporting the clinicians in data production and quality control. Methods on how to maintain and improve the network activities are extended to additional medical domain, e.g., Neurodevelopmental Disorders, in a project entitled “Troubles Envahissants du Dévelopment - Information System” (TEDIS).

References

[1] Décret n° 2002-1198 du 23 septembre 2002 relatif aux conditions techniques de fonctionnement des établissements de santé qui exercent l'activité de traitement de l'insuffisance rénale chronique par la pratique de l'épuration extrarénale et modifiant le code de la santé publique | Legifrance. Paris-France. [Internet]. Available from : https://www.legifrance.gouv.fr/eli/decret/2002/9/23/SANH0222824D/jo/texte.

[2] Gliklich, R.; Dreyer, N.; Leavy, M. Registries for Evaluating Patient Outcomes: A User’s Guide, 3rd ed.; Agency for Healthcare Research and Quality: Rockville, MD, USA, 2014.

[3] Kodra Y, Weinbach J, et al. Recommendations for Improving the Quality of Rare Disease Registries. Int J Environ Res Public Health. 2018 Aug 3;15(8).

[4] Agence de la Biomédecine. Saint Denis – France. [Internet]. Available from : https://www.agence-biomedecine.fr/?lang=fr.

[5] Société francophone de néphrologie dialyse et transplantation. Paris – France. [Internet]. Available from: http://www.sfndt.org/sn/index.php.

[6] U.S. Renal Data System. 2017 Researcher’s Guide to the USRDS Database. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2017. [Internet]. Available from : https://www.usrds.org/2017/rg/2017_USRDS_Researcher_Guide.pdf.

[7] European Dialysis and Transplant Association (ERA-EDTA) Registry. [Internet]. Available from: https://www.era-edta-reg.org/index.jsp?p=2.

[8] National Institute of Statistics and Economic Studies Panorama - Île-de-France » [Cited on 2017 Nov 16]. Paris-France. [Internet]. Available from: https://www.insee.fr/fr/statistiques/2018915.

[9] ASSOCIATION DES NEPHROLOGUES D'ILE-DE-FRANCE (ANEDIF). [Cited on 2005 April 16]. Paris-France. [Internet]. Available from: http://www.journal-officiel.gouv.fr/association/index.php.

[10] Ben Saïd M, Simonet A,et al.. A dynamic Web application within an n-tier architecture: a Multi-Source Information System for end-stage renal disease. Stud Health Technol Inform. 2003;95:95-100.

[11] Teetor P. R Cookbook. Sebastopol, (CA). O’Reilly; 2011.

[12] Xie Y. Dynamic Documents with R and Knitr, Second edition. Chapman & Hall/CRC The R series; Boca Raton , FL, 2016.

Address for correspondence

Mohamed Ben Said, Biostatistics Unit, Necker University Hospital. 149 rue de Sèvres. 75015, Paris. France.

Ph: +33-1-44495588. Mails: mohamed.bensaid@aphp.fr , mohamed.ben-said@ext.paris-descartes.fr.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190467

Standardizing Data from the Dead

Shamsi Daneshvari Berry[a] and Heather J.H. Edgar[b,c]

[a]Department of Health Informatics and Information Management, University of Mississippi Medical Center, Jackson, MS, USA
[b]Department of Anthropology, University of New Mexico, Albuquerque, NM, USA
[c]Office of the Medical Investigator, University of New Mexico, Albuquerque, NM, USA

Abstract

A database of full-body CT scans and associated lifestyle and health data from decedents who underwent an autopsy at the Office of the Medical Investigator (OMI) is under construction. The dataset has 68 metadata fields containing data from the OMI's database and interviews with next of kin. Some metadata fields could be mapped to existing standards, but the majority of fields required some modifications to current standards or the creation of new standards.

Keywords:

Vocabulary, controlled; imaging, three-dimensional; forensic anthropology

Introduction

Informatics is most commonly focused on using data from the living to improve the health of the living. The research presented here takes a novel approach, using data from the dead to create a database that can be used to improve the health of the living. Medical examiner data is a great resource for informaticians due to a number of factors, the most important of which is known cause of death.

The Office of the Medical Investigator (OMI) is a statewide, centralized medical examiner's office for the State of New Mexico. Any individual who is found dead with an unknown cause of death or who has died in a sudden, violent, untimely, or unexpected manner is routed to the OMI for a possible autopsy. In 2010, the majority of cases were deaths from natural causes (25%) or accidental deaths (35%), followed by suicides (17%), deaths from unknown causes (13.5%), and homicides (9.5%) [1]. The makeup of those cases closely mirrored the ethnic and racial composition of the state [2].

The Center for Forensic Imaging at the OMI was awarded a grant from the National Institute of Justice in 2010 to evaluate whether postmortem computed tomography (CT) scans could supplement or supplant a traditional autopsy. Between 2010 and 2017, this work produced over 15,000 whole-body 3D CT images.. The trove of image information is enormous; however, the scans were not associated with any organized metadata about lifestyle, health, or cause of death.

In 2014, a study was performed to determine a minimum data set using a modified Delphi method [3]. Researchers from a wide variety of fields (anthropology, medicine, forensics, informatics, epidemiology, biomedical research, and dentistry) used an iterative process to select 59 variables that they believed to be essential to making the CT scans useful for researchers in multiple fields.

The National Institute of Justice awarded a grant in 2016 (2016-DN-BX-0144) to create a database of deidentified CT scans along with associated lifestyle, health, and cause of death data. The associated information includes nine additional variables that are derived from the investigator's research. The data for all variables are derived from both VAST (the OMI's primary database, used to track the investigation of causes and manners of death) and phone interviews with next of kin.

Data in VAST is primarily stored in unstandardized free-text fields. The lack of standardization limits the ability to retrieve information effectively and necessitates a significant time investment to clean the data before it can be used. In its current form, as data complexity increases, data recovery becomes less accurate.

The New Mexico Decedent Image Database (NMDID) is available at NMDID.unm.edu.

Methods

Once the metadata fields were operationalized, each field was searched as a term in the Unified Medical Language System (UMLS) [4] to determine which data standards already existed for that concept. Each data standard was investigated and its ability to capture the appropriate level of data was analyzed. For example, some data was coming from next of kin and could not be expected to be at as technical a level as other data coming from medical records in the VAST database.

Vocabulary standards were selected, modified if necessary, and implemented in the database. For those metadata fields where no known vocabulary standard existed, the authors determined a new standard for this particular data set.

Results

After determining which variables needed to be combined, which needed to be separated into two or more fields, and adding those from the author's research, we had 68 metadata variables in three areas: census, health, and circumstances of death.

There were three categories of vocabulary standards for metadata capture:

1. An existing data standard that required no modification;
2. A data standard that existed but required modifications;
3. No appropriate data standard existed, so a new standard was created.

18 of the 68 metadata fields had existing standards that could be applied without modification. Standards were derived from multiple sources including LOINC (including trial codes), SNOMED CT, internal standards specific to the OMI, ISCO, ICD-10, and RxNorm [4].

Five of the 68 metadata fields required a modification from an existing standard. These standards were primarily from LOINC (including trial codes) [4].

The remaining 45 metadata fields did not have an appropriate data standard to draw upon. This included simple metadata fields like year of death or length of time at current occupation,

which was easily described as {yyyy}. However, creating standards for other fields was more difficult; for instance, when working with medical diagnoses, a balance had to be struck between knowledge of next of kin and medical detail.

Conclusion

Vocabulary standards are of the utmost importance in the creation of databases in order to ensure interoperability. As such, we searched medicine, forensics, anthropology, and other fields for standards to apply to our new database. 26% of the metadata fields could be mapped directly to existing standards. This included fields such as smoking status, sex, gender, and marital status. Seven percent required a modification of the standard to be applied to this particular database. Among those modified were birthplace and drug usage. The majority of the metadata fields (66%) required new standards to be developed. Many of the fields used in medical examiner data are not yet standardized, as the data historically have been in a free-text format.

Medical examiner data is an underused resource for public health data, because, unlike vital statistics data, it has generally been unavailable to the research public. However, medico-legal data has several advantages. For instance, it is associated with lifestyle and health data unlike vital statistics, which, while easily available and standardized, are not very rich. The NMDID is an entirely new resource that will provide over 15,000 CT scans and their associated metadata to researchers. NMDID sets a precedent, and perhaps some new standards, for the use of medico-legal data from other resources worldwide.

Acknowledgements

Funded by National Institute of Justice 2016-DN-BX-0144. Statements made are solely the responsibility of the authors.

References

[1] New Mexico Office of the Medical Investigator, *Annual Report 2013*, 2013.
[2] U.S. Census Bureau, *Modified Race Data 2010*, Washington, DC, 2010.
[3] S. Berry, *Metadata Determination for a Cadaveric Collection*, Masters Thesis, University of New Mexico, 2014.
[4] O. Bodenreider, The Unified Medical Language System (UMLS): integrating biomedical terminology, *Nucleic Acids Res* **32** (2004), D267-270.

Address for correspondence

Shamsi Daneshvari Berry is the corresponding author. She can be contacted at srberry@umc.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190468

Linked Open Data in the Biomedical Information Area: A Keywords Analysis

Stefano Bonacina

Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

Abstract

The objective of this paper was to determine the extent of the usage of "Linked Open Data" within biomedical literature. Applying PRISMA statement for literature reviews, forty-six papers were included in the analysis and keywords identified. Keywords have been classified according to MeSH categories, when possible. Twenty-three keywords had a frequency > one, 146 keywords had a frequency equal to one. Two MeSH categories were recurring. Future work includes applying association rules learning to keywords.

Keywords:

Semantic Web, Review Literature as Topic.

Introduction

Based on protocols defined by the World Wide Web Consortium (W3C), Semantic Web allows data sharing among applications and application-application cooperation, even without human intervention [1].

In the context of Semantic Web, the term "Linked data" is about linking and sharing resources of data (e.g., datasets, knowledge bases) generated and exposed by heterogeneous authors (e.g., governmental departments, research institutions, industry) [1]. When linked resources of data, e.g., heterogeneous datasets connected each other, are exposed through open access, the "Linked Open Data" paradigm is fulfilled, according to the "Linking Open Data" community project [2]. Operatively, "Linked Data" resources are represented with the Resource Description Framework (RDF) [3], a family of standardised data formats (RDF is explained briefly in the next section).

In 2016, we published a review paper that gave a view of research relating to "Linked Open Data" (LOD) in the area of health- and clinical-care [4]. Widening the scope, and including publications from the last two years, the objective of this paper is an attempt to determine the extent of the usage of linked open data within biomedical literature. The main questions guiding this review were as follows,

1. To what extent "Linked Open Data" is used within Biomedical Information research?
2. How could it be represented in a synthesised way?

The rest of the paper is organized as follows. The next section is a short background on Semantic Web technology, and RDF data format. Then, the methods used in the research and the obtained results are presented. Finally, some concluding remarks end the paper.

Semantic Web

Semantic Web stack is a set of protocols and computer languages, organised according to a layered structure, where communication is allowed between adjacent layers [5]. Semantic Web stack has been implemented to achieve data sharing among applications. In the stack, three main computer languages are: the Web Ontology Language (OWL) [6], the Resource Description Framework (RDF) Language [3], and the SPARQL Protocol And RDF Query Language - SPARQL language [7], which has a recursive acronym. OWL and RDF are for specifying ontologies, which are collections of data and meta-data exposed to the web and shared among computer applications [1,6]. Similarly to a simple sentence in English (which is composed by a Subject a Predicate, and an Object), a RDF triple - the basic element of this language - is composed by a Subject, a Predicate, and an Object. An example is, "The patient is suffering from diabetes", where "The patient" is the Subject, "is suffering from" is the Predicate, and "diabetes" is the Object. Then, , a set of triples is represented by a graph, where the nodes of the graph are the Subjects and the Objects, and the edges are the Predicates. Triples are stored according to the Extensible Mark-up Language (XML) [3].

Methods

To answer the first research question, a review of the literature has been performed, following the different phases of a systematic review (PRISMA Statement), described in [8]. Then, to answer the second research question, MeSH hierarchy has been used to aggregate concepts represented by the keywords associated to each paper. The following sub-sections describe the applied methods in detail.

Search strategy

The author searched online bibliographic databases for relevant papers. Specifically, PubMed/Medline, Scopus, Web of Science (MEDLINE), Web of Science (All the other Indexes), and IEEE Xplore - Digital Library were enquired, using "Linked Open Data" as main search term. Because Scopus, Web Of Science (All the other Indexes), and IEEE Xplore - Digital Library are not specific to biomedical literature, the searches on these databases were limited to the following subject areas: Biochemistry, Genetics and Molecular Biology, Medicine, Pharmacology, Toxicology and Pharmaceutics, Psychology, Immunology and Microbiology, Neuroscience, and Nursing. In this way, the search was limited to papers in the Biomedical Information area.

Inclusion and exclusion criteria

Inclusion criteria were as follows: 1 - Journal articles in English published from 1st January 2009 to 13th August 2018 (the date of searches); 2 - Papers published on journals related to Linked Open Data within the Biomedical Information Domain; 3 - Abstract and full text pdf are available; and 4 - Papers should describe Linked Open Data solutions within the Biomedical Information Domain.

The exclusion criteria were complementary to the inclusion criteria. In addition, duplicated items were considered and deduplicated. The final selection of papers was performed by screening title, abstract, and through reading the full text, when more details were needed.

Data extraction

To answer the second research question, keywords associated to each paper were extracted. Keywords are descriptors of the content of a publication, and in the present research, they have been preferred to the title, as the titles cannot reflect the contents of a publication [9]. Keywords are defined by the authors, even according to a specified vocabulary (e.g., Medical Subject Headings-MeSH). Publications indexed in Medline database are also described by MeSH terms. Publications in IEEE are indexed according to IEEE descriptors.

Keywords describing the papers selected from the previous phase of the research have been collected in a MS Excel spreadsheet and then processed according to the following steps. First, keywords on the full text pdf have been considered; then, if they were not included, the journal web page describing the paper has been considered, and the keywords available there collected. After that, if the keywords were not available on the web site, MeSH terms have been considered for the publications available on Medline. With this process all the papers had a set of words associated to them.

The phrase “Linked open data” and combinations of those words have been excluded, as the main query in the search phase was that one (i.e., to retrieve papers having LOD as main topic), duplications have been removed, and acronyms have been used to replace their explanations (e.g., RDF instead of “Resource Description Framework”). After that, the entire list has been alphabetized and the frequency of each keyword counted, and stored in the spreadsheet. Finally, a MeSH term and a MeSH category have been connected with each keyword, when possible.

Results

The search phase provided 181 potential articles, and following the four-phase flow diagram [8], 46 have been included in the analysis. Due to space limitation, the references of the 46 included papers are not included in the present manuscript; however, they are available upon request. The analysis of the 46 papers gave 265 keywords, including duplications. After processing the keywords according to the methods explained in the previous section, 225 keywords had been obtained, including duplicates. 169 distinct keywords were extracted; 23 had a frequency from 2 to 15 (maximum frequency), while 146 had a frequency equal to one. The most recurrent keyword was “RDF” (15 times), followed by “Semantic web” (9 times), “Electronic Health Records” appeared 4 times, and “Unified Medical Language System” – UMLS appeared 3 times. The main MeSH categories matched by the keywords were "Information Science Category" (45 keywords) and "Analytical, Diagnostic and Therapeutic Techniques and Equipment Category" (12). Unfortunately, it was not possible to associate 80 keywords to a MeSH term (and category).

Conclusions

The objective of this paper was to determine the extent of the usage of linked open data within biomedical literature. Applying the PRISMA statement for literature reviews, forty-six papers have been included in the analysis and their keywords identified. Twenty-three keywords had a frequency greater than one (and the maximum frequency was 15), 146 keywords had a frequency equal to one. "Information Science Category" and "Analytical, Diagnostic and Therapeutic Techniques and Equipment Category" of MeSH were the main MeSH categories keywords belong to. Future work includes applying association rules learning (shopping basket analysis) to the keywords.

Acknowledgements

This research has been supported by the Health Informatics Centre (HIC), Institutionen för Lärande, Informatik, Management och Etik (LIME), Karolinska Institutet, Stockholm, Sweden.

References

[1] T. Berners-Lee, J. Hendler, O. Lassila, The Semantic Web, *Scientific American* 284 (2001), 34-43.
[2] C. Bizer, T. Heath, T. Berners-Lee, Linked data, the story so far, *International Journal on Semantic Web and Information Systems (IJSWIS)* 5 (2009), 1-22.
[3] G. Klyne, J.J. Carroll, editors, Resource Description Framework (RDF): Concepts and Abstract Syntax. [Online, 2003], W3C Proposed Recommendation 15 December 2003. Available at http://www.w3.org/TR/rdf-concepts/ Last Access: 22 November 2018.
[4] S. Bonacina, Linked Open Data in Health and Clinical Care. A Review of the Literature, *European Journal for Biomedical Informatics – EJBI* 12 (2016), en2-en11..
[5] T. Berners-Lee, WWW past & future. [Online, 2003]. Available at http://www.w3.org/2003/Talks/0922-rsoc-tbl/. Last Access: 22 November 2018.
[6] W3C OWL Working Group, OWL 2 Web Ontology Language Document Overview (Second Edition). 11 December 2012. W3C Recommendation. URL: http://www.w3.org/TR/ owl2-overview/ Last Access: 22 November 2018.
[7] E. Prud'hommeaux, A. Seaborne, editors, SPARQL Query Language for RDF. W3C Recommendation 15 January 2008. [Online, 2008]. Available from: http://www.w3.org/TR/rdf-sparql-query/ Last Access: 22 November 2018.
[8] A. Liberati, D.G. Altman, J. Tetzlaff, C. Mulrow, P.C. Gøtzsche, J.P.A. Ioannidis, et al, The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration, *BMJ* 339 (2009), b2700.
[9] R.G. McGee, J.C. Craig, What is being published? A word cloud of titles from the Journal of Paediatrics and Child Health, *J Paediatr Child Health*, 48 (2012), 452.

Address for correspondence

Stefano Bonacina, M.Sc. | Ph.D. | Senior Researcher in Health Informatics | Health Informatics Centre | Department of learning, informatics, management and ethics | Karolinska Institutet | Tomtebodavägen 18a, 171 77 Stockholm, Sweden.
Phone: +46 8 524 83346 Email: stefano.bonacina@ki.se
Web site: http://ki.se/HIC

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190469

Oral e-Health: Definition of Essential Attributes of Oral Health for the Information Record in Primary Care

Renata Dutra Braga[a], Panel of Specialists in Oral Health (Definer Group)[1b], Fábio Nogueira de Lucena[c], Rejane Faria Ribeiro-Rotta[d]

[a] *Postgraduate Program in Health Sciences, Universidade Federal de Goiás, Goiânia, Goiás, Brazil*
[b] *Researchers and staff, Universidade Federal de Goiás, Goiânia, Goiás, Brazil (School of Dentistry)*
[c] *Institute of Informatics, Universidade Federal de Goiás, Goiânia, Goiás, Brazil*
[d] *School of Dentistry, Universidade Federal de Goiás, Goiânia, Goiás, Brazil*

Abstract

Standardization is essential for information sharing among different health care institutions. Our objective was to identify the essential oral health attributes to include in an electronic health record for primary care. This action research study utilized a Definer Group, which selected attributes as a mind map, into four main pillars: Data Collection, Diagnosis, Care Plan and Evaluation. This research applied the practice of knowledge leveling, favoring the interaction of dental specialties and identification of attributes.

Keywords:

Oral Health; Primary Health Care; Dental Records.

Introduction

There is no health care without information. "The collapse of comprehensive health care and the inadequate use of financial resources are directly associated with a lack of access to readily available and reliable information, as well as rework and lack of support for clinical decision-making" [1].

Clinical decision-making based on fragmented and impaired information has a direct impact on the prognosis and outcome of health care treatment. Longitudinal monitoring and multiprofessional health care are also compromised and even impossible to perform [2].

Primary health care is the main gateway for the individual in the Brazilian Unified Health System (SUS). This reinforces the importance of a ready and trustable information access system, since this level of health care aims to develop comprehensive care that impacts the general health status, the autonomy of people, and the determinants and health conditions of the collectivities [3].

Standardization is essential for information sharing among different health care institutions and health professionals. This requires a national effort in e-health, with large investments, participation of various segments and health actors, which several countries have carried out [3]. The e-health is defined as "the combined use of electronic information and communication technology in the health sector for clinical, educational, research, and administrative purposes, both at the local site and across wide geographic regions" [4]. In Brazil, guidelines and actions established in the e-health vision [5] already reveal the prioritization of the subject by Brazilian government. However, what data should be recorded?

Despite the advance of e-health in the SUS [5], most of the information provided by its various systems relates to administrative management, to feed health indicators that provide partial or skewed information by registry instruments. Clinical information focusing on the individual, sufficient for clinical decision-making, is not yet available in an integrated and longitudinal format.

The development of the interoperable Electronic Health Record (EHR), in particular for primary care, will require a great effort. It must involve different health care professionals working to define the information model, aiming to ensure the completeness and interdisciplinary that primary health care needs [6].

In 2016, the researchers' group of this study proposed an informational model that meets the principle of integrality in health care, with a comprehensive approach [6]. This model intends to be the starting point for the different primary care professions to define, with greater granularity of details, the set of essential information to be registered. Dentistry makes up a set of health professions that work in primary care, and based on the information model mentioned, the objective of this study was to identify the essential attributes on oral health to be contemplated in a multiprofessional electronic health record for primary care.

Methods

This observational, exploratory, cross-sectional and action research study was approved by the Ethical Committee at Universidade Federal of Goiás, Brazil (protocol 2,206,915).

It included five steps: a) Selection of the experts' panel in dentistry (Definer Group - DG); b) Action research - knowledge leveling workshop; c) Individual activity by specialty (comparison of the health record used by the dental specialties with the defined semiological tree); d) collective activity (proposal of Oral e-health as a mind map format); e) selection of attributes for validation and modeling.

The DG was composed of five dental specialties, representing the areas of Stomatology, Periodontics, Collective Health, Pediatric Dentistry, and Endodontics. These specialists fulfilled the inclusion criteria and agreed to participate in the study, comprised of three males and three females [7].

Of the eight invitations sent, two professionals did not accept for the unavailability of time. In terms of professional training, all participants were masters, with publications and dissertations defended in one of the following areas: Semiology, Stomatology, Endodontics, Periodontics, Public Health, or Pediatric Dentistry. Of these, 83% completed their doctorate degree, 17% had their doctorate in progress, and 66% had at least one specialization.

[1] Panel of Specialists in Health - Definer Group, see the Appendix A.

All participants have been teaching, performing research, or other extension and oral health care activities.

The DG performed action research, by means of knowledge leveling workshops on health information collection, mediated by professors from the area of Semiology in Dentistry [8].

The result of the collective workshops, face-to-face, and distance meetings allowed building a mind map, using the FreeMind tool. This strategy resulted in a model, the Oral e-Health, integrated to the multiprofessional information model.

Results and Discussion

The mind map, a product of more than 50 hours of meetings, contained four pillars: Data Collection, Diagnosis, Care Plan and Evaluation. A total of fifteen face-to-face workshops were held with the DG, in addition to approximately twenty hours of individual meetings.

In this first step of defining Oral e-health, the DG focused its work on the data collection pillar of the multiprofessional information model [6], a priority for the continued development of the integration of dental health attributes to the general health of the individual in the other pillars (Diagnosis, Care Plan and Evaluation).

The mind map, a product of the workshops performed with the DG, resulted in the modification, detailing, and inclusion of several items related to Data Collection. The complete mind map version is available in FreeMind format at https://github.com/professorarenatabraga/Oral-e-Health.

The mapping of this set of attributes to the construction of an electronic health record is an important activity; perhaps, it is one of the highest priorities, which requires the involvement and collaboration between professional with diffferent areas of knowledge [3].

This action research was interdisciplinary, since the DG performed the investigation of the essential attributes of each dental specialty simultaneously with the collective knowledge leveling of semiological concepts; this has contributed to each professional understanding the importance of attributes of the other specialty, their overlaps, the need for exclusion or inclusion of terms, and how to present and organize the information used in their routine clinical practice.

The profile of the professionals that composed the DG, determined by the criteria adopted in the selection, ensured the scope of primary care, without losing the detail and flexibility necessary for evidence-based clinical decision-making in a multiprofessional context [1,8].

The use of the mind map allowed the DG an easy navigation, as well as an extended and at the same time detailed view of the set of attributes that was being delineated [9]. The mind map included the essential attributes for oral health in the scope of primary care, integrated with those of general health, and will be the basis for the elaboration of an informational model, taking into account the guidelines and actions proposed by the Ministry of Health [5].

The attributes presented in the mind map, besides serving as a basis for the elaboration of the information model, also constitute the basis for the construction of archetypes, an independent approach of technology. The use of archetypes and terminology standards is decisive for the promotion of the desired level of interoperability of health information record systems and their consequences: search and sharing of health information with quality, and improvements in health care, like in Norway, Germany, New Zealand and Australia [3].

Conclusions

The main contribution of this study was to increase the level of detail, completeness and pertinence of the essential attributes for oral health primary care, presenting a methodological approach to do it. The intervention of the action research was crucial to leveling semiological knowledge, favoring the interaction of different dental specialties on establishing essential information for an oral health register. The workshops allowed a reflection on the essentiality of the information that each specialty required individually and how they could be rearranged collectively to the individuals' general health register.

Specialists interaction as a Definer Group was fundamental for identification of Dentistry´s concepts, in the primary health care context, that will require a formal clinical modeling using the standards adopted by the Brazilian Ministry of Health.

Acknowledgements

To CAPES Foundation (Coordination for the Improvement of Higher Education Personnel) by the scholarship funding. The participation and support of all the individuals during this study are thankfully acknowledged.

References

[1] G. Kopanitsa, Integration of hospital information and clinical decision support systems to enable the reuse of electronic health record data, *Methods Inf. Med.* **56** (2017), doi:10.3414/ME16-01-0057

[2] E. Kalenderian, J.D. Halamka, H. Spallek, An EHR with Teeth, *Appl. Clin. Inform.* **7** (2016) 425–429, doi:10.4338/ACI-2015-09-LE-0124

[3] M.Z. Huang, C.J. Gibson, A.L. Terry, Measuring Electronic Health Record Use in Primary Care: A Scoping Review, *Appl. Clin. Inform.* **9** (2018) 15–33.

[4] Y.J. Hong, J. Kim, Analysis of Interactive E-Health Tools on United Arab Emirates Patient Visited Hospital Websites, *Healthc. Inform. Res.* **25** (2019) 33–40, doi:10.4258/hir.2019.25.1.33

[5] M. da S. Brasil, Estratégia e-Saúde para o Brasil., (2017) 80, Retrieved from http://bit.ly/2GQ2YEi.

[6] R.D. Braga, F.N. Lucena, Panel of Specialists in Health, R.F. Ribeiro-Rotta, A multiprofessional information model for Brazilian primary care: Defining a consensus model towards an interoperable electronic health record, *Int. J. Med. Inform.* **90** (2016) 48–57.

[7] C.G. Rodrigues, M.A. Moraes, J.M. Sauer, R.A.K. Kalil, E.N. de Souza, Nursing diagnosis of activity intolerance: Clinical validation in patients with refractory angina., *Int. J. Nurs. Terminol. Classif.* **22** (2011) 117–122, doi:10.1111/j.1744-618X.2011.01182.x

[8] R.F. Toledo, P.R. Jacobi, Action research and education: Sharing basic knowledge and community empowerment to face problems, *Educ. Soc.* **34** (2013) 155–173, doi:10.1590/S0101-73302013000100009

[9] H. Kashfi, J. Robledo-Jr, *Towards a Case-Based Reasoning Method for openEHR-Based Clinical Decision Support*. In Proceedings 3rd Int. Work. Knowl. Represent. Heal. Care, (2011), 182–195.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190470

P-Hacking Lexical Richness Through Definitions of "Type" and "Token"

K. Bretonnel Cohen[a], Lawrence E. Hunter[a], and Peter S. Pressman[b]

[a]Computational Bioscience Program, University of Colorado School of Medicine, Aurora, Colorado, USA
[b]Department of Neurology, University of Colorado Hospitals, Anschutz Medical Campus, Aurora, Colorado, USA

Abstract

"P-hacking" is the repeated analysis of data until a statistically significant result is achieved. We show that p-hacking can also occur during data generation, sometimes unintentionally. We use the type-token ratio to demonstrate that differences in the definitions of "type" and "token" can produce significantly different results. Since these terms are rarely defined in the biomedical literature, the result is an inability to meaningfully interpret the body of literature that makes use of this measure.

Keywords:

Language, vocabulary

Introduction

There is a growing awareness that many published scientific results are not reproducible. Statistical practices are one of the most common factors blamed for this "reproducibility crisis," as it has come to be known [1]. There are many aspects of the contributions of statistical malpractice to the reproducibility crisis, including an increasingly recognized practice known as p-hacking. *P-hacking* is the act of repeatedly re-analyzing data until a statistically significant result is achieved [1; 2]. One method of p-hacking is varying aspects of the hypothesis tests themselves until a positive finding is reached; another is reselecting subsets of the data. However, there are other ways that p-hacking can be done, without "cherry-picking" hypothesis tests *or* data – as this paper shows, it can also be done at the data generation phase. This kind of p-hacking is especially pernicious for at least four reasons: (1) it leaves no trace in the form of questionable statistical assumptions that might otherwise be noticed by a reviewer, (2) it leaves no trace even when using publicly available data sets, since the data itself is unchanged, (3) it is not difficult to justify to oneself, and (4) publication practices make it quite easy. The approach in question is especially relevant to the computational bioscience community because it is easily illustrated with a measure that is widely used in neuroscience, psychiatry, developmental psychology, and neurology: the type-token ratio.

We illustrate the issues with the concept of "lexical richness" because its calculation seems very straightforward but is quite nuanced in practice. Informally, *lexical richness* refers to the quality, variability, and sophistication of the vocabulary of a speaker or a text. It is usually defined by some form of the *type-token ratio,* calculated as the number of distinct words (*types*) divided by the total number of words (*tokens*). It is commonly used in several biomedical fields [3]. Despite the prevalence of this measure in the scientific literature, it is unclear how to interpret the literature using the type-token ratio because papers typically do not define what they are counting as types and tokens, and there is no consensus definition for either.

Publication bias in the literature on diagnostic use of measures of lexical richness may promote p-hacking and unreliable results. We know that the tendency is not to be able to publish negative results [4]. Do people sometimes modify those "simple" decisions perfectly innocently, without actually realizing their linguistic consequences, until they get a positive finding on a subsequent statistical hypothesis test—a positive finding that would in fact be a negative finding if they made different preprocessing decisions? It is not difficult to do so, because "type" and "token" are generally not defined when used in the literature or elsewhere. We present modeling results that explore the implications of failures to consistently define those variables. If we find that modifying the definitions of "type" and "token" can affect whether or not statistically significant differences are found in the type-token ratio, then it should be clear that these terms *must* be defined when using them. If, in fact, they are *not* typically defined, then there is a problem in our field, and we should be aware of it and take action by being explicit about those definitions.

Materials and Methods

The materials are drawn from the CRAFT corpus of biomedical journal articles [5]. We modeled two different definitions of *type* and three different definitions of *token.* Definitions of "type" unavoidably interact with definitions of "token", which can result in different classifications of the same term depending on precedence orderings, so we processed all data sources using 10 random orderings of random subsets of five sets of type and token definitions. We used the conservative (nonparametric) Wilcoxon signed rank test to look for statistically significant differences between the distributions of type-token ratios in the outputs. The null hypothesis is that there is no effect of different definitions of the variables. For our definitions, we used observations from three comprehensive analyses of tokenization [6-8]. See the code available on GitHub for how we handled phrasal verbs (e.g. *have_to* versus *have to*), negative clitics (e.g. *do n't* versus *don't*), and repetitions. Type definitions applied here include normalized or unnormalized digit types or letter cases, punctuation inclusion in or exclusion from the token counts, and different treatments of hyphens and underscores. (See [9] for how contentious any of these decisions can be.) There are interactions between any definition of type or token, so we ran different combinations of the various definitions of each and their orderings, all randomized, for 10 combinations of definitions and orderings. Type-token ratio was calculated as

the count of types divided by the count of tokens for the complete text.

Analysis of the resulting distributions follows the approach of [10] (see also [8; 11]). Figure 1 shows the distributions of the type-token ratios for 10 randomly selected permutations of type definition, token definition, and order of application. Results cluster into groups with two widely differing magnitudes, one just below a type-token ratio of 0.5 (1, 3, 5, and 10) and the other around 0.2 (2, 4, 6, 7, 8, and 9); clearly, the definitions of type and token can have an enormous effect on the magnitude of the type-token ratio. In no case are the type-token ratios normally distributed, meaning the type-token ratio is vulnerable to Type I errors from parametric hypothesis tests.

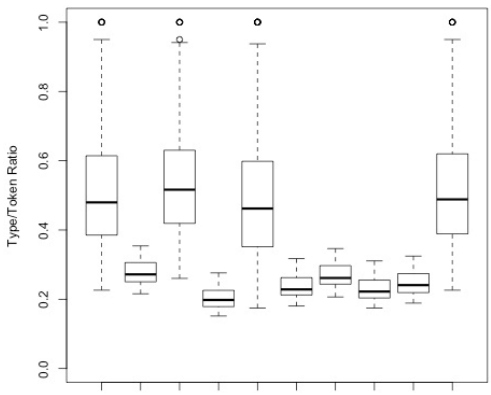


Figure 1 – 10 random permutations of subsets of the definitions of type and token. Each data point is the type-token ratio for one paper. Each column represents one permutation.

The distributions of the permutations with the highest and lowest median type-token ratios in the low-magnitude group were significantly different ($p < 0.001$); within the high-magnitude group, however, they were not. We also tested all pairs of permutations within the two groups. Table 1 shows that more than 40% of the definition set pairs did not yield statistically significant differences, while nearly 60% did, suggesting that there are many ways that statistically significant differences can be found (or not) based only on the different definitions of "type" and of "token". Like typical p-hackers, we did no multiple testing correction.

The complete code base and outputs of all steps of the analysis are available on GitHub.

Table 1 – Number of pairwise differences between sets of definitions of type and token, divided into groups with high- and low-magnitude type-token ratios.

Group	Pairs	Significantly different (%)
High	7	0 (0%)
Low	16	13 (81%)
All pairs	23	13 (57%)

Discussion

These results demonstrate that there are many "minor" preprocessing decisions that can be the difference between having statistically significant results and not, independent of any actual differences in the underlying data.

Conclusions

The precise definitions of "type" and "token" are so important that, if they are absent, research findings based on the type-token ratio are essentially uninterpretable. In fact, precise definitions of these terms are rarely given in the biomedical literature — even in work on tokenization [11]! We have shown that type-token ratios can be significantly different based only on differences in how "type" and "token" are defined. These results suggest that the relevant literature should be read with caution. Talking about his initial work on this problem. J. Simmons, the originator of the term "p-hacking," says that "we realized entire literatures could be false positives" [12]. If research using measures of lexical richness is to avoid turning out that way, future work in the field should always give the precise definitions of "type" and "token" that were used.

References

[1] M.L. Head, L. Holman, R. Lanfear, A.T. Kahn, and M.D. Jennions, The Extent and Consequences of P-Hacking in Science, *PLOS Biology* **13** (2015), e1002106.

[2] D.V.M. Bishop and P.A. Thompson, Problems in using p-curve analysis and text-mining to detect rate of p-hacking and evidential value, *PeerJ* **4** (2016), e1715.

[3] G. Fergadiotis, H.H. Wright, and T.M. West, Measuring lexical diversity in narrative discourse of people with aphasia, *Am J Speech Lang Pathol* **22** (2013), S397-408.

[4] C. Wilson, D. Kerr, A. Noel-Storr, and T.J. Quinn, Associations with publication and assessing publication bias in dementia diagnostic test accuracy studies, *Int J Geriatr Psychiatry* **30** (2015), 1250-1256.

[5] K.B. Cohen, K. Verspoor, K. Fort, C. Funk, M. Bada, M. Palmer, and L. Hunter, The Colorado Richly Annotated Full Text (CRAFT) Corpus: Multi-Model Annotation In The Biomedical Domain, in: *Handbook of Linguistic Annotation*, 2016.

[6] K. Tomanek, J. Wermter, and U. Hahn, A reappraisal of sentence and token splitting for life sciences documents, *Stud Health Technol Inform* **129** (2007), 524-528.

[7] D. Trieschnigg, W. Kraaij, and F. de Jong, The influence of basic tokenization on biomedical document retrieval, in: *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2007, pp. 803-804.

[8] N.P. Cruz Diaz and M. Maña López, An Analysis of Biomedical Tokenization: Problems and Strategies, in: *Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis*, Association for Computational Linguistics, Lisbon, Portugal, 2015, pp. 40-49.

[9] É. Née, *Méthodes et outils informatiques pour l'analyse des discours*, Presses universitaires de Rennes, 2017.

[10] B. Habert, G. Adda, M. Adda-Decker, P.B. de Marëuil, S. Ferrari, O. Ferret, G. Illouz, and P. Paroubek, Towards tokenization evaluation, in: *Proceedings of LREC*, 1998, pp. 427-431.

[11] Y. He and M. Kayaalp, A Comparison of 13 Tokenizers on MEDLINE, *The Lister Hill National Center for Biomedical Communications* **48** (2006).

[12] S. Dominus, When the Revolution Came for Amy Cuddy, in: *The New York Times Magazine*, New York Times, New York, 2017.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190471

Mapping Medication Metadata from the ABDA Data Model to an OpenEHR Medication Archetype: A Qualitative Analysis

Tobias Bronsch[a], Ruwen Böhm[b], Claudia Bulin[b], Björn Bergh[a], Björn Schreiweis[a]

[a] *Institute for Medical Informatics and Statistics, University Hospital Schleswig-Holstein and Kiel University, Kiel, Germany*
[b] *Institute of Experimental and Clinical Pharmacology, University Hospital Schleswig-Holstein and Kiel University, Kiel, Germany*

Abstract

Integrating data from various source systems to gain knowledge and meaningful data about patients for care and research is challenging. This work demonstrates how medication knowledge data from the database of the Federal Union of German Associations of Pharmacists (ABDA) can be used for storing and annotating medicinal products in an openEHR medication archetype.

Keywords:

Electronic Health Records, Datasets as Topic, Health Information Exchange

Introduction

Medical data are reused for both patient care and biomedical research, not only within the boundaries of healthcare institutions, but very often also across multiple sites. However, data are often not findable, accessible, interoperable and reuseable (FAIR Data Principles [1]), as they are missing common terminologies for semantic interoperability. The German Federal Ministry for Education and Research (BMBF; ger: Bundesministerium für Bildung und Forschung) is tackling these issues with the German Medical Informatics Initiative (MII) by funding university hospitals to implement Data Integration Centers to make data available and interoperable on a national and international level [2]. University Hospital Schleswig-Holstein (UKSH) joined HiGHmed, one of the four MII-projects, and is implementing the HiGHmed use case cardiology [3]. For this use case among other data, medication data need to be integrated.

At UKSH, the management and documentation of medication data is based on names and identifiers of a de-facto-standard database provided by the Federal Union of German Associations of Pharmacists (ABDA) [4].

The objective of this work is to show the extent of overlap between the ABDA metadata and an openEHR medication archetype. Our aim is to provide a first step for storing and annotating medication data in openEHR archetypes using ABDA information.

Methods

ABDA Database

ABDA provides a comprehensive drug database with information on all medicinal products on the German market and its contents are constantly updated every 14 days. It comprises particularly details on active and inactive ingredients using different naming conventions, pharmacological classification systems, pharmaceutical and clinical instructions, summary of product characteristics as well as package inserts [4].

OpenEHR

OpenEHR describes a platform approach enabling open electronic health record (EHR) architectures using archetypes which are shared nationally and internationally [5].

In order to integrate data of source systems into openEHR-based systems, data need to be mapped and transformed.

Mapping of ABDA Metadata Items to an OpenEHR Archetype

For the mapping of the ABDA data model the tables ‘medicinal products’ (FAM_DB), ‘substances’ (STO_DB) and ‘names of substances’ (SNA_DB) were considered. In total 79 ABDA metadata items were used for the mapping.

Two medical informaticians (TB and BS) separately analyzed the 79 metadata items and mapped them to an openEHR archetype. Inter-rater reliability for TB and BS was calculated afterwards. Next, they compared their results per item and discussed disagreements to agree upon a mapping per item. In cases where the discussions did not lead to an agreement, a pharmacologist (RB) was included in the discussion and helped resolve the mapping.

Results

The ABDA metadata were mapped to the openEHR medication archetype (namely openEHR-EHR-CLUSTER.medication.v1).

The independent mapping of 79 ABDA metadata items to the archetype leads to initial agreements on 54 out of 79 ABDA metadata items (68.4 %) with disagreements for 25 ABDA metadata items. 26 ABDA metadata items could not be mapped to the archetype. Inter-rater-reliability is $p_o = \frac{54}{79} \approx 0.684$. Dicussions on the disagreements could resolve 23 out of 26.

However, 36 out of 79 ABDA items (45.6 %) were mapped to the generic free-text item ‘description’ of the archetype for integrating structured ABDA information.

Discussion

Mapping structured ABDA metadata items to free-text openEHR items results in loss of information. Therefore, more structured items that fit to structured ABDA items may be necessary in openEHR archetypes.

Instead of integrating ABDA information into openEHR archetypes, openEHR patient medication data could also be linked to the ABDA database by using a unique identifier from ABDA and reference that identifier together with the version of ABDA that was used.

This could be done using a unique identifier from ABDA to identify a specific medicinal product or substance in the openEHR medication archetype item 'Name' as an openEHR DV_CODED_TEXT (e.g. 'Tamoxifen beta 20'). This item is specified by an openEHR CODE_PHRASE (e.g. terminology_id = 'ABDA2016.FAM_DB.Key_FAM' and code_string = ' 3340739900') using a terminology mapping (openEHR TERM_MAPPING).

The pros and cons of that alternative need to be further evaluated.

Related Work

Other work has been done using openEHR archetypes to represent medication information before, like Chen et al., who showed that using openEHR archetypes for managing chemotherapy information and representing chemotherapy guidelines in openEHR is possible [6]. Marco-Ruiz et al. built an archetype-based data warehouse, but are focusing their information model on laboratory data [7].

Medication information has previously been transformed and mapped to other data models. E.g. HL7 Fast Healthcare Interoperability Resources (FHIR) was used to implement a system storing structured medication information retrieved via an NLP pipeline from unstructured documentation [8]. Sinha et al. built a clincial decision support system for safe opioid prescription based on FHIR [9].

Limitations

This work has to be considered a preliminary step for integrating medication data in an openEHR environment.

In addition, the ABDA database is specific to Germany. Although it contains both foreign pharmaceutical products and substance names this database might not be sufficient for international purposes.

Since the effort presented in this work has to be conducted once and only future changes of ABDA metadata need to be included, an automated mapping system was not considered. Structural changes to ABDA happen once per one year or every second year. The mapping does not imply regulatory or ethical challenges, due to metadata being mapped and not contents.

Conclusions

The selected metadata items concerning medication information can be mapped from the ABDA data model to an openEHR medication archetype. Based on these results our next steps are the integration of medication information in an openEHR Clinical Data Repository.

Acknowledgements

This work was funded by the German Federal Ministry for Education and Research (BMBF) as part of grant no. 01ZZ1802T.

References

[1] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A. t Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao and B. Mons. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data **3**(2016): 160018.

[2] S.C. Semler, F. Wissing, and R. Heyder, German Medical Informatics Initiative, *Methods Inf Med* **57** (2018), e50-e56.

[3] B. Haarbrandt, B. Schreiweis, S. Rey, U. Sax, S. Scheithauer, O. Rienhoff, P. Knaup-Gregori, U. Bavendiek, C. Dieterich, B. Brors, I. Kraus, C.M. Thoms, D. Jager, V. Ellenrieder, B. Bergh, R. Yahyapour, R. Eils, HiGHmed Consortium and M. Marschollek, HiGHmed - An Open Platform Approach to Enhance Care and Research across Institutional Boundaries, *Methods Inf Med* **57** (2018), e66-e81.

[4] ABDA, ABDA Datenbank, http://abdata.de/datenangebot/abda-datenbank/ last access: 16.11.2018.

[5] S. Garde, P. Knaup, E. Hovenga, and S. Heard, Towards Semantic Interoperability for Electronic Health Records - Domain Knowledge Governance for openEHR Archetypes, *Methods Inf Med* **46** (2007), 332-343.

[6] R. Chen, P. Georgii-Hemming, and H. Ahlfeldt, Representing a chemotherapy guideline using openEHR and rules, *Stud Health Technol Inform* **150** (2009), 653-657.

[7] L. Marco-Ruiz, D. Moner, J.A. Maldonado, N. Kolstrup, and J.G. Bellika, Archetype-based data warehouse environment to enable the reuse of electronic health record data, *Int J Med Inform* **84** (2015), 702-714.

[8] N. Hong, A. Wen, F. Shen, S. Sohn, S. Liu, H. Liu, and G. Jiang, Integrating Structured and Unstructured EHR Data Using an FHIR-based Type System: A Case Study with Medication Data, *AMIA Jt Summits Transl Sci Proc* **2017** (2018), 74-83.

[9] D. Delamarre, G. Bouzille, K. Dalleau, D. Courtel, and M. Cuggia, Semantic integration of medication data into the EHOP Clinical Data Warehouse, *Stud Health Technol Inform* **210** (2015), 702-706.

Address for correspondence

Tobias Bronsch

email: tobias.bronsch@uksh.de.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190472

Using Big Data Techniques to Improve Prostate Cancer Reporting in the Gauteng Province, South Africa

Cassim N[a,b], Mapundu M[c], Olago V[c,d], George JA[b,e] and Glencross DK[a,b]

[a] *Department of Molecular Medicine and Haematology, University of the Witwatersrand, Johannesburg, South Africa*
[b] *National Health Laboratory Service (NHLS), Johannesburg, South Africa*
[c] *School of Public Health, Faculty of Health Sciences, University of Witwatersrand, Johannesburg, South Africa*
[d] *National Health Laboratory Service (NHLS), National Cancer Registry (NCR), Johannesburg, South Africa*
[e] *Department of Chemical Pathology, Faculty of Health Sciences, University of Witwatersrand, Johannesburg, South Africa*

Abstract

Prostate cancer (PCa) data is of public health importance in South Africa. Biopsy data is recorded as semi-structured narrative text that is not easily analysed. Our study reports a pilot study that applied predictive analytics and text mining techniques to extract prognostic information that guides patient management. In particular, the Gleason score (GS) reported in a number of formats were extracted successfully. Our study reports that predominantly older men were diagnosed with PCa reporting a high-risk GS (8-10). Where cell differentiation was reported, 64% of biopsies reported poor differentiation. The approaches demonstrated in our study should be extended to a larger dataset to assess whether it has the potential to scale up to the national level.

Keywords:

Prostate cancer, Gleason score, Risk, Cell differentiation

Introduction

Prostate cancer (PCa) is an important non-communicable disease (NCD) in South Africa with a reported age-standardised incidence rate of (ASIR) of 67.9 per 100 000 in 2012 [1]. Local studies have reported that African men present with advanced and aggressive PCa reducing the opportunity for remission [2]. Local guidelines indicate that patient management is directly linked to Gleason score (GS) grading, percentage and numbers of cores as well as ancillary information such as seminal vesicle involvement and peri-neural invasion [3]. Prostate biopsy results are stored as semi-structured narrative text that cannot be easily analyzed. The GS is captured as follows; (i) major score, e.g. 4 and (ii) minor score, e.g. 3. This equates to a 4+3=7 GS. This data is captured in multiple formats depending on reporting practices at each laboratory. For example, a GS of 3+4=7 could be reported as follows; (i) major 3 and minor 4 = 7, (ii) 3+4=7, (iii) (major pattern 3 + minor pattern 4) = 7, (iv) 3+4=7 and (v) (3+4)=7. The number of cores, GS and cell differentiation provides valuable insights to assess late presentation and poor prognosis [3]. The objective of our study was to investigate the use of big data analytics (text mining) to extract meaningful data from narrative prostate biopsy results.

Methods

The retrospective descriptive study design was used to analyse laboratory data between 2006 and 2016 for men 30 years and older. The sample population of 1000 cases was randomly selected from prostate biopsies with an adenocarcinoma histological finding determined using Systematized Nomenclature of Medicine (SNOMED) morphology codes [4]. Convenience sampling was used. Local guidelines were used to categorise PCa risk using the GS as follows; (i) GS 2-6: low-risk disease (LRPCa), (ii) GS 7: intermediate-risk disease (IRPCa) and (iii) GS 8-10: high-risk disease (HRPCa) [3]. An example of a fictitious biopsy narrative report is provided:

> EPISODE NUMBER: ABC1234 SPECIMEN DETAILS: PROSTATE BIOPSIES. CLINICAL DETAILS: THE PATIENT IS 66-YEAR-OLD MALE WITH AN ENLARGED PROSTATE AND A PSA LEVEL OF 20.8. PROSTATE CORE BIOPSIES WERE SUBMITTED TO EXCLUDE CARCINOMA OF THE PROSTATE. MACROSCOPY: SEVEN CORE BIOPSIES ARE RECEIVED, RANGING IN LENGTH FROM 25MM TO 10MM. PATHOLOGICAL DIAGNOSIS: PROSTATE CORE BIOPSIES, WITH REPRESENTATION OF THE SEMINAL VESICLE: INVASIVE MODERATELY DIFFERENTIATED PROSTATIC ADENOCARCINOMA. GLEASON SCORE = MAJOR 3 + MINOR 4 = 7 THERE IS PERINEURAL INVASION IDENTIFIED. NO LYMPHOVASCULAR SPACE INVASION IS PRESENT IN THE SECTIONS EXAMINED. THE TUMOUR IS PRESENT IN 2 OUT OF 7 CORE BIOPSIES WITH AN APPROXIMATE BULK OF 20%.

Figure 1: Fictitious prostate biopsy narrative report

This study applied text mining techniques and specifically made use of regular expressions to address the complex relationships among the data [5], including structured, unstructured and semi-structured data [6]. Prostate biopsy reports data were extracted and loaded to the Spyder Integrated Development Environment (IDE) which uses the Python programming language [7]. Spyder IDE offers features such as advanced editing, debugging, profiling, data exploration and interactive execution [8-10]. The data was first pre-processed and cleaned using tokenisation and stop word removal. Stemming was used to derive the root word. Using feature generation, the bag of words was developed and n-grams generated. Feature selection was used to develop a vector space by selecting a subset of features from the biopsy reports. Regular expressions representative of the GS target feature such

as “gleason”, “Gleason”, “GLEASON”, “Gleeson”, etc. were used to identify the target feature. Lastly, the required data was then extracted for each biopsy and produced as an output extract that was analysed and visualised. The output variables included; (i) episode number (ABC1234), (ii) age (66), (iii) cell differentiation type (moderate), (iii) number of cores (7), (iv) type on cellular invasion (Perineural) and (v) GS (3+4=7). The ages extracted from the biopsy report were correlated against the manually captured values. The data reported includes; (i) GS, (ii) percentage of biopsies by cell differentiation and (iii) distribution of the number of cores submitted. Age were categorised as follows; (i) <50, (ii) 50-59, (iii) 60-69, (iv) 70+ and (v) “Age not stated”.

Results

The episode number was extracted for all biopsies (100%). The age was extracted for 893 biopsies (89%). An age was captured for 877/1000 (88%) of biopsies compared to 956/1000 (91%) in the LIS. There were 17 biopsies where different ages were recorded (2%); 3/17 likely transcription errors, e.g. 6 recorded instead of 60. The majority of biopsies were requested for men ≥60 years (86%).

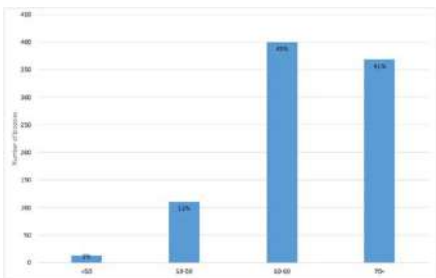


Figure 2: Distribution of the extracted age

Of 1000 biopsies, 286 reported cell differentiation characteristics (29%). Poorly differentiated adenocarcinoma was reported for 64% of biopsies followed by 29% for moderate differentiation.

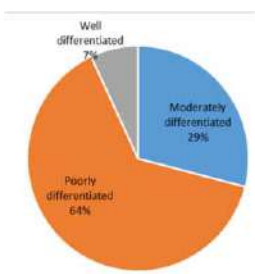


Figure 3: Cell differentiation distribution

The GS was extracted for all 1000 biopsies. A GS of 3+3=6 and 5+4=9 was reported for 18% of biopsies each (total of 26%). This was followed by the 4+3=7 GS (16%). Half of the patients reported a high-risk GS (8-10: 47%).

Conclusion

The use of predictive analytics holds tremendous potential especially in a resource-poor setting where documented clinical data is poor or absent. The prognostic information could be used to categorise patients into risk groups. The accrued laboratory

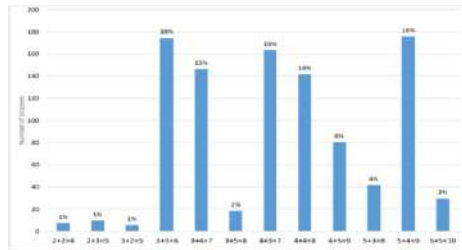


Figure 4: Distribution of Gleason score risk category

data could enable an understanding of PCa differentiation and GS by unlocking clinical information captured in the narrative reports. This could enable real-time cancer data analysis (as opposed to cancer registry reporting that is delayed). The approach reported in our study should be extended to a larger dataset to assess the ability to scale up to the national level.

Acknowledgements

The authors would like to thank the Academic Affairs, Research and Quality Assurance (AARQA) department for providing access to the data.

References

[1] International Agency for Research on Cancer, *Cancer Fact Sheets: Prostate Cancer*, International Agency for Research on Cancer, Geneva, Switzerland, 2012.

[2] H.A. Le Roux, R.J. Urry, B. Sartorius, and C. Aldous, Prostate cancer at a regional hospital in South Africa: we are only seeing the tip of the iceberg, *S Afr J Surg* **53** (2015), 57-62.

[3] D. Anderson *et al.*, *South African Prostate Cancer Guidelines*, South African Urological Association, Pretoria, South Africa, 2017.

[4] U.S. National Library of Medicine, *SNOMED CT International Edition*, U.S. National Library of Medicine, Bethesda, MD, USA, 2016.

[5] X. Wu, X. Zhu, G. Wu, and D. Wei, Data Mining with Big Data, *IEEE Trans Knowl Data Eng* **26** (2014), 97-107.

[6] A. Oussous, F.Z. Benjelloun, A.A. Lahcen, and S. Belfkih, Big Data technologies: a survey, *J King Saud Univ Comp Inf Sci* **30** (2017), 431-448.

[7] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*, O'Reilly, Cambridge, MA, USA, 2009.

[8] V. Korde and C.N. Mahender, Text classification and classifiers: a survey, *Int J Artif Intel Appl* **3** (2012), 85-99.

[9] M. Hofmann and A. Chisholm, *Text Mining and Visualization: Case Studies Using Open-Source Tools*, CRC Press, Boca Raton, FL, USA, 2016.

[10] C.D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, New York, NY, USA, 2008.

Address for correspondence

Naseem Cassim
Room 3B10, Faculty of Health Sciences
University of the Witwatersrand
7 York Road, Parktown, 2193, South Africa
Email: naseem.cassim@wits.ac.za

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190473

Evaluating the Performance of a Terminology Search Engine Using Historical Data

José Castaño[a], Hee Park[a], Pilar Ávila[a], David Pérez[a], Hernán Berinsky[a], Laura Gambarte[a], Carlos Otero[a], Daniel Luna[a]

[a]*Departamento de Informática en Salud, Hospital Italiano, Buenos Aires, Argentina*

Abstract

Clinical terms are noisy descriptions typed by healthcare professionals in Spanish language in the electronic health record system (EHR). Thus, an evaluation of terminology search engine that extends SNOMED CT and an approach that uses historical data of clinical terms is described. We show how to measure precision and recall using historical search data, and we show how the performance of the search engine can be improved significantly using the technology available in the search engine.

Keywords:

Systematized nomenclature of medicine, Search engine, Electronic health records

Introduction

Some electronic health records (EHR) implementations allow users to introduce free text descriptions even in structured data entries. Free text descriptions enable more expressiveness, ease of use and flexibility to physicians. Description terms must be encoded to be mapped to concepts in a controlled vocabulary. The Hospital Italiano de Buenos Aires (HIBA) has a Spanish interface vocabulary [1] where each term is mapped via a direct relation or using compositional post-coordinated expressions to SNOMED CT.

An enumeration of desirable characteristics [2] has been stated for an interface terminology including normalization, completion, spell correction and matching. These are related to string representation and similarity matching problems.

As far as we know, there is no established methodology to evaluate the results of querying a large terminology server. We propose an information retrieval-based methodology to evaluate its performance.

Methods

The HIBA EHR initially allowed healthcare professionals to type free text entries in its problem list. Later, an interface terminology for physicians was developed where they could choose available descriptions for a given problem or a procedure. Some examples of description terms that map to the same concept are the following:

- Neoplasia maligna de pulmón – cáncer
- Neoplasia malignant of lung – cancer of the lung

Compilation of Historical Data

We used the information that has been logged in our system. We analyzed a period of 5 years (2012-17) depicted in Table 1.

Table 1– Data Set Description

Description	Quantity
Number of Registers	2,679,848
Number of Concepts	54,280
Number of Users	4,719
Number of Unique Queries	146,412

An analysis of this data, considering user behavior, shows that most users have recorded less than 500 concept instances (registers) and that 90% of the users recorded less than 1500 registers. Most users are probably medical residents that have been at the hospital for less than 3 years; therefore, there are not many registers per user. This is reasonable also because the HIBA is a university hospital. We could also observe that most users use very few concepts.

Evaluation: Precision and Recall in Information Retrieval

Evaluation is a critical component of information retrieval (IR) [3]. We use precision at *n* (*P@n)*, i.e. at a given cut-off rank [3]. Information retrieval in the medical domain has a strong tradition [4,5], and it has been a difficult domain [6]. In a classical information retrieval model, more than a single document are usually relevant for a given query. This is not the case here: a healthcare professional is looking for a single description term that depicts the problem that has to be registered in the EHR. Therefore, for any number of returned results, there is most of the time only one *relevant* result, as shown in the following examples where the first element in each pair is the query and the second element the selected description:

- (CPS, control de salud)
- (GME, adult health examination)

Evaluation data-set and measures

We collected a number of tuples consisting of (typed term, selected term, corresponding concept). The output of the system to be evaluated is a list of ranked concepts, and each concept is depicted by the corresponding canonical description term.

Given there is only one relevant result chosen, recall at *n* (*R@n*) measures how many times the right concept is returned within the *n* first ranked results. Therefore *R@1* tells us how many times the correct result was the first ranked one and *R@10* tells us how many times the correct result was within the 10 first

ranked results. In this context Precision at *n* (*P@n*) will tell us the same information that *R@n*. We propose here an alternative *P@n* measure (related metrics based on semantic distance have been proposed [7]). We cannot decide, based on historical data, whether a description term (a concept) is *relevant* for a given query, but we do know which term was chosen by the user. We can calculate the semantic distance between two given concepts based on the hierarchical structure of an ontology [8]. We propose 1-Precision distance, considering 1-relevant results also those concepts at a distance ≤ 1 from the relevant concept for a given query.

Results

We built concept, specialty/service area, and user models based on logged data corresponding to 10 years of patient health records, using as features, description terms, concept-id, specialty, user and the corresponding vectors based on the frequency of use.

A large combination of features combinations was tested using Elastic Search (https://www.elastic.co). Tests were performed over of 100,000 records. We used stemming, indexing of common abbreviations, stop words, a spell-checker and automatically detected synonyms. We also made a careful selection of the set of indexed terms, filtering low quality infrequent terms. It was rather surprising that specialty and user information did not improve much the performance. The following Table 2 shows some of the recall results which showed larger differences in the performance for some feature combinations. Best results are in column **n-gram full** (n-gram based search, vs word and word with spell checker) where the largest number of features were used. It can be seen that there is no significant difference with the next column (**n-gram part**) that uses fewer features.

Table 2– Experiments with different features using historical frequencies

#Results	Word	WordSpell	n-gram f	n-gram p
R@1	87.77	88.35	**90.11**	90.16
R@3	92.03	92.18	**94.33**	94.20
R@5	93.19	93.5	**95.35**	95.28
R@10	94.12	94.55	**96,32**	96.24

We also compared the current implementation deployed in the EHR, which we designate as (A), versus the best combination of features (n-gram full) in Table 2), which we designate as (B) in Figure 1. The current implementation uses Lucene off-the-shelf similarity metrics and spell corrector. All the terms at the HIBA thesaurus are indexed. Figure 1 shows that there is a significant improvement in the performance even if the metrics for the current implementation are good, from an information retrieval perspective. They also show that both alternatives soon reach a plateau @5 with little improvement afterward.

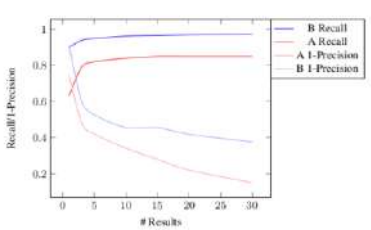


Figure 1–Precision and Recall for (A) Current vs. New (B) Implementation

Discussion

We have presented a methodology to evaluate the performance of a terminology server based on the information retrieval notions of query and relevant concept using historical data.

There is not a single factor contributing to improving the performance, some of them are idiosyncratic to the HIBA interface vocabulary (such as decisions on what description terms need to be indexed). Weight schemas combinations provided by the search engine, use of n-grams, stop-word combinations, use of abbreviations, and other features enabled a considerable improvement in the performance reaching a 90% recall at the first ranked results.

Conclusion

The use of an information retrieval engine customized to noisy user-generated queries was able to deal with string issues stated as desiderata for a terminological server. The use of historical data and associated term weights showed an improvement in the performance of the previous system. However, we were not able to build a proper model for individual users or groups of users (services or specialties). A fine-grained analysis of the results might enable improvements in the performance from a user perspective.

References

[1] Luna D, Lopez G, Otero C, Mauro A, Casanelli CT, de Quirós FGB. Implementation of interinstitutional and transnational remote terminology services. In: *AMIA Annual Symposium Proceedings*. vol. 2010. American Medical Informatics Association; **2010**. p. 482.

[2] Chute CG, Elkin PL, Sherertz DD, Tuttle MS. Desiderata for a clinical terminology server. In: *Proceedings of the AMIA Symposium. American Medical Informatics Association.* (**1999**) p. 42.

[3] Manning CD, Raghavan P, Schütze H. Introduction to Information Retrieval. New York, NY, USA: Cambridge University Press; 2008.

[4] Hersh W, Buckley C, Leone T, Hickam D. OHSUMED: an interactive retrieval evaluation and new large test collection for research. In: SIGIR94. Springer; **1994**. p. 192–201.

[5] Voorhees EM, Hersh WR. Overview of the TREC 2012 Medical Records Track. In: TREC; **2012**.

[6] Koopman B, Zuccon G. Why assessing relevance in medical IR is demanding. In: *Medical Information Retrieval Workshop at SIGIR* 2014. biomedical test collection. Healthcare informatics research. **2012**;18(1):65–73.

[7] Maynard D, Peters W, Li Y. Metrics for evaluation of ontology-based information extraction. In: International world wide web conference. Edinburgh, UK; **2006**. p. 1–8.

[8] Gan M, Dou X, Jiang R. From ontology to semantic similarity: calculation of ontology-based semantic similarity. The Scientific World Journal. **2013**.

Address for correspondence

José M. Castaño
Ingeniería de Software
Hospital Italiano de Buenos Aires
jose.castano_at_symbol_ hospitalitaliano.org.ar.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190474

Ontology-Driven Real World Evidence Extraction from Clinical Narratives

L. Chiudinelli[a,b], M. Gabetta[c], G. Centorrino[c], N. Viani[a,d], C. Tasca[b], A. Zambelli[b], M. Bucalo[c], A. Ghirardi[b], N. Barbarini[c], E. Sfreddo[b], C. Tondini[b], R. Bellazzi[a], L. Sacchi[a]

[a] *Laboratory for Biomedical Informatics "Mario Stefanelli", University of Pavia, Italy*
[b] *Hospital "ASST Papa Giovanni XXIII", Bergamo, Italy*
[c] *BIOMERIS (BIOMEdical Research Informatics Solutions), Pavia, Italy*
[d] *Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK*

Abstract

Unstructured clinical notes contain a huge amount of information. We investigated the possibility of harvesting such information through an NLP-based approach. A manually curated ontology is the only resource required to handle all the steps of the process leading from clinical narrative to a structured data warehouse (i2b2). We have tested our approach at the Papa Giovanni XXIII hospital in Bergamo (Italy) on pathology reports collected since 2008.

Keywords:

Natural Language Processing, Biomedical Ontologies, Data Warehousing

Introduction

i2b2 is a data warehouse supporting clinical and biological research [1]. Its structure contains both data (clinical events such as diagnoses, therapies and laboratory tests, as well as genetic data [2]) and metadata (the terminologies which the data refer to, hierarchically represented). C*oncepts* and *modifiers* are essential elements of the *metadata* structure; the latter are useful to expand and differentiate the meaning of the former. To increase the value of clinical narratives as sources of information, Natural Language Processing (NLP) pipelines are run, which are able to extract structured data from unstructured sources. The extracted information can be saved as i2b2 *observations*. Many existing methods can process English narratives, but the same research in other languages is quite limited; for this purpose, we have designed and developed an ontology-based NLP pipeline [3]. Our ontology follows the OWL (Web Ontology Language) language [4] which allows the definition of elements and relations between them; moreover, its elements are enriched with information on how to perform the NLP tasks and to build the i2b2 taxonomy and data.

Methods

The main goal of the developed system is to maintain a single resource to configure all the processes between a repository of clinical narratives and a data warehouse that enables querying the information extracted from them.

OWL Model

The ontology is the core of the system; it guides the information extraction from the reports, defines the i2b2 *metadata* structure and helps in the storage of the extracted data as i2b2 *observations*. The first step to building a new ontology is defining the entities. The most important Entities are in the class hierarchy. The second step is to create object property entities that link two different classes. Annotation properties are instead entities that can be attached to a class. The two main types of classes in our system are called "Event" and "Attribute".

Workflow

1. Building up the OWL Ontology

The design of an ontology is a complex process in which the needs and knowledge of different stakeholders must be taken into account. An important task is also to organize this knowledge, where possible, to standard terminologies, such as SNOMED-CT, ICD9CM, LOINC and ATC. The ontology development involves many (re)design-and-feedback iterations in order to leverage the outputs of the clinicians without abandoning the standards. Along with a close interaction with clinicians, especially in the first phases of the process, a preliminary automatic analysis of the sample corpora of narratives is useful to recognize usage of some elements and try to change the classes accordingly. A technique that is suitable to be used for this purpose is extracting the most used N-grams in the texts.

2. From OWL to i2b2 Taxonomy

Once the ontology is ready, it can be processed to build the *metadata* taxonomy in i2b2. In most cases events and attributes will be saved respectively as *concepts* and *modifiers*. The taxonomy building process is governed by object properties: *subClassOf* contains the parent class entity and is used to maintain the original hierarchy both for events and attributes; *isModifier* tells which event the attribute will refer to in the form of a *modifier*. The remaining information required to populate the *metadata* tables are retrieved from annotation properties of the classes.

3. From OWL to NLP Extraction

Our NLP pipeline is ontology-based: it uses some of the properties of the ontology elements to extract events and attributes from the text. The first step consists in searching the events in the text. For this first task, annotation properties are used to point out which regular expressions must be used to find each event, or to focus the process to specific parts of the analyzed texts. The second step is to search attributes that refer to the events just found; the *hasModifier* object property is used to identify them. As happens with events, attributes

have annotation properties which drive the extraction of the terms.

4. From NLP Extraction to i2b2-Data

A final ETL (Extract, Transform and Load) task is required to store the NLP output to the i2b2 data warehouse. The new *observations* are aligned with the i2b2 taxonomy created at point 2, in order to be retrieved in the form of *concepts* and *modifiers* of belonging. Each extracted event is stored as an i2b2 *observation*; similarly, the linked attributes will lead to the creation of additional *observations* with the associated *modifiers* as well as with the *concepts* of the event.

Results

The Papa Giovanni XXIII, a general hospital located in Bergamo (Italy), has more than 200,000 pathology reports collected from 2008 (to 2018). The methodology described in the paper has been applied to these documents as a test case. In particular, the reports that have undergone this procedure are pathology reports describing breast cancer cases. The reports are composed mainly of three sections: the list of specimens received, macroscopic description of the specimens and diagnosis conclusions.

OWL Model

Important elements described in the pathology reports have been listed among the events in the ontology: the type of specimen analyzed, the associated diagnosis and other findings. These events have been further specified with child events and linked attributes.

Workflow

1. Building up the OWL Ontology: Breast Cancer Case Scenario

BioPortal PATHLEX [5] was used as a starting point defining various entity and attribute classes. Analyzed anatomic pathology reports were focused on breast cancer, so the specimens and diagnoses are relative to this domain. The final list of possible diagnoses was therefore not very long, discriminating between benign and malign tumor. The specimens group was a broader and more heterogeneous one, including biopsies and various surgical resections. Moreover, N-gram extraction was used to support this phase. As mentioned, the process was conducted based on repeated interactions with clinicians. For this preliminary task 20 reports were used as training examples for the creation of the ontology.

2. From OWL to i2b2 Taxonomy: Preparing i2b2 Metadata for Future Observation

The breast cancer ontology's hierarchical structure and event-attribute links are faithfully represented in the i2b2 taxonomy thanks to the object properties *subClassOf* and *isModifier*. For instance: "/Event/Predictive_Prognostic_Factors/Ki67" *subClassOf* "/Event/Predictive_Prognostic_Factors/" allows the Ki67 leaf to be correctly placed into the i2b2 hierarchy.

3. From OWL to NLP Extraction: Information Extraction Task on Pathology Reports

Our NLP pipeline is the extension of a previous pipeline used in the cardiology domain [3]. One of the key features of the approach is a single resource (i.e. the owl ontology in protege) allowing the governance of different processes interconnected with each other.

For example, the "Ki67" test is found thanks to the associated regular expression that considers different aliases. When the event is found the system looks at the associated attributes (percentages). A validation was performed on a small subset of 34 randomly selected documents to evaluate the NLP extraction performance. After the information extraction the NLP output was exported in a human readable text file, aligned with the original text of the reports; domain experts were trained to manually recognize errors in the extraction. The measured performance demonstrated a good quality (>90%) in precision, recall and F1 score [6].

4. From NLP Extraction to i2b2-Data

After the extraction the last step is to import specimens, diagnoses and other findings into the i2b2 database. This step is governed by the same ontology used in the NLP information extraction task and in the i2b2 taxonomy creation. As a practical rule, all *observations* imported in i2b2 have been associated with the date of their corresponding reports, while the associated visit ids were the ones reported in the documents.

Conclusions

We started the ontology curation steps by integrating local domain knowledge with an existing ontology from BioPortal. Raw reports were analyzed with a Natural Language Processing (NLP) pipeline that extracts structured data from unstructured narrative sources. These steps were in the anatomy pathology ward of the hospital 'Papa Giovanni XXIII' in Bergamo in the context of breast cancer.

References

[1] S.N. Murphy, G. Weber, M. Mendis, V. Gainer, H.C. Chueh, S. Churchill, and I. Kohane, Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2), *J. Am. Med. Informatics Assoc.* **17** (2010).

[2] S.N. Murphy, P. Avillach, R. Bellazzi, L. Phillips, M. Gabetta, A. Eran, M.T. McDuffie, and I.S. Kohane, Combining clinical and genomics queries using i2b2 – Three methods, *PLoS One*. **12** (2017).

[3] N. Viani, C. Larizza, V. Tibollo, C. Napolitano, S.G. Priori, R. Bellazzi, and L. Sacchi, Information extraction from Italian medical reports: An ontology-driven approach, *Int. J. Med. Inform.* **111** (2018).

[4] G. Antoniou, and F. van Harmelen, Web Ontology Language: OWL, Springer Berlin Heidelberg, 2004.

[5] C. Daniel, D. Booker, B. Beckwith, V. Della Mea, M. García-Rojo, L. Havener, M. Kennedy, J. Klossa, A. Laurinavicius, F. Macary, V. Punys, W. Scharber, and T. Schrader, Standards and specifications in pathology: image management, report management and terminology., *Stud. Health Technol. Inform.* **179** (2012).

[6] N. Viani, L. Chiudinelli, C. Tasca, A. Zambelli, M. Bucalo, A. Ghirardi, N. Barbarini, E. Sfreddo, L. Sacchi, C. Tondini, and R. Bellazzi, Automatic Processing of Anatomic Pathology Reports in the Italian Language to Enhance the Reuse of Clinical Data., *Stud. Health Technol. Inform.* **247** (2018).

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190475

Expansion of EHR-Based Common Data Model (CDM)

Wona Choi[a], Soo Jeong Ko[a], Hyuck Jun Jung[a], Tong Min Kim[a], Inyoung Choi[a]

[a] *Department of Medical Informatics, College of Medicine, The Catholic University, Seoul, Republic of Korea,*

Abstract

We expanded and constructed a Common Data Model (CDM) based on hospital EHR to enable analysis and comparison of Adverse Drug Reactions(ADRs) integrated with external organizations with different data structures. This is significant in that it is possible to conduct joint research, analysis, and comparisons among institutions with the same type of CDM constructed, and provide the basis for conducting the same research simultaneously on various data sources.

Keywords:
Electronic Health Records, Observational Study, Database

Introduction

Traditionally, we have relied on the information obtained from clinical trials on drug side effects and post-marketing individual patient case reports to identify potential safety issues of marketed drugs [1; 2]. In fact, Korea has been surveilling drug side effects through the spontaneous ADRs reporting system.

However, in addition to the recognition that the spontaneous reporting system is not the only means that is effective in monitoring drug side effects [3; 4], and there is a growing interest in the monitoring of ADRs using hospital EHR.

Most hospitals in Korea use their own EHR system, and many Korean codes for diagnosis, medication, and treatment are not compatible with international coding systems [5]. Due to the differences in the data structure, format, and terminology used in these individual data sources, it takes a lot of time to analyze data and it is difficult to compare the results of studies using different databases [1].

The purpose of this study is to expand and build a common data model based on EHR to enable systematic analysis of different databases.

Methods

Data Source

The Catholic University of Korea Seoul St. Mary's Hospital is a tertiary university hospital with 1,356 beds and over 10,000 outpatient visitors per day. The hospital has a Clinical Data Warehouse (CDW) that extracted EHR data such as patient visits, examination, diagnosis, prescription and nursing needed for clinical research. We would like to convert the data stored in CDW of Seoul St. Mary's Hospital to CDM format.

K-CDM(Korean Common Data Model)

The Korean Common Data Model (K-CDM) has been defined in 2017 by the Ministry of Food and Drug Safety [6] through the optimization and localization of Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) [7] and Sentinel Common Data Model (Sentinel CDM) [8]. K-CDM includes 8 tables (Person, Visit, Procedure, Condition, Observation, Drug, Measurement, Vital signs).

Code Mapping

The in-hospital codes in the EHR are mapped according to the international standard terminology system.

For the 584 drugs in the hospital, containing the designated 288 components, the component and in-hospital code were mapped at 1:N. The items including combination in the same component name were mapped as RxNorm and ATC by separating the single agent and the complexing agent. All in-hospital diagnosis codes for diagnosis and treatment and the in-hospital codes for 138 designated tests were mapped at 1:N to SNOMED-CT and LOINC, respectively.

Extraction, Transformation, Loading (ETL) Process

The ETL process involves the entire process of taking data from a database system and moving to another database system[5]. We created SQL statements for each table in the K-CDM and extracted, transformed, and loaded the patient, medication, diagnosis(condition), and test information(Measurement) according to the K-CDM format from the CDW(*figure 1*).

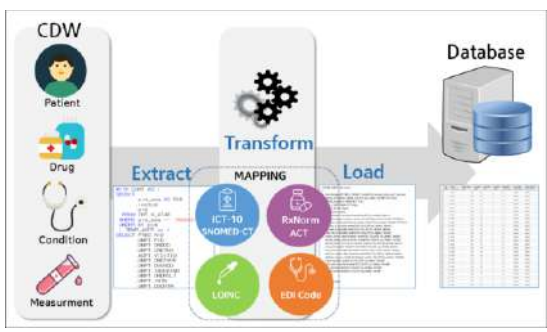


Figure 1 – ETL Process

Database

We used PostgreSQL, an open source object-relational database management system (ORDBMS), which is used as a medical information database to load data [9].

Data Analysis

Allopurinol is a kind of xanthine oxidase inhibitor that is commonly used as a preventive and therapeutic agent in

diseases that uric acid is excessively produced such as gout, uric acid nephrolithiasis, intracellular erythropoiesis, chemotherapy of lymphoma, Lesch-Nyhan Syndrome, chronic renal failure [10; 11].

Based on pervious study [12], data extraction and analysis were performed to confirm the association of allopurinol and TSH levels, which were not reflected in the Korean authorization.

Results

We extracted EHR data of 3,212,915 patients from Seoul St. Mary's Hospital from January 1, 1997 to December 31, 2017, transformed them into K-CDM format and completed the loading. The items used with in-hospital codes and domestic codes were mapped with the international standard terms, and data quality was verified using ATLAS and ACHILLES.

A Patients group (5,148 people) and a control group (20,581 people) were extracted from the established DB.

Table 1 – Number of Data

	Patient Group	Control Group
Number of group	5,148 (100%)	20,581 (100%)
Male	1,845 (35.8%)	7,377 (35.8%)
Female	3,303 (64.2%)	13,204 (64.2%)
Taking allopurinol		
Yes	172 (1.5%)	207 (1.0%)
No	4,976 (96.7%)	20,374 (99.0%)
Age-yr	56.06 ± 14.15	55.88 ± 13.93

Conditional logistic regression model was used to analyze allopurinol and the risk of elevated TSH levels (odds ratio, OR). As a result, the odds ratio was 3.46 times (95% CI: 2.91-4.26).

Discussion & Conclusion

We expanded and constructed the common data model based on hospital EHR to enable analysis and comparison of adverse drug reactions integrated with external organizations without releasing hospital EHR data to outside. This is significant in that it is possible to conduct joint research, analysis, and comparisons among institutions with the same type of CDM constructed, and perform the same research on various data sources simultaneously.

In fact, a multicenter study has been conducted on the subject of our study (risk of elevated thyroid stimulating hormone levels in allopurinol), and we hope to have meaningful results.

Our study has the following limitations. By applying a domestic common data model named K-CDM, part of OMOP CDM format was applied, which is widely used internationally. This can lead to limitations in international research, and supplementation is necessary in the future. In addition, governance for international standard medical terminology should be actively and specifically made at the national level, and many studies using medical data that are being accumulated infinitely in each hospital should be conducted in the future.

Acknowledgements

This research was supported by a grant from Korea Institute of Drug Safety and Risk Management in 2018.

References

[1] S.J. Reisinger, P.B. Ryan, D.J. O'hara, G.E. Powell, J.L. Painter, E.N. Pattishall, and J.A. Morris, Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare databases, *Journal of the American Medical Informatics Association* **17** (2010), 652-662.

[2] Managing the risks from medical product use. Creating a risk management framework, in: *Report to the FDA Commissioner from the Task Force on Risk Management Rockville, MD*, 1999.

[3] R.S. Evans, S.L. Pestotnik, D.C. Classen, S. Bass, and J. Burke, Prevention of adverse drug events through computerized surveillance, in: *Proceedings of the Annual Symposium on Computer Application in Medical Care*, American Medical Informatics Association, 1992, p. 437.

[4] A.K. Jha, G.J. Kuperman, E. Rittenberg, J.M. Teich, and D.W. Bates, Identifying hospital admissions due to adverse drug events using a computer-based monitor, *Pharmacoepidemiology and Drug Safety* **10** (2001), 113-119.

[5] D. Yoon, E.K. Ahn, M.Y. Park, S.Y. Cho, P. Ryan, M.J. Schuemie, D. Shin, H. Park, and R.W.J.H.i.r. Park, Conversion and data quality assessment of electronic health record data at a Korean tertiary teaching hospital to a common data model for distributed network research, **22** (2016), 54-58.

[6] Development of a common data model for adverse drug reactions using hospital EMR, in, Ministry of Food and Drug Safety, Korea, 2017.

[7] https://www.ohdsi.org/data-standardization [accessed on March 22, 2019]

[8] https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model [accessed on March 22, 2019]

[9] https://www.postgresql.org/, postgreSQL [accessed on March 22, 2019]

[10] J.S. Cameron and H.A.J.B.m.j. Simmonds, Use and abuse of allopurinol, **294** (1987), 1504.

[11] S.H.S. Jung Hee Kim, Dong Won Lee, Soo Bong Lee, Ihm Soo Kwak, Case Reports : A Case of Allopurinol Hypersensitivity Syndrome with Esophageal Ulcer Bleeding, *Kidney Res Clin Pract* **25** (2006), 669-673.

[12] F. Perez-Ruiz, S.P. Chinchilla, J. Atxotegi, I. Urionagüena, A.M. Herrero-Beites, and M.A.J.R.i. Aniel-Quiroga, Increase in thyroid stimulating hormone levels in patients with gout treated with inhibitors of xanthine oxidoreductase, **35** (2015), 1857-1861.

Address for correspondence

Inyoung Choi, PhD, Assistant Professor

222, Banpo-daero, Seocho-gu, Seoul, 06591, Republic of Korea

Department of Medical Informatics, College of Medicine, The Catholic University, Seoul, Republic of Korea, iychoi@catholic.ac.kr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190476

SmartCRF: A Prototype to Visualize, Search and Annotate an Electronic Health Record from an i2b2 Clinical Data Warehouse

Sébastien Cossin[ab], Luc Lebrun[ab], Niamkey Aymeric[ab], Fleur Mougin[ab], Mathieu Lambert[c], Gayo Diallo[ab], Frantz Thiessard[ab], Vianney Jouhet[ab]

[a]Bordeaux Hospital University Center, Pôle de santé publique, Service d'information médicale, Unité Informatique et Archivistique Médicales, F-33000 Bordeaux, France
[b]Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, team ERIAS, UMR 1219, F-33000 Bordeaux, France
[c]Bordeaux Hospital University Center, Pôle spécialités médicales, Service de médecine post-urgences

Abstract

Clinical information in electronic health records (EHRs) is mostly unstructured. With the ever-increasing amount of information in patients' EHRs, manual extraction of clinical information for data reuse can be tedious and time-consuming without dedicated tools. In this paper, we present SmartCRF, a prototype to visualize, search and ease the extraction and structuration of information from EHRs stored in an i2b2 data warehouse.

Keywords:

Electronic Health Records, Information Storage and Retrieval, User-Computer Interface

Introduction

A significant part of data daily produced in Electronic Health Records (EHRs) is either in an unstructured (free text) or semi-structured (forms) format [1]. Bordeaux University Hospital is no exception to this observation. Information extraction (IE) for secondary use of clinical data consists in transforming heterogeneous data in an EHR to a structured format of a case report form (CRF). A CRF is a specialized document used in clinical research to collect standardized information about a patient for further statistical analysis [2]. So far, researchers needed to transcribe the data from an EHR to a CRF manually. With the ever-increasing amount of information in patients' EHRs, this task has become tedious and both time and cost-consuming [3]. RAVEL [4], a previous research project carried on in Bordeaux has shown the importance of search engine and data visualization tools to retrieve information in EHRs.

The objective was to develop an interface to speed-up information extraction (IE) task for researchers.

Methods

Bordeaux University Hospital deployed an i2b2 [5] data warehouse for secondary use of medical data. Multiple data sources were integrated such as claims data, lab tests, drug prescription and dispensing, discharges summaries and radiology reports. On the top of i2b2, we've developed two tools for patient centered information retrieval and exploration:

- **i2b2 webclient timeline plugin**: This plugin was developed using D3.js library. The plugin interacts with the i2b2 data warehouse to retrieve data. A simple free text search engine was added in order to find relevant observations.
- **Standalone prototype for data exploration and CRF data capture**. Unstructured data of an electronic health record is first normalized, tokenized and lemmatized with Stanford NLP tools and TreeTagger [6] for French language. Then, noun phrases are extracted with linguistic methods and regular expression [7]. These extracted terms are then indexed in ElasticSearch™ for autocompletion and information retrieval. A web interface was developed with the Shiny package [8] of the R programming language to visualize information in an EHR.

Results

I2b2 webclient timeline plugin

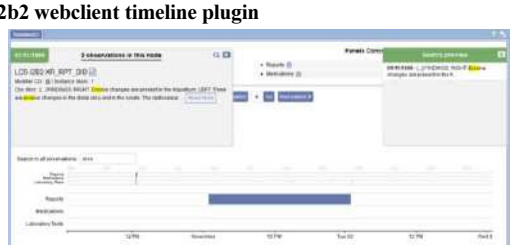


Figure 1 Search in the i2b2 timeline plugin

Figure 1 is a screenshot of the i2b2 timeline plugin. Selected observations were filtered using the search engine with the term "eros". The timeline shows a preview of the records that matched the query. Detail can be obtained by clicking on the preview panel. The term searched is highlighted. The timeline can also manage numerical values. The plugin is available under a GPL-3.0 license[1].

Standalone prototype for data exploration and CRF data capture

An EHR overview is displayed with a timeline (figure 2). The user can zoom in and zoom out the timeline. Each element on the timeline is clickable to show its content. Every element of a same data source has the same background color and different elements within a data source have different colors icons. For example, each ICD-10 category has a specific icon to represent the disease, disorder or symptom.

[1] https://github.com/vianneyJouhet/TimelineD3

Figure 2 - Timeline of a single patient's EHR. Each element is clickable to display its content. Each background color denotes a data source (claims data, lab tests, discharge summaries and medical forms).

The search engine suggests several noun phrases when a user starts typing (figure 3). Autocomplete finds terms present in the EHR and speeds up human-computer interactions.

Figure 3 - The search box. Only the sentence containing the word is displayed in the results and the term being queried is highlighted.

A wordcloud of symptoms / diseases is displayed by default (figure 4). Each term weight corresponds to its Term Frequency-Inverse Document Frequency (TF-IDF) value. A click on a term triggers a search query in the search box. Special focuses on structured data are available: a sunburst displays all ICD-10 (International Classification of Diseases, 10th revision) codes and a scatterplot shows numerical values of selected lab tests. Each sub-element of interest, like a lab test result or an ICD-10 code can be added to the timeline. The idea is to have an adaptive interface where user could decide what he wants to display for a specific IE task. For example, if the user must enquire about a past history of anemia for a set of patients he can choose that hemoglobin test and ICD-10 codes of anemia must appear on the timeline by default and the wordcloud must show terms found in unstructured data related to this disease.

To extract information, an annotation module lets the user select a textual content or a code and stores it in a database.

The prototype is open-source[2] and can be tested online[3]. The prototype was tested by medical users that provided valuable feedback to improve the interface and human-computer interactions. A software version based on this prototype is currently developed to be fully integrated in an i2b2 module.

Our next objective will be to connect the interface directly to REDCap (Research Electronic Data Capture) [9], an electronic CRF application, used in our hospital to store data of clinical studies. The REDcap export module allows the user to export data in specific formats for different analysis software.

[2] https://github.com/scossin/SmartCRF

[3] http://www.smartcrf.fr/

Figure 4 - The wordcloud. The bigger the term, the higher the TF-IDF value of that term. A click on a term triggers a search.

Conclusion

In this article, we present a clinical information extraction prototype to speed-up clinical research and facilitate data reuse in EHRs. Our next objective will be to develop an industrialized version fully integrated in our i2b2 data warehouse and to evaluate its performance in terms of usability and precision/recall for data extraction tasks.

References

[1]P. Raghavan, J.L. Chen, E. Fosler-Lussier, and A.M. Lai, How essential are unstructured clinical narratives and information fusion to clinical trial recruitment?, *AMIA Summits Transl. Sci. Proc.***2014** (2014) 218–223.

[2]S. Bellary, B. Krishnankutty, and M.S. Latha, Basics of case report form designing in clinical research, *Perspect. Clin. Res.***5** (2014) 159–166. doi:10.4103/2229-3485.140555.

[3]Y. Matsumura, A. Hattori, S. Manabe, T. Takeda, D. Takahashi, Y. Yamamoto, T. Murata, and N. Mihara, Interconnection of electronic medical record with clinical data management system by CDISC ODM, *Stud. Health Technol. Inform.***205** (2014) 868–872.

[4]F. Thiessard, F. Mougin, G. Diallo, V. Jouhet, S. Cossin, N. Garcelon, B. Campillo, W. Jouini, J. Grosjean, P. Massari, N. Griffon, M. Dupuch, F. Tayalati, E. Dugas, A. Balvet, N. Grabar, S. Pereira, B. Frandji, S. Darmoni, and M. Cuggia, RAVEL: retrieval and visualization in ELectronic health records, *Stud. Health Technol. Inform.***180** (2012) 194–198.

[5]S.N. Murphy, M.E. Mendis, D.A. Berkowitz, I. Kohane, and H.C. Chueh, Integration of clinical and genetic data in the i2b2 architecture, *AMIA Annu. Symp. Proc. AMIA Symp. AMIA Symp.* (2006) 1040.

[6]H. Schmid, Improvements In Part-of-Speech Tagging With an Application To German, in: Proc. ACL SIGDAT-Workshop, 1995: pp. 47–50.

[7]S.S. Kathait, S. Tiwari, A. Varshney, and A. Sharma, Unsupervised Key-phrase Extraction using Noun Phrases, *Int. J. Comput. Appl.***162** (2017) 1–5.

[8] Shiny, http://shiny.rstudio.com/ (accessed November 25, 2018).

[9]P.A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, and J.G. Conde, Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support, *J. Biomed. Inform.***42** (2009) 377–381. doi:10.1016/j.jbi.2008.08.010.

Address for correspondence

sebastien.cossin@chu-bordeaux.fr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190477

Detecting Child Autism Using Classification Techniques

Md Delowar Hossain, Muhammad Ashad Kabir

School of Computing and Mathematics, Charles Sturt University, NSW, Australia

Abstract

Autism spectrum disorder (ASD) is a brain development disorder that restricts a person's communication abilities and social interaction capabilities from natural growth. In this paper, we have applied various supervised classification techniques to detect the presence of child autism. Our findings show that the Sequential Minimal Optimization (SMO) classifier performs best to detect ASD cases with the highest accuracy and minimum execution time and error rate. We also identify the most dominant features in dectecting child autism.

Keywords:

Autism Spectrum Disorder, Supervised Machine Learning, Child

Introduction

Autism spectrum disorder (ASD) is considered as a neurodevelopment disorder that hinders daily communication and social behavior from natural growth [1]. Data mining classification techniques can be applied to detect ASD cases. The main objectives are to reduce the diagnosis time in order to get quicker access to healthcare services, improve diagnosis accuracy, and finding the highest ranked features of ASD.

There are number of prior studies that have applied classification techniques on adult ASD data. For example, K. Basu *et al* [2] analyzed the Adult Autism screening data set using supervised data mining techniques and showed that Support Vector Machine (SVM) classification technique outperforms other classifiers. Later, Brian McNamara *et al* [3] classified the same dataset by applying decision tree and random forest classifiers. They pre-processed the dataset by removing the records with missing categorical instances and less significant variables, before applying the classifiers. They found that random forest outperforms over decision tree classifier.

Compared to the aforementioned research, in this paper, we apply supervised learning techniques on child ASD data. To the best of our knowledge, our study is the first one that applied supervised learning techniques on ASD dataset of children aged 4 to 11 years. Our goal is to analyze the dataset using existing classification techniques and classify them in one of the two categories: "children having ASD" or "children not having ASD". We compare the performance of various methodologies to determine the best classification technique for this dataset. Our results show that Sequential Minimal Optimization (SMO) classifier performs best amongst all the supervised classifiers. We also analyze the dataset to find out dominant features that cause ASD based on the answers given to the ASD questionnaire.

Methods

In this section, we present our approach for detecting ASD cases using various classification techniques. It includes several steps such as dataset exploration, data preprocessing, and classification.

Data Set Exploration

We used the dataset from UCI Machine Learning Repository [4]. The child dataset contains ten binary features (A1_Score to A10_Score), two numeric features - age, results, and categorical variables such as gender, ethnicity, jaundice status, family member having PDD (Pervasive Developmental Disorder), country of residence of the person who answered the survey, used the screening app before, age description, and ASD class. The dataset contains 292 records and 19 attributes after selecting the screening type.

Data Preprocessing

In order to simplify our analysis, we discarded less significant variables such as used_app_before, country_of_res, and age_desc [2, 3]. By analysing the "results" feature, we found that result score =>7 indicates ASD positive and score <7 indicates non-ASD. So, we excluded this attribute before classification to avoid predefined situation where the output is already known. We removed all of the records containing missing values. We found 43 ethnicity entries missing in the dataset, so we removed those records. We also noticed "age" was missing for one record. So, we replaced this missing value with the median value of age. Finally, our observations on final datasets count are shown in Table 1.

Table 1- Final Child Dataset

Female	Male	ASD_Class
38	85	No (123)
36	90	YES (126)
Total number of cases:		249

Results

We analyzed the child dataset and applied 28 supervised classification techniques of different groups such as Rules, Bayes, Function, Lazy, Meta (Decision Tree used a base algorithm) and Tree. We applied pruning for Tree base algorithm. We used WEKA data mining tool and applied 10-fold cross validation. We found the classifiers from "Function" such as Multilayer Perceptron, Simple Logistic, SMO, classifiers from "Meta" group such as Iterative Classifier Optimizer, LogitBoost, and Real Adaboost, and LMT (classifier from Tree group) result in 100% accuracy, precision, recall, and F-measure (Table 2). We also determined the dominant features for child Autism by analyzing the LogitBoost supervised classifier and performing weight-based analysis. We found the following are the most dominant features in detecting child autism:

- *A4_Score*: S/he finds it easy to go back and forth between different activities
- *A10_Score*: S/he finds it hard to make new friends
- *A8_Score*: When s/he was in preschool, s/he enjoys playing games involving pretending with other children

Here, *A4_Score* is related to child's intellectual disability. *A10_Score and A8_Score* are related to social interaction abilities. Autistic child find it difficult to make new friends and sometimes they enjoy playing games involving pretending with other children. This implies that difficulty with social interaction is a key symptom of ASD. Answering "yes" to these questions are dominant contributors to the model prediction.

Table 2- Classifier Performance Statistics for the child dataset

Classifier	Accuracy	Precision	Recall	F-measure
ZeroR	50.60	?*	0	?*
OneR	78.31	0.82	0.71	0.76
PART	89.95	0.90	0.88	0.89
ByesNet	96.78	0.97	0.95	0.96
Naïve Bayes	97.59	0.98	0.96	0.97
Naïve Byes Updateable	97.59	0.98	0.96	0.97
LibSVM	96.38	0.98	0.94	0.96
Multilayer Perceptron	**100**	**1**	**1**	**1**
Simple Logistic	**100**	**1**	**1**	**1**
SMO	**100**	**1**	**1**	**1**
IBK	91.96	0.972	0.86	0.91
KStar	90.76	0.99	0.82	0.89
LWL	76.70	0.788	0.72	0.75
Bagging	85.94	0.87	0.82	0.85
Classification Via Regression	92.77	0.92	0.93	0.92
Iterative Classifier Optimizer	**100**	**1**	**1**	**1**
LogitBoost	**100**	**1**	**1**	**1**
Multi class Classifier	98.79	0.99	0.98	0.98
Multi class Classifier Updateable	99.19	1	0.98	0.99
Random Committee	91.16	0.89	0.93	0.91
Real Adaboost	**100**	**1**	**1**	**1**
Hoeffding Tree	97.59	0.98	0.96	0.97
J48	90.76	0.91	0.89	0.90
LMT	**100**	**1**	**1**	**1**
NBTree	95.18	0.96	0.93	0.95
Random Forest	96.78	0.97	0.95	0.96
Random Tree	80.72	0.83	0.76	0.79
SysFor	86.34	0.85	0.87	0.86

Note. *no result produced

We observed different types of error during the classification time such as mean absolute error, root mean square error, relative absolute error, and root relative squared error. These errors are the outcome of the difference of our predicted model and observed data. We determined which classifier has the least error or is totally free of errors.

From Table 3, we can see that the best performing classifier is SMO which implements John Platt's sequential minimal optimization algorithm for training a support vector classifier.

Table 3- Comparison of Classifier's Error and Execution Time

	Mean Absolute Error	Root Mean Square Error	Relative Absolute error	Root relative squared error	Execution Time (sec)
Multi-layer Perceptron	0.005	0.026	1.102	5.357	1.22
Simple Logistic	0.061	0.117	12.20	23.50	0.24
SMO	**0**	**0**	**0**	**0**	**0.06**
Iterative Classifier Optimizer	0.098	0.156	19.73	31.22	0.3
LogitBoost	0.098	0.156	19.73	31.22	0.1
Real Adaboost	0.069	0.13	13.90	25.99	0.03
LMT	0.061	0.117	12.20	23.50	0.14

It breaks the problem into a series of smallest possible sub problems, which are then solved analytically. As a result, this classifier has least execution error (0 in this case) and execution time (0.06 second).

Conclusions

In this study, we classified ASD child dataset using supervised classification techniques and found that SMO classifier performs best in terms of accurate prediction of child ASD. Additionally, SMO is also free from classification errors, suggesting that it aligns well with 10-Fold cross validation. By analyzing our experimental results, we also identified the most dominant features that significantly contribute to detecting child autism.

References

[1] P. Bolton, H. Macdonald, A. Pickles, P. Rios, S. Goode, M. Crowson, A. Bailey and M. Rutter, A Case-Control Family History Study of Autism, *The Journal of Child Psychology and psychiatry* 35(5) (1994), 877-900

[2] K. Basu, Machine learning approaches to the classification problem for autism spectrum disorder, *GitHub*, 2018. https://github.com/kbasu2016/Autism-Detection-in-Adults/blob/master/report.pdf (accessed February 23, 2018)

[3] B. McNamara, C. Lora, D. Yang, F. Flores and P. Daly, Machine Learning Classification of Adults with Autism Spectrum Disorder, *RPubs,* 2018. http://rpubs.com/brianmcnamara/ASD_Classification (accessed March 2, 2018)

[4] F. Thabtah, Autistic Spectrum Disorder Screening Data for Children Data Set, https://archive.ics.uci.edu/ml/datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++# (accessed Feb 5, 2018)

Address for correspondence

Md Delowar Hossain,
Email: mailtodelowar@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190478

TBench: A Collaborative Work Platform for Multilingual Terminology Editing and Development

Panpan Deng, Yujing Ji, Liu Shen, Junlian Li, Huiling Ren, Qing Qian, Haixia Sun

Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing, China

Abstract

Terminology facilitates consistent use and semantic integration of heterogeneous, multimodal data within and across domains. This paper presents TBench (Termilology Workbench) for multilingual terminology editing and development within a distributed environment. TBench is a web-service Java tool consisting of two main functionalities that are knowledge construction (i.e.extended model based on ISO25964, batch reusing and constructing multilingual concept hierarchy and relationships) and collaborative control in order to achieve custom extensions, reuse, multilingual alignment, integration and refactoring.

Keywords:

Controlled Vocabulary, Software, Architecture

Introduction

Many terminology, taxonomy and ontology editing tools have been developed, and some are world-famous, such as Protégé [1]. However, many domain experts, even terminology experts, still think that there are no suitable tools for them to build what they want. In addition to economic factors, complexity of operations and cognitive burden may be two important reasons. The former is usually manifested as new hierarchy trees and relationship instances, which can only be created one-by-one. Other lingual terms can only be created as non-preferred terms, are unable to reuse related thesauri and ontologies flexibly, have difficult to extend attributes and relationships for knowledge units, and so on. The latter is mainly manifested as lack of understanding of the meaning of user interface function labels. Motivated by these shortcomings, TBench was designed and developed.

Methods

TBench is based on a classical three-tier architecture (Figure 1), structured through a data layer, a functional layer and an interface layer.

Data Layer Design

Data Description Format Design

According to SKOS and ISO25964 [2], TBench has a customizable data model including core and extended models for all knowledge units (versions, thesaurus, concept group, concept, term, note, etc.) respectively. Each terminology system or top concept tree can define its own extension model on term type, hierarchical and semantic relationships, descriptions of attributes, etc.

Data Storage Format and Data Format Converter Design

All data are stored in an Oracle relational database. To facilitate users conversions of data between different formats (TXT, JSON, XML and RDB) and semantic description models (SKOS, OWL and RDF), data format converters are also designed.

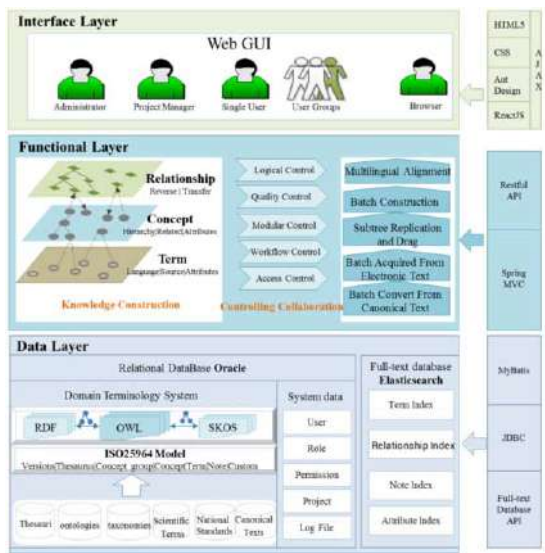


Figure 1- The Architecture of TBench (Termilology Workbench)

Functional Layer Design

The functional layer consists of two main parts: knowledge construction and collaborative control. In addition to basic editing functions, batch construction and multilingual mechanisms are also designed and developed to support cross-language knowledge construction easily and efficiently.

Batch Construction by Flexible Reuse and Operation

Multiple ways are designed for reuse and batch construction, inculding:

1. Using automatic batch conversion techniques, canonical text can be batch converted to hierarchical structure, sibling concepts, non-preferred terms and semantic relationships.

2. Through grammatical rules, concepts and relationships from electronic text are quickly acquired and integrated with a clear workflow.
3. For reuse and integration, fast replication and simple drag operations of subtrees within or between terminology systems are designed and implemented.
4. Some objects having common features can be batch created using 'indication' relationships with the same domain.

Multilingual Mechanism Design

In order to create a terminology system containing multilingual terms, each concept has a preferred term in every language. To provide the culture feature, all objects (concepts, relationships, attributes, etc.) of TBench have a 'language' description. Exact, inexact and partial equivalence mappings among preferred terms of the same concept in different languages can also be created to reveal cultural differences. For instance, the mapping between 'aircraft' and '飞机' is inexact.

Controlling Collaboration Between Users

A role-based access control (RBAC) model was adjusted to support flexible configuration between users, roles, permissions, and task state automatic transitions by using direct permissions setting and the introduction of resource work states. Locking and unlocking were used to resolve knowledge unit conflicts during real-time collaborative editing.

Interface Layer Design

Different Tasks, Different Workflows, and Different User Interfaces

To provide multi-user personalized usability, customizable user interfaces were also designed. Users can choose the function modules, data items in modules, and system labels. For instance, menus, buttons and attributes can be described in natural language depending on the user.

Results

TBench has been developed (Figure 2) in Java™. In addition to the data converter, all the above designs have been implemented. It is being used for cross-institutional constuction of the Chinese Clinical Terminology System (CSCT), consisting of 14 categories of 207,000 concepts organized hierarchically, and 65 semantic relationships.

Conclusions

TBench, a web-service tool, can be used for collaborative building and maintaining of controlled vocabularies, thesauruses, classifications, taxonomies and ontologies. It improves several key points, such as customizable data models and interfaces for different users, batch construction operations for relationships and concept hierarchies, and so on. The reliability of TBench is being tested and verified with CSCT. Crowdsourcing of construction will continue to be studied in the near future.

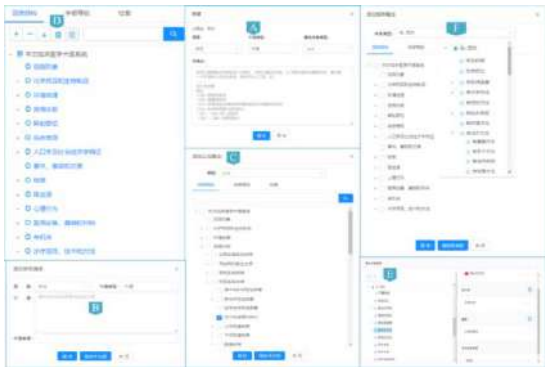


Figure 2 – TBench (Terminology Workbench) user interface. (A) Concept creation interface. It can automatically batch convert from canonical text and configuration language, term type, and relationship type. (B) Non-preferred terms batch creation interface. (C) Upper concept configuration interface. Selecting multiple upper concepts is an effective way to establish poly-hierarchical relationships. (D) Tree structure interface enables users to create, delete, copy, cut, paste and drag a concept or subtrees. (E) Relationship configuration interface enables users to customize various new relationships with their domain and range. (F) Relationship instances with the same domain can be batch created by choosing multiple concepts as a range.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (grant #2016YFC0901901), Medical and Health Technology Innovation Program of The Chinese Academy of Medical Sciences (grant #2017-I2M-3-014).

References

[1] Musen MA. The Protégé Project: A Look Back and a Look Forward. *AI Matters*. 2015, 1(4):4–12.

[2] International Standards Organization (ISO). ISO 25964 - 1:2011 Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 1: Thesauri for information retrieval, 2011.

[3] Abburu S, Babu G S. Survey on Ontology Construction Tools[J]. *International Journal of Scientific & Engineering Research, IV* (2013), 1748-1752

Address for Correspondence

Corresponding author : Haixia Sun
Email: sun.haixia@imicams.ac.cn

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190479

AdhereR: An Open Science Approach to Estimating Adherence to Medications Using Electronic Healthcare Databases

Alexandra Dima[a], Samuel Allemann[a,b], Dan Dediu[c]

[a] *EA 7425 Health Services and Performance Research, Université Claude Bernard Lyon 1, Lyon, France,*
[b] *Pharmaceutical Care Research Group, University of Basel, Basel, Switzerland,*
[c] *Laboratoire Dynamique du Langage, Université Lumière Lyon 2, Lyon, France*

Abstract

Adherence to medications is a key performance indicator and behavioral outcome in healthcare. Electronic healthcare databases represent rich data sources for estimating adherence in both research and practice. To build a solid evidence base for adherence management across clinical settings, it is necessary to standardize adherence estimation and facilitate its appropriate use. We present the recent development and oportunities offered by AdhereR, an R package for visualisation of medication histories and computation of adherence.

Keywords:

Medication Adherence; Medical Records; Reproducibility of Results

Introduction

Taking medication as prescribed (medication adherence) is essential for treatment effectiveness. Current research estimates total annual costs from suboptimal medication adherence of up to 50000 USD per patient [1]. Effective adherence management in routine care requires appropriate measurement and interpretation of medication use in real-life settings. For long-term treatments, medication use involves multiple prescription and dispensing events, recorded in administrative or clinical databases. These data have been successfully used to detect delays or interruptions in medication acquisition, suggesting lower use or interruption of treatment. However, there have been substantial variations in algorithms to estimate adherence, and minor differences in algorithms have been shown to impact results [2]. This raises concerns over the validity of adherence evidence and highlights the need to standardize analysis and encourage reproducible science. We therefore developed functions to facilitate reproducible adherence calculations in the statistical environment R [3].

Methods

We reviewed current guidelines for adherence definition and methodological literature on calculating adherence from electronic healthcare data (EHD). We developed several flexible functions in a new R package (AdhereR) and training materials to facilitate all stages of data analysis, from data preparation to visualisation and calculation of estimates. The 3-component consensus taxonomy for medication adherence [4] was used to build separate functions for initiation, persistence, and implementation.

Results

Currently, AdhereR is released as an R package (library). Its core functionality is written in pure R and is optimised for various use scenarios, being able to effectively scale up from the analysis and displaying of a few patients on a consumer-grade laptop to the batch processing of millions of records on parallel heterogeneous compute clusters. It can use data stored in various formats, from "flat files" in CSV, Excel, SPSS, Stata or SAS files to large relational databases using SQL or other architectures such as Hadoop.

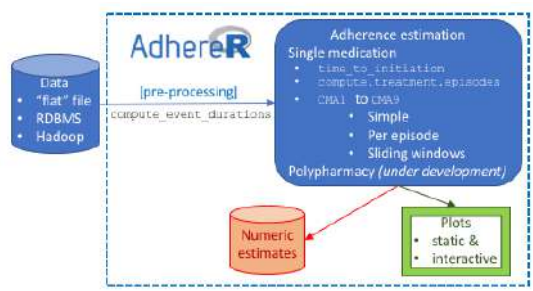


Figure 1– Schema of the AdhereR Package Functions, Input and Output.

Didactically, we can distinguish three main classes of functionalities implemented by AdhereR (Figure 1). First (and optionally), prescription, dispensation and hospitalisation data can be pre-processed to extract the type of information used by the subsequent steps, namely the patient unique ID, the date and the duration of each event, and possibly its medication class and dosage (compute.event.durations). Second, these data can be used for the estimation of initiation (time_to_initiation), (non-)persistence (compute.treatment.episodes), and various types of implementation estimates: simple, per-episode, and sliding-window Continuous Medication Availability (CMA) functions. Among the "simple" numeric CMAs, eight (CMA1–CMA8) are implemented from the literature and one (CMA9) original; these differ in the assumptions they make about the underlying processes, possibly being appropriate for different scenarios and sometimes resulting in drastically different estimates. Moreover, we have developed two "temporal" CMAs that produce a time series of CMA estimates, one based on the idea that there are separate treatment episodes (defined in various ways), the other defining regularly spaced (and possibly overlapping) sliding windows of fixed duration. These

functions are currently available for single medications, and equivalent functions for polypharmacy are under development.

Finally, AdhereR can produce publication-quality static plots of the data (and possibly of adherence estimates) or interactive graphs which react in real-time to the user's input (currently using Shiny, HTML5, CSS and JavaScript, which can be run in a standard web browser). These allow data exploration and may be used to examine the influence of analysis assumptions on the adherence estimates for individual medication history, thus guiding the expert validation of parameter choice for adherence calculations. Figure 2 shows an example plot for one of the implementation functions (CMA9) calculated over an observation window of 6 months within a follow-up window of 2 years. In the interactive plotting function (plot_interactive_cma), parameters can be modified to select various time windows and CMA functions and observe their effect on the CMA value. Further information is available in the extensive help and vignettes included in the package, as well as online at www.adherer.eu.

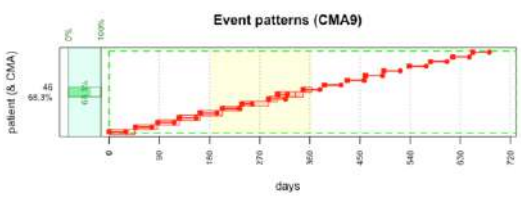


Figure 2– Visualisation of a Patient's Medication Event Patterns (Red Bars) across 730 Days (x-axis), with CMA9 over 182 Days (Yellow Area) as 68.3%. Horizontal lines represent the calculated period covered by the supply prescribed or dispensed during an event.

It is important to note that such numeric estimates and plots of adherence represent only part of a broader, integrated approach to addressing the root causes of non-adherence, which must include specific training of the healthcare professionals in interpreting these indicators, in interaction with the patients. Issues of data quality need to be acknowledged, and the interpretation of the results must provision for various sources of error and missing data. Moreover, we must stress that there is no single 'best measure' of adherence that fits every clinical setting, research question, or individual patient. Therefore, the selection of adherence measures and plotting must consider the specificities of each particular case.

Conclusions

AdhereR represents a proof-of-concept for reproducible and transparent research on EHD. The package is under intense development on three mains axes: first, we are developing a set of functions that compute the adherence to polypharmacy, which represents an important and complex set of use scenarios; second, we are exploring advanced longitudinal analyses that better reflect the temporal dynamics of adherence; finally, we are improving the technology powering interactive plotting, aiming towards better performance on thin clients with a reduced need for costly server-side processing.

Acknowledgements

Authors declare no conflict of interest. AdhereR has been supported by the European Commission (FP7 n°282593; MSCA-IF n°706028), the Swiss Science Foundation (P2BSP3_178648), and IDEXLYON (ANR-16-IDEX-0005).

References

[1] R.L. Cutler, F. Fernandez-Llimos, M. Frommer, C. Ben rimoj, and V. Garcia-Cardenas, Economic impact of medication non-adherence by disease groups: a systematic review, BMJ Open 8 (2018), e016982.

[2] W.M. Vollmer, M. Xu, A. Feldstein, D. Smith, A. Waterbury, and C. Rand, Comparison of pharmacy-based measures of medication adherence, *BMC Health Services Research* **12** (2012), 155.

[3] A.L. Dima and D. Dediu, Computation of adherence to medication and visualization of medication histories in R with AdhereR: Towards transparent and reproducible use of electronic healthcare data, *PloS One* **12** (2017), e0174426.

[4] B. Vrijens, S. De Geest, D.A. Hughes, K. Przemyslaw, J. Demonceau, T. Ruppar, F. Dobbels, E. Fargher, V. Morrison, and P. Lewek, A new taxonomy for describing and defining adherence to medications, *British Journal of Clinical Pharmacology* **73** (2012), 691-705.

Address for correspondence

Alexandra Dima, alexandra.dima@univ-lyon1.fr.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190480

Machine Learning Methods to Predict Lung Cancer Survival Using the Veterans Affairs Research Precision Oncology Data Commons

Nhan V Do[a,b], Jaime C Ramos[a], Nathanael R Fillmore[a], Robert L Grossman[c], Michael Fitzsimons[c], Danne C Elbers[a], Frank Meng[a,b], Brett R Johnson[a], Samuel Ajjarapu[a], Corri L DeDomenico[a], Karen E Pierce-Murray[a], Robert B Hall[a], Andrew F Do[a], Kelly Gaynor[a], Peter L Elkin[d], Mary T. Brophy[a,b]

[a]*MAVERIC, Department of Veterans Affairs Medical Center, Boston, MA, USA*
[b]*Boston University School of Medicine, Boston, MA, UA*
[c]*University of Chicago, Chicago, IL, USA*
[d]*University at Buffalo, Buffalo, NY, USA*

Abstract

We completed a pilot study to guide the development of the VA Research Precision Oncology Data Commons infrastructure as a collaboration platform with the greater research community. Our results using a small subset of patients from the VA's Precision Oncology Program demonstrate the feasibility of our data sharing platform to build predictive models for lung cancer survival using machine learning, as well as highlight the potential of target genome sequencing data.

Keywords:

Machine Learning, Lung Neoplasms, Precision Medicine

Introduction

The Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC) launched the Research for the Precision Oncology Program (RePOP) and the Precision Oncology Data Repository (PODR) in 2016 to acquire the knowledge necessary to improve care for Veterans with cancer. In the fall of 2017, MAVERIC entered a partnership with the University of Chicago to establish a VA Research Precision Oncology Data Commons (RePODC) to provide greater access to both VA data and robust computing resources to the research community external to the VA. We conducted a pilot study to guide the development of the RePODC infrastructure for data sharing and to demonstrate the feasibility of building predictive models with both clinical and genomic data. .

Methods

The exploratory dataset contained 60 instances, each of them a patient record containing a preselection of 153 features based on targeted genome sequencing, an extract from the VA's comprehensive clinical information on each patient, as well as basic demographic data. Following previous work by Lynch et al [1], a small number of candidate features were selected based on their potential relationship with survival time, including both continuous and categorical variables, for a total of 16 attributes. Among the selected features we included the presence of tumor variants for CDKN2A, KRAS and TP53, the number of primary tumors, tumor stage and location, histological findings, exposure to radiation protocols, and the sequence of radiation with surgery. Logistic regression and random forest algorithms were used to predict survival status at 24 months after diagnosis. Ten-fold cross validation was used to assess accuracy.

Results

Cross-validated accuracy was 0.886 for for both the logistic regression and random forest classifiers. In order of importance, the random forest algorithm identified primary tumor location, age, surgery primary site and the variant CDKN2A as the most relevant features to predict survival stage, with higher scores than other features of known importance like stage or histology.

Conclusions

Our results, using a small subset of patients from the VA Research in Precision Oncology Data Commons, demonstrate the feasibility of using this data to build predictive models for lung cancer survival using machine learning methods, and particularly highlight the potential of targeted tumor sequencing data for building predictive models of survival. This work suggests that by expanding access to VA data, we enable use of modern machine learning methods and advanced analytics to contribute to scientific knowledge and improve Veterans' healthcare.

Acknowledgements

This material is based upon work supported by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, Cooperative Studies Program. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.

References

[1] C.M. Lynch, B. Abdollahi, J. D. Fuqua, A. R. de Carlo, J. A. Bartholomai, R. N. Balgemann, V. H. van Berkel, H. B. Frieboes, Prediction of lung cancer patient survival via supervised machine learning classification techniques, *International Journal of Medical Informatics* **108** (2017), 1-8.

Address for correspondence

Nhan Do, MD
150 S. Huntington Ave
Boston, MA 02130; nhan.do@va.gov

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190481

Semantic Data Integration Service for eHealth Applications

Arthur Domingues, A. Sousa Neto, A. Freitas, S.D. Silva, A. Félix, F.M. Mendes Neto

Abstract

Technologies for health have been receiving considerable attention with the popularization of devices for internet access. The Internet can be seen as a repository of knowledge due to its large amount of available information; however, on the other hand, in the midst of this vast amount of content, there is information either scientifically inaccurate or incomplete. This work presents a semantic integration service to provide information of diabetes from medical databases to eHealth applications.

Keywords:
Content; Information Retrieval; Search Engine.

Introduction

The term health education, according to the World Health Organization [1], refers to a set of practices that contribute to the development of autonomy of the population regarding the maintenance of their health. Technology applied in health education narrows the relationship between health professionals and patients, in addition to renewing the understanding of health practices [2]. Besides, with the growing expansion of the Internet and information available on it, it is essential that the individual seek this knowledge to become an active member in the management of their health, thus sharing a responsibility that was previously exclusive of the health professionals.

A survey conducted by the Britsh United Provident Association (BUPA) in 2011, illustrated that 86% of Brazilians seek health content on the internet, but only 25% confers the source. The research illustrates that the absence of Internet content quality standards can pose a great risk in the health area, both among health professionals and among patients[3].

In this regard, systems that integrate and make available specialized database information appear as solutions to the problem of the lack of credibility of data sources. Data integration combines information residing in different sources and provide the user with a unified view of that data [4].

It is in this scenario of technology used to health information retrieval that the present work is grounded and developed. The work presents a service to integrate information of diabetes from medical databases to eHealth applications. The purpose of the service is to enable eHealth systems to provide reliable content to the reality of their users, seeking a better and greater knowledge of health and, consequently, improvements in the quality of life.

Methods

The integration service was developed in PHP language. The data captured by the service is stored in a MySQL database. The architecture of the proposed integration service consists of three components: Search Interface, Crawler and Indexer.

Search Interface works like as mediators between the requests and indexed documents [5]. The search interface is the communication channel between eHealth applications and the integration service. This component has the function of receiving the search parameters such as quantity and content types, and also returning the corresponding contents.

Crawlers are generally deployed to retrieve documents for search services [6]. Crawler is responsible for periodically visiting the databases automatically, in order to collect content. Crawler captures the XML file for content, performs tag read, and stores the information in service database. The contens are provides by medical specialists and health professionals databases. Contents are described by metadata and made available in the form of scientific articles, videos and electronic records on health and human well-being. The bases chosen were: Brazilian Virtual Health Library, Medline and Webmd. Contents collected by the crawler are sent to the indexer.

In search and retrieval services, indexers works like as collections of data sources that use some technique to rank information [7]. Indexer component store lists of the contents of the databases. After crawler finds contents, indexer use an ontology to store and rank the retrieved contents.

The ontology was described in OWL language through the PROTEGÉ tool. Ontology has integrated into the indexer, by OWL-API. Through the API was possible to indexer create and manipulate ontologies. The terms and concepts of this ontology are widespread by the Brazilian Society of Diabetes and American Diabetes Association. The ontology is formalized by terms referring to diabetes, such as: causes, symptoms and treatments. The ontology accomplishes the classification of the meta data, and returns to the indexer an index of semantic relation of the contents with the diabetes. The indexer returns to the search interface a list with the classified contents. In Figure 1, the architecture is represented graphically.

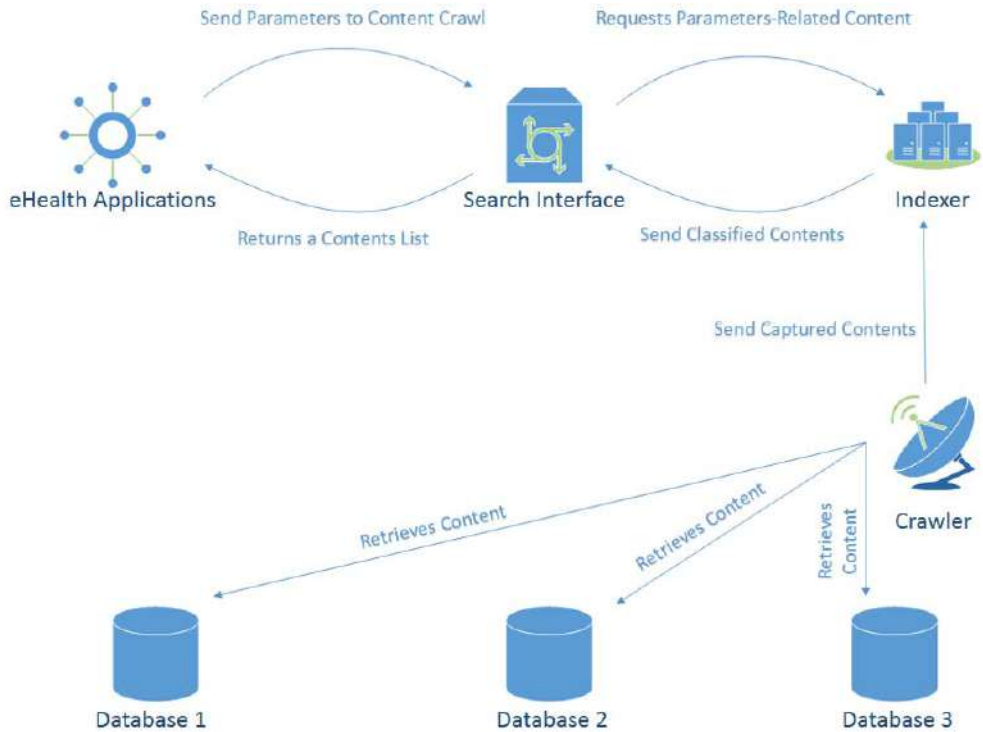


Figure 1. Architecture of Integration System

Conclusions

This work presented an integrated service with information from health databases. The purpose of the integration service is to ensure that eHealth systems provide its users correct and scientifically relevant content in the context of diabetes.

The service works as a middleware between the eHealth systems and the health databases. The service components perform all the process of reading the request of the applications, verification of corresponding information and semantic relation of the results with the parameters.

In order to evaluate the results obtained, the service will be integrated to MobiLEHealth. The MobiLEHealth is as a recommendation and informal learning system in the context of Health 2.0 that performs the monitoring of web content and social media accessed or generated by users, through their interactions [8].

References

[1] World Health Organization. World report on ageing and health. World Health Organization, 2015.

[2] Santos, Camila Tahis. "TECNOLOGIAS VOLTADAS PARA EDUCAÇÃO EM SAÚDE: O QUE TEMOS PARA A SAÚDE DOS IDOSOS?." Anais do Seminário Tecnologias Aplicadas a Educação e Saúde 1.1 (2015).

[3] Britsh United Provident Association (2011). BUPA Health Pulse. Online Health: Untangling the Web. London: BUPA.

[4] A. Y. Halevy. Answering queries using views: A survey. Very Large Database J., 10(4):270–294, 2001

[5] Tao, C., Song, D., Sharma, D., & Chute, C. G. (2013). Semantator: semantic annotator for converting biomedical text to linked data. Journal of biomedical informatics, 46(5), 882-893.

[6] Dong, H., Hussain, F. K., & Chang, E. (2010, November). A framework enabling semantic search in health service ecosystems. In Semantics Knowledge and Grid (SKG), 2010 Sixth International Conference on (pp. 235-242). IEEE.

[7] Mancini, F., Sousa, F. S., Teixeira, F. O., Falcão, A. E. J., Hummel, A. D., da Costa, T. M., ... & Pisa, I. T. (2011). Use of Medical Subject Headings (MeSH) in Portuguese for categorizing web-based healthcare content. Journal of biomedical informatics, 44(2), 299-309.

[8] Mendes Neto, FM, da Costa, AAL, Shadow, EL, Moreira, JDC, Valentine Medeiros, RA, Zapater, JJS, ... & Flowers, CD (2014). Content's Personalized Recommendation for Implementing Ubiquitous Learning in Health 2.0. IEEE Transactions Latin America, 12 (8), 1515-1522.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190482

Identifying Patients with Significant Problems Related to Social Determinants of Health with Natural Language Processing

David Dorr[a], Cosmin A. Bejan[b], Christie Pizzimenti[a], Sumeet Singh[a], Matt Storer[a], Ana Quinones[a]

[a] *Oregon Health & Science University, Portland, OR, USA;*
[b] *Vanderbilt University, Nashville, TN, USA*

Abstract

Social and behavioral factors influence health but are infrequently recorded in electronic health records (EHRs). Here, we demonstrate that psychosocial vital signs can be extracted from EHR data. We processed structured and unstructured EHR data using expert-driven queries and Natural Language Processing (NLP), validating results through structured annotation. We found that although these vital signs are present in EHRs, with 681 structured entries identified for psychosocial concepts, NLP identified a nearly 90-fold increase in patients.

Keywords:

Social Determinants of Health, Natural Language Processing

Introduction

Social determinants of health (SDH) – the social circumstances that shape health risks and outcomes, such as behaviors, social connections, and housing – comprise a large percentage of overall health and well-being across the lifespan. In the United States, medical professionals are increasingly being asked to consider pertinent social determinants in their care, and to act on them when appropriate. However, psychosocial assessments are infrequent, incomplete, and are often not captured as discrete data elements within the electronic health record (EHR).

One strategy to identify patients with SDH needs is to extract these concepts from narrative text. In prior studies, we found that primary care practices caring for persons with multimorbidity both recorded social and behavioral constructs as narrative text [1,2]. Here, we investigate how best to detect key psychosocial variables in a broad dataset by adapting a data-drive process developed by Bejan et al [3].

Methods

Overview

The phenotypic profiles for four key psychosocial vital signs were extracted from EHR data using queries and notes using lexical associations expanded by expert input. Then, for each psychosocial vital sign, we manually reviewed the retrieved charts.

Data Set & Study Design

We extracted and standardized into a common data model EHR data from 2 hospitals, 30 primary care practices and 90 specialty clinics in Oregon representing 3.28M patients with a mix of ages and insurance types. From these individuals, 22% had multimorbidity defined by two of 21 chronic illnesses common in risk scores. From this set, we extracted a subset from 2017 data; these 358k patients had 89.2M notes during 2017.

We selected 4 psychosocial vital signs in this study: chronic stress, social isolation, financial insecurity; and one replication – housing insecurity or homelessness – from previous work by Dr. Bejan [3]. We used lexical association measurements and expert input to identify seeds, or word expressions, that were likely to be used in narrative notes to represent the 4 psychosocial vital signs; then, we ran an information retrieval system over the entire dataset using the identified seeds as query input, validated the retrieved results, and modified the seeds in an iterative process.

Query expansion based on lexical association

Starting with all the notes from the data extract, the method first selected all the documents with at least 1 seed word. Next, the content of these documents was tokenized, the tokens were converted to lowercase, and the low frequency tokens and punctuation marks were discarded. After this preprocessing step, each word w occurring in the context of a seed was ranked according to the following formula:

$$score(w) = \log \frac{LAM(w,S)}{(1+f(w))^{\lambda}}$$

Here, $f(w)$ represents the frequency count of w, λ is a parameter that controls the degree to which the inverse word frequency of w should weight in the relevance score, and LAM denotes a metric of lexical association between w and the seed words in S. The query expansion based on lexical association for extracting the final query terms was performed on the clinical notes from the Vanderbilt EHR [3]. The lexical association measures used in our experiments include the χ^2 test.

To assess the relevance of patients to a psychosocial vital sign, we calculated the similarity of the corresponding patient notes to the psychosocial vital sign query by the standard term frequency–inverse document frequency (TF-IDF) weighted cosine metric such that the most relevant patients to the query were ranked on top of the retrieved patient list. Notably, each patient in our framework was represented as a meta-document that included all the patient notes. To weight the query term i in the meta-document of patient j, we computed the TF-IDF weighting scheme as follows:

$$w_{i,j} = tf_{i,j} \cdot \log \frac{N}{df_i}$$

where $tf_{i,j}$ is the number of occurrences of the term i in the meta-document of patient j, df_i is the number of patients whose corresponding meta-documents contain the term i, and N is the total number of patients in the data extract.

Annotation

Three assessors who were either faculty or research assistants collaborated to design an annotation process that assessed the accuracy of the results. For each concept, a specific set of instructions and a data collection template were created. These

were tested with 5-10 retrieved patients, then revised. For the formal collection, 30 retrieved patients (out of the 600 most relevant) were selected by one annotator, with a third each coming from highest, medium, and lowest relevance. A Likert scale of 0-5 was used for confidence of whether the concept was present; a score of 3 or greater was deemed a true positive. These were randomized and another annotator completed the manual chart review. For a subset of 25%, two independent annotations were done and compared, with disagreements resolved by consensus.

Query reformulation based on relevance feedback

Query reformulation was completed based on the initial annotations. We adopted this approach to improve each search by iteratively refining the 4 psychosocial vital sign queries. For this, the assessors reviewed the chosen notes to manually detect recurrent keywords. The most prevalent and relevant keywords from each run were selected for each reformulation. During this process, the selection of terms to be checked for negation was also performed. Once this was completed, the annotators completed the extraction again, repeating the same process. Reformulation was considered complete after 4 cycles or when precision was greater than .90.

Analysis

We describe the general demographics of included patients. We provide average and 95% confidence intervals for relevance and inverse patient frequency (IPF). We used the manual chart review as true positives (TP) and calculated precision based on TP/all retrieved meta-documents.

Results

In all, there were 358,244 patients with 86 million notes. Their average age was 43.0 ± 25.5 and 55.1% were female. Overall, 82.6% were Caucasian, 3.9% were Asian, and 3.4% Multiracial. They had 2.3 outpatient encounters on average. Overall, there were 358,244 patients with 89.2 million notes. This approach matched between 375 and 54,950 of patients across the concepts. The IPF range was wide, from 1.9 – 7.5 in chronic stress to 6.5-12.7 in financial insecurity. Relevance scores for the first 600 patients ranged from 12.3 to 267.3. Initial precision varied; homelessness was replicated on the first evaluation, with a patient level precision of .93. Other concepts ranged from .60-.68. For the iterations, each improved significant in IPF or relevance. Precision, our primary outcome, achieved levels of > .90 in all except social isolation. Our preliminary error analysis of social isolation identification indicated that most of the false positives were generated due to occurrences of negated social isolation expressions in patient notes. For that concept, once negation was resolved, it achieved target precision. Average scores improved as well, with initial seeds ranging from 2.75-2.9/5 and final seeds from 3.8 to 4.9/5. Higher relevance scores were more associated with matches, indicating reasonable calibration.

Conclusions

We were successful in implementing an iterative query approach using lexical associations, achieving our goal precision of > .90. NLP produced a nearly 90-fold increase in patient retrieval relative to querying structured data elements. By querying across a dataset for rarer conditions, we may more quickly allocate resources to needs. This method demonstrates a significant improvement in detection of persons who may be eligible for care coordination or social services to address these concerns.

References

[1] A.R. Quinones, S. Markwardt, S. Thielke, O. Rostant, E. Vasquez, and A. Botoseneanu, Prospective Disability in Different Combinations of Somatic and Mental Multimorbidity, *Journals of Gerontology. Series A: Biological Sciences and Medical Sciences* 73 (2018), 204-210.

[2] A.R. Quinones, K. Ramsey, J.T. Newsom, and D.A. Dorr, Racial and ethnic differences in clinical outcome trajectories for care managed patients, *Medical Care* 52 (2014), 998-1005.

[3] C.A. Bejan, J. Angiolillo, D. Conway, R. Nash, J.K. Shirey-Rice, L. Lipworth, R.M. Cronin, J. Pulley, S. Kripalani, S. Barkin, K.B. Johnson, and J.C. Denny, Mining 100 million notes to find homelessness and adverse childhood experiences: 2 case studies of rare and severe social determinants of health in electronic health records, *Journal of the American Medical Informatics Association* 25 (2018), 61-71

Address for correspondence

Christie Pizzimenti

email: pizzimen@ohsu.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190483

Automated Reports for Monitoring and Improving Data Quality in a Translational Research Network

Esther E Schmidt[a,b], Corinna Eichelser[a,b,c,d], Bernd Ahlborn[a,b,e], Dietmar Keune[f], Esmeralda Castaños-Vélez[g], David Juárez[a,b], Martin Lablans[a,b]

[a] *German Cancer Consortium (DKTK), Heidelberg, Germany,*
[b] *German Cancer Research Center (DKFZ), Heidelberg, Germany,*
[c] *Department of Internal Medicine III, Klinikum der Universität München (KUM), Germany,*
[d] *Department of Medicine III, Klinikum rechts der Isar (MRI), TUM, Germany,*
[e] *Institute for Medical Information Sciences, Biometry and Epidemiology (IBE), Ludwig-Maximilians-Universität München (LMU), München, Germany,*
[f] *Clinical Cancer Registry, Charité Comprehensive Cancer Center, Berlin, Germany,*
[g] *Experimental Neurology, Charité – Universitätsmedizin Berlin, Berlin, Germany*

Abstract

Standardised, automated quality reports were generated at three pilot locations of the decentralized translational research network DKTK with separated local data warehouses (LDW), for assessing syntactic conformity against common data element definitions deposited in a central metadata repository (MDR). Deviations in the LDW were categorised, and locally corrected. Comparisons of reports from two time points confirm a major improvement in data quality in terms of syntactic conformity, an essential prerequisite for network-wide data integration.

Keywords:

Common Data Elements, Metadata, Data Accuracy

Introduction

The German Consortium for Translational Cancer Research (DKTK) [1] is a network of more than 20 institutions at nine locations, pursuing the aim of efficiently transferring results from cancer research into the clinical practice. The DKTK implements a federated data storage concept, interconnecting LDW with so-called bridgeheads [2]. In order to allow comparisons between the partners' data on clinical cases and biomaterial, the network has agreed on a common data set (MDS) with defined data elements and their permitted values, which has been deposited in a central metadata repository (MDR) [3,4]. An essential prerequisite for the success of data integration processes within decentral research networks is high data quality [5,6]. For the assessment of data quality (here: syntactic conformity [7] with the MDR) we implemented a quality report generator (QR-generator), which can be installed at every network location, in order to perform an alignment of data in the LDW against the centrally defined data definitions in the MDR [8,9].

Methods

The QR-generator installed at a given network location produces quality reports in Excel format as described in [8,9]. Briefly, this report lists all values recorded for each data element in the LDW, highlighting each value that is not concordant with the respective definition specified in the MDR. In addition, the report provides statistical calculations for each data element with regard to completeness and syntactical correctness. Upon production of a report, all errors are manually categorised, compared with network partners' errors, and suggestions are given for solutions to be implemented by the site. Then a second report is produced and compared with the previous one, in order to assess changes in data quality. Though the reports provide detailed mismatch information on a per-patient basis to the local sites, the results described here are summarized on a per-data-element basis for data protection reasons.

Results

QR-generators were installed at each of the nine DKTK locations. In a pilot study, sets of quality reports have been produced at three network locations, allowing the sequential assessment of data quality over a period of 8-10 months.

At the Charité in Berlin, the LDW in the bridgehead serves as the primary tumour documentation system. Quality reports have been produced in March and December 2017 (104.000 and 137.000 patients in the LDW, respectively). The following error categories have been identified: deviations in the source system (including the usage of special codes that are not conformant with the German cancer coding system); minor syntactic errors such as differences in capitalisation; missing mappings of local values to the MDS; technical errors, e.g. with regard to date definitions. Error correction strategies included re-documentation in the source system under controlled conditions based on SOPs; adjustment of the technical interfaces to correct systematical deviations as well as mappings at the level of data elements or their values.

Out of 52 data elements assessed, 21 were syntactically correct from the beginning (Figure 1). Two data elements were fully corrected directly in the source system. Other data elements, showing capitalisation errors or additional text following a systematic pattern, could be corrected by adjustments to the technical interface, improving MDR conformity considerably, in six cases to more than 95%. For three data elements, an additional adjustment of the mapping at value level led to complete MDR conformity.

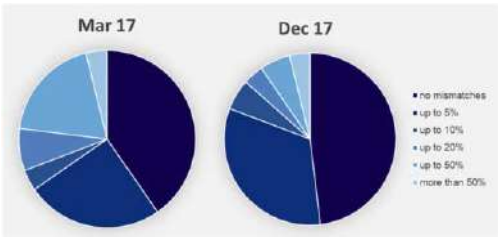


Figure 1 - Assessment of 52 data elements at the Charité, Berlin, at two time points

In Munich, there are two university medical centers hosting their own bridgeheads, namely the MRI, and the KUM. Both institutions provide an LDW in the bridgehead purely dedicated to making data available to the DKTK network. Data from the primary tumour documentation systems are transferred to the LDW via "extract-transform-load" (ETL) processes. ETL processes are well suited to handle systematic deviations. The major challenge here consists in the identification of the correct mappings between the primary systems and the LDW. The error categories identified were: missing data elements (not found); non-MDR-conformant value mappings (mismatch); correct element mapping with some remaining individual errors (partial match). Error correction strategies relied mainly on adjusting the local ETL processes, as well as centrally provided adjustments to the technical interface.

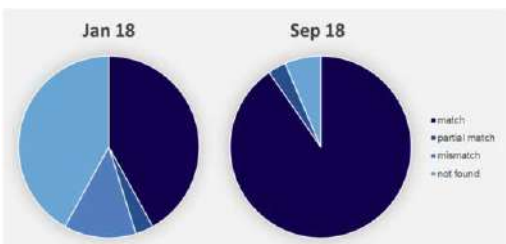


Figure 2 - Assessment of 62 data elements at the MRI at two time points

At the MRI, out of 62 data elements assessed in January 2018, 26 showed a perfect syntactic match (Figure 2). Another 26 data elements were not found in the LDW, while 10 data elements had mismatches or partial matches. At re-assessment in September 2018, the vast majority of data elements (90%) were fully MDR-conformant, with only 6 elements not being found or showing partial matches.

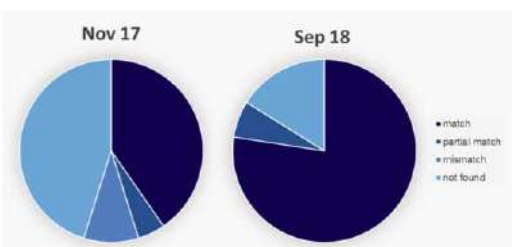


Figure 3 - Assessment of 62 data elements at the KUM at two time points

At the KUM, the data element status at the beginning (November 2017) was very similar to the MRI (Figure 3). Here, 48 out of 62 data elements (77%) showed full MDR-conformity at re-assessment in September 2018, while 14 elements were not found or had partial matches. Five of the data elements are missing due to the fact that the biobank could not yet be integrated into the LDW for organisational reasons.

Conclusions

Standardised quality reports represent suitable tools for the assessment and continuous improvement of data quality at DKTK locations. Moreover, they facilitate the exchange and consolidation of data within the network and help the local documentation systems in identifying and correcting errors. Given the dynamic of newly added data elements and growing patient numbers, the monitoring of data quality represents a continuous challenge.

Acknowledgements

This research is funded by the German Cancer Consortium (DKTK).

References

[1] Klinische Kommunikationsplattform (CCP) im DKTK, https://dktk.dkfz.de/de/klinische-plattformen/ueber-die-ccp/about-ccp [accessed on April 1, 2019].

[2] M. Lablans, E.E. Schmidt, and F. Ückert, An Architecture for Translational Cancer Research As Exemplified by the German Cancer Consortium, *JCO Clinical Cancer Informatics* **1** (2017), 1–8.

[3] A.A. Sinaci, and G.B. Laleci Erturkmen, A federated semantic metadata registry framework for enabling interoperability across clinical research and care domains, *J Biomed Inform* **5** (2013), 784–794.

[4] Metadata Repository (MDR) des DKTK, https://mdr.ccp-it.dktk.dkfz.de [accessed on April 1, 2019].

[5] M.G. Kahn, T.J. Callahan, J. Barnard, A.E. Bauck, J. Brown, B.N. Davidson, H. Estiri, C. Goerg, E. Holve, S.G. Johnson, S.-T. Liaw, M. Hamilton-Lopez, D. Meeker, T.C. Ong, P. Ryan, N. Shang, N.G. Weiskopf, C. Weng, M.N. Zozus, and L. Schilling, A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data, *EGEMS (Washington, DC)* **1** (2016), 1244.

[6] I. Fortier, P.R. Burton, P.J. Robson, V. Ferretti, J. Little, F. L'Heureux, M. Deschênes, et al., Quality, quantity and harmony: The DataSHaPER approach to integrating data across bioclinical studies, *International journal of epidemiology* **5** (2010), 1383–1393.

[7] Batini C, Scannapieco M (2016) *Data and Information Quality: Dimensions, Principles and Techniques,* Springer, Cham.

[8] D. Juarez, E.E. Schmidt, M. Lablans, K. Senghas, and F. Ückert, Report generator for the external validation of conformity of local clinical data to a global metadata repository (MDR), *62. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V. (GMDS)* (2017).

[9] D. Juarez, E.E. Schmidt, S. Stahl-Toyota, F. Ückert, and M. Lablans, A generic method and implementation to evaluate and improve data quality in distributed research networks (submitted), (2019)

Address for correspondence

Dr. Esther Schmidt, DKFZ Heidelberg, Germany
e.schmidt@dkfz-heidelberg.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190484

SNOMED CT Coding and Analytics of *in vitro* Diagnostics Observations

Maël Le Gall[a,b], René Vachon[a], Raphaël Petit[c], Xavier Gansel[a]

[a] *System Development – Knowledge Engineering, bioMérieux, Centre C. Mérieux,5 Rue des Berges, Grenoble, France*
[b] *Master BioInformatique Modélisation et Statistique, Université de Rouen Normandie, Normandie Université, France*
[c] *System Development – Software Development, bioMérieux, 3 route de Port Michaud, 38390 La Balme-les-Grottes, France*

Abstract

This work investigates the capability of SNOMED CT to encode microbiology laboratory data with the goal of fully describing multidrug resistance and breakpoint assignment by specimen.

Keywords:
Systematized Nomenclature of Medicine; Microbiology

Introduction

Antimicrobial resistance (AMR) is transforming the treatment of common infectious diseases. There is a strong international call for actions [1; 2] which include improved data sharing to support secondary usage of medical data [1].

Clinical microbiology laboratory data plays a key role in the fight against AMR [3]. Recording lab data without systematized nomenclature, such as LOINC & SNOMED CT interoperability standards, creates a burden for primary and secondary users of the data [4]. Initiatives in several countries demonstrate the added value of nomenclature standards to support national AMR surveillance [5]. Adopting standards by *in vitro* diagnostic (IVD) systems manufacturers should help microbiology labs report data to local, national and international AMR surveillance systems.

bioMérieux's (IVD solutions provider for the diagnosis of infectious diseases) VITEK®2, VITEK® MS, ETEST® & VIDAS® systems, have [6] >99% LOINC coverage and 90% coverage of nominal SNOMED CT observation values and 60% of SNOMED CT ordinal values. We have identified actions to help close the remaining gaps.

The work presented here describes (1) how we plan to complete the SNOMED CT encoding of our microbiology laboratory data using analytics ; (2) analyze the capability to SNOMED CT to represent and support analysis of MultiDrug Resistant Organisms (MDRO) [7] and (3) initiate SNOMED CT mediated laboratory data analytics.

Methods

Biological specimen

Specimen used in bioMérieux MYLA®, ARGENE® CONNECT systems are grouped according to their implied semantic and business role. Those groups were manually mapped to SNOMED CT (sub-) hierarchies concepts (table 1). The mapping was guided by the HL7 SPM segments.

Table 1– SNOMED CT hierarchies for each group & link to HL7 SPM segment

Group	SNOMED CT hierarchy or sub-hierarchy	HL7 (SPM)
What	*105590001 |Substance (substance)|*	4, 5
Where	*442083009 |Anatomical or acquired body structure (body structure)|*	8, 9
How	*71388002 |Procedure (procedure)|*	7
Why	Out of the scope	--
Status	*362981000 |Qualifier value (qualifier value)|*	24

Antibiotic

Antibiotics supported by bioMérieux VITEK®2, ETEST® & ATB™ antibiotic susceptibility testing systems were semi-automatically mapped to SNOMED CT subtypes of *410942007 |Drug or medicament (substance)|* using the process described in figure 1.

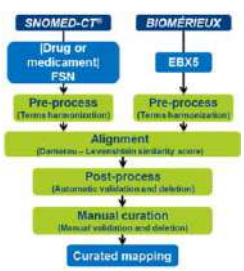


Figure 1

Workflow of antibiotics mapping to SNOMED CT

SNOMED CT mediated labs data analytics

The specimen analytics were performed using the OWL version of SNOMED CT integrated in the Jena Fuseki triple store. The endpoint analysis, Figure2, used SPARQL queries.

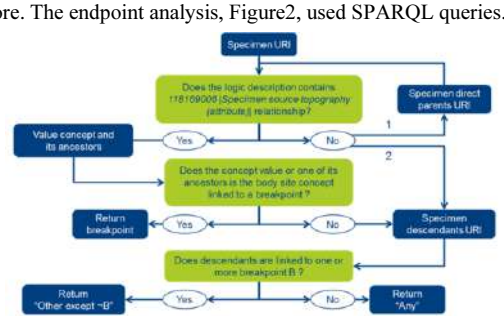


Figure 2– Workflow to run the appropriate sample preparation workflow

Representing MDRO in SNOMED CT

MRDO may be defined through rules in the form of <species name> and ((<susceptibility testing result> to <drug name>) or (<test result> to < Susceptibility test name >)). We used "methicillin resistant *Staphylococcus aureus*" (MRSA) as a test case to examine SNOMED CT's ability to described MDRO by:

1. Pre-coordination of the concept model.
2. Post-coordination of concepts and SNOMED CT compositional grammar expressivity.

SNOMED CT

We used the July 2018 International release of SNOMED CT for all our work.

Results

A total of 107 internal specimen concepts were mapped to SNOMED CT with a 70% (75/107) coverage. 19 additional mappings are under validation by an expert biologist. We used 358 antibiotics form 3 antibiotic susceptibility testing systems; SNOMED CT mapping gave an 88% (316/358) coverage.

Our data analytics use case is based on drug clinical breakpoints [8; 9]. Those are key to interpret antibiotic susceptibility testing results. Breakpoints are applicable for microbial isolates originating, in the case of *Streptococcus pneumoniae* from 'non meningitis' (understood here as 'pneumonia') or 'meningitis' or 'neither pneumonia nor meningitis' specimens origin [9]. We defined 'pneumonia' breakpoint origin as "a specimen obtained from a body site subtype of *20139000 |Structure of respiratory system (body structure)|*" and 'meningitis' breakpoint as "a specimen obtained from a body site subtype of *1231004 |Meninges structure (body structure)|*". Under the working assumption that a lab may use other specimen than in the above mentioned mapping ; we considered all SNOMED CT subtype of *123038009 |Specimen (specimen)|*. Breakpoints were identifed through the *118169006 |Specimen source topography (attribute)|* relationship that tie specimen to body structure. All cases allowed to identify the breakpoints to be applied.

"Methicillin resistant *Staphylococcus aureus*" analysis, as an archetype of MDRO, shows that MDRO concepts themselves do exist in SNOMED CT. Nevertheless, they exist as primitives, and thus do not embark definition supporting future analytics. It appears that (1) some concepts needed to post-coordinate MDRO rule do also exist and (2) that compositional grammar syntax allows building some MDRO rules but not all. The limitation lies in both the SNOMED CT concept model (implemented in the MRCM) and in EL++ that do not support the exclusive disjunction 'XOR' as described in the SNOMED CT Editorial Guide.

Conclusions

Although, SNOMED CT is able to encode biological specimen and antibiotics for the bioMérieux systems analyzed, it does not contain some of the higher-level microbiological concepts required for tracking antimicrobial resistance concepts

Our mediated analytics shows that one can use the SNOMED CT concept model to identify high level ARM concepts if a biological specimens origin is available so clinical breakpoints can be identified. We will continue to investigate this aspect based on additional relationships and usage of LOINC codes to refine and improve our analysis.

In the future we would like to model the microbial phenotypes MRSA and MDRO using pre or post-coordinated analysis of SNOMED CT concepts. Additionally, we would like to supplement our analysis by including the LOINC code mappings. Those do conflict with the SNOMED CT concept model and EL++ expressivity.

Acknowledgements

This work as support by bioMérieux and Master BioInformatique Modélisation et Statistique, Université de Rouen Normandie, Normandie Université

References

[1] J. Emberger, D. Tassone, M.P. Stevens, and J.D. Markley, The Current State of Antimicrobial Stewardship: Challenges, Successes, and Future Directions, *Current Infectious Disease Reports* **20** (2018), 31.
[2] WHO, Antimicrobial resistance: global report on surveillance, in, World Health Organization, 2014.
[3] P.-E. Fournier, M. Drancourt, P. Colson, J.-M. Rolain, B.L. Scola, and D. Raoult, Modern clinical microbiology: new challenges and solutions, *Nature Reviews Microbiology* **11** (2013), 574-585.
[4] R. Flatman, Terminology, units and reporting - how harmonized do we need to be?, *Clin Chem Lab Med* (2018).
[5] X. Gansel, M. Mary, and A. van Belkum, Semantic data interoperability, digital medicine, and e-health in infectious disease management: a review, *European Journal of Clinical Microbiology & Infectious Diseases* (2019).
[6] M.l. Gal, R. Vachon, B. Cellière, M. Dante, F. Holder, M. Mary, M. Olleon, and X. Gansel, Semantic laboratory data interoperability: Encoding microbiology tests and test results using LOINC ® and SNOMED-CT ® from an IVD manufacturer perspective, in: *ASM microbe 2018,* A.S.f. Microbiology, ed., American Society for Microbiology, Atlanta, 2018.
[7] A.-P. Magiorakos, A. Srinivasan, R.B. Carey, Y. Carmeli, M.E. Falagas, C.G. Giske, S. Harbarth, J.F. Hindler, G. Kahlmeter, B. Olsson-Liljequist, D.L. Paterson, L.B. Rice, J. Stelling, M.J. Struelens, A. Vatopoulos, J.T. Weber, and D.L. Monnet, Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance, *Clinical Microbiology and Infection* **18** (2012), 268-281.
[8] CLSI, M100 Performance Standards for Antimicrobial Susceptibility Testing - 28th Edition, in: *CLSI standards*, CLSI, USA, 2018.
[9] EUCAST, Clinical Breakpoints, in, European Society of Clinical Microbiology and Infectious Diseases, , 2018.

Address for correspondence

Xavier Gansel ; xavier.gansel@biomerieux.com ; +33 624 01 43 30
bioMérieux, Centre C. Mérieux - 5 rue des Berges - 38000 Grenble - FRANCE

ORCID https://orcid.org/0000-0001-7148-5461

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190485

CONCERN Factorial Design Survey (FDS) Methods Test: Using REDCap as a Survey Platform

Jose P. Garcia, Jr.[a], Sarah A. Collins[c,d], Kenrick D. Cato[d], Suzanne Bakken[c], Haomiao Jia[d], Min J. Kang[a,b], Christopher Knaplund[d], Kumiko O. Schnock[a,b], Patricia C. Dykes[a, b]

[a] Brigham & Women's Hospital, Boston, MA;
[b] Harvard Medical School, Boston, MA;
[c] Columbia University, Department of Biomedical Informatics, New York, NY;
[d] Columbia University, School of Nursing, New York, NY

Abstract

We assessed the feasibility of using REDCap as a factorial design survey (FDS) platform. REDCap lacks randomization and automation functionality, requiring the development of a workaround. A template survey was created containing all vignettes, copied for each survey instance and edited to hide unwanted content. REDCap configuration required three hours for forty-two surveys. The utilized "copy-and-hide" workaround was successful, providing quasi-automation and reasonable labor-time. Additional strategies are planned using REDCap's Data Dictionary and other survey software.

Keywords:

Survey Methods, Clinical Decision Support Systems

Introduction

Previous research shows that the effectiveness of clinical decision support (CDS) systems decreases when clinicians are presented with a high frequency of alerts that are not perceived as clinically relevant, leading to overriding or ignoring alerts [1]. Factorial design is an experimental design that contains two or more factors, which consist of discrete levels, whose experimental conditions take on all possible combinations of levels across all factors, referred to as *vignettes*. For example, a study containing three factors, each with two levels, has a total number of six possible combinations (2 x 2 x 2 = 6), or six vignettes. A factorial design survey (FDS) does not study the effects of individual factors, but enables researchers to draw conclusions about the significance of the vignettes. It allows researchers to study clinician decision-making, such as which combination of statistically significant factors are perceived as clinically significant [2]. In turn, this helps yield more generalizable conclusions. In the context of our Communicating Narrative Concerns Entered by RNs CONCERN study, vignettes provide scenarios that can be indicative of patient deterioration in clinical documentation, and we are performing FDS to determine the perceived clinical relevance of these scenarios for our CDS model.

REDCap is a software tool for survey generation and electronic data capture, but it has limitations when serving as an FDS platform. Although REDCap contains subject randomization capability, it lacks survey field randomization and automated field population that can help relieve the burden of this complex survey design on the researcher. This case study outlines a workaround developed to circumvent REDCap's limitations and to facilitate its use to run a factorial design survey.

Methods

A pilot survey trial was conducted using a workaround, referred to as "copy-and-hide". Four factors, each with two levels, were used in the pilot (Table 1). We sought to test surveys containing four of the sixteen total vignettes (Table 2). To assign which vignettes were tested in each survey, block randomization was performed off-platform. Each vignette was numbered (1-16), and a function in Excel was created that randomly assigned four numbers out of the 16 respectively to four surveys. A template survey containing fields for all sixteen vignettes was created in REDCap's Online Designer. Every field in the template survey containing a vignette had the action tag "@HIDDEN-SURVEY" applied, which hides fields from survey participants' view. For every consecutive test survey created, the template survey was copied, and the fields containing the assigned four vignettes to be tested in that survey had their action tags removed. This allowed participants to see and score these four vignettes in their survey and left the remaining twelve hidden from view.

Table 1– Pilot Factors and Levels

Factor	Levels (count)	Factor Level Values
Factor 1. Vital sign frequency	2	≤ 22 sets of vital signs in the last 48 hours > 22 sets of vital signs in the last 48 hours
Factor 2. Note / comment frequency	2	≤ 5 comments/notes in the last 48 hours > 5 comments/notes in the last 48 hours
Factor 3. Highlighted Oxygen saturation comment/note	2	Yes No
Factor 4. Highlighted blood pressure comment/note	2	Yes No

Table 2- Sample Vignettes

Vignette Number	Vignette Content
1	Patient chart shows the following: • ≤ 22 sets of vital signs in the last 48 hours • ≤ 5 comments/notes in the last 48 hours • A highlighted Oxygen saturation comment/note • A highlighted blood pressure comment/note
8	Patient chart shows the following: • ≤ 22 sets of vital signs in the last 48 hours • > 5 comments/notes in the last 48 hours • No highlighted Oxygen saturation comment/note • No highlighted blood pressure comment/note
11	Patient chart shows the following: • > 22 sets of vital signs in the last 48 hours • ≤ 5 comments/notes in the last 48 hours • No highlighted Oxygen saturation comment/note • A highlighted blood pressure comment/note

Results

Forty-two surveys were created and disseminated to forty-two participants in three hours. Thirty-one responses were returned. Extracted data table was sized 43 by 821 cells.

Conclusions

The significance of this work is that we identified limitations of available survey software for performing a factorial design survey. Factorial design is an experimental design that necessitates multiple iterations of a survey, where its components' material and order change from iteration to iteration. REDCap's lack of randomization and automation functions are limitations. The "copy-and-hide" method was a successful workaround in that it provided our team a reproducible method with reasonable labor time (~4.5 minutes/survey). However, the reliance on manual input in this method introduces increased potential for human error when scaled up. Additionally, REDCap cannot differentiate between hidden and unhidden cells when a user extracts data, which created a large spreadsheet with numerous empty cells for the hidden fields that were not scored during our pilot. To sacle up this method for FDS with hundreds of surveys would yield a cumbersome data workbook for researchers. A second pilot test is planned utilizing REDCap's Data Dictionary, which is a completely off-platform survey designer that we hope will decrease labor-time further by allowing us to rapidly create smaller, specialized surveys which we can upload onto the software and streamline data extraction by decreasing the number of empty cells in the spreadsheet. Additionally, we plan to run an FDS pilot using other popular survey software and compare results.

Acknowledgements

This study was funded by the National Institute of Nursing Research (NINR): 1R01NR016941-01, Communicating Narrative Concerns Entered by RNs (CONCERN): Clinical Decision Support Communication for Risky Patient States. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] A. S. Kesselheim, K. Cresswell, S. Phansalkar, D. W. Bates, and A. Sheikh, Clinical decision support systems could be modified to reduce 'alert fatigue' while still minimizing the risk of litigation, *Health Affairs* 30(12) (2011), 2310-2317. doi: 10.1377/hlthaff.2010.1111

[2] BJ Taylor. Factorial surveys: using vignettes to study professional judgement, *British Journal of Social Work* 36(7) (2006), 1187-1207. doi:10.1093/bjsw/bch345

Address for correspondence

Jose Pedro Garcia, Jr.

Brigham and Women's Hospital: Division of General Internal Medicine and Primary Care

1620 Tremont Street

Boston, MA 02120

Phone: (617) 278-0638

Email: jgarcia41@bwh.harvard.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190486

Maxwell®: An Unsupervised Learning Approach for 5P Medicine

Joël Gardes[1], Christophe Maldivi[1], Denis Boisset[1], Timothée Aubourg[1,2], Nicolas Vuillerme[2,3], Jacques Demongeot[2,3]

[1] Orange Labs, Meylan, France
[2] Université Grenoble Alpes, AGEIS, EA 7407, La Tronche 38700, France
[3] Institut Universitaire de France, Paris, France

Abstract

In the 5P medicine (Personalized, Preventive, Participative, Predictive and Pluri-expert), the general trend is to process data by displacing the barycenter of the information from hospital centered systems to the patient centered ones through his personal medical records. Today, the use of artificial intelligence for supporting this transition shows real limitations in its implementation in operational practice, both at the level of patient care, but also in the general daily life of the health professional, because of the medico-legal imperatives induced by the promises of the '5P medicine'. In this paper, we propose to fill this gap by introducing an original artificial intelligence platform, named Maxwell, which follows an unsupervised learning approach in line with the medico-legal imperatives of the '5P medicine'. We describe the functional platform characteristics and illustrate them by two examples of clustering in genomics and magnetic resonance imaging.

Keywords:

Machine Learning, Magnetic Resonance Imaging, Genomics

Introduction

Recent advancements in predictive medicine are now partly due to the implementation of artificial intelligence algorithms based on the intensive exploitation of big data collected from new medical technologies [1-3]. Usually, these algorithms can be divided into two distinct categories, namely (1) neural networks and deep learning, performing and fast, but opaque (black box type), with a traceability difficult to explain to the 4 worlds of the concerned health ecosystem, medical, para-medical, forensic and familial, and (2) classical methods playing with distances or proximities to the centroids of classes, more easily explicable, but sometimes slower. Despite major and continuous breakthroughs in both of these domains, it is interesting to note that most of algorithms fail to be successfully implemented in practice. Indeed, the medico-legal context imposes these algorithms to satisfy fundamental legal and ethic criteria not usually taken into account. In this paper, we propose a new unsupervised learning approach implemented through a "classical" decision platform named "Maxwell®" that addressed three of these criteria, namely (1) traceability in decision process, (2) consideration of the role of the human expert in the decision-making and (3) need of a feedback to the patient for increasing his empowerment. We describe first the learning characteristics of the method, and then illustrate its functioning by two examples, that are the clustering in (1) genomics, and (2) magnetic resonance imaging. Note that detailed description will be given in upcoming publications.

Methods

Maxwell's Functions and Dynamic of Processes

Maxwell is a research result in domain of data clustering leaded at Orange Labs® (France) that aims to produce an agnostic and unsupervised system for clustering and classifying heterogeneous content. Two initial postulates were defined to guide our work: (1) the ability to start building clusters with a short number of content and (2) no recourse to initial labeling.

Calculus' Principles

Depending on the algorithm's step, two types of calculus are used: (1) similar content clustering using a normalized compression distance [4-6], and (2) content classification in existing clusters. To enable incremental classification, both of these processes can be iteratively chained.

Clustering Principle

Clustering is based on building neighborhood graphs. It is initiated by computing triangles formed with the current point in a distance matrix, the first neighbor of this current point and the first neighbor of this second point. Then, surfaces of these triangles are computed as well as basic statistics such as mean, standard deviation and quantiles, are used as parameters for tuning the selectivity of the clustering process. Thus, the result is a metric space segmented in sub-graphs containing clusters.

Classification Principle

During the classification step, the system is focused on edges of sub-graphs. Classification of a new content is done by searching the nearest centroid neighbor and deciding inclusion with respect to existing cluster nodes distances to centroid. Note that unclassified content is stored in singleton clusters.

Traceability and Role of Human Expert

Cluster management can be done either by the user or by the system itself. In fact, user visualizes the clusters produced throughout the clustering/classification process. He can select elements of a cluster, or the cluster itself to be deleted. This gives the traceability of Maxwell and destroying cluster from an automatic threshold procedure or from a user decision is near of a reinforcement learning by a "punishment/reward" system. On the other side, the system will compute all consequences of a user action on clusters distribution: singletons recall (due to its unclassified added content or to a cancelling by user or machine) and new content addition, inserted in a queue in order to detect new clusters. Finally, user can also validate clusters by locking and prohibiting their destruction. This allows also the control of clusters consistency. Locked clusters are named by user in order to obtain semantic classes.

Results

Genomics

Figure 1 illustrates a classification obtained from about 120 genomes coming from different phyla (bacteria, archæa, giant viruses and virophages). The obtained clusters reveal 3 distinct classes, which present remarkable properties: (1) all the Pandora giant viruses belong to the same class 1, (2) all the genomes participating to the same "infectious" ecosystem are in the same class 2: Amœba Acanthamœba can be infected by two giant viruses, Mimivirus and Mamavirus. Both have their virophage, respectively Sputnik and Zamilon. All the corresponding genomes belong to the same class. (3) Furthermore, Acanthamœba can be parasitized by the bacteria Legionella pneumophila. It belongs to singleton class 3, but can join class 2, when compression distance between genomes is relaxed.

Figure 1 – Three Clusters Within a Set of 120 Genomes

Hence, Maxwell puts in the same cluster all the genomes participating to the same infectious ecosystem, which allows for predicting susceptibility of human to infectious agents.

BRATS Images

BraTS image dataset is a manually indexed set of brain MRI images, used for image processing challenges on MRI images. Our sample contains 1500 blood stroke 2D images (horizontal sections at different levels). Figure 2 illustrates some results using the Maxwell platform.

Conclusion

The Maxwell platform constitutes a good alternative to the deep learning approach using neural networks, being more comprehensible in the medical ecosystem both on the patient side (family, care-givers and patient), on the medical side (physicians and paramedics) and on the social side (psycho-social workers). It is not strictly antinomic with the deep learning in the sense that some parts of the classification process (in particular the refinement of some ambiguous clusters) can be reprocessed by neural networks.

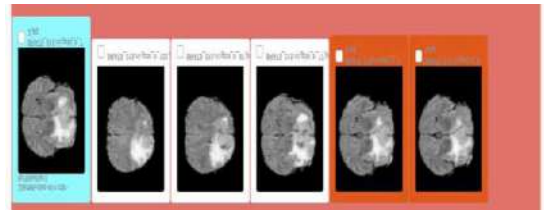


Figure 2 – Traceability: modified cluster of BRATS_015 class, before recalculation (best representative in light blue), images added in orange. All thumbnails can be selected by user to be deleted, if a better representative image is selected.

References

[1] J. Demongeot, A. Elena, M. Jelassi, S. Ben Miled, N. Bellamine Ben Saoud, C. Taramasco. Smart Homes and Sensors for Surveillance and Preventive Education at Home: Example of Obesity. Information; 2016(7):50.

[2] C. Taramasco, T. Rodenas, P. Fuente, R. Munoz, R. Olivares, V. Ibuquerque, J. Demongeot. A Novel Monitoring System for Fall Detection in Older People. IEEE Access; 2018(6):43563-43574.

[3] C. Taramasco, F. Martinez, E. Fredes, J. Demongeot. A Novel Low-Cost Sensor Prototype for Nocturia Monitoring in Older People. IEEE Access; 2018(6):52500-52509.

[4] M. Li, P.M.B. Vitanyi. An Introduction to Kolmogorov Complexity and its Applications. Springer-Verlag 3rd Edition; 2008.

[5] R. Cilibrasi, P.M.B. Vitanyi. Clustering by compression, IEEE Trans. Inform. Theory; 2005(51):1523–1545.

[6]. AR. Cohen, P.M.B. Vitanyi. Normalized compression distance of multisets with applications. IEEE Transactions on Pattern Analysis and Machine Intelligence; 2015(37):8.

Address for correspondence

Joël Gardes: joel.gardes@orange.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190487

Assessing the Concordance of Clinical Classification Criteria for Lupus Between Electronic Health Records and a Physician Curated Registry

Theresa L. Walunas [a], Anika S. Ghosh [a], Jennifer A. Pacheco [a], Kathryn L. Jackson [a], Anh H. Chung[a], Daniel L. Erickson [a], Karen Mancera-Cuevas [a], Rosalind Ramsey-Goldman [a], Abel N. Kho[a]

[a] Northwestern University Feinberg School of Medicine, Chicago, IL, United States of America

Abstract

We developed a computable phenotype for systemic lupus erythematosus (SLE) based on the Systemic Lupus International Collaborative Clinics clinical classification criteria set for SLE. We evaluated the phenotype over registry and EHR data for the same patient population to determine concordance of criteria detected in both datasets and to assess which types of structured data detected individual classification criteria. We identified a concordance of 68% between registry and EHR data relying solely on structured data.

Keywords:
Electronic health records, phenotype, systemic lupus erythematosus.

Introduction

Clinical classification criteria are often used to understand the clinical presentation of diseases with complex or varied presentations. Electronic health records (EHR), now commonly used in the course of routine health care throughout the United States provide a rich source of data [1], including diagnoses, lab results and medication use that can be used to enhance understanding of the development, manifestation and treatment of complex disease over time.

Our team developed a rules-based algorithm to identify patients with Systemic Lupus Erythematosus (SLE), a complex systemic autoimmune disease that affects multiple organ systems and has diverse manifestations that develop over time [2]. This computational phenotype is based on the Systemic Lupus International Collaborating Clinics (SLICC) classification criteria for SLE [3].

We evaluated our SLICC-based computational phenotype in a physician-validated registry and the medical record data for the same patient population: 1) to assess the concordance of criteria detection in these two distinct datasets; and 2) to better understand whether medical records can be used as a substitute for manual chart abstraction in the identification of clinical classification criteria for a complex autoimmune disease.

Methods

Established in 1991, the Chicago Lupus Database (CLD) is a physician validated registry of 1,052 patients with possible or definite lupus according to the revised 1982 American College of Rheumatology classification criteria [4][5]. The CLD has laboratory data, symptoms and paitent demographics based on each known visit. If a patient was referred, previous history information from the notes are documented. The data is entered in MEDLOG [6] which is then compiled into the CLD.

The Northwestern Medicine Electronic Data Warehouse (NMEDW) is the primary data repository for of all the electronic health records of patients who receive care within the Northwestern Medicine system. Established in 2002, the NMEDW contains records for over 3.8 million patients. Patients in the CLD consented to allow their medical records to be used for research and their medical records can be found in the NMEDW using a medical record number (MRN) stored in the CLD. Review medical records in the NMEDW was approved by the Northwestern University IRB.

We identified 878 patients who had definite lupus according the SLICC criteria [3], as defined by meeting at least one clinical and one immunolgical criteria and meeting 4 or more criteria overall. After removing patients who did not have medical records in the NMEDW, there were 818 patients remaining. To ensure sufficient depth of data for analysis by our algorithm, we removed any patients who had less than four encounters documented in the NMEDW, reducing the cohort size to 472. Finally, we assessed our the full algorithm over the patient medical record data. Only 408 of those patients who also met the SLICC criteria for definite lupus in the CLD also met the criteria for definite SLE based on their medical record data.

The SLICC clinical criteria-based algorithm for detection of SLE was run over each patient in the cohort and each of the 17 individual criteria was scored for presence or absence in the CLD and the NMEDW. A combination of ICD9/10 codes and laboratory results were used to determine whether each individual criterion was satisfied.

A criterion was considered discordant when it was present in either the CLD or NMEDW (but not both) for a given patient record. A KNIME (3.4.2) workflow [7] was developed to process each patient data set and determine the number of concordant and discordant criteria. For discordant criteria, we determined whether the criterion was present in the CLD or the NMEDW. For each criterion, we used McNemar's test to see if the NMEDW and CLD results were different. A p value $< .05$ was used to determine significant difference. Descriptive statistics were calculated using SAS software version 9.4 (SAS Institute)Results

Table 1 describes the basic demographics of the cohort we assessed that met the SLICC classification criteria for definite SLE in the CLD and NMEDW. Our cohort is predominantly female, white, and had an average disease onset age of 30 years. The gender distribution and age of onset is consistent with previous studies demonstrating the predominance of SLE among women that develops relatively early in life [8]. The racial and ethnic distribution is consistent with the patient

population receiving treatment within the Northwestern Medicine System.

Table 1— Cohort Demographics (N=408)

Sex	**% of Cohort**	
Female	92%	
Male	8%	
Race/Ethnicity	**% of Cohort**	
Caucasian	48.1%	
African American	30.5%	
Hispanic	12.4%	
Asian	7.8%	
Other	1.2%	
Age	**Mean Years**	**SD**
Current	50 years	13.48
At Diagnosis	30 years	9.69

To further understand the concordance between the registry and the medical record data, we determined the number of present and absent criteria that were concordant as well as the average number of criteria detected per patient in each data set. The results of this assessment are shown in Table 2. Within the concordant criteria, on average, there were 5 that were concordant and present, 7 that were concordant but absent. When evaluating the criteria set over the CLD, on average, we detected 8 criteria per patient, while we detected 7 criteria per patient using their medical record data. A paired t-test for significance resulted in a p-value of .59, indicating that the results are not significantly different.

Table 2— Overall Concordance between Registry and Medical Record Data for Patients Meeting SLICC Criteria (out of 17 total).

Number of Criteria Per Patient	**Mean**	**Median**	**SD**
Overall Concordant Criteria	11.6	12	1.7
Present Concordant Criteria	4.9	5	1.8
Absent Concordant Criteria	6.7	7	2.0
Criteria Detected in CLD	7.9	8	2.1
Criteria Detected in EDW	7.3	7	2.1

Discussion

We developed a clinical classification criteria-based computational phenotype to identify patients with SLE in a physician validated SLE registry and in a large medical record data repository and assessed concordance of the overall algorithm and individual criteria that comprise the algorithm. For those patients who satisfied the classification criteria for definite lupus in both the registry and the medical record data set, there were, on average, 12 concordant criteria out of 17, 5 of which were concordant and present in both the CLD and NMEDW, 7 of which were concordant and absent in both the CLD and NMEDW. When we assessed concordance for individual criterion, we found that concordance was highest for criteria that were based on laboratory data. The highly discordant criteria were primarily clinical criteria detected with diagnosis codes that may not always be documented as part of routine clinical care or may be documented in locations within the medical record, such as physician notes, that are not easily queried using simple structured data elements, such as arthritis, oral ulcers, and serositis.

Conclusion

Using a computational phenotype for SLE based on the SLICC clinical classification criteria, we demonstrated an overall high concordance of 68% between physician validated registry information and data found in patient medical records suggesting medical record data can be used to supplement manual chart abstraction for the application of clinical classification criteria to patients with complex disease.

Acknowledgements

Research reported in this publication was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health under Award Number R21AR072263. Development of the CLD database were supported by grant P60 AR064464 and P30 AR072579.

References

[1] Rasmussen LV. The electronic health record for translational research. *J Cardiovasc Transl Res*. 2014;07(06):607–614.

[2] Walunas et al. A Comparison of the American College of Rheumatology (ACR) and Systemic Lupus International Collaborating Clinics Classification (SLICC) Criteria to Detect Patients with Systemic Lupus Erythematosus (SLE) in Electronic Health Record (EHR) Data. Poster presented at 2018 ACR/ARHP Annual Meeting; October 24, 2018; Chicago, IL

[3] Petri M, Orbai A-M, Alarcón GS, et al. Derivation and Validation of Systemic Lupus International Collaborating Clinics Classification Criteria for Systemic Lupus Erythematosus. *Arthritis and rheumatism*. 2012;64(8):2677-2686. doi:10.1002/art.34473.

[4] Tan, EM, Cohen, AS, Fries, JF, Masi, AT, McShane, DJ, Rothfield, NF, et al., The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1982;25:1271-7

[5] Hochberg, MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997;40:1725.

[6] Information about Medlog. Retrieved from: http://med- logsystems.com/medinfo.htm

[7] Berthold M.R. et al. (2008) KNIME: The Konstanz Information Miner. In: Preisach C., Burkhardt H., Schmidt-Thieme L., Decker R. (eds) Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg

[8] Dall'Era M. Systemic lupus erythematosus. In: Imboden JB, Hellman DB, Stone JH. (Eds). Current Rheumatology Diagnosis and Treatment. 3rd ed. New York, NY:McGraw-Hill; 2013.

[9] Gliklich RE, Dreyer NA, Leavy MB, editors. Registries for Evaluating Patient Outcomes: A User's Guide [Internet]. 3rd edition. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014 Apr. 15, Interfacing Registries With Electronic Health Records.Available from: https://www.ncbi.nlm.nih.gov/books/NBK208625/

[10] Bruce, I. N., et al. (2015). "Factors associated with dam- age accrual in patients with systemic lupus erythematosus: results from the Systemic Lupus International Collaborat- ing Clinics (SLICC) Inception Cohort." Ann Rheum Dis **74**(9): 1706-1713.

Address for correspondence

Theresa L. Walunas: 625 N Michigan Ave, 15th Floor, IL, 60611; t-walunas@northwestern.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190488

Identification of Cancer Survivors Living with PTSD on Social Media

Nur Hafieza Ismail[a], Ninghao Liu[a], Mengnan Du[a], Zhe He[b], Xia Hu[a]

[a] *Department of Computer Science & Engineering, Texas A&M University, College Station, Texas, USA,*
[b] *School of Information, Florida State University, Tallahassee, Florida, USA*

Abstract

The trauma of cancer often leaves survivors with PTSD. Tweets posted on Twitter usually reflect the users' psychological state, which is convenient for data collection. However, Twitter also contains a mix of noisy and genuine tweets. The process of manually identifying genuine tweets is expensive and time-consuming. Thus, we propose a knowledge transfer technique to filter out unrelated tweets. Our experiments show that our model outperforms the baselines.

Keywords:

Cancer survivors; stress disorders, post-traumatic; social media

Introduction

Being diagnosed with cancer often causes psychological distress due to painful treatments and the trauma of the experience [1]. Treated cancer patients may find themselves at risk of recurrence, which is often more traumatic than the initial diagnosis. This psychological impact can cause post-traumatic stress disorder (PTSD), a significant concern in clinical oncology [2]. Untreated PTSD patients are more likely to adopt health-risk behaviors. The lack of quantifiable data about PTSD is one of the main obstacles to making reliable diagnoses and providing effective treatment [1]. The fast growth of social media provides new opportunities for investigating mental health issues, since the simplicity of social media allows its users to easily share their daily thoughts and feelings. We propose a deep neural network (DNN)–based technique to create a labeled dataset using cancer free–related keywords and PTSD features. Our primary contributions can be summarized as follows: (1) formally defining the problem of extracting relevant tweets, (2) presenting and training a framework for identifying cancer survivors with PTSD based on tweets, and (3) evaluating the model's performance by producing a label with associated probability for new tweets.

Methods

Here, we briefly introduce our proposed framework for identifying cancer survivors living with PTSD on social media. First, we extracted patterns of words used by sufferers from conventional studies on depression. Second, the extracted patterns were used to identify tweets from the cancer survivors dataset that contain PTSD indicators. We will detail the process of our proposed framework in three subsections: (1) feature extraction, (2) knowledge transfer, and (3) the convolutional neural network (CNN) architecture.

Feature Extraction

This process requires specific extraction methods to be applied to the input data in order to create an accurate prediction model. Previous work related to depression detection using social media data has employed a crowdsourcing approach to identify tweets associated with mental illness [3] which we label as PTSD-positive, while the rest are labeled as PTSD-negative. The process continues by combining PTSD-positive and -negative datasets. The goal of this process is to understand the differences in linguistic style between the two groups. LIWC, a text analysis tool, is employed to assess the tweets for proportions of words used from several linguistic categories in the psychological expression lexicon. The comprehensive list produced by the tool automatically presents the words most frequently used by people with mental illness.

Knowledge Transfer

In this work, 'knowledge transfer' is defined as the process that uses the output from the feature extraction step to develop a labeled PTSD dataset. Diverse epidemiological samples have shown that most PTSD sufferers also have depression. This comorbidity reflects overlapping symptoms between the two disorders [4]. Therefore, even though there is no existing PTSD lexicon, we can use the depression lexicon as a proxy to identify PTSD-related tweets. We collected our raw dataset using "cancer" as a keyword through Twitter's Application Programming Interface (API) over a period of four months from October 2017 to January 2018. Then, we conducted the extraction process in two steps using two sets of keywords. First, we created the cancer survivor dataset using related hashtags and terms such as "cancer survivor", "cancer free", "i had cancer", "post-cancer", "survive from cancer", and "free from cancer" . Second, we used the depression lexicon taken from [3] to filter out tweets that contain no PTSD signals to create a ground truth dataset. We also added the word "PTSD" to filter out irrelevant tweets. This technique saved a lot of time and cost in creating a reliable dataset for training the CNN.

CNN Architecture

Our proposed CNN adopts one convolutional layer to produce results for tweet classification. We trained the CNN with the embedding layer. It requires specification of the vocabulary size, the size of the real-valued vector space, and the maximum length of words in the input. For convolutional feature maps, we used 100-dimensional word embedding for

text representation. 32 filters were applied by referring to the conservative setting for word processing, with a kernel size of 8 and a rectified linear activation function. This was followed by a pooling layer in which the filters generated feature maps and reduced the output by half. The end layer used a sigmoid activation function to output a value between 1 and 0 (positive and negative) for the tweets based on the concatenation of the previous vectors. Then, the extracted model was saved for later evaluation.

Experiment

The PTSD-positive dataset represents the diagnosed group, while PTSD-negative is the control group. For the diagnosed group, we retrieved tweets from users who publicly stated that they survived cancer and had PTSD. To construct the PTSD-negative group, we made use of tweets from the 'Twitter User Gender Classification' dataset on the Kaggle website. This dataset contains random tweets not limited to any specific topic. Both groups have the same amount of tweets (10k), which allows us to create balanced datasets. The data preparation phase consists of three tasks: (1) splitting the dataset into training (80%) and test (20%) sets, (2) cleaning the dataset to remove punctuation, stop words, and numbers, and (3) defining a vocabulary of preferred words from a training dataset by keeping only tokens that appear at least five times. All models were trained over ten epochs. The Adam optimizer implementation of stochastic gradient descent was used.

Results

To evaluate our proposed model, we chose three baseline models that are capable of handling text datasets: multilayer perceptrons (MLP), CNN n-gram, and recurrent neural networks (RNN). Our results indicate that CNN can effectively identify cancer survivors with PTSD. Experimental results in Table 1 show that the accuracy of CNN reaches 98.5%. Due to the stochastic nature of DNNs, we ran the experiments multiple times to get a reasonably accurate result.

Table 1 – Experiment Results

Method	Accuracy (%)
CNN	**98.5**
MLP	95.5
CNN N-gram	98.3
RNN	96.9

Our simple CNN-based prediction system assigns probability values for cancer survivors being either PTSD-positive or PTSD-negative. Sample tweets and our method's assignments are displayed in Table 2.

Table 2 – Predictions for New Tweets

Tweet: "I have had more difficulty post cancer than during my active treatment. To me it is a neverending path (hate the word journey)."
POSITIVE (99.96%)
Tweet: "I hate myself, I don't feel like living anymore."
NEGATIVE (99.99%)
Tweet: "I got a gold MacBook that only use for music and homework..Still keep it in apple box."
NEGATIVE (99.95%)

The system is consistently able to classify tweets correctly. For example, the second tweet is a statement with negative sentiment that is not related to the cancer survival; the system classifies it as PTSD-negative. Assigning the right label with high probability is essential for accurate diagnosis. To the best of our knowledge, this is the first work that deploys an extracted model of cancer survivors living with PTSD into a prediction system that is capable of evaluating new tweets. The outcome shows high potential for a low-cost text classification technique that may also apply to other medical conditions that can affect patients' mental health.

Conclusions

We have demonstrated that Twitter can be used to identify PTSD among internet users, presenting a CNN prediction framework that produces promising results in cancer survivors with a PTSD diagnosis. Social media users with a history of cancer who suffer from depression will benefit from the prediction framework; for instance, it can act as an alarm system by detecting the presence of depression based on users' posts. However, our model was trained solely on text-based tweets. In the future, we plan to explore the utility of other data modalities for uncovering PTSD indicators, such as audio and images, to improve our diagnoses. In addition, we hope to identify developing conditions such as suicidal thoughts and side effects of PTSD treatment.

Acknowledgements

This work is supported in part by the Ministry of Higher Education Malaysia and University Malaysia Pahang (UMP).

References

[1] J.M. Salsman, S.C. Segerstrom, E.H. Brechting, C.R. Carlson, and M.A. Andrykowski, Posttraumatic growth and PTSD symptomatology among colorectal cancer survivors: a 3-month longitudinal examination of cognitive processing, *Psychooncology* **18** (2009), 30-41.

[2] M.L. Stuber, A.E. Kazak, K. Meeske, L. Barakat, D. Guthrie, H. Garnier, R. Pynoos, and A. Meadows, Predictors of posttraumatic stress symptoms in childhood cancer survivors, *Pediatrics* **100** (1997), 958-964.

[3] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, Predicting depression via social media, in: *Seventh international AAAI conference on weblogs and social media*, 2013.

[4] J.D. Flory and R. Yehuda, Comorbidity between post-traumatic stress disorder and major depressive disorder: alternative explanations and treatment considerations, *Dialogues in clinical neuroscience* **17** (2015), 141.

Address for correspondence

Xia Hu, 330B H.R. Bright Building, College Station, TX 77843-3112
Email: hu@cse.tamu.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190489

Facilitating Clinical Trial Recruitment by Recommending Cost-Efficient Medical Exams

Bibo Hao [a], Shouyu Yan [b], Eryu Xia [a], Shilei Zhang [a], Jing Mei [a]

[a] *IBM Research – China, Beijing, China*
[b] *IBM Watson Health, Beijing, China*

Abstract

Clinical trials are key and essential processes for researchers to develop new treatments as well as evaluate their effectiveness and safety, whilst more than half of all clinical trials experience delays, which leads to a considerable amount of cost. In this paper, we present a cost-effective framework to reduce the time and monetary cost in the stage of recruiting and screening eligible clinical trial participants. By leveraging patients' observed conditions and the cost of medical examinations, the proposed framework uses collaborative filtering techniques to predict the utilized cost for the to-do medical examinations and then rank patients and medical examinations. The preliminary experiment results indicate that the framework is promising to reduce the cost spent on medical examinations by three quarters or even more and accelerate the recruitment process in the screening stage.

Keywords:
Eligibility Determination, Cost Savings, Machine Learning.

Introduction

Clinical trials, aiming at answering specific questions about the safety and efficacy of biomedical or behavioral interventions, are essential processes in medical research and innovation, in which human participants are recruited into experiments or observations. Unfortunately, recruitment is often slower or more difficult than expected, which leads to considerable monetary, time, and opportunity cost [1].

Faced with clinical trial recruitment are challenges from many aspects: fewer eligible patients than expected, fewer patients agreeing to participate than expected, etc. [2] When recruiting clinical trial subjects, it is common to see that some condition values of the patients are not available in EMRs. In a clinical trial determining the eligibility of patients with unobserved conditions is therefore difficult. To clarify, it is needed for additional medical examinations to determine the eligibility, while other patients with the history of these medical examinations, are not always eligible. Unobserved conditions can turn out to be eligible or not eligible.

Given patients and their observed/unobserved conditions, to maximize the probability of matching the patients to meet the eligibility criteria and minimize the cost of medical examinations, which patients shall be firstly selected to conduct medical examinations, and which medical examinations shall be conducted first? In this study, we formulate the problem definition and then use a synthetic dataset and a mockup clinical trial to simulate the recommending cost-efficient medical examinations scenario.

Methods

Problem Definition

Given a clinical trial that requires a set of clinical trial eligibility criteria, which corresponds to n conditions: $\boldsymbol{C} := \{C_1, C_2, \dots, C_n\}$, each condition requires one or more medical examinations to check patients' eligibility, and the cost of the evaluation is known as $\boldsymbol{E} := (e_1, e_2, \dots, e_n)^T$.

Given a patient cohort of m clinical trial participant candidates $\boldsymbol{P} := \{P_1, P_2, \dots, P_m\}$, we denote P_i's condition value on C_j as $v_{i,j}$, and we use $\boldsymbol{V}$ to denote this patient to medical examination matrix. We let the data types of the values in the matrix to be binary values or real numbers, since we can split nominal values into several binary values. The values in $\boldsymbol{V}$ can be extracted from EMRs, or they can be missing because the patients have not yet finished specific medical examination, in other words, the matrix $\boldsymbol{V}$ is sparse.

With the above conditions, from the cost perspective, we are to solve the problem of: which patients in $\boldsymbol{P}$ shall be reached first for medical examination to check eligibility and given a patient P_i, which medical examination shall be conducted first to reduce wastage.

The Cost-Efficient Recommendation Framework

To maximize the utilization of the cost used for medical examinations and accelerate clinical trial subjects' recruitment process, we propose a cost-efficient recommendation framework to measure the utilized cost for a specific medical examination to each participant candidate when the condition value is unobserved. By leveraging this framework, we can 1) rank the medical examinations for a given participant candidate to reduce the potential wastage for the individual, and 2) rank the participant candidates to reduce the potential wastage for a group of patient candidates.

Introducing Utilization and Utilized Cost

To model the utilization and utilized cost, we take two aspects into consideration in this framework:

- Cost of each medical examinations required to check the eligibility criteria, aforementioned as $\boldsymbol{E}$, which are already known and constant values for all patients.
- Patients' probabilities satisfying eligibility criteria given their unobserved conditions.

We then propose a method to measure the *utilized cost* for patients to conduct medical examinations. The *utilized cost* function of conducting a medical examination is designed as below: Given a medical examination C_j, for a patient P_i, we let the *utilized cost* $Z_{i,j} := -\ln(q_{i,j}) \cdot e_j$

where $q_{i,j}$ is the estimated probability ($q \in [0,1]$) that $v_{i,j}$ can satisfy C_j. Higher $q_{i,j}$ results to lower utilized cost $Z_{i,j}$.

Intuitively, if $v_{i,j}$ is already observed to satisfy eligibility criteria, then no cost is needed to be spent on C_j for P_i, whilst if $v_{i,j}$ is observed to dissatisfy eligibility criteria, then any other cost spent for P_i is of no utilization. So here we design the value of $q_{i,j}$ to represent the model' estimation of the probability that $v_{i,j}$ will satisfy C_j after medical examination.

If it is already observed that $v_{i,j}$ satisfy C_j or not satisfy C_j, the values of $q_{i,j}$ are *1* and *0,* respectively. If $v_{i,j}$ is unobserved yet, $q_{i,j}$ need to be estimated based on observed values in $\boldsymbol{V}$.

Estimating the Probabilities of Patients to Satisfy Criteria

As mentioned, for observed values $v_{i,j}$ in $\boldsymbol{V}$, the correspondent values of $q_{i,j}$ can be calculated as *1* or *0* based on eligibility criteria C_j. Given the sparsity of $\boldsymbol{V}$, there are also missing values in $\boldsymbol{Q}$ corresponds to unobserved values in $\boldsymbol{V}$. Therefore, we propose a method to predict those missing values in $\boldsymbol{Q}$, based on the observed values.

Based on the intuition of "patients of similar latent status will have similar conditions and phenotype", we introduce the *Collaborative Filtering* (CF) technique to predict those missing values in $\boldsymbol{Q}$. The CF technique is already widely used in the e-commerce recommender systems and has shown its values in predict users' interest to new items based on existing feedbacks. In this study, we applied two popular methods of CF: the User-based Similarity Model (UBSM) [3], and the Singular Value Decomposition (SVD) algorithm [4], to approximate patients' probabilities satisfying eligibility criteria.

Ranking Patients by Utilized Cost

After using the CF technique to approximate the matrix $\boldsymbol{Q}$, we can compute the overall utilized cost Z_i expected for each participant candidate P_i: $Z_i = \sum_j Z_{i,j}$. Patient candidates with lower utilized cost values should therefore be reached first for the screening and medical examinations. And for a specific candidate P_i, medical examinations corresponding to smaller $Z_{i,j}$ values should be conducted first.

Evaluating the Cost Efficiency

Here we use this metric to measure the efficiency of the methods – specifically, we calculate the cost spent on medical examination to recruit *K* eligible participant.

Experiment

Dataset and Pre-Processing

To simulate the process of recruiting clinical trial participants, we used the open-source *Synthea Patient Generator* [2] software to generate synthetic patients and synthetic electronic healthcare record dataset. In our experiment, the simulated cohort includes 7,889 and their 846,073 condition values in 136 different LOINC codes – 21.14% of the condition values are not yet observed. To simplify the experiment, we removed all 20 nominal features except for sex (binarized to 1 or 0).

Experiment Setup

Given the above cohort, we set up a mocked clinical trial and corresponding eligibility criteria to simulate the screening process, and the corresponding medical examination costs are estimated by searching the internet. We first selected the patients who have the above condition values observed, so that we can take them as the ground truth to evaluate our method. Secondly, to simulate the unobserved values, we randomly erased 25% of condition values related to eligibility criteria to empty values. Apart from the two CF methods, we also set up a baseline method, which simply conducts all correspondent medical examinations to participant candidates for unobserved values.

Experimental Results

Listed in Table 1 are the evaluation results of the baseline and two CF techniques.

Table 1 – Number of patients conducted examinations in a simulated clinical trial recruitment screening setting

Method	Cost on medical examination to recruit K eligible participants ($)			Number of participant candidates who conducted medical examinations		
	K=50	K=100	K=200	K=50	K=100	K=200
Baseline	3382.0	6587.0	13357.0	208	391	780
UBSM	78.0	310.0	3031.0	21	55	209
SVD	9.0	204.0	2856.0	5	55	207

Conclusions

In this paper, we adopt the collaborative filtering technique to resolve the unobserved patient condition values in screening clinical trial participants and improve the cost efficiency of the medical examinations. The proposed method also provides a framework to reach clinical trial participant candidates in a proactive way, which is promising to reduce the time cost in the screening and recruitment stage.

References

[1] S.T. Frantz, Why are clinical costs so high?, *Nature Reviews Drug Discovery*. **2** (2003) 851–851. doi:10.1038/nrd1250.

[2] A.M. McDonald, R.C. Knight, M.K. Campbell, V.A. Entwistle, A.M. Grant, J.A. Cook, D.R. Elbourne, D. Francis, J. Garcia, I. Roberts, and C. Snowdon, What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies, *Trials*. **7** (2006). doi:10.1186/1745-6215-7-9.

[3] J.S. Breese, D. Heckerman, and C. Kadie, Empirical Analysis of Predictive Algorithms for Collaborative Filtering, *ArXiv:1301.7363 [Cs]*. (2013). http://arxiv.org/abs/1301.7363 (accessed November 25, 2018).

[4] G.H. Golub, and C. Reinsch, Singular value decomposition and least squares solutions, *Numerische Mathematik*. **14** (1970) 403–420. doi:10.1007/BF02163027.

[5] J. Walonoski, M. Kramer, J. Nichols, A. Quina, C. Moesel, D. Hall, C. Duffett, K. Dube, T. Gallagher, and S. McLachlan, Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record, *Journal of the American Medical Informatics Association*. **25** (2018) 230–238. doi:10.1093/jamia/ocx079.

Address for correspondence

Bibo Hao: IBM Research – China
Building 28, ZPark, Beijing, 100193, China
haobibo@cn.ibm.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190490

Using FAIR Metadata for Secondary Use of Administrative Claims Data

Christian Haux, Petra Knaup

Institute for Medical Biometry and Informatics, University of Heidelberg, Heidelberg, Germany

Abstract

The FAIR principles require the reporting of rich metadata. However, when researchers use data for secondary use from external data owners, the FAIR principles require a different implementation as if the researchers would describe their own data. In this paper, we specify how FAIR metadata can be implemented for secondary data analyses and provide a suggestion for relevant metadata.

Keywords:
Metadata, Data Collection, Data Curation

Introduction

The FAIR (Findable, Accessible, Interoperable and Reusable) principles [1] aim to achieve high reusability of data that were used or produced by researchers. By providing metadata, data become interpretable and easier to be found and used by other researchers [2]. Therefore, some funding agencies require to implement the FAIR principles in data management plans, to ensure that data from publicly funded research projects can be used further by other researchers [3].

Because the FAIR principles itself do not provide or require a specific implementation that defines the FAIRness of data, several metrics and implementations exist [2,4]. For example, researchers can use the GO-FAIR metadata to assess the FAIRness of their datasets [5]

The FAIR principles and metadata can be well applied to data sets that are produced by researches and that can be made available by researchers. However, if researchers use data from external data owners for secondary analyses, they often do not have data sovereignty. This applies, for example, when administrative claims data are used by researchers for secondary analyses. The datasets that are provided to researchers, further called analysis datasets, are often generated for a specific research purpose. Therefore, the FAIR principles require a different implementation and should be used to describe the data that were used for analysis, how to acquire the data and under which license the data can be used [1,6].

In this paper, we analyze how the FAIR metadata can be implemented when researchers use administrative claims data from external data owners to conduct secondary analysis. We further provide suggestions for metadata that are needed when administrative claims data are used by researchers for secondary use.

Methods

The FAIR principles [1] and the GO-FAIR metadata layers [5] as well as the "REporting of studies Conducted using Observational Routinely-collected health Data" (RECORD) [7] and "Strengthening the Reporting of Observational Studies in Epidemiology" (STROBE) [8] statements as existing guidelines for using routinely collected data were analyzed.

The requirements for the GO-FAIR metadata layers were compared with the results of a qualitative content analysis of data usage agreements from a secondary analysis of routinely collected administrative claims data [9]. It was determined, which FAIR metadata can be provided by researchers that used claims data for secondary analysis. and recommendations for metadata were developed.

Results

The results of the qualitative content analysis revealed that data owners of administrative claims data put different conditions on the release of their data for research. For reasons of data minimization, data owners generate specific analysis datasets for researchers instead of grant access to their complete database. The analysis datasets only contain the attributes and entities that are needed for the research question of the researchers. Further, the agreements often stated that the analysis datasets have to be deleted after the end of the analysis by the researchers.

Therefore, it must be distinguished whether the analysis datasets or the data source of the data owner is described with FAIR metadata. Whereas the description of the analysis datasets provides metadata to published analyses, The description of the data source should provide general metadata about the data of the data owner and how other researches can access it.

Our analysis resulted in the following concept for the GO-FAIR metadata to describe both the analysis datasets and the data source:

(i) The **Data Repository Metadata** should be used to describe data provenance of the data source from that the analysis data sets were generated from. Data provenance includes a "description of the origins of a piece of data and the process by which it arrived in a database" [10]. For the secondary use of routine data, this means to create an audit trail of the data that includes the point of data entry at the data consumer, data extraction and processing, and research output. To assess the representativeness of claims data, data owners should share descriptions of the composition of the policyholders and available ranges. Descriptive statistics should be used to provide general information about the data. If data from multiple data owners are used, such a description should be provided for each data source.

(ii) If the data repository is organized in groups, such information should be provided for each group as **Catalogue Metadata**.

(iii) The **Dataset Metadata** describes the analysis datasets that were provided by a data owner. Researchers should provide a description of the datasets they received from a data owner. The FAIR principles further require that the data are assigned a globally unique and persistent identifier to make them findable and citable. However, the qualitative content analysis showed that analysis datasets often have to be deleted by the researchers after the intended use. In some cases, the data owner retains the analysis datasets. Therefore, it might be problematic to apply a persistent identifier to the analysis datasets. Instead of assigning the identifier to the analysis datasets, it can be assigned to the data source in general.

(iv) The **Distribution Metadata** describes how data were made available for researchers.

Researchers should provide a description of how they accessed the analysis data sets and which license was associated with the data. For example, it can be described if the analysis data sets were transmitted to the researchers or if the data access was only possible in the premises of the data owner.

Because licensing for secondary use of routine data is often regulated individually, it can differ between data owners and usage scenarios. For this reason, secondary users can not provide general information about licensing. They rather should provide a description of the license under which they got data provided by the data owner. Qualitative content analysis can be used to compare usage requirements from multiple data owners.

(v) The **Data Record Metadata** layer contains information about the structure and the content of the data set. Therefore, data owners should provide data descriptions or codebooks that contain information about the columns, attributes, and variables of the data sets. If data linkage across multiple analysis data sets is planned, it should be analyzed which attributes in one data set the map to corresponding attributes in other data sets. A matching of attributes in the data sets can be achieved with a spreadsheet-based harmonization table [11]. This approach can also be used to reveal data conflicts in the data.

Privacy regulations require that data owners remove sensitive information to protect the privacy of individuals before data are shared with researchers. This can cause the problem, that in research projects, that involve the usage of data from multiple data owners, each data owner may apply different sanitization methods. Therefore, a description of how each data owner sanitized the data should be provided, as well as an assessment of the data quality and if data linkage across multiple data sets was possible, despite the sanitization.

Conclusions

Making claims data permanently available is difficult because licenses with data owners often require that the data are destroyed after the end of an analysis.

In this paper, we specified how FAIR principles and GO-FAIR metadata can be implemented for secondary data analyses and provided a suggestion for relevant metadata. The suggestions can be used to describe how researchers accessed and used data for their analysis. The implementation can be used by researchers that want to provide information about which data they used for their studies and how other researchers can access these data.

References

[1] M.D. Wilkinson, M. Dumontier, I. Aalbersberg, et al., The FAIR guiding principles for scientific data management and stewardship, *Sci. Data.* **3** (2016).

[2] L.O.B. da Silva Santos, M.D. Wilkinson, A. Kuzniar, et al., *FAIR data points supporting big data interoperability*, in: M. Zelm, G. Doumeingts, J.P. Mendonça (Eds.), Enterprise Interoperability in the Digitized and Networked Factory of the Future, ISTE, London, 2016.

[3] European Commission Directorate-General for Research & Innovation, *Guidelines on FAIR data management in Horizon 2020*, [cited 2018 Oct 16]. Available from: URL: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

[4] M.D. Wilkinson, S.-A. Sansone, E. Schultes et al., A design framework and exemplar metrics for FAIRness, *Sci. Data.* **5** (2018).

[5] GO FAIR, *GO FAIR metadata*, [cited 2018 Nov 2]. Available from: URL: https://www.go-fair.org/technology/go-fair-metadata/.

[6] A. Burgun, E. Bernal-Delgado, W. Kuchinke et al., Health data for public health: towards new ways of combining data sources to support research efforts in Europe, *Yearb. Med. Inform.* **26** (2017) 235–240.

[7] E.I. Benchimol, L. Smeeth, A. Guttmann, et al., The reporting of studies conducted using observational routinely-collected health data (record) statement, *PLoS Medicine.* **12** (2015).

[8] E. von Elm, D.G. Altman, M. Egger, et al., The strengthening the reporting of observational studies in epidemiology (strobe) statement: Guidelines for reporting observational studies, *Int. J. Surg*. **12** (2014) 1495–1499.

[9] C. Haux, M. Ganzinger, F. Gabel, et al., Qualitative content analysis of data usage agreements for data aggregation in a European wide oral health care project, in: E. Grill, M. Müller, U. Mansmann (Eds.), Health – exploring complexity: an interdisciplinary systems approach HEC2016, Springer Science+Business Media, Dordrecht, 2016: p. S75.

[10] P. Buneman, S. Khanna, and W.C. Tan, Why and where: A characterization of data provenance, *Database Theory - ICDT 2001, 8th International Conference, London, UK, January 4-6, 2001, Proceedings*. **1973** (2001) 316–330.

[11] C. Haux, O. Kalmus, A.-L. Trescher, et al., Detecting and resolving data conflicts when using international claims data for research, *Stud. Health. Technol. Inform*. **247** (2018) 1–5.

Address for correspondence

Christian Haux, Institute for Medical Biometry and Informatics, Heidelberg University, Im Neuenheimer Feld 130.3, 69120 Heidelberg, Germany; Christian.Haux@med.uni-heidelberg.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190491

Exploring the Discrepancies in Actual and Perceived Benefits of Dietary Supplements Among Obese Patients

Zhe He[a], Laura A. Barrett[a], Rubina Rizvi[b], Seyedeh Neelufar Payrovnaziri[a], Rui Zhang[b]

[a] School of Information, Florida State University, Tallahassee, Florida, USA,
[b] Institute for Health Informatics and Department of Pharmaceutical Care & Health Systems, University of Minnesota, Minneapolis, Minnesota, USA

Abstract

Dietary supplements (DSs) have gained increased popularity for weight loss due to its availability without prescription, relatively low price, and ease of use. Consumers with limited health literacy may not adequately know the benefits and risks associated with DSs. In this project, we found a knowledge gap between reported benefits of major DSs by adults with obesity in the National Health and Nutrition Examination Survey 2003-2014 and those reported in existing DS knowledge databases.

Keywords:
Obesity; Dietary Supplements; Nutrition Surveys

Introduction

Obesity, a complex chronic medical condition having multifactorial etiology, is on the rise not only in the US but also around the globe [1]. In the US, the prevalence of obesity among the 18 years or older population has increased from 33.7% in 2007-2008 to 39.6% in 2015-2016; whereas, the prevalence of severe obesity in adults has increased from 5.7% in 2007-2008 to 7.7% in 2015-2016 [2]. Many patients who are overweight and obese consider using dietary supplements (DSs) for weight loss. Currently, there are a number of available options out there, either to simply maintain a healthy weight or to actually treat people who are overweight/obese and are at high risk of weight-related comorbidities; however, the vast majority of treatment around lifestyle changes often fail due to noncompliance resulting from various factors e.g., poor motivation, lack of time, gaps in knowledge/awareness, and lack of strong and prolonged commitment to observe the actual results [3]. On the other hand, most of the pharmacological and surgical procedures are associated with a substantial health risk at a high cost [4].

The popularity of complementary alternative medicine, especially the use of dietary supplements for weight management, has gained much popularity. Aside from the health benefits resulting from weight loss, there are various other reasons (several of them being misconceptions) for people to turn to DSs for losing weight and/or maintaining a healthy weight, such as DSs being a natural product, effective, fast acting, associcated with minimal side effects, easily available, and at relatively low price. Often consumers swtich to DSs in frustration resulting from failures in previous weight loss attempts following strict diet and exercise regimens.

Interestingly. earlier studies have revealed that use of DSs is more commonly preventive with an aim to maintain and improve overall health, rather than being therapeutic to treat obesity. In this project, we aim to gain a better understanding of the use of DS products among obese people as well as their perceived benefits of these products. We used the combined National Health and Nutrition Examination Survey (NHANES) data from 2003-2014 to answer two research questions (RQ): RQ1: What are the perceived benefits of DS usage among obese patients? RQ2: Is there a knowledge gap between their perceived benefits and those reported in existing DS knowledge bases?

Methods

NHANES is a continuous cross-sectional health survey conducted by the National Center for Health Statistics [5]. It evaluates a stratified multistage probability sample of the non-institutionalized US population. We first extracted the demographic, examination, and dietary data from NHANES for survey years 2003 – 2014 (6 survey cycles). To strengthen the analytical power of the study, survey data from multiple survey cycles were combined for the following analyses. Inclusion criteria for the cohort include: (1) BMI $\geq$ 30 kg/m^2, and (2) age $\geq$ 18. This left us with a cohort of 11,959 participants. DS use was pulled for this cohort. Total and individual DS use was available for all survey cycles although detailed data was inconsistent for years 2003-2004 and 2007-2008. These inconsistencies caused minor issues with data processing but not with the data validity. This data was used for the demographic data. DSs used for specific reasons were grouped into types based on product information for the analysis.

Data Analysis

We first created a profile of obese adults in the cohort with respect to gender, age, race, and household income. Using the 2007-2014 data with the needed variables, we assessed the major perceived benefits of the DSs used by obese adults in the cohort, stratified by specific DS type. In our previous study [6], we learned that the Natural Medicine Comprehensive Database (NMCD) [7] is the most comprehensive resource, providing DS information that is reliable, clinically relevant, and evidence-based. Hence, in this study, we compared the reported benefits of obese adults in NHANES with NMCD aiming to identify the knowledge gap.

Results

Out of 11,959 survey participants, 5,421 (45.3%) self-reported taking at least one DS. The maximum number of DS used by an individual was 20. Survey participants were taking a total of 8,057 DSs , out of which they took 5,591 (69.4%) DSs on their personal will, and they took the remaining 2,466 (30.6%) DSs on their clinicians' advice. We looked at individual reasons for DS use between the years 2007-2014. Participants were given a list of top five reasons for DSs intake where they could choose 1 or more options (Table 1). We also looked at DS use reasons '*For Weight Loss*' and '*To Maintain Blood Sugar/Diabetes*'. These reasons were matched to specific DS type. Overall, this same group (2007-2014) used 6,929 DSs.

Table 1– Reason for DS use matched to type of DS

Reason (Total Responses)	Highest Type	# of responses for this DS type	% of total responses for this reason
General Overall Health[a] (5,313)	MVMM[c]	1,677	41.0%
Bone and Joint Health[b] (1,569)	Calcium / Bone/ Joint	834	57.7%
To Supplement Diet/Food Not Enough (941)	MVMM	462	49.1%
Heart Health/Cholesterol (787)	Omega-3	410	52.1%
To Get More Energy (737)	MVMM	316	42.9%
For Weight Loss (184)	Botanical[d]	55	29.9%
To Maintain Blood Sugar/Diabetes (125)	Botanical[d]	40	32.0%

a: Includes: To prevent health problems (813), to improve my overall health (2170), to maintain health/to stay healthy (1734), and to prevent colds/boost immune system (596). 4090 total responses once duplicative response were removed.

b: Includes: For healthy joints/arthritis and for bone (522) health/build strong bones/osteoporosis (1047). 1446 total responses once duplicative responses were removed.

c: Multivitamins/multiminerals

d: DS classified as a botanical if it is part of a plant, tree, shrub, herb, etc.

In addition, we investigated if the information provided in the existing knowledge base aligns with the reported use of a particular DS. We found consistency between the reported use of a particular DS and its use/effectiveness as provided in the existing knowledge base (NMCD) for improving certain conditions, e.g. general overall health, bone/joint health, heart health/cholesterol, and as a dietary supplement. In fact, we found useful information about the primary use of a DS in addition to its other common uses, e.g. use of calcium for bone and joint health other than to improve general health. For the remaining three conditions, i.e., getting energy, losing weight, and maintaining blood sugar, we found that consumers were taking DSs indiscriminately without sufficient, current, and scientific knowledge on how a particular DS impacts the human body, e.g. the use of MVMM and Vitamin B-complexes for weight loss and/or maintaining blood sugars among diabetic patients, rather than the intended role of a diet supplement in people with restricted diets.

Discussion and Conclusions

In this study, we used the NHANES data to assess the use and perceived benefits of dietary supplements among obese adults. Demographics clearly play a role in DS use. We found that females were more likely than males to use DSs. With respect to age, the older the respondent was, the more likely he or she used at least one DS. These results conclude that there is lack of adequate knowledge about specific DS use and their resulting benefits, esepcially among consumers who are taking DSs for obesity and/or related conditions. Even though they consider obesity as a health concern, they are not using DSs for actual weight loss. This includes the use of MVMM and Vitamin B complexes for losing weight and/or maintaining blood sugar among diabetic patients rather than their actual dietary supplment role for people with restricted diets. We found that consumers were taking DSs indiscriminately without sufficient, current, and scientific knowledge on how a particular DS actually impacts the human body e.g., use of calcium for heart health, despite a considerable number of risks [8]. Overall, we found NMCD to be reasonable in finding the relevant information for most ingredients/products despite few challenges. This study pinpoints the gaps of the perceptions of dietary supplements among the general public, thereby informing informaticians of the opportunities for developing informatics tools to disseminate the "accurate and factual" knowledge to the consumers.

Acknowledgments

This study was partially supported by the National Center for Complementary & Integrative Health Award (#R01AT009457) (PI: Zhang).

References

[1] C. Pozza and A.M. Isidori, What's Behind the Obesity Epidemic, in: Imaging in Bariatric Surgery, Springer, 2018, pp. 1-8.

[2] C.M. Hales, C.D. Fryar, et al., Trends in Obesity and Severe Obesity Prevalence in US Youth and Adults by Sex and Age, 2007-2008 to 2015-2016, JAMA 319 (2018), 1723-1725.

[3] E. Burgess, P. Hassmen, et al., Determinants of adherence to lifestyle intervention in adults with obesity: a systematic review, Clin Obes 7 (2017), 123-135.

[4] K. Fujioka, Current and emerging medications for overweight or obesity in people with comorbidities, Diabetes Obes Metab 17 (2015), 1021-1032.

[5] CDC, National Health and Nutrition Examination Survey Data: U S Department of Health and Human Services, Centers for Disease Control and Prevention.

[6] R.F. Rizvi, T.J. Adam, et al., Comparing Existing Resources to Represent Dietary Supplements, AMIA Jt Summits Transl Sci Proc 2017 (2018), 207-216.

[7] Natural Medicine Comprehensive Database (NMCD), https://naturalmedicines.therapeuticresearch.com [accessed November 6, 2018]

[8] X. Wang, H. Chen, et al., Dietary calcium intake and mortality risk from cardiovascular disease and all causes: a meta-analysis of prospective cohort studies, BMC Med 12 (2014), 158.

Address for correspondence

Zhe He, PhD. School of Information, 142 Collegiate Loop, Tallahassee, Florida, USA 32306-2100. Email: zhe@fsu.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190492

Patient-Pivoted Automated Trial Eligibility Pipeline: The First of Three Phases in a Modular Architecture

Paul M. Heider[a], Stéphane M. Meystre[a]

[a] *Biomedical Informatics Center, Medical University of South Carolina, Charleston, South Carolina, USA*

Abstract

Automated extraction of patient trial eligibility for clinical research studies can increase enrollment at a decreased time and money cost. We have developed a modular trial eligibility pipeline including patient-batched processing and an internal webservice backed by a uimaFIT pipeline as part of a multi-phase approach to include note-batched processing, the ability to query trials matching patients or patients matching trials, and an external alignment engine to connect patients to trials.

Keywords:

Eligibility Determination, Natural Language Processing, Machine Learning

Introduction

The success or failure of clinical research studies depend heavily on patient enrollment. Achieving sufficient levels of enrollment is an expensive endeavor in terms of both time and money [1]. Unfortunately, patients are commonly not enrolled in relevant trials because of their doctor's lack of awareness rather than the patient failing to meet eligibility criteria [2]. Natural Language Processing (NLP) and machine learning (ML) can be used to automatically extract relevant evidence from a patient's electronic health record (EHR) regarding their eligibility to participate in a clinical research study. Importantly, NLP and ML can glean these details from the structured fields (easily accessible through common database queries) and the unstructured clinical notes, which often lay untouched. Automated eligibility criteria extraction has been shown to significantly decrease enrollment time [3].

Building off our prior experiences developing an eligibility criteria extraction engine for breast cancer trials [4] and competing in the 2018 n2c2 Shared Task on Cohort Selection [5,6], we have found a gap in the discussion and documentation of trial eligibility approaches, especially with respect to balance between component reuse vs. customization to accommodate an ever-changing list of trials. We are in the first phase of three to evaluate and distribute a modular architecture for extracting trial eligibility criteria from EHRs.

Modularity comes in various forms and abstraction levels in the architecture. First, within the clinical research domain, new trials are constantly introduced and old trials are phased out. As such, the architecture needs to accommodate a formalism for describing criteria and their relationship to trials to allow trials to be easily added and removed from the monitoring cycle without heavy re-engineering of the core application.

Second, the constant turnover of trials requires that the extraction engine be able to accommodate new concepts. As such, there must exist modularity in the engine itself to allow for the easy extension of the core application.

Third, treating the patient as an individual is the ultimate goal of care but documentation of that care occurs in more quantal units. As such, the architecture needs to be able to integrate together extractions from individual notes (i.e., these quantal units) and structured data into a single picture of eligibility.

We propose that, just as automated trial eligibility surveillance can decrease enrollment time and costs over fully manual efforts, trial eligibility surveillance systems balancing a stable core application against the three needs described above decrease development time and costs over fully bespoke applications. For the first phase, we have focused on developing towards the first two needs, as described below.

Methods

Figure 1 shows the high-level architecture of our initial phase implementation. At the heart of our architecture is an Apache UIMA pipeline [7]. UIMA (Unstructured Information Management Architecture) is designed to be a highly modular system for processing documents. With uimaFIT, specific modules can be activated for a document on-the-fly based on general program arguments or properties of the document itself. This two-fold modularity allows weaving together off-the-shelf components developed as part of other clinical data-oriented information extraction tools (e.g., from cTAKES [8]) with custom modules. While pre-existing tools like cTAKES are a good stepping-off point, such generalized systems are missing some required algorithms and concept extractors. We can easily add, remove, or upgrade concept extraction modules to match the evidential needs of new trials, as per our second need above. Overall, the system extracts eligibility criteria from clinical notes and aligns those criteria with trials for optimal pairing of patient to trial. (Details can be seen in Figures 2 and 3, below.)

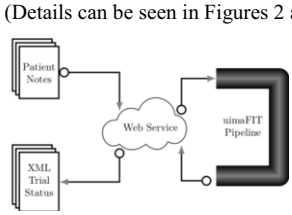


Figure 1 – High-Level Architecture of a Service-Based Pipeline Treating Patients as the Primary Pivot

The most significant deviation from a standard UIMA pipeline is our use of the OMOP Common Data Model (CDM) [9] as our underlying data model for eligibility criteria. The two implications of this formalism are found in Figure 2. First, we

developed a module to standardize all other concept representations into a OMOP CDM representation.

Second, we developed eligibility criteria aggregator modules for each trial which programmatically filter documents just as OHDSI's ATLAS tool allows users to define cohorts. The parallels between these implementations is intended to eventually allow for migration of definitions between the two. Quarantining the eligibility determination logic for each trial into its own module allows us to easily add and remove modules to match the flow of trials, as per our first need above.

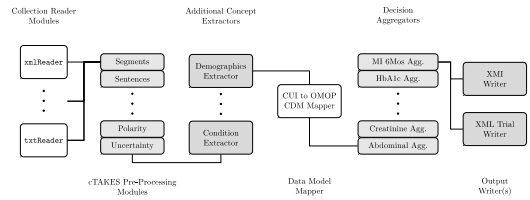


Figure 2 – Six Conceptual Classes of Pipeline Modules in Use

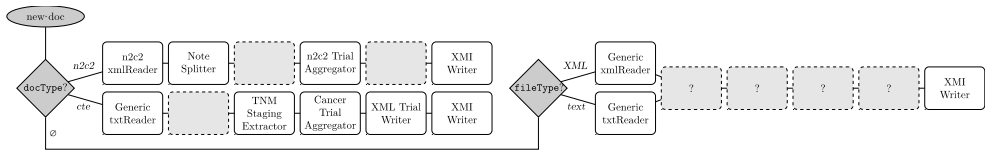


Figure 3 – Configurable paths through uimaFIT. Light Gray modules are not used by default but can be parametrically called.

A webservice sits in front of our UIMA NLP application to allow users to batch process patient records. As an early and partial implementation of our system, we currently only support treating a document as a self-contained representation of a patient. A patient's eligibility is determined based on the contents of a single document rather than on the accumulation of evidence over time.

Results

A simple webservice was developed in Java using the Spring Java framework. It accepts a single document or a batch of documents such that each document contains the entire collection of notes associated with a given patient.

General guidance on processing documents can be passed to the engine via the parameters of *document-type* and *annotators-list*. The former is used when the document fits squarely into a predefined limited set of known documents types (e.g., 'cte' type documents are processed as per the cancer trial eligibility pipeline we developed prior to this work). The latter allows the user to specify additional modules to run on the documents beyond the standard core pipeline. Figure 3 depicts a simplified view into several common flows through the pipeline.

Due to the uncertain nature of how long processing takes, the webservice does not hang while waiting to return results. The user must re-query the webservice with a provided batch-key (which uniquely identifies a batch) to get the break-down of eligibility by trial and criteria per patient in a JSON file.

Discussion

The next phase of development will be to accumulate extracted information about a patient across notes and structured data and improving the pipeline's flexibility. Webservice query parameters cannot yet fully control the system. Finally, trial eligibility criteria are programmatically determined. Ideally, this alignment between patient information and trial would be done externally and be less rigidly defined (e.g., through a spreadsheet, which we have experimented with through OHDSI's ATLAS, or using machine learning algorithms).

We have consolidated the development wisdom from separate applications built for similar tasks within the domain of eligibility surveillance. The shared components will reduce development and upkeep costs and will help us better clarify our abstract representations of trial criteria, as experimented with using OHDSI's ATLAS tool.

Acknowledgements

Our work is supported in part by pilot research funding, Hollings Cancer Center's Cancer Center Support Grant P30 CA138313, by NIH/NCATS 5UL1TR001450-03, and by the SmartState endowment.

References

[1] N.S. Sung, W.F. Crowley, Jr., M. Genel, et al., Central challenges facing the national clincial research enterprise, *JAMA* **289** (2003) 1278-1287.

[2] C.P. Somkin, L. Ackerson, G. Husson, et al., Effect of medical oncologists' attitudes on accrual to clinical trials in a community setting, *J Oncol Pract* **9** (2013) e275-e283.

[3] L. Penberthy, R. Brown, F. Puma, and B. Dahman, Automated matching software for clinical trials eligibility: Measuring efficiency and flexibility, *Contemp Clin Trials* **31** (2010) 207-217.

[4] S.M. Meystre, P.M. Heider, Y. Kim, et al., Automatic trial eligibility surveillance: Pilot study focused on breast cancer, In: *AMIA 2019 Informatics Summit*, San Francisco, CA, USA, (forthcoming).

[5] N2C2: National NLP Clinical Challenges, (n.d.), https://n2c2.dbmi.hms.harvard.edu/ (accessed November 19, 2018).

[6] P.M. Heider, Y. Kim, A.K. AlAbdulsalam, et al., Hybrid approaches for automated clinical trial cohort selection, In: 2018 N2C2 Shared-Task and Workshop, San Francisco, CA, USA, 2018.

[7] D. Ferrucci and A. Lally, UIMA: An architectural approach to unstructured information processing in the corporate research environment, *Nat Lang Eng* **10** (2004) 327-348.

[8] G. Savova, J. Masanz, P. Ogren, et al., Mayo clinic clinical text analysis and knowleddge extraction system (cTAKES): Architecture, component evaluation and applications, *JAMIA* **17** (2010) 507-513.

[9] G. Hripcsak, J.D. Duke, N.H. Shah, et al., Observational health data sciences and informatics (OHDSI): Opportunities for observational researchers, *Stud Health Tehnol Inform* **216** (2015) 574-578.

Address for Correspondence

Paul Heider at heiderp@musc.edu.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190493

Technology in the Determination of People Health Level: Design of a Computational Tool

Sara Herrero Jaén[ab], Marta Fernández Batalla[ac], Alexandra González Aguña[ad], Adriana Cercas Duque[ae], José Mª Santamaría García[ac], Sergio Martínez Botija[f], Niurka Vialart Vidal[g], Sylvia Claudine Ramírez Sánchez[g], Daniel Flavio Condor Camara[g]

[a] Research Group MISKC, University of Alcalá (Spain). [b] Hospital Universitary Severo Ochoa (SERMAS), Spain. [c]Centro de Salud Meco (SERMAS), [d]Hospital Universitary Príncipe de Asturias (SERMAS), [e] Hospital Universitary La Zarzuela (SANITAS) [f] Lapicero Blanco [g] International Network of Latin American and Caribbean Informatics Nursing

Abstract

The health concept has evolved throughout history. The people health level is determined by the perception that each individual has of it. It is a dynamic process over time, so the variations can be see from one moment to another. In this way, knowing the health of the patients you care for will facilitate decision-making in the treatment of care. To know the level of health of the people, a technological tool is presented that calculates the people health level through the Health Variables and Nursing Outcomes Classification (NOC) labels.

Keywords:

Health Care Technology, Population Health, Health

Introduction

The present study starts from the following research aim: Design an application for calculate the people health level based on the health variables.

Today in Spain, the so-called "Health Indicators" are used to establish the health level of a population. These indicators can be found in the health indicators document “Evolution of the indicators of the state of health in Spain and its magnitude in the context of the European Union” [1]. However, no tools have been found that determine the level of health of people, understanding that health goes beyond the absence of the disease, as proposed by WHO [2].

The objective of the research presented in this article is to design a computational tool that determines the level of health of a person at a given time. For this, a study has been carried out in which the logical formalization of the WHO Health concept takes place. This formalization allows to represent all those potential health states of a person [3].

Methods

The present study used deductive type methodology and was developed in four phases:

1. Extraction and representation of knowledge.
2. Design and development of the prototype tool.
3. Specification of Soft-ware Engineering Requirements through the Standard “IEEE Recommended Practice for Software Requirements Specification ANSI/IEEE 830-1998”.
4. Verification of the tool by group of experts of the MISKC Research Group, Alcalá University.

Results

Definition of Health Variables

The first result obtained derives from the extraction and description of the variables that constitute the concept of Health proposed by the WHO. This set of 11 variables has been called health variables.

- Physical Functioning: Functioning of the apparatuses and corporative systems of the person. It has a reversible character.
- Mental Functioning: In order to determine the mental functioning, the state of consciousness, spacetime orientation, behavior and language will be valued. It is reversible.
- Social Functioning: To determine the social functioning of a person will be assessed the ability to communicate and interact with other people. It is reversible.
- Confort Status: The welfare of the person, their tranquility, personal security, adaptation to the environment will be valued.
- Material Resources: Set of material goods needed to live [4].
- Time Resource: Temporary availability to carry out your care [4].
- Sings Presence: Sign measurable and valuable by the healthcare professional.
- Symptoms Presence: Manifestations of each person.
- Physical Condition: Sequel of irreversible character in the functioning of the body.
- Mental Condition: Sequel of irreversible character in mental functioning.
- Social Condition: Sequel of irreversible character in social functioning.

Tool Design

The computational tool will collect information about the health status of the person through NOC indicators. For its development, each one of the 11 variables of health were correlated with a label of the Taxonomy of NOC Result Criteria (Nursing Outcomes Classification) [5].

Health Variables	NOC Related
Physical Functioning	Personal Health Status (2006)
Mental Functioning	Personal Health Status (2006)
Social Functioning	Personal Health Status (2006)
Confort Status	Estado de comodidad (2008)
Material Resources	Health Belief: Resources Perception (1703)
Time Resource	Health Belief: Resources Perception (1703)
Signs Presence	Knowledge: Disease Process(1803)
Symptoms Presence	Knowledge: Disease Process(1803)
Physical Condition	Personal Health Status (2006)
Mental Condition	Personal Health Status (2006)
Social Condition	Personal Health Status (2006)

Figure 1 – Relationship between Health Variables and NOC Related

In this way, when relating the variables of health with NOC labels, they are related to their scale of measurement, being said scales those used to determine the people health level.

IEEE 830 Requirements Specification Interface

The application (app) is intended for all healthcare professionals. Then, this is the last of the results, the Specification of Software Engineering Requirements through the Standard "IEEE Recommended Practice for Software Requirements Specification ANSI/IEEE 830-1998", through the developed interface. The start screen of the app is shown in Figure 2.

Figure 2–Principal APP Interface

Once the user has accessed the app, an information screen appears in which clicking on each one of the health variables will show its definition. Finally, Figure 3 shows what the health calculation tool would look like, in which health variables are identified along with the Likert-type scale based on the NOC taxonomy.

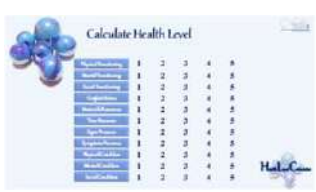


Figure 3 – Health Level Calculator

Discussion

7 out of 10 Spaniards consider that their state of health is good[6]. These data are based on the results obtained after the National Health Survey, in which the section of the health status module asks the population of the type: "Would you say that your health has been very good, good, regular, bad, very bad?", "Do you have any chronic or long-term illness or health problem?" or "What type of problem is the cause of its difficulty in carrying out the activities that people usually do?" [7]. However, no tools have been found that determine the level of health of people who go beyond the impact that health problems have on the individual.

Conclusions

The analysis of the data obtained will draw a graph where one can observe how the level of health of each participant is modified throughout a certain process. Determining the level of health of a given population facilitates the establishment of assessment criteria and health management, allowing prediction and prioritization of different health care strategies.

Acknowledgements

The present study was born within the MISKC University Research Group (Management about Information and Standard Knowledge of Care) of the University of Alcalá, in relation to the International Network of Computer Nursing (RIEI).

References

[1] Ministry of Health, Consumption and Social Welfare, Health indicators 2017. Evolution of state indicators of health in Spain and its magnitude in the context of the European Union ,(2017). https://www.mscbs.gob.es/estadEstudios/estadisticas/inforRecopilaciones/indicadoresSalud.htm.

[2] World Health Organitation, World Health Organitation. (2018). Available: http://www.who.int/suggestions/faq/es/.

[3] S. Herrero Jaén, Formalization of the concept of Health through logic: impact of formal language in Health Sciences, *ENE revista de Enfermería,* vol. 10, nº 2, (2016).

[4] M. Fernández Batalla, The person in the continuum of care: Formalization of the Basic Care Variables involved in the health trajectory, Alcalá de Henares: Doctoral Thesis, University of Alcalá, 2018.

[5] S. Moorhead, M. Johnson, M. Maas y E. Swanson, Nursing Outcomes Classification (NOC), 5ª ed., Elsevier, 2013.

[6] Ministry of Health, Consumption and Social Welfare, .Health in data. 2018. https://www.mscbs.gob.es/estadEstudios/sanidadDatos/home.htm.

[7] Ministry of health, consumption and social welfare, National Health Survey 2017. Adults Questionnaire. 2017. https://www.mscbs.gob.es/estadEstudios/estadisticas/encuestaNacional/encuesta2017.htm.

Address for correspondence

Sara Herrero Jaén. Research Group MISKC, University of Alcalá, Alcalá de Henares, Madrid (Spain).

Email: sara_herrerojaen@yahoo.es

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190494

Network-Based Prediction of Major Adverse Cardiac Events in Acute Coronary Syndromes from Imbalanced EMR Data

Pengwei Hu[a], Eryu Xia[a], Shochun Li[a], Xin Du[b], Changsheng Ma[b], Jianzeng Dong[b,c], Keith C.C. Chan[d]

[a] *IBM Research, Beijing, China*
[b] *Department of Cardiology, Beijing Anzhen Hospital, Capital Medical University, National Clinical Research Center for Cardiovascular Diseases, Beijing, China*
[c] *Department of Cardiology, The First Affiliated Hospital, Zhengzhou University, Zhengzhou, Henan, China*
[d] *Department of Computing, The Hong Kong Polytechnic University, Hong Kong*

Abstract

The low proportion and the rapid evolvement of major adverse cardiac events (MACE) present challenges for predicting MACE by machine learning models. In this paper, we propose a method to predict MACE from large-scale imbalanced EMR data by using a network-based one-class classifier. It only used the reliably known MACE samples to establish the hyperspherical model. Experiments show that our model outperforms the state-of-the-art models.

Keywords:
Acute coronary syndrome, machine learning, algorithms

Introduction

Acute coronary syndrome (ACS) is a group of diseases, resulting from a sudden decrease of blood flow in the coronary arteries and cardiac tissues leading to cardiac insufficiency or death. Given the fatal impact of MACE on ACS patients, researchers have consciously collected a large number of resident admit notes at the early stage of hospitalization, making the risk prediction of MACE a resoluble data analysis problem. In this state, two main methods are developing around this problem. The more traditional MACE prediction method is to use the risk score model. For example, TIMI [1], GRACE [2] and other tools introduce the factor score model into pre-selected risk factors to realize MACE risk prediction. Such approaches have a few drawbacks, they only apply to small sample evaluations, and they have strict requirements on data integrity, that make them difficult to deploy effectively in various real environments. A recent class of approaches advocates the use of supervised learning algorithms to directly model large-scale EMR data, and models can predict new cases of ACS patients in different environments [2]. In particular MACE, as a rare event, brings label imbalance to the prediction work. For example, only 3.5% of ACS patients in the data set we collected have MACE. However, existing machine learning models that have been improved to address imbalances still have several limitations that cannot be avoided for low level performance, resampling training data, and the false negatives of negative samples in the actual case data. Therefore, we present a Network-based Support Vector Data Description, called N-SVDD, to learn a similarity network of ACS patients as a new network representation and training of diagnosed MACE samples by Support Vector Data Description.

Methods

Cohort Construction

The multicenter, retrospective cohort study was conducted in 38 urban and rural hospitals in three provincial-level regions of China. Adult inpatients who were diagnosed with ACS at the time of death or discharge were recorded in the data set. We tooke an intermittent acquisition mode to record ACS patients corresponding to hospitals of different scales. A total of 26,986 selected patient samples were included in this study. Overall, there were 955 MACE records, only 3.5% of the total sample. The patient data were collected from the information system of each hospital and reviewed by investigators with professional knowledge and experience in the process of summarizing, to establish a clinical archive set. Our previous study [3] has assessed the association between all available patient variables and MACE. On this basis, we screened out 41 variables as train features; including age, demographics, diabetes mellitus, personal history, living habits, ACS type, Killip class, etc. as Figure 1 shows. The missing values were re-estimated by multiple imputations, the predictive mean matching was used for continuous variables. We performed one-hot encoding on categorical variables with more than two classes. Also, to standardize the continuous values, we removed the mean and scaled the unit variance.

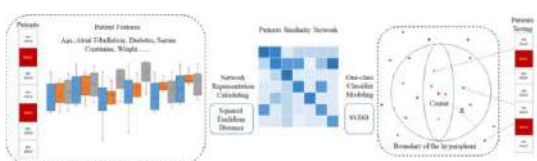


Figure 1–Flow diagram of N-SVDD.

N-SVDD (Network-SVDD)

The procedure of the proposed N-SVDD is shown in Figure 1. N-SVDD converts the patient features to the left of the figure into the similarity network representation in the middle of the figure, and then sends it to the one-class classification model training to the right of the figure, and predicts whether the new patient will suffer from MACE. Suppose that we have *n* patient samples and *m* feature variables, then we transform the original patient features into a similarity network based on these associations. A patient similarity network is represented as a

graph $G = \{V, E\}$ with a two-element tuple, where $V = \{u_1, u_2, \dots, u_n\}$ is a set of vertices representing n patients in a network, and the $E = \{e_{ij}\}$ is a set of edges. E is made up of the similarity of the paired elements in V. To achieve a similarity matrix of patients, we assess nodes similarity by the use of Squared Euclidean distance as it is good at evaluating the interrelationship between continuous data [4]. Support Vector Data Description (SVDD) [5-6] is a sophisticated one-class classification model whose goal is to determine whether a test sample should be classified as a specified category. Typically, patient variables of one class samples contain only feature information which is expressed by such samples. Instead, the sample trained by N-SVDD already contains many more representations by introducing the patient-patient similarity network. Suppose that there are k ACS patients labeled MACE, all the N-SVDD used samples should be represented as $X = [x_1, x_2, \dots, x_k]^T \in L^{n \times n}$, where $k < n$. Given the n network representations, such as for the training features, SVDD first maps the new input representation with a nonlinear transformation to a new feature space via a mapping function $\varphi(.)$. Then, it determines a sphere with a minimal volume containing all or most of the mapped patients in the new space. Let a center α and radius R express the hypersphere, the SVDD by using a minimum radius of hypersphere to describe the input samples, makes the minimum volume of hypersphere surround the patient samples as much as possible. When testing whether an ACS patient will have MACE or not, we calculate the distance from the test object z to the center of the hypersphere. When the distance is smaller than the radius R, the test object z is considered to have MACE.

Results

Due to that our ACS dataset has a high class imbalance, we use a 5-fold cross-validation for testing. Since the reliable positive samples are too small this may lead to no positive training samples if randomly dividing the dataset. We separate positive and negative samples and divide both into each fold equally. At present, many studies are subtly designed for predicting the MACE in ACS patients. We selected four comparison approaches; including two traditional models: SVM and Random Forest, and two models especially proposed for imbalanced data: SVM-BS and SVM-SMOTE [7]. The prediction task of unbalanced data cannot evaluate the prediction performance with common standards, and the evaluation metric must be able to weigh many true negative results. The area under the ROC curve (AUROC) has been proven to be able to avoid the influence of a large number of negative samples well and has become the evaluation standard for unbalanced tasks [8]. We compared the best AUROC scores of the N-SVDD with the above approaches.

Table 1– The average AUROC of proposed N-SVDD and comparison methods for the four imbalanced patient datasets.

	Data Set			
Method	**A (8098)**	**B 11980)**	**C (6890)**	**All (26968)**
N-SVDD	**0.713**	**0.739**	**0.751**	**0.760**
SVM-BS	0.659	0.671	0.662	0.673
SVM-SMOTE	0.701	0.705	0.701	0.703
SVM	0.568	0.575	0.562	0.577
Random Forest	0.692	0.697	0.677	0.695

Table 1 shows the AUROC scores of N-SVDD on the datasets of three provinces (A, B, C) and the consolidated dataset (All). As the results look like those shown above, the prediction accuracy of the proposed N-SVDD is superior compared to all methods on four datasets.

Conclusions

In this paper, we have proposed a novel machine learning approach called N-SVDD to predict MACE of ACS. It is a unique approach in the sense that it represents an attempt to make use of one-class network representations to predict the potential MACE of ACS patients. The experimental results show that N-SVDD can predict MACE accurately. The proposed model is also used in comparison with the state-of-the-art algorithms on four datasets, with approximately 7 to 13 % improvement of AUC values over the compared methods. The prediction accuracy of N-SVDD enables it to be used in clinical MACE prediction. N-SVDD can be a promising, intelligent tool for assisting in decision making related to the ACS treatment. It has good potential to improve the efficiency and effectiveness of automated clinical diagnosis.

Acknowledgements

This work was supported by grants from the National Science Foundation of China: 81530016 to Changsheng Ma, and the Beijing Municipal Commission of Science and Technology: D151100002215003 and D151100002215004 to Jianzeng Dong.

References

[1] E.M. Antman, M. Cohen, P.J.L.M. Bernink, C.H. McCabe, T. Horacek, G. Papuchis, B. Mautner, R. Corbalan, D. Radley, and E. Braunwald, The TIMI Risk Score for Unstable Angina/Non–ST Elevation MI: A Method for Prognostication and Therapeutic Decision Making, *JAMA* 284 (2000), 835-842.

[2] D. Hu, Z. Huang, T.M. Chan, W. Dong, X. Lu, and H. Duan Acute Coronary Syndrome Risk Prediction Based on GRACE Risk Score. *Stud Health Technol Inform*. 245 (2017), 398-402.

[3] Y. Li, X. Du, T. Chen, X. Li, G. Xie, and C. Ma, A Risk Model to Predict In-hospital Major Adverse Cardiac Events in Acute Coronary Syndromes Patients. *J Am Coll Cardiol* 71 (2018), A250.

[4] B. Wang, A.M. Mezlini, F. Demir, M. Fiume, Z. Tu, M. Brudno, B. Haibe-Kains, and A. Goldenberg, Similarity network fusion for aggregating data types on a genomic scale, *Nature Methods* 11 (2014), 333.

[5] D.M.J. Tax and R.P.W. Duin, Support Vector Data Description, Machine Learning 54 (2004), 45-66.

[6] H. Pengwei, C.C.C. Keith, and H. Yanxing, Predicting Drug-Target Interactions Based on Small Positive Samples, *Current Protein & Peptide Science* 19 (2018), 479-487.

[7] Z. Huang, T.-M. Chan, and W. Dong, MACE prediction of acute coronary syndrome via boosted resampling classification using electronic medical records, *Journal of Biomedical Informatics* 66 (2017), 161-170.

[8] J. Huang and C.X. Ling, Using AUC and Accuracy in Evaluating Learning Algorithms, *IEEE Trans. on Knowl. and Data Eng.* 17 (2005), 299-310.

Address for Correspondence

Pengwei Hu, IBM Research, Software Park, Haidian District, Beijing, China, 100193. E-mail: upwei@cn.ibm.com.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190495

Prediction of Synergistic Drug Combinations by Learning from Deep Representations of Multiple Networks

Pengwei Hu[a], Shochun Li[a], Zhaomeng Niu[b]

[a]*IBM Research, Beijing, China*
[b] *Rutgers Cancer Institute of New Jersey, New Brunswick, USA*

Abstract

Drug combination therapy can improve drug efficacy, reduce drug dosage, and overcome drug resistance. Many studies have focused on predicting synergistic drug combinations. However, existing methods fail to consider the heterogeneous characteristics of drugs fully, and it is difficult to identify effective drug combinations. Therefore, we propose a new integrated prediction model based on deep representations by integrating information from multiple domains to accurately and effectively predict drug combinations.

Keywords:

Heuristics; Deep Learning; Drug Therapy, Combination

Introduction

It is common for providers to prescribe a single drug that is precisely targeted to a diagnosis. However, in some cases, single drug use limits the efficacy of the drug and the condition's resistance to that drug [1]. One application of drug combination that covers a large number of people is the treatment of chronic diseases. In general patients with chronic diseases are older, with multiple diseases, and need two or more doses of drugs in daily treatment [2]. This type of synergistic drug combinations aims to cover all therapeutic targets and reduce the off-target effect.

Similarly, cancer treatment is also a difficult task that needs to be solved with polypharmacy. The identification of effective combinations of drugs is crucial to finding effective treatment for drug-resistant cancer. Recently, regulators have treated drug combinations as new, special drugs for the treatment of various cancers [3-4]. Although the combination of drugs is effective, it is not feasible for doctors to screen all possible drug combinations when prescribing. In addition, it is extremely time consuming to find effective compound drugs through biological experiments.

Many studies have shown that multiple biomedical variables are associated with patients' responses to drugs [5]. With the emergence of various related databases and the development of EMR systems, integrating various biomedical information to predict effective drug combinations will be an ideal clinical tool for medication treatment. Currently, machine learning methods based on drug features are the main methods used to predict synergistic drug combinations. One study has proposed a new prediction method for synergistic drug combinations, combining molecular and pharmacological information [6]. Further, more sophisticated algorithms have integrated a series of related properties such as drug-drug interactions, protein-protein interactions, and pathways together, and then use random forests with feature selection methods to predict synergistic drug combinations [7]. Ding et al. [8] introduced more related features of drugs, converted feature expressions into similarity networks, and integrated multiple networks to achieve the purpose of prediction with ensemble learning. Unfortunately, these methods directly concatenate the features, failed to keep the network structure information, and performed network embedding after concatenating.

To address these challanges, we propose a deep representation model to fuse multiple similarity networks for synergistic drug combination prediction called DSDC. The experimental results presented here indicate that our method outperformed the state-of-the-art approaches on the real-world dataset.

Methods

Figure 1 displays the the procedure of the proposed DSDC. We first constructed the similarity networks for various properties that could be calculated according to object features. Then we applied the network integration method to fuse the multiple similarity network to one network. Deep Auto-Encoder was used to conduct network embedding, and the classifier used the network representation to train the prediction model.

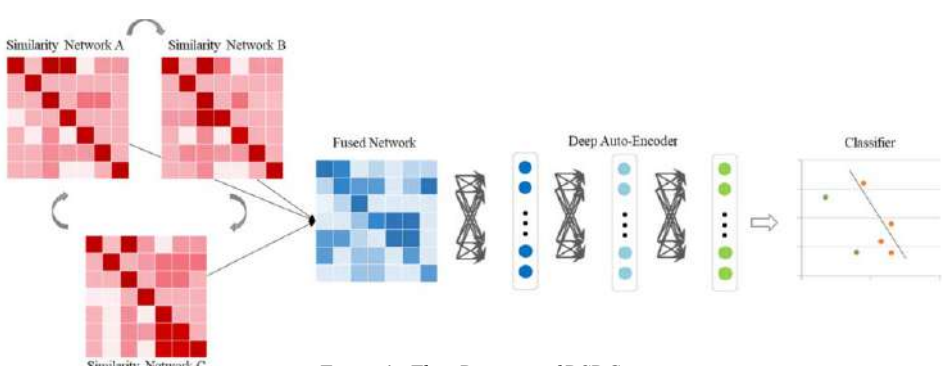


Figure 1– Flow Diagram of DSDC

Materials

Synergistic Drug Combination

The drug combination information used in this study was obtained from a widely used database, DCDB (http://www.cls.zju.edu.cn/dcdb). The validated drug combinations recorded (1363 records) by database records were collected from the FDA, clinicalite.gov, and PubMed literature. To obtain various types of measurements, we used the methods of Bai et al. [9] to collect transporters, targets, pathways, enzymes, and side effects of drugs from multiple data sources for the drug similarity network learning.

Similarity Network Learning

Suppose we have n samples and m measurements; this means there are m networks for one kind of drug properties. Let $G = \{V, E\}$ represent one kind of network, where $V = \{u_1, u_2, \dots, u_n\}$ is a set of n nodes representing all the drugs in the network, and the $E = \{e_{ij}\}$ is the edge set containing the edges between pairwise drugs, and their values represent how similar these nodes are. The edge weights are constructed by an $n \times n$ similarity matrix N.

Network Integration

Since the influence of each attribute on the target is different, a weighted combination strategy was used to integrate the similarity in order to adjust the deviation of each similarity and balance multiple networks. In order to achieve an integrated network from multiple networks, DSDC used m weight parameters to control the weight for each network. The integrated similarity network IS can be defined as follows: $IS = \frac{\sum_{v=1}^{m} W_v N_v}{\sum_{v=1}^{m} W_v}$, $v = 1,2,3, \dots, m$.

Deep Representation and Prediction

Then, we imported Stacked Auto-Encoder as an unsupervised learning model to get new deep representations from the integrated network. One effective classifier was used to predict whether a given drug-target interaction was positive or not according to the gold standard dataset. Finally, Support Vector Machine (SVM) was used as a classifier to train the new deep representations.

Results

The experimental results are presented in Table 1. We use 5-fold cross-validation, where each fold leaves out 20% of the positive and negative samples for testing. At present, many studies are proposed for predicting the synergistic drug combinations. We further considered several important studies of feature-based and network-based models which have been implemented in the prediction of synergistic drug combinations. We selected three studies to include two feature-based machine learning approaches (Ensemble model and SVM) and a method using the drug similarity network. When we compared the the DSDC score using these methods against the best AUROC scores, DSDC performed better.

Table 1– The average AUROC of proposed DSDC, NLLSS, Ensemble model, original SVM

Methods	NLLSS [10]	Ensemble model[9]	SVM	DSDC
AUROC	0.783	0.801	0.652	0.825

Conclusions

This study shows that the network integration method can effectively use the network representation of various properties to predict the potential SDC. In the context of clinical treatment, providers can quickly identify which drugs play an auxiliary role according to the prescribed primary drug. In the future, if we further combine the personal information of patients with the properties of drugs, we hope to provide more reliable synergistic drug combinations.

References

[1] M. Molokhia and A. Majeed, Current and future perspectives on the management of polypharmacy, *BMC Fam Pract* **18** (2017).

[2] K.K. Viktil, H.S. Blix, T.A. Moger, and A. Reikvam, Polypharmacy as commonly defined is an indicator of limited value in the assessment of drug-related problems, *Br J Clin Pharmacol* **63** (2007), 187-195.

[3] K. Han, E.E. Jeng, G.T. Hess, D.W. Morgens, A. Li, and M.C. Bassik, Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions, *Nat Biotechnol* 35 (2017), 463-474.

[4] C. Robert, B. Karaszewska, J. Schachter, P. Rutkowski, A. Mackiewicz, D. Stroiakovski, M. Lichinitser, R. Dummer, F. Grange, L. Mortier, V. Chiarion-Sileni, K. Drucis, I. Krajsova, A. Hauschild, P. Lorigan, P. Wolter, G.V. Long, K. Flaherty, P. Nathan, A. Ribas, A.-M. Martin, P. Sun, W. Crist, J. Legos, S.D. Rubin, S.M. Little, and D. Schadendorf, Improved Overall Survival in Melanoma with Combined Dabrafenib and Trametinib, *N Engl J Med* **372** (2015), 30-39.

[5] W. Reiter, P. Stieber, C. Reuter, D. Nagel, C. Cramer, H. Pahl, and A. Fateh-Moghadam, Prognostic value of preoperative serum levels of CEA, CA 19-9 and CA 72-4 in gastric carcinoma, *Anticancer Res* **17** (1997), 2903–2906.

[6] X.M. Zhao, M. Iskar, G. Zeller, M. Kuhn, V. van Noort, and P. Bork, Prediction of drug combinations by integrating molecular and pharmacological data, *PLoS Comput Biol* **7** (2011), 1-7.

[7] L. Chen, B.-Q. Li, M.-Y. Zheng, J. Zhang, K.-Y. Feng, and Y.-D. Cai, Prediction of effective drug combinations by chemical interaction, protein interaction and target enrichment of KEGG pathways, *Biomed Res Int* **2013** (2013), 1-10.

[8] P. Ding, R. Yin, J. Luo, and C.-K. Kwoh, Ensemble Prediction of Synergistic Drug Combinations Incorporating Biological, Chemical, Pharmacological, and Network Knowledge, *IEEE J Biomed Health Inform* **23** (2019), 1336–1345.

[9] L.-Y. df, H. Dai, Q. Xu, M. Junaid, S.-L. Peng, X. Zhu, Y. Xiong, and D.-Q. Wei, Prediction of Effective Drug Combinations by an Improved Naïve Bayesian Algorithm, *Int J Mol Sci* **19** (2018).

[10] X. Chen, B. Ren, M. Chen, Q. Wang, L. Zhang, and G. Yan, NLLSS: Predicting Synergistic Drug Combinations Based on Semi-supervised Learning, *PLoS Comput Biol* **12** (2016).

Address for correspondence

Corresponding author: Pengwei Hu

IBM Research, Software Park, Haidian District, Beijing, China, 100193.

hupwei@cn.ibm.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190496

Study on Patient Similarity Measurement Based on Electronic Medical Records

Yanqun Huang[a,b], Ni Wang[a,b], Honglei Liu[a,b], Hui Zhang[a], Xiaolu Fei[c], Lan Wei[c], Hui Chen[a,b]

[a] *School of Biomedical Engineering, Capital Medical University, Beijing, China*
[b] *Beijing Key Laboratory of Fundamental Research on Biomechanics in Clinical Application, Capital Medical University, Beijing, China*
[c] *Xuanwu Hospital, Capital Medical University, Beijing, China*

Abstract

A comprehensive scheme of patient similarity based on different types of patient features and the corresponding similarity measurements was proposed. Patient similarity was used in building a predictive model where training samples similar to the index patient were selected instead of randomly selected samples. The predictive models using the proposed patient similarity measurement outperformed those using Euclidean distance based similarity and those not using patient similarity.

Keywords:

Patient Similarity, Electronic Medical Records, Disease Prediction

Introduction

Recently, predicting the onset of a certain disease for individuals applying machine learning technology based on patient similarity became an important and prevalent research in personalized healthcare. Since patients similar with a certain patient may provide similar information on disease history and treatment history which may facilitate building effective predictive models, accurate identification of similar patients became a fundamental and crucial step [1], and there has been rapidly growing research of calculating and applying patient similarity for personalized analytics [1-4]. However, a series of questions have emerged with the development of patient similarity research, such as which feature(s) will be more reasonable for computing the patient similarity, and how to integrate the feature similarities into the patient similarity?

In this study, we conducted an exploratory study on the Patient Similarity Measurement (PSM) where the similarity at the patient level was calculated directly from four similarities at the feature level. The proposed PSM was then used to select similar patients for predicting the disease status of an index patient to assess the effectiveness of the proposed PSM.

Methods

Data set

Data used in this study was extracted from the Electronic Medical Records (EMR) system of a tertiary hospital in Beijing, China, from 2014 to 2016, covering 104,482 patients with over 140,000 hospital admissions. Each record included demographics (age and gender), up to 11 disease diagnoses, and laboratory tests. In order to build predictive models for the diabetes status for a certain person, 5,000 patients with diabetes were randomly selected from the overall 27,803 patients with diabetes in the dataset. Together with another 5,000 randomly selected patients without diabetes, they composed the study sample set for the subsequent study.

Calculation of the patient similarity

Four patient features included in this study included age, gender, disease diagnosis and laboratory test. Similarities at the feature level were first calculated according to the feature's data type. Age similarity S_A was set as the ratio of the minimum to the maximum of two age, and gender similarity S_G was defined as 1 if the two patients have the same gender and 0 otherwise. Laboratory test similarity S_L was defined as one minus the Euclidean distance between two sets of laboratory tests.

$$S_D(P_i, P_j) = 1 - \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

Disease diagnosis similarity S_D was computed by one minus the Jaccard distance (Eq. 1), where A and B were the disease diagnoses sets (represented by the ICD-10 codes) of patient P_i and P_j. Patient similarity was the weighted sum of those similarities at the feature level.

Feature combination scheme

We used the task-guided method to identify the optimal combination of four feature similarities. There were three main variants for the PSM calculation, including the code scheme for disease diagnosis which may impact the Jaccard similarity, the number of laboratory tests which may impact the Euclidean distance, and the weights assigned to each feature similarity.

- ICD-10 codes truncated by the first three or four letter and numbers (named ICD-T3 and ICD-T4, respectively) were used in this study. Additionally, Clinical Classification Software (CCS) code scheme was used in this study, where ICD-10 codes were grouped into 259 categories (encoded as 1–259) which may be more meaningful clinically.
- 77 most common and regular laboratory tests were selected as the full set of laboratory tests. Among them, 45 and 55 laboratory tests with significant difference between patients with and without diabetes ($P<0.01$ and $P<0.05$, respectively) were grouped into Lab subset A and B, respectively.
- Three typical weight allocation schemes for the feature similarities were chosen taking the clinical importance into account. Scheme Ⅰ hypothesized that all the four features supported the same contribution to the patient similarity, giving the equal weights of 0.25 for them. Schemes Ⅱ (weights of 0.1, 0.1, 0.4, and 0.4, respectively) and Ⅲ (weights of 0.1, 0.1, 0.6, and 0.2, respectively) paid more attention to the laboratory tests and disease diagnoses, which maybe more important for the clinical diagnoses.

In total, we designed 27 PSM calculation schemes, involving three disease code systems, three laboratory test sets, and three weight allocation schemes.

Application of the patient similarity measurement

PSM was used to build predictive models for diabetes status. The study samples were randomly split into three parts: i) an optimization set with 3,000 samples for determining the optimal patient similarity. They were further split randomly into a subset of 1,000 and 2,000 samples for seeking the optimal PSM; ii) a training set with 6,000 samples for building predictive models (logistic regression models in this study), and iii) a validation set with 1,000 samples for validating.

For a given patient in the validation set, k randomly selected training samples or k most similar training samples identified by the proposed PSM and the Euclidean distance based PSM were used to build the predictive model for that patient, respectively. The Euclidean distance based PSM was defined as one minus the Euclidean distance between two patients, represented by age, gender (1 for male and 0 for female), laboratory test (77 items in total), and disease was represented by an vector $X = \{x_d | x_d \in \{0,1\}, d = 1, \ldots, n\}$ where n was the number of all possible disease recorded in the dataset, and x_d=1 if the patient had the disease and 0 otherwise.

For logistic regression, log-likelihood loss function was used to measure the predictive error (Eq. 2). The predictive performance was evaluated by the area under the Receiver Operating Characteristic (ROC) curve (AUC), and the average of AUCs (AAUC) was used as the comprehensive index to compare the performance when k changed.

$$\mathcal{L}(\hat{y}, y) = -y * log\, \hat{y} - (1 - y) * \log(1 - \hat{y}) \qquad (2)$$

where y was the true value and ŷ was the predict value.

Results

The optimal scheme for calculating the patient similarity

Under all the possible combinations of disease code schemes (CCS, ICD-T3, and ICD-T4), laboratory test set (Lab subsets A, Lab subsets B and full set), and weights allocation schemes (Schemes , , and), the AAUC of predicting diabetes status for 1000 samples in the optimization set ranged from 0.7334 to 0.8521, and the optimal PSM scheme was the combination of ICD-T3 for disease diagnosis, Scheme for weights allocation and laboratory tests of Lab subset A (Figure 1).

Performance comparison

The predictive performance of the models built on k (from 60 to 1800 step by 60) training samples selected by different methods were shown in Figure 2. Models based on the proposed PSM could obtain higher performance with less training samples, and the corresponding AAUC was 0.86, which was higher than those (0.80 and 0.83, respectively) on training samples selected randomly and by Euclidean distance based PSM.

Discussion

There are increasing studies on patient similarity. The commonly used features for computing the similarity involved clinical data, diagnoses, medications, procedures, and lab results [1], and vital signs like heart rate and blood pressure [4]. Though the similarities were computed only with age, gender, laboratory test and disease diagnosis in this study, they were helpful in building and improving predictive models. For the specific predictive task in this study, the models built on the similar samples identified by the proposed PSM could get higher performance than those built on randomly selected samples. We also found the proposed PSM was most useful when there was little training data available.

In the proposed PSM, computation methods used for computing similarities varied from feature to feature, instead of a single measure like cosine similarity [4]. Our experimental results showed that predictive models built with the help of the proposed PSM had a higher performance than those built with a single Euclidean similarity, suggesting that a comprehensive scheme of similarity measurements was necessary.

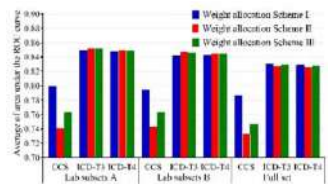


Figure 1 – Average areas under the ROC curve for all the 27 PSM calculation schemes. CCS: Clinical Classification Software; ICD-T3 or ICD-T4: ICD-10 code truncated by the first three or four letter and numbers.

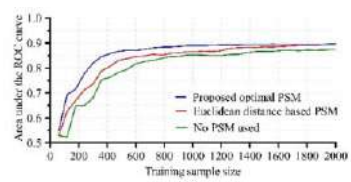


Figure 2 – Performance of predictive models built on training samples selected by using different methods.

However, other clinical features like bio signals, procedures, and diagnostic imaging were also important in similarity computation [1], which were not available in our current study. In addition, the proposed optimal task-guided PSM scheme may be limited in cross-domain applications.

Conclusions

This study proposed a patient similarity measurement scheme which the optimal combination of feature similarities was determined by experiments and the proposed patient similarity showed a higher performance than others. It could be expected that the proposed method will be helpful in EMR applications.

References

1. E. Parimbelli, S. Marini, L. Sacchi and R. Bellazzi, Patient similarity for precision medicine: A systematic review. *J Biomed Inform* **83** (2018), 87-96.
2. K. Ng, J. Sun, J. Hu and F. Wang, Personalized predictive modeling and risk factor identification using patient similarity. *Amia Summits Transl Sci Proc* (2015), 132-136.
3. J. Sun, F. Wang, J. Hu and S. Edabollahi, Supervised patient similarity measure of heterogeneous patient records. *Acm Sigkdd Explorations Newsletter* **14** (2012), 16-24.
4. L. Joon, D.M. Maslove, J.A. Dubin and E.S. Frank, Personalized mortality prediction driven by electronic medical data and a patient similarity metric. *Plos One* **10** (2018), e0127428.

Address for correspondence

Correspondence to: Hui Chen, PhD, School of Biomedical Engineering, Capital Medical University, No.10, Xitoutiao, YouAnMen, Fengtai District, Beijing, P.R.China, 100069. E-mail: chenhui@ccmu.edu.cn. Tel: 861083911545

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190497

Successful Implementation of Terminology Binding in Hong Kong Hospital Authority

Karissa Hung, Maggie Lau, Vicky Fung

Health Informatics Section, Hospital Authority, Hong Kong Special Administrative Region

Abstract

Terminology binding has been adopted in Hong Kong Hospital Authority to link terminology component to clinical information models. This linkage facilitates structured data capturing and streamlines clinical workflow. With information models and data representation standards in place, data interoperability and data integration can be maintained for seamless patient care delivery.

Keywords:
Health Information Exchange, Interoperability, Terminology

Introduction

In Hong Kong Hospital Authority (HA), clinical information beyond diagnoses and procedures can be captured by electronic Generic Clinical Documentation (GCD) forms with customized questions and answers. The HA constructs Information Architecture (IA) with "Form", "Entity" (form questions), and "Data Value" (form answers) under clinical information model. The GCD forms are easy tools which were designed to fit clinical workflow for organized data capturing and reviewing. It helps to capture comprehensive cross-domain-data on one screen within a few clicks that enables flexible documentation. Clinicians can synthesize information at a glance for fast and precise decision [1].

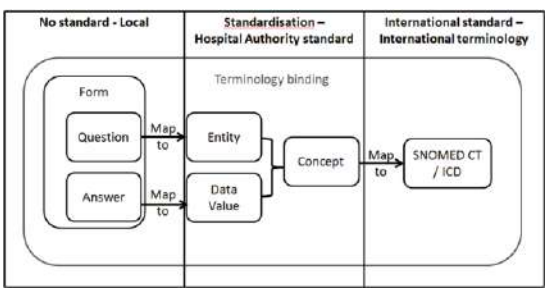


Figure 1– Illustration of terminology binding used in Hong Kong Hospital Authority

As advocated globally, HA has implemented standardization of clinical records using international standards to prepare for the data interoperability needs of the territory-wide electronic health record [2]. Thus, HA has adopted terminology binding for data standardization to facilitate clinical workflow by linking clinical information model and clinical terminology, and achieving semantic interoperability [3].

For the terminology component, HA uses Hong Kong Clinical Terminology Table (HKCTT), a medical dictionary developed for HK use, as a standard clinical terminology. It includes but not limited to diagnoses, procedures, substances, and pharmaceutical product concepts. Each concept was mapped to international terminologies, including SNOMED CT, International Statistical Classification of Diseases (ICD), etc.

Method

This paper demonstrates the initiation of applying terminology binding of diagnosis and procedure terms in HA corporate forms. Locally, we customized form questions and answers. Behind the scene, we implemented standardization by mapping the questions and answers to reusable and standardized "Entity" and "Data value" under IA (Figure 1).

With the application of terminology binding, the clinical terms captured by form questions and answers were mapped to structured HKCTT terms with international terminologies (see Table 1). Therefore, linkages were established between terminology component and information model artifact. This implementation achieved the real-time automation of clinical terminology binding from front-end to international standard without negative impact on clinical workflow. Furthermore, different questions and answers across the interface would generate the same concept as long as they mean the same.

Afterward, the form data were sent to other systems and healthcare organizations for patient care, for public health reporting including Hong Kong Department of Health then to the World Health Organisation (WHO), and for statistical analysis, resource planning, and research.

The screen in Figure 2 shows how structured procedure terms were mapped to clinical form. The forms, questions, and answers refer to the specified structured data mapping to international terminology.

Figure 2– Electrophysiology Study & Catheter Ablation Form

Result

The procedure capture rate from the mentioned form increased by 31.5% from 2017 to 2018 after terminology binding implementation in five months (see Table 2). Clinical

information was standardized with no negative impact on clinical workflow as reported by clinical users.

Table 1 – The journey from customized form question and answers to HKCTT concepts mapped to International Terminologies. Form name: EP Study & Catheter Ablation Form. Question / Entity description: Energy Source(s) / Energy source used in mapping and catheter ablation

Answer (Data Value)	HKCTT Concept	ICD9CM mapped by HKCTT Concept	SNOMED CT mapped by HKCTT Concept
Cryoablation	Transluminal cryoablation of heart lesion for arrhythmia	37.34 Excision or destruction of other lesion or tissue of heart, other approach	Is a = 233161001 Cryoablation operation for arrhythmia (procedure) Surgical approach = 261449002 Transluminal approach (qualifier value)
RFA	Radiofrequency ablation - heart lesion	37.34 Excision or destruction of other lesion or tissue of heart, other approach	Is a = 447954001 Radiofrequency ablation of lesion of heart (procedure)
RFA (irrigation)	Radiofrequency ablation of lesion of heart with saline irrigated electrode	37.34 Excision or destruction of other lesion or tissue of heart, other approach	Is a = 447954001 Radiofrequency ablation of lesion of heart (procedure), primitive

Despite the small sample size, standardization was achieved, and clinical documentation was facilitated by mapping structured HKCTT concepts to SNOMED CT from customized form questions and answers. Terminology binding facilitates clinical workflow, reinforces information models and data representation standards, and enhances data interoperability and data integration.

Clinical data, not limited to diagnoses and procedures data, could be captured in a faster way than inserting terms individually. In the binding model, the captured data were comprehensive since the data capturing became convenient. This implementation served as a prelude to clinical decision support by reviewing form data. With the improved clinical documentation, duplicated procedures on patients could be avoided, and the treatment given could be more precise. Moreover, structured and standardized terminologies could be captured, stored, and retrieved across forms with a user-friendly interface. Data sharing, retrieval, and analysis could be promoted. Furthermore, this linkage can contribute to clinical research and resource planning for better population health.

Table 2– Comparison of Procedure capture rate before and after implementation of terminology binding

	Procedure Reported Date (Year/Month)	The total number of the listed procedure reported	Average procedure captured per month
Before terminology binding implementation	2017-05	34	35.58
	2017-06	33	
	2017-07	39	
	2017-08	40	
	2017-09	50	
	2017-10	37	
	2017-11	31	
	2017-12	38	
	2018-01	27	
	2018-02	36	
	2018-03	35	
	2018-04	27	
After terminology binding implementation	2018-05	35	46.8 (↑31.52%)
	2018-06	39	
	2018-07	43	
	2018-08	64	
	2018-09	53	
	Total	330	

Conclusion

Terminology binding has been implemented in HA successfully that enabled standardized clinical data capturing and smoothened clinical workflow. Accurate data interpretation could be ensured no matter how data were being captured at front-end. Information models and data representation standards were maintained by linking the terminology component to our information model. Data interoperability and data integration were enhanced. This binding model helped to improve population health.

Reference

[1] J.L. Schnipper, J.A. Linder, M.B. Palchuk, J.S. Einbinder, Q. Li, A. Postilnik, and B. Middleton, "Smart Forms" in an Electronic Medical Record: documentation-based clinical decision support to improve disease management, *J Am Med Inform Assoc* **15** (2008), 513-523.

[2] European Commission. Interoperability & stadardisation connecting eHealth services. (2018).

[3] A. Moreno-Conde, D. Moner, W.D. Cruz, M.R. Santos, J.A. Maldonado, M. Robles, and D. Kalra, Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis, *J Am Med Inform Assoc* **22** (2015), 925-934

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190498

Extending Achilles Heel Data Quality Tool with New Rules Informed by Multi-Site Data Quality Comparison

Vojtech Huser[a], Xiaochun Li[b], Zuoyi Zhang[b], Sungjae Jung[c], Rae Woong Park[c], Juan Banda[d], Hanieh Razzaghi[e], Ajit Londhe[f], Karthik Natarajan[g]

[a] National Library of Medicine, MD, USA, [b] Indiana University, IN, USA, [c] Ajou University School of Medicine, Suwon, South Korea, [d] Georgia State University, GA, USA, [e] Children's Hospital of Philadelphia, PA, USA, [f] Janssen Research & Development, NJ, USA, [g] Columbia University, NY, USA

Abstract

Large healthcare datasets of Electronic Health Record data became indispensable in clinical research. Data quality in such datasets recently became a focus of many distributed research networks. Despite the fact that data quality is specific to a given research question, many existing data quality platform prove that general data quality assessment on dataset level (given a spectrum of research questions) is possible and highly requested by researchers. We present comparison of 12 datasets and extension of Achilles Heel data quality software tool with new rules and data characterization measures.

Keywords: *data quality, observational study*

Introduction

Data quality is an important pre-requisite for research on Electronic Health Record (EHR) data. In recent years, several efforts and tools emerged that perform data quality assessment (DQA).[1] Another important trend is that research is increasingly conducted using distributed research networks. Such networks often provide tools to their data partners that lower the barrier to join or participate within the network and help with data preparation or analysis execution.

The Achilles tool from the Observational Health Data Sciences and Informatics Consortium (OHDSI) is one such tool that performs data characterization and includes an Achilles Heel part that contains rules for checking data quality (DQ). The Achilles tool has been first deployed in October 2014 (version 1.0) with several updates (versions 1.1 to 1.6) during a period from 2014 to 2018. Since 2016, the web-based user interface part of Achilles was incorporated into the OHDSI Atlas tool, which is a new interface that integrates into one interface several previously developed OHDSI tools.

In developing the Achilles tool, the OHDSI consortium actively encourages researchers to submit requests for new data quality checks or insightful data visualizations that would extend the tool's utility. The Achilles' software repository receives numerous inputs (in a form of Github issues) that identify such new DQ measures or checks. In addition to this ongoing feedback, European EMIF research network conducted a formal survey of the tool that indicated the need for new features.

This study describes a set of extensions of the Achilles tool based on a comparison of data quality indicators of several healthcare datasets.

Methods

The study had two goals. The first goal was to compare data quality characteristics across datasets. Informed by this comparison, the second goal was to extend Achilles with new features and new data quality rules that would improve the assessment of data quality generated by the tool. This study includes a larger set of exported dataset metadata compared with a previous study done by our team, that only focused on Achilles Heel output messages.

The Achilles tool currently generates over 170 measures. However, many healthcare dataset administrators are not permitted to share such comprehensive set of dataset indicators. To be able to conduct our comparison, we designed a smaller set of measures generated by Achilles pre-computations that includes only measures that were deemed acceptable by the dataset administrators.

To maintain a data aggregation privacy-preserving principle for our comparison, our study used a small-cell count threshold of 11+ patients per aggregated count. Achilles tool allows suppressing pre-computations that result in small counts of patients (or small counts of providers, or healthcare events). This filtering is done either when Achilles pre-computations are executed, but if it was not done during the Achilles pre-computation phase, our methodology enforced it again during when site data extract generation. The R package for our study (called DataQuality) is open source and available on the Github platform at https://github.com/OHDSI/StudyProtocolSandbox/tree/master/DataQuality.Actual input data for the study consisted of the following: (1) subset of Achilles analyses converted to ratios (for example, ratio of persons with at least one visit by visit type); (2) Achilles derived measures (for example, percentage of unmapped source data concepts by domain) and (3) an approximate size of the dataset (for example, <10k, 10-99k,100k-1M, 1-5M,5-9M and >10M; exact size of populations is masked into a dataset size category). Sample input data (for a synthetic SynPuf OMOP dataset) is available at Github.

Each dataset was assigned a meaningless identifier to facilitate the comparison. The purpose for this dataset masking is the fact that data quality comparisons can lead to withdrawal of a data partner from a research consortium (or an analysis project) if a particular partner's dataset is identified as having low quality data. Masking was done to avoid this outcome and to focus on advancing the methodologies for DQA. For the same reason, neither a list of individual datasets is provided. We plan to

destroy individual site aggregated data used as input at 6 months after the publication of the study results. To protect the sites, only masked and isolated combined comparisons are reported in this article. Non-aggregated, single dataset DQ data are never posted publically.

Determination of goodness of fit or "data fitness" is highly dependent on the research question being asked. This phenomenon was described earlier and is sometimes referred to a task-dependence nature of DQA.[2] A dataset that only contains inpatient events and data may not be appropriate for general research questions (e.g., descriptive study of a course of a disease); however, it may be sufficient for a subset of research questions (e.g., inpatient-only research questions).

One can conclude that without knowing the specific research question context, any data quality assessment is impossible to pre-empt. This requirement for specifying research question context up-front makes development of general DQA tools almost impossible. However, existing DQA tools and efforts indicate that some general DQA rules indeed exist.

To partially overcome this problem ("data fitness for what?"), we assumed that the dataset being assessed represents lifetime record of general population and the tool should perform DQA for a wide range of possible research questions ("*general* data fitness for a wide *range of research questions*"). Once a general DQA analysis is done, a researcher with a specific research analysis can simply ignore DQA messages that do not apply to his/her context. (e.g., ignore messages about lack of eye doctor's visit and eye care data if data about vision care are not essential for his/her research question).

Results

A total of 12 datasets were compared in the study; however due to use of prior Achilles versions by some sites, comparison of some newly implemented measures are made on data from datasets that implemented at least Achilles version 1.4 at the time of our study data extraction.

Version 1.6 of Achilles contains a total of 44 data quality rules (also called data quality checks). A total of 12 rules are *model conformance rules* that check adherence to the CDM specification. For example, a model conformance rule may require that provider specialty column contains only concepts that are indeed specialties. The remaining 22 rules are *data quality rules* that check for data completeness, data plausibility or other data quality problems. Such rules can be considered model-independent and should be portable to other data models, such as Sentinel model or PCORNet. The pooled dataset of all Achilles Heel messages from all datasets consisted of 546 messages. Median number of Heel messages for a single dataset was 51 with a median of 25 for errors, 22 for warnings and 4 for notification. Poster will show evaluation of severity of each rule violation by computing median record counts for each rule. The second goal of our study was to add new functionality (either new DQ rules or new DQ measures to Achilles) based on availability of data about multiple CDM datasets. The results are divided into multiple sections according to the data domain of the new rule and will be included in the poster.

(1) **Empirical rules:** Comparing selected dataset parameters and computing 90th or 10th percentile and using them as benchmark thresholds.

(2) **Data density rules:** We considered data density at three levels (*concepts per person* as a number of distinct measurements per person (e.g., count of 2 measurements per person, such as cholesterol and hematocrit). This comparison aims at "data breadth"; *records per person* as total number of all measurement records per person (e.g., count of 8 tests, such as 3 LDL cholesterol and 5 hematocrit measurements). This comparison aims at "data depth"; records per visit as a data density measure on a visit level. Because visits with no measurements occur, the per visit ratio measure can be below 1. However, for a ratio looking at clinical notes (if in scope for the dataset), it may be reasonable to expect at least one note per visit.

(3) **Minimum-data patients:** For many research question, at least one data point in a given clinical data domain (such as medications) is required for any meaningful analysis. For example, for analyzing event prevalence, using a proper denominator and determining the size of the relevant population can significantly affect the reported measure. We determined empiric thresholds for existing Achilles DQA measures that count number of patients with at least one event in a clinical data domain. (e.g., patients with at least one visit, patients with at least 1 diagnosis and 1 medication.

(4) **Unmapped data:** OMOP CDM allows storage of data that is not fully semantically mapped to standard concepts (for example, drug exposure data may include data rows that have a value of 0 ('No matching concept') in drug_concept_id while the yet-to-be-mapped local code is stored in drug_source_value). We introduced measures computing unmapped data and threshold rules for several domains, such as Conditions, Procedures or Drug Exposure.

Discussion and Conclusion

Our current method for picking an empiric threshold is using a fixed threshold (e.g., 10th percentile). Future methodology revision may alter this approach for each considered DQ measure. Another limitation is our primary focus on OMOP CDM sites. Our extension to Achilles rule knowledge base, however, point to what data measures are required by each rule and whether a rule is terminology dependent. We compared data quality indicators across several datasets. We arrived at empirical values that could be used as thresholds for several DQA measures. The study resulted in several new data quality checks being added to Achilles.

We would like to thank Ritu Khare, Taha Abdul, Chris Knoll and Martijn Schuemie. VH work was supported by the Intramural Research Program of the National Institutes of Health (NIH)/ National Library of Medicine (NLM)/ Lister Hill National Center for Biomedical Communications (LHNCBC).

References

[1] H. Estiri and K. Stephens, DQe-v: A Database-Agnostic Framework for Exploring Variability in Electronic Health Record Data Across Time and Site Location. , eGEMs (Generating Evidence & Methods to improve patient outcomes) 1 (2017).

[2] N.G. Weiskopf and C. Weng, Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research, J Am Med Inform Assoc 20 (2013), 144-151.

Address for correspondence

vojtech.huser@nih.gov

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190499

Web-Based Visualization of MeSH-Based PubMed/MEDLINE Statistics

Maria I. Restrepo[a,b,*], Mary C. McGrath[a,b,*], Indra Neil Sarkar[a], Elizabeth S. Chen[a]

[a] *Center for Biomedical Informatics, Brown University, Providence, RI, USA*
[b] *Center for Computation & Visualization, Computing and Information Services, Brown University, Providence, RI, USA*
[*] *Equally contributing authors*

Abstract

Statistical analysis of Medical Subject Headings (MeSH) descriptors to improve biomedical literature search is an active research area. Existing tools have limited interactive visualizations that are accessible to researchers investigating how their hypotheses compare to trends in the research literature. We present a web application that computes and provides an interactive visualization of basic frequencies and co-occurrence statistics of MeSH descriptors associated with a PubMed query.

Keywords:

Data Mining, Data Display, Medical Subject Headings

Introduction

Biomedical literature mining is an active area of research for enhancing information retrieval and knowledge discovery. The exponential growth [1] of publications available in PubMed/MEDLINE in the last decades fueled the development of many biomedical literature mining systems. Lu has presented a comprehensive survey [2] reviewing 28 such tools.

Of the 28 entries in Lu's review, only 14 continue to be available. A public list [4] of web applications for querying, mining, and visualizing results from PubMed has been created and made available as a complement to this study. From this list, it is apparent that most available frameworks focus on querying, with only a few having visualization capabilities.

The importance of providing visualizations to help literature exploration and understanding was leveraged in more recent work. Zhang *et al*. proposed PubMedMiner [3], a software package developed to mine and visualize associations of MeSH descriptors. In this study, a tool is presented that builds on the foundations developed in the original PubMedMiner package[3]. We enhance the methods with statistical measures and visualizations. The developed web-based tool follows best practices to provide visualizations that are interactive and require no software installation. The functionality is demonstrated through queries to identify mental health comorbidities for suicide.

Methods

Figure 1 provides a summary of the workflow and technology stack for the web-based tool. A user can access the tool from a web browser to explore sample studies or submit a new query. The query is sent to the backend server using a Representational State Transfer (REST) Application Programming Interface (API). The backend checks if the query has been performed before; if so, cached statistics are returned. Otherwise, the statistics are computed and returned. The results are presented as interactive visualizations.

Statistical Methods

Two types of statistical methods are presented to the user: *(1) Univariate statistics:* Total counts for the 50 most common MeSH descriptors related to the input query are provided; and *(2) Covariate statistics:* Pairwise statistics are based on occurrence or co-occurrence matrices. The tool includes three types of covariate statistics:

1. *Pearson's Correlation Coefficient:* Pearson's correlation coefficient is computed to investigate the statistical relation between the random variables (MeSH terms). In particular, the coefficient measures the strength and direction of the linear relationship of the two variables.
2. *Pointwise Mutual Information (PMI)*: This measure quantifies the distance between the joint probability of the variables and their individual distributions. If the variables are independent, then PMI is equal to zero.
3. *Association Rules and Frequent Itemsets:* An association rule is defined as a relation of the form X $\rightarrow$ Y, where X is a set of MeSH and Y is one MeSH. A frequent itemset is defined as X$\cup$Y. We use an implementation that is a variation of the Apriori and the Eclat algorithms [5,6].

Visualization Methods

Interactive Bar Plot: Univariate statistics are presented using an interactive bar plot that displays counts and names on hover.

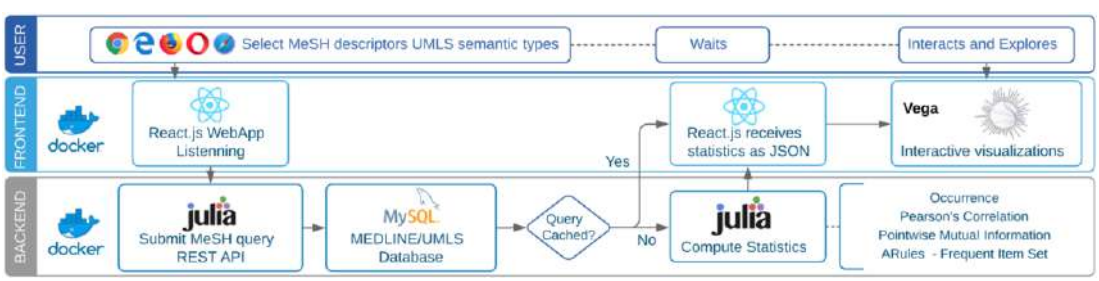


Figure 1 – Workflow and Technology Stack of the Web Application

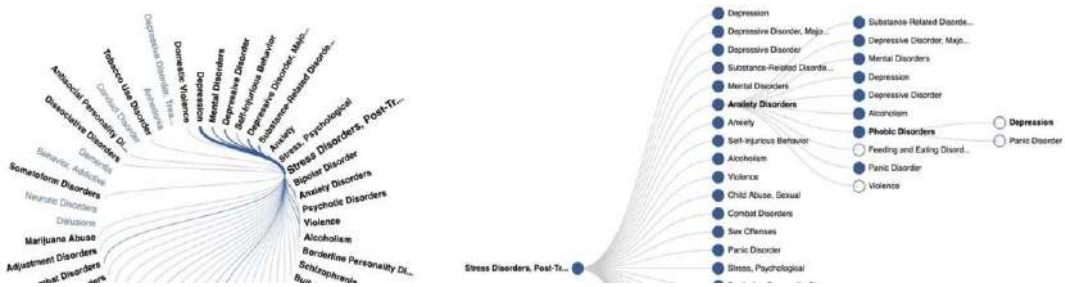


Figure 2 – Co-occurrence Chord Diagram and Expanded Tree of Frequent-Set Items for Case Study – MeSH: "Suicidal Ideation" and UMLS Semantic Type: Mental or Behavioral Dysfunction

Interactive Matrix View: Pearson's correlation coefficient and PMI are presented to the user as interactive matrices. The color ranges according to the strength and direction of the measure.

Co-Occurrence Chord Graph and Interactive Tree: The interactive chord graph displays co-occurrence as well as frequent itemsets in an intuitive and simple manner as shown in Figure 2. The conditional co-occurrence matrix is given in the interactive chord graph. On selection of a node, the adjacent tree displays the frequent itemsets. A button is provided to search PubMed/MEDLINE for these terms, turning the visualization into a novel way to query PubMed.

Technology Stack

Figure 1 illustrates the technology stack and how the technologies interact. Technologies used in this work, which were chosen to support maintenance and reproducibility, include: Julia [6], React.js [7], Vega [8], and Docker [9].

Results

Web App

The web app developed in this work is available at https://bcbi.brown.edu/pubmedminer. The results are presented within the context of a case study to demonstrate how the tool may support researchers and clinicians to explore relationships between concepts imputed from biomedical literature.

Case Study

Mental health (MH) conditions affect 1 in 5 people in the United States [10]. These conditions are among the factors contributing to suicide. Studies have identified specific MH conditions as risk factors; however, there is a need to better understand the interactions of MH comorbidities on suicidal thoughts and behaviors (STB) [11–13].

PubMed could be used to characterize and discover connections between risk factors such as MH comorbidities for STB. We explore the query for *"Suicidal Ideation"* filtered by the UMLS semantic type *"Mental or Behavioral Dysfunction"*.

Figure 2 shows the chord graph and expanded tree view corresponding to the frequent itemset {Stress Disorders, Post-Traumatic, Anxiety Disorders, Phobic Disorders, Depression} resulting from the query. To the best of our knowledge, this is the first time a chord diagram combined with an interactive tree is used to explore frequent itemsets.

Visualizations of the article counts of the 50 most frequent MeSH descriptors and Pearson's correlation and PMI matrices are available on the accompanying web site.

Conclusions

This study presents a web application that provides interactive visualizations of the univariate and covariate statistics of MeSH descriptors in biomedical literature. The technology stack used in this project is intended to be of high performance, composable, reproducible, open source, and be extendable.

Acknowledgements

This work was supported in part by U54GM115677 from National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] L. Hunter, and K.B. Cohen, Biomedical language processing: what's beyond PubMed?, *Mol. Cell*. **21** (2006) 589–594.
[2] Z. Lu, PubMed and beyond: a survey of web tools for searching biomedical literature, *Database*. (2011).
[3] Y. Zhang, I.N. Sarkar, and E.S. Chen, PubMedMiner: Mining and Visualizing MeSH-based Associations in PubMed, *AMIA Annu. Symp. Proc.* (2014) 1990–1999.
[4] M.I. Restrepo, M.C. McGrath, and E.S. Chen, bcbi/awesome-pubmed-webapps: v0.1.0, (2018). doi:10.5281/ZENODO.1493578.
[5] P. Stey, and M.I. Restrepo, bcbi/ARules.jl: v0.0.1, (2018). doi:10.5281/ZENODO.1485983.
[6] J. Bezanson, A. Edelman, S. Karpinski, and V.B. Shah, Julia: A Fresh Approach to Numerical Computing, *SIAM Rev.* **59** (2017) 65–98. doi:10.1137/141000671.
[7] React.js. https://reactjs.org/ (accessed November 19, 2018).
[8] Vega. https://vega.github.io/vega/ (accessed November 19, 2018).
[9] Docker. https://www.docker.com/ (accessed November 19, 2018).
[10] https://www.cdc.gov/mentalhealth/learn/index.htm, (n.d.).
[11] J. Fawcett, Diagnosis, Traits, States, and Comorbidity in Suicide: The Neurobiological Basis of Suicide, in: Neurobiol. Basis Suicide, 2012. doi:10.1002/pen.760140611.
[12] M.K. Nock, I. Hwang, N.A. Sampson, and R.C. Kessler, Mental disorders, comorbidity and suicidal behavior: Results from the national comorbidity survey replication, *Mol. Psychiatry*. (2010). doi:10.1038/mp.2009.29.
[13] Y. Veisani, F. Mohamadian, and A. Delpisheh, Prevalence and comorbidity of common mental disorders and associations with suicidal ideation in the adult population, *Epidemiol. Health*. (2017). doi:10.4178/epih.e2017031.

Address for correspondence

Elizabeth S. Chen, PhD

Email: liz_chen@brown.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190500

Development of Patient State Model to Overview Clinical Registry Database

Masamichi Ishii[a], Kengo Miyo[a], Takehiro Sugiyama[b], Mitsuru Ohsugi[b], Kohjiro Ueki[b]

[a] Healthcare Intelligence Technology Center, NCGM (National Center for Global Health and Medicine), Shinjuku-ku, Tokyo, Japan,
[b] Diabetes Research Center, NCGM, Shinjuku-ku, Tokyo, Japan

Abstract

By applying the Bayesian network method to the clinical registry database J-DREAMS (Japan Diabetes compREhensive database project based on an Advanced electronic Medical record System), we have developed a reference model that summarizes the exploration of the patient group's state and facilitates a bird's-eye view. This visualization method would help registry researchers to screen the registry database.

Keywords:

Data Aggregation, Bayesian , Registries

Introduction

Our institution NCGM (National Center for Global Health and Medicine) is promoting J-DREAMS (Japan Diabetes compREhensive database project based on an Advanced electronic Medical record System) [1], which is a clinical registry database project in collaboration with the Japan Diabetes Society, aiming to register between 100,000 and 200,000 patients. The development of artificial-intelligence systems using clinical data collected by J-DREAMS is expected.

The J-DREAMS database consists of 50 tables (over 400 columns) and has over 50,000 patients' records, making it difficult to get an intuitive grasp of the patient group's state for each particular topic. Previously, if one wanted to discover a particular patient group's status one was researching, they were required to apply basic statistics on each clinical question. It was necessary to repeat the statistical analysis according to the increasing number of clinical questions [5].

Accordingly, as a solution we propose to apply the Bayesian network method [2-4] to summarize the registry database [6-9]. We believe that this visualization method would help registry researchers to screen the entire database easily without the need for a prior understanding of the database schema and mutual relations between its tables.

Methods

We propose a patient group's state model, expressed in terms of Bayesian networks, as a method to utilize accumulated multi-institutional clinical data for general purpose effectively. The features of the model are as follows:

1. Extraction of observation items: Extract observation items from the columns of J-DREAMS registry tables (Figure 1, left side).
2. Bayesian network node definition: Define observation items as nodes in Bayesian networks.
3. DAG (Directed Acyclic Graph) definition: Define hierarchical relationships between nodes in terms of DAG (Figure 1, right side).

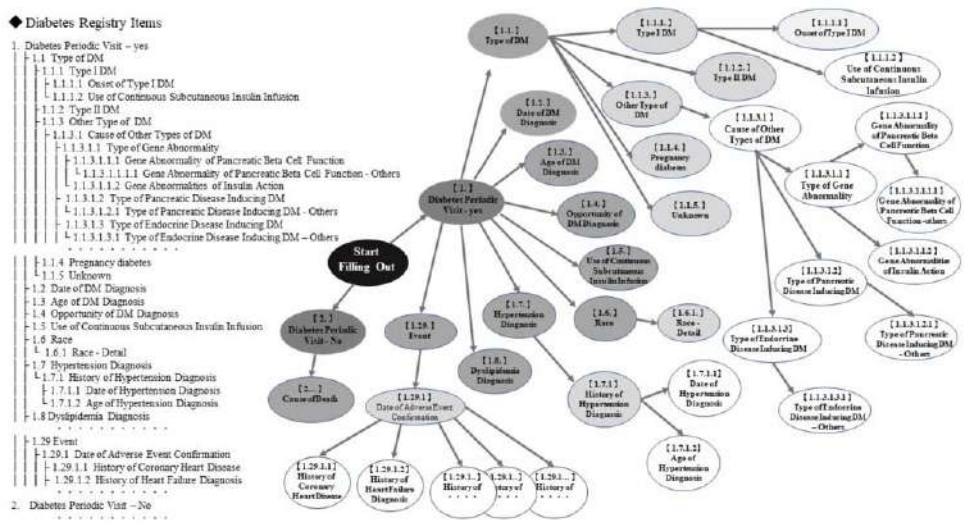


Figure 1. J-DREAMS observation items and Bayesian network DAG definition (example)

For example, the observation items of J-DREAMS are gradually detailed and they have up to nine-level hierarchy. In that case, define the direction of the arrow from the upper item node to the one-step lower one.

When calculating the posterior probability from the given conditional probability in Bayesian statistics, the direction of the arrow and the causal relationship between the nodes do not necessarily have to coincide, since it does not depend on the direction of the arrow.

4. Extraction of real data: Extract real data corresponding to the observation items from the clinical registry database.
5. Preparation of learning data: Prepare learning data from the extracted data by cleansing them (e.g., abnormal value judgment), calculating new feature quantity (phenotyping), and converting numerical data into categorical variables (discretization).

Link the lab test data that satisfies a specific time-series condition to each case record. For example, extract the latest result data of HbA1c (within the past one month), data after 90 days, data after 180 days, etc., when the case report date is taken as the reference date from the HbA1c of the timeline.

6. Modeling (Bayesian network generation): Calculation of conditional probability distribution
7. Evaluations by diabetes specialists: The generated model expresses the entire registry database.

If one gives conditional probabilities to the model and re-calculates the posterior probabilities, they could interpret a particular patient group's status. In this way, evaluations are carried out by diabetes specialists.

Results

A prototype of the patient group's state model is developed by using the Bayesian networks method. The relationships between the registry items are expressed with the DAG model, such as a parent-child relationship between two nodes. Accordingly, the entire clinical registry database can be visualized.

Comments from diabetes specialists:

- Compared to the previous basic descriptive statistics, the registry database schema can be visualized, and the registry data can be easily overviewed.
- Before analysis, it is possible to grasp the bias of collected data (e.g., missing values, abnormal values, etc.).
- By changing the conditional probability setting, we can search the registry database in an exploratory manner and screen the registry.

Discussion

An inference is possible by giving a prior probability distribution, executing the probability propagation method, and calculating the posterior probability.

Before considering the analysis axis for data analysis, it is possible for researchers to grasp the number of cases and trends for each search condition.

By facilitating database search, we have made it easy for experts to provide their awareness and feedback from their tacit knowledge.

Conclusions

We showed that a clinical registry database could be visualized by constructing a patient state model with Bayesian network architecture, and it could be helpful for constructing a clinical knowledge base.

Acknowledgments

This research was supported by Japan Agency for Medical Research and Development under Grant Number JP17lk1010010.

References

[1] R. Kominami, K. Miyo, T. Sugiyama, M. Ohsugi, and K. Uemi, Development of "Japan Diabetes compREhensive database project based on an Advanced electronic Medical record System" and various issues to unite data from multiple hospital, *36th Joint Conf Med Inform* (2016), 292-293.

[2] J. Pearl, Bayesian networks: a model of self-activated memory for evidential reasoning, *Proc Cog Sci Society* **7** (1985), 329-334.

[3] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Francisco, CA, 1988.

[4] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*, MIT Press, Cambridge, MA, 2009.

[5] W. Chen, B.K. Armstrong, R. Zheng, S. Zhang, X. Yu, and M. Clements, Cancer burden in China: a Bayesian approach, *BMC Cancer* **13** (2013), 458.

[6] M. Soto-Ferrari, D. Prieto, and G. Munene, A Bayesian network and heuristic approach for systematic characterization of radiotherapy receipt after breast-conservation, *BMC Med Inform Decis Mak* **17** (2017), 93.

[7] W. Yin, S. Garimalla, A. Moreno, M.R. Galinski, and M.P. Styczynski, A tree-like Bayesian structure learning algorithm for small-sample datasets from complex biological model systems, *BMC Sys Biol* **9** (2015), 49.

[8] P. Berchialla, C. Scarinzi, S. Snidero, and D. Gregori, Comparing models for quantitative risk assessment: an application to the European registry of foreign body injuries in children, *Stat Meth Med Research* **25** (2016), 1244-1259.

[9] J.A. Forsberg, J. Eberhardt, P.J. Boland, R. Wedin, and J.H. Healey, Estimating survival in patients with operable skeletal metastases: an application of a Bayesian belief network. *PLoS ONE* **6** (2011), e19956.

Address for correspondence

Masamichi Ishii

NCGM (National Center for Global Health and Medicine)

Address: 1-21-1 Toyama, Shinjuku-ku, Tokyo 162-8655, Japan

Email: masaishii@hosp.ncgm.go.jp

Phone: +81(3)3202-7181

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190501

Risk of Acute Myocardial Infarction in Patients with Rheumatic Arthritis: A National-Wide Population-Based Cohort Study

Md. Mohaimenul Islam[a,b], Taimina Nasrin Poly[a,b], Hsuan-Chia Yang[a,b], Yu-Chuan (Jack) Li[a,b]

[a] *Graduate Institute of Biomedical Informatics, Taipei Medical University, Taipei, Taiwan*
[b] *International Center for Health Information Technology (ICHIT), Taipei Medical University, Taipei, Taiwan*

Abstract

We performed a cohort study to quantify the association between rheumatic arthritis (RA) and acute myocardial infarction (AMI) risk. ICD-9 was used to identify AMI and RA patients, and the Cox proportional hazards model with adjusted confounding factors was used to quantify the risk. The overall risk of AMI for RA patients was an aHR of 1.05 (95% CI 1.01–1.09). We found RA was associated with an increased risk for AMI.

Keywords:

Acute Myocardial Infarction, Rheumatic Arthritis, Cohort Studies

Introduction

Rheumatic disease is characterized by chronic and abnormal inflammatory reactions of the immune system against body tissues which increases the chance of joint destruction and physical disability. Nowadays, the prevalence of Rheumatic disease has increased leading to higher substantial social costs like the increased risk of work-related disability. Contrarily, acute myocardial infarction (AMI) is one of the primary causes of frequent hospital admissions and mortality around the globe [1]. A patient with AMI and multiple diseases could affect both treatment and prognosis. Over the past several decades, acute myocardial infraction has become a significant public health problem worldwide.

Rheumatic arthritis (RA), in fact, has appeared as an independent risk factor for developing AMI. Several biological studies have shown that inflammatory mediators in RA patients are linked to prothrombotic factors and endothelial dysfunction which are associated with developing AMI. Moreover, the previous epidemiological studies have reported that rheumatic arthritis is associated with a higher risk of acute myocardial infarction.

However, the extent to which rheumatic arthritis may confer risk of acute myocardial infarction remains uncertain. We therefore performed a cohort study to quantify the magnitude of the association between rheumatic arthritis and acute myocardial infarction risk.

Method

This population-based cohort study was conducted within a cohort of patients registered with the National Health Insurance Bureau that contributes to the National Health Insurance Research and development Database (NHIRD). This is a large database which collects information on patient demographics (with encrypted patient identification numbers, birthdates, and sex), inpatients or outpatients claim data, and pharmacy records. The NHIRD database covers over 99% of the 23 million Taiwanese population who are registered with the National Health Insurance program [2,3]. Additionally, NHIRD records information on the medications prescribed, laboratory and diagnostic test data, dates of visits, lengths of hospitalization, and diagnoses based on the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). The NHIRD database, however, has been widely used for research to make valuable clinical decisions and health care policies.

In our current study, a 3-year dataset was retrieved and analyzed which included all the patients that visited the hospital between January 1, 2009 and December 31, 2011. All participants who were aged 20 years and above, and visited for rheumatic arthritis, were considered for inclusion in this study. However, the index date is defined as the date a primary diagnosis of rheumatic arthritis was made between January 1, 2009 to December 31, 2011. The diagnoses of rheumatic arthritis in outpatients and inpatients were also identified by validated ICD-9 Codes (International Classification of Disease, Clinical Modification, Ninth Revision [ICD-9-CM] codes 714.xxx).

Four participants who received no diagnosis of rheumatic disease were randomly frequency matched according to sex, age, and baseline year to each patient diagnosed with rheumatic disease. Patients who developed AMI (ICD-9-CM code 410) during the study period were also identified. The patients with a diagnosis of AMI before the index date were excluded from our study. The only two demographic variables measured were sex and age.

Additionally, to reduce confounding factors, we also evaluated different comorbidities associated with AMI. Baseline pre-existing diagnoses of cardiovascular disease (CVD) prior to the Rheumatic disease comorbidities included CVDs, which was defined as patients who were assigned to any one of the procedural codes for coronary artery bypass grafts (ICD-9-CM code 361.xx), coronary stents (ICD-9-CM code 36.xx), chronic obstructive pulmonary disease (COPD) (ICD-9-CM codes 490–492, 494, 496, A324), diabetes (ICD-9-CM codes 250), hypertension (HTN) (ICD-9-CM codes: 401–405), and hyperlipidemia (ICD-9-CM codes: 272).

We performed a Chi-square test to determine categorical variables for assessing the distribution differences between the two cohorts. We also drew the Kaplan-Meier method to estimate cumulative incidence curves of AMI for the cohorts and the log-rank test was performed to estimate the difference between the curves. Furthermore, the Cox proportional hazards regression model with adjusted potential confounding factors was used to estimate the hazard ratios (HRs) and confidence

intervals (CIs). Data management and analysis were performed using SAS, and a value of $P<0.05$ was considered significant.

Results

A total of 2512 patients were included in the final analysis; a total of 394 cases of Atrial Fibrilation were used in the analysis. The mean age of rheumatic patients was 72.93 and non-rheumatic patients was 50.64. The age range of rheumatic, and non-rheumatic patients was 20 to above 80. A significant difference exists in the distribution between RA patients and controls in gender and age (Table 1).

However, several comorbidities were more prevalent among RA patients than controls, including hypertension (69.11% vs. 57.53%; $P<0.001$), diabetes (10.83% vs. 12.34%; $P=.23$), hyperlipidemia (5.65% vs. 10.34%; $P<0.001$), CHF (2.23% vs. 2.0%), and COPD (1.27% vs. 1.12%; $P=0.72$). After adjusting the covariates, the overall risk of AMI for RA patients was an aHR of 1.05 (95% CI 1.01–1.09). Women had a lower AMI incidence with an aHR of 0.84 (95% CI 0.66–1.07).

Table 1: Demographic characteristics and comorbidity in patients with or without rheumatic disease

	Rheumatic (n=1256)	**Non-Rheumatic (n=1256)**	**p-value**
Gender	Male=916 (72.93)	Male=632 (50.64)	<0.001
Age			
20-29	16 (1.28)	77 (6.9)	
30-39	39 (3.12)	133(10.86)	
40-49	110 (8.79)	202 (16.49)	
50-59	295 (23.58)	305 (24.90)	
60-69	348 (27.82)	248 (20.24)	
70-79	309 (24.70)	176 (14.37)	
80++	134 (10.71)	84 (6.84)	
Diabetics	136 (10.83)	154 (12.37)	0.23
CHF	28 (2.23)	25 (2.0)	0.69
Hypertension	868 (69.11)	718 (57.53)	<0.001
COPD	16 (1.27)	14 (1.12)	0.72
Hyperlipidemia	71 (5.65)	129 (10.34)	<0.001

Conclusion

In this population-based study, we found that rheumatic disease patients were not associated with an increased risk for acute myocardial infarction though AMI is strongly associated with hypertension and COPD patients. Although it is an observational study from a large database, our study was unable to prove causality. All of the findings only relied on the EMR database and did not examine categories related to the severity of the disease with AMI. Furthermore, the findings could be varied due to some confounding factors we did not include such as medication history, other comorbidities, duration of RA, etc. Therefore, additional studies with more variables are required to establish whether rheumatic disease is actually responsible for developing AMI or not. Physicians should routinely screen RA patients to avoid unwanted consequence effects.

References

[1] P. Asaria, P. Elliott, M. Douglass, Z. Obermeyer, M. Soljak, A. Majeed, and M. Ezzati, Acute myocardial infarction hospital admissions and deaths in England: A national follow-back and follow-forward record-linkage study, *The Lancet Public Health* **2** (2007) e191-e201.

[2] C.C. Wu, Y.Y. Yu, H.C. Yang, P.A. Nguyen, T.N. Poly, M.M. Islam, U. Iqbal, H.A.A. Khan, Y.C. Wang, Y.T. Cheng, Y.C. Li, and W.S. Jian, Levothyroxine use and the risk of breast cancer: A nation-wide population-based case-control study, *Archives of Gynecology and Obstetrics* **298** (2018) 389-396.

[3] H.C. Yang, P.A.A. Nguyen, M. Islam, C.W. Huang, T.N. Poly, U. Iqbal, and Y.J. Li, Gout drugs use and risk of cancer: A case-control study, *Joint, Bone, Spine* **85** (2018) 747-753.

Address for Correspondence

Yu-Chuan (Jack) Li, MD, PhD, FACHI, FACMI, Graduate Institute of Biomedical Informatics, Taipei Medical University, Taiwan

Email: jaak88@gmail.com / jack@tmu.edu.tw

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190502

Analysis of Usage of Term Weighting Algorithm for Mapping Health Procedures into the Unified Terminology of Supplemental Health (TUSS)

Eluizio H. Saraiva Barretto[a], Diogo F. da Costa Patrao[b], Márcia Ito[a,c]

[a] *Instituto de Pesquisa Tecnológica de São Paulo, São Paulo, SP, Brazil,*
[b] *A. C. Camargo Cancer Center, São Paulo, SP, Brazil,*
[c] *Informar Saude-Grupo Bem, São Paulo, SP, Brazil*

Abstract

TUSS is a Brazilian health procedure standard used by the supplementary health providers. Currently, there is no available mapping between TUSS and other standards. In this paper, we analyze performance of two term weighting algorithms when classifying TUSS procedure description. The TF-IDF classified 99% of chapters, 89% of groups, and 33% of subgroups; Doc2Vec classified 65%, 43%, and 33%, respectively, showing that those algorithms can support creation of an accurate mapping between those procedure standards.

Keywords:

Health Information Interoperability, Current Procedural Terminology, Natural Language Processing

Introduction

Healthcare interoperability is defined as enabling access to patient information independently of the system or location that holds the record. Several methods, such as text processing, full text search, association rules, ontologies and taxonomies, are used to achieve it. Some methods are based on term weighting, which detects the word relevance given a specific context [1]. The adoption of terminologies and standards facilitates the manipulation of electronic records across heterogeneous systems, enabling system interoperability.

In Brazil, the lack of standardization and the use of multiple terminologies among health systems hinders interoperability. The challenge is to develop methods to make new standards compatible with others. *Terminologia Unificada da Saúde Suplementar* (Unified Terminology of Supplemental Health - TUSS) [2], one of the most important Brazilian health procedures standards, is used by the supplementary health system, mainly for claims. However, there is no available mapping between this coding system and others, e.g. the SIA/SIH SUS coding, used by the public health system in Brazil, which hinders any effort in database integration. The public healthcare and legacy systems require ways to communicate with this standard, which calls for methods of conversion and classification to enable its adoption.

A systematic review was performed to describe the state of the art on this subject. Several of the selected papers involved usage of SNOMED [1,3-8], others the UMLS [1,4-6,8], proving great interest and importance of these terminologies. Some works propose automated methods [5,6,8,9], others are based on ontology mapping [1,3,4], use keywords in their methods [5-7], and only two use term weighting [8,9]. Despite the great diversity of works in the literature related to interoperability, however, the usage of ontologies or taxonomies associated with term weighting has not been reported to date.

We propose an approach for classifying TUSS' medical records using term weighting and taxonomy. We applied two term weighting algorithms, namely TF-IDF and Doc2Vec, and compare their accuracy when classifying TUSS' records.

Methods

TUSS' records are hierarchically organized in chapters, groups and subgroups, as a taxonomy. We used the OWL format to represent TUSS' codes and hierarchical organization. We evaluated the relevance of each description in the TUSS terminology procedures using the mentioned term-weighting techniques. We developed vector models, assigning weights for each identified term, considering their relevance within the hierarchy specified by the taxonomy. Each model uses distinct algorithms, such as TF-IDF, and Doc2Vec, to allow comparisons and analysis of their accuracy in predicting (ability of the method of hitting the conversion, expressed as a percentage). In order to validate the mechanism proposed, we used Table 22 of the TUSS standard, which includes 5,769 codes and description of health procedures and events for the December 2015 edition. These codes are divided into 6 chapters, 16 groups and 38 subgroups. Chapter, group, and subgroup identifications are used to classify each procedure description through the composition of the TUSS code. We created three models for each of the proposed algorithms, with 90%, 85%, and 66% of the total data mass (5,769 procedures) as the model training set and 10%, 15%, and 33% as the validation set. The first experiment randomly selected 90% of the procedures descriptions to be used in the training of the model. In this step of the Doc2Vec algorithm, each procedure is transformed into a vector according to its terms, being positioned by the algorithm according to its similarities and its chapters, groups and subgroups. Then, the remaining 10% of the data is classified according to the proximity of the vectors of the trained model. The classification obtained by the algorithm is compared with the current classification of the TUSS standard and, thus, the accuracy rate of the model is calculated. Similarly, for the TF-IDF algorithm, a portion of the procedures are randomly selected so that the algorithm calculates the repetition and weight of each word involved in the procedures, thus creating a term weighting model. The remaining portion of the data is used for model validation. Each procedure description is broken in tokens and according to the weighting model of these terms, a classification of the procedure is obtained and the accuracy rate of the calculated model.

Results

The results show that the model using the 90% / 10% ration between training set and validation set obtained the best hit rates for all classification levels of the Doc2Vec algorithm, matching 65.16% of the chapters using one hundred dimensions and 3,000 iterations. For the groups, the best accuracy rate was 43.32% with two hundred dimensions and 4,500 iterations. For subgroups, we obtained 33.44% of hits with the same group configuration. The algorithm TF-IDF obtained the best accuracy with the 85% / 15% division, with 99.08% of the chapters, 89.38% of the groups and 33.26% of the subgroups. Figure 1 presents the summary with the best results of each algorithm for different levels of taxonomy.

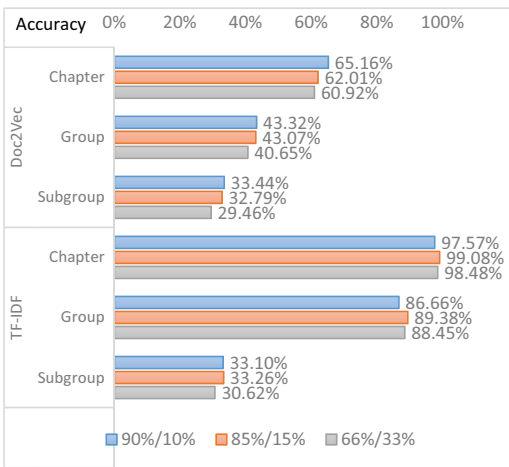


Figure 1 – Accuracy of Doc2Vec and TF-IDF Models

The TF-IDF performed better in classifying groups, yielding accuracy of 89%, while Doc2Vec scored 43%. However, for the classification of subgroups the two algorithms obtained lower accuracy rate, with the TF-IDF reaching the maximum 33.26% and the Doc2Vec slightly higher with 33.44%. In turn, Doc2Vec achieved its best performance for all levels of the taxonomy by dividing the data in the proportion of 90% for training and 10% for validation.

Discussion

The experiment results show that the greater the generalization of the level of taxonomy, the greater the accuracy. The classification of chapters obtained the best classification rates followed by groups and subgroups. Conversely, the more specific the hierarchy of taxonomy, the worse the algorithm performance. It is noted, at the more specific levels, that the proximity between the descriptions of the procedures is related to the decrease of the levels of accuracy for both algorithms. Although the mechanism of this work does not identify a TUSS code in its entirety, it can be concluded that the use of taxonomy associated to machine learning algorithms was effective in classifying a procedure. Since TUSS' codes have eight digits, composed by numerals of its chapter, group, subgroup and followed by a service sequence number and a checker digit, the first five elements of the code have been identified. The service number does not have a specific logic, it is only the sequence of creation of each procedure. A limitation of this work concerns the mass of data used, based exclusively on Table 22 of TUSS medical procedures. Other data sources can be used to obtain trained models with a wider variety of terms or in different domains of knowledge. This work was restricted to compare the TF-IDF and Doc2Vec algorithms. Both work with vector word mapping. As future work, other algorithms can be analyzed, integrated and compared, seeking to improve accuracy, especially at more specific levels of taxonomy where the results of these algorithms were not satisfactory.

Conclusions

We were able to discuss the challenges and importance of achieving interoperability between distinct health terminologies, and how machine learning and the weighting of terms can help achieve this goal. We showed that two algorithms with similar functionalities can present complexities and different results in their use. We also presented the relevance of the hierarchical structuring of the information, in the form of taxonomy, in order to facilitate its classification. Finally, we presented a mechanism for classifying TUSS terminology chapters and groups. We also demonstrated that, in this specific context, the TF-IDF algorithm was more efficient than the Doc2Vec algorithm and may be helpful when constructing mappings between two different code systems.

References

[1] V.D. Mea, et al. A web-based tool for development of a Common Ontology between ICD11 and SNOMED-CT. *Proceedings of 2014 IEEE International Conference on Healthcare Informatics*, ICHI 2014, (2014), 144-148.

[2] ANS. In: http://www.ans.gov.br/prestadores/tiss-troca-de-informacao-de-saude-suplementar date accessed: 10/28/2018.

[3] A. Ryan. Towards semantic interoperability in healthcare: ontology mapping from SNOMED-CT to HL7 version 3, *Australian Computer Society*, (2006), 72:69–74.

[4] L. Marenco, R. Wang, P. Nadkarni. Automated Database Mediation Using Ontological Metadata Mappings. *J Amer Med Info Assoc.* (2009), 16(5): 723-737.

[5] J. Patrick, Y. Wang, P. Budd. An automated system for conversion of clinical notes into SNOMED clinical terminology. *Proceeding of the fifth Australasian Symposium on ACSW frontiers*, ACSW'07, (2007), 68:219-226.

[6] Y. Wang, J. Patrick. Mapping Clinical Notes to Medical Terminology at Point of Care, *Proceedings of Current Trends in Biomedical Natural Language Processing*, BioNLP (2008), 102-103.

[7] S. Yuwen, X. Yang. Conversion between SNOMED and CDA. *Proceedings of 2010 International Conference on Biomedical Engineering and Computer Science*, ICBECS, (2010), 165-168.

[8] P. Ruch, et al. Automatic medical encoding with SNOMED categories. *BMC Med Inform Decis Mak*, (2008), 8(Suppl 1): S6.

[9] P. Chen, A. Barrera, C. Rhodes. Semantic analysis of free text and its application on automatically assigning ICD-9-CM codes to patient records. Proceedings of the 9th IEEE International Conference on Cognitive Informatics, ICCI, (2010), 68–74.

Address for correspondence

Márcia Ito

Informar Saude – Grupo Bem

Rua Josef Kryss, 319 – 01140-050

Sao Paulo/SP – Brazil e-mail: m.ito@uol.com.br

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190503

The Development of an Electronic Phenotyping Algorithm for Identifying Rhabdomyolysis Patients in the MID-NET Database

Rieko Izukura[a], Tadashi Kandabashi[a], Yoshifumi Wakata[b], Chinatsu Nojiri[a], Yasunobu Nohara[a], Takanori Yamashita[a], Atsushi Takada[a,c], Jinsang Park[a], Yoshiaki Uyama[c], Naoki Nakashima[a]

a. Medical Information Center, Kyushu University Hospital, Fukuoka, Japan
b. Hospital Information Center, Tokushima University Hospital, Tokushima, Japan
c. Pharmaceuticals and Medical Devices Agency, Tokyo, Japan

Abstract

We aimed to develop rhabdomyolysis (RB) phenotyping algorithms using machine learning techniques and to create subphenotyping algorithms to identify RB patients who lack RB diagnosis. Two pattern algorithms, one with a focus on improving predictive value and one focused on improving sensitivity, were finally created and had a high area under the curve value of 0.846. Although we were unable to create subphenotyping algorithms, an attempt to detect unknown RB patients is important for epidemiological studies.

Keywords:

Algorithms, Machine learning, Rhabdomyolysis

Introduction

Phenotyping algorithms that automatically identify patients diagnosed with a particular disease based on electronic health records (EHRs) have been gradually increasing [1][2]. The Medical Information Database Network (MID-NET) is a large-scale, anonymized database that contains more than 4 million (estimated to reach 10 million in the future) patients' structured EHRs. The MID-NET stores laboratory test values differently than other domestic large databases [3]. The MID-NET was constructed primarily by the Ministry of Health, Labour, and Welfare and Pharmaceuticals and Medical Devices Agency for the early detection of adverse drug events. With a nationwide use of the MID-NET that began in Japan in 2018, further development of phenotyping algorithms using various MID-NET data is required.

Rhabdomyolysis (RB) is a well-known drug-induced adverse event. However, there are no useful phenotyping algorithms that automatically detect RB patients via the MID-NET. Therefore, we aimed to develop RB phenotyping algorithms using machine learning (ML) techniques and to create sub-phenotyping algorithms to identify RB patients who lack RB diagnosis.

Methods

Population

We retrospectively analyzed 93,106 inpatient cases in Kyushu University Hospital between January 1, 2013 and December 31, 2016.

Development of the Phenotyping Algorithm for RB

We developed an algorithm to detect RB. This algorithm is described in the next three steps as shown in Figure 1.

Step 1: Creating the Main Phenotyping Algorithm

1-1. Creation of an initial RB phenotyping algorithm using International Classifications of Diseases 10th revision (ICD10) codes (M6289, G720-G722, and G729) **[RB algorithm A].**

> **RB algorithm A**
> (Rhabdomyolysis [M6289]) or (Myopathy [G720-G722 or G729])
> Index date: Diagnosed date based on a 30-day interval

1-2. Data extraction using from MID-NET and random sampling

1-3. Review by two physicians per case and calculation of the weighted kappa coefficient (κ-coefficient) as a summary of the coincidence degree between each pair of physicians

1-4. Modification and improvement of RB algorithm A using ML techniques
We estimated a predictive model for RB using the highest area under the curve (AUC), the predictive value, or the sensitivity using the gradient boosting decision tree (GBDT)
[mRB algorithm A'].

Step 2: Creating the Subphenotyping Algorithm

2-1. Creation of an initial phenotyping algorithm to detect RB cases without RB algorithm A and instead using creatine kinase (CK), creatine kinase-MB (CK-MB), and troponin-T values [**RB algorithm B]**.

The CK-MB or troponin-T value was established to exclude cardiovascular diseases with CK elevation, such as myocardial infarction.

> **RB algorithm B**: **a NOT b**
> **a.** CK over 10-fold increase from reference value[*1]
> **b.** ([CK-Mb over 10% increase in CK] or [Troponin-T more than 0.014 μg/dL[*2]]) during the same time period as **a**
> Index date: Abnormal laboratory results based on a 30-day interval
> *[1] Japanese Committee for Clinical Laboratory Standards (JCCLS)-based reference value. *[2] Hospital-based reference value

2-2. Data extraction (same as step 1)

2.3. Physician review (same as step 1)

2-4. Modification and improvement of RB algorithm B' (same as step 1) **[mRB algorithm B']**

Step 3: Creating the Final Phenotyping Algorithm

We integrated (mRB algorithm A' [step 1]) and (mRB algorithm B' [step 2]) to create the final phenotyping algorithm.

Results

Data Extraction and Random Sampling

The RB algorithm A identified 111 cases, and 3,231 cases were extracted by the RB algorithm B. We randomly sampled 100 cases per algorithm.

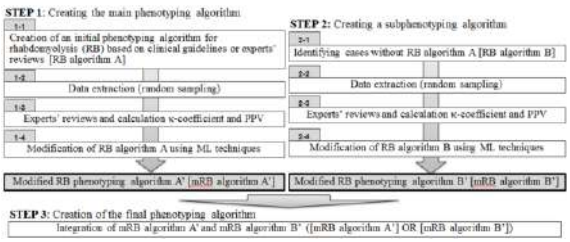


Figure 1. *Workflow of phenotyping algorithm development. Step 1 aimed to improve the phenotyping algorithm based on disease diagnosis data, and Step 2 aimed to identify RB patients who were not extracted in Step 1.*

Expert Review Results

The κ-coefficient for the review of 2 physicians was 0.50.

Results of RB algorithm A

Thirty-four cases were identified as true-positive, and the PPV was 34.0%. Few of the cases with a myopathy diagnosis code were true-positive. Thus, we excluded the myopathy diagnosis codes (G720-722 and G729) from the algorithm during the modification step.

Results of RB algorithm B

Eighteen true-positive cases were identified, and the PPV was 18.0%. Half of the true-negative cases (n = 55) caused temporary increases in CK values because of surgery. Thus, we excluded "surgery implementation within three days from the index date" during the next step.

Algorithm Modification

A total of 1,619 explanatory variables for the GBDT were extracted from 200 cases. These consisted of the following structured-data codes or values in the MID-NET: Disease diagnosis, hospitalization number, pharmaceutical products, laboratory test values, surgery implementation, medical intervention in Japanese administrative and claims, patient's age and gender.

Modification of RB algorithm A

The top 5 principle variables that contributed to the model, in order, were CK, aspartate aminotransferase, lactate dehydrogenase, M6289 (ICD10 code), and serum chloride (AUC = 0.846). According to the GBDT, all variables without CK was less than 20% of the gain rate. Thus, we focused on the CK value for algorithm modification considering PPV and sensitivity within 200 cases.

The modified phenotyping algorithm is shown below:

Modification of RB algorithm B

According to the GBDT, the AUC was 0.622. We could not improve the RB algorithm B. Eventually, (mRB algorithm A') was accepted as the final RB phenotyping algorithm.

mRB algorithm A'

An algorithm based on improving PPV

(PPV = 92.9%, Sensitivity[*1] = 4.6%)

[RB ICD10 code (M6289) on DPC] or ([RB ICD10 code (M6289) on HIS] and [CK value over five-fold increase compared to reference value])

An algorithm based on improving sensitivity

(PPV = 36.8%, Sensitivity[*1] = 83.7%)

(ICD10 code [M6289] on DPC) **OR** (ICD10 code [M6289] on HIS) **and** [CK more than reference value[*1]]) **OR** ([RB algorithm B] **not** [Surgery implementation within three days before the index date])

*[1] Sensitivity = estimated sensitivity in all possible cases

Discussion

We aimed to develop an RB phenotyping algorithm following a three-step procedure. Improvement of the initial phenotyping algorithm based on disease diagnosis was successful and resulted in a high AUC. However, we could not modify the initial algorithm based on the laboratory results because of the low AUC value. When creating the initial phenotyping algorithm, we should have reconsidered the following: the time interval for excluding myocardial disorders based on CK-MB or troponin-T values and the exclusion of cases that temporary increased CK caused by surgical procedures. However, we were able to identify previously unknown RB cases that were not diagnosed as RB. Eventually, we created two types of RB algorithms: one that focused on improving the PPV and one that focused on improving the sensitivity. These algorithms should be used differently depending on the study purpose.

Conclusions and Limitations

We created two types of new RB phenotyping algorithms. These algorithms should be used differently depending on the study purpose. Although these algorithms were created using small samples, we will validate these samples at other hospitals this year.

Acknowledgements and Ethics

This study was supported by the Japan Agency for Medical Research and Development Grant Number 17mk0101088h0001. This study was approved by the ethics review committee of Kyushu University (No. 30-423).

References

[1] J.C. Kirby et al., PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability, *JAMIA* **23** (2016), 1046-1052.

[2] C. Shivade et al., A review of approaches to identifying patient phenotype cohorts using electronic health records, *JAMIA* **21** (2014), 221-230.

[3] Pharmaceuticals and Medical Devices Agency, available at https://www.pmda.go.jp/safety/mid-net/0001.html

[4] H. Martin, Drug induced rhabdomyolysis, *Curr Opin Pharmacol*, **12(3)** (2012), 335–339.

Address for correspondence

Izukura Rieko, RN, PhD

ochamame0830@med.kyushu-u.ac.jp Tel: (+81)92-642-588

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190504

Evaluating a Clinical Decision Support System for Drug-Drug Interactions

Amin Jalali[a], Paul Johannesson[a], Erik Perjons[a], Ylva Askfors[b], Abdolazim Rezaei Kalladj[b], Tero Shemeikka[b], Anikó Vég[b]

[a] Department of Computer and Systems Sciences at Stockholm University, Stockholm, Sweden,
[b] Health and Medical Care Administration, Stockholm County Council, Sweden

Abstract

Janusmed is a clinical decision support system, developed by the Stockholm County Council that supports physicians in identifying drug-drug interactions. To determine how Janusmed is used in and affects the clinical practice, an evaluation study is currently being carried out that analyzes multiple data sources through descriptive statistics. The study focuses on how Janusmed affects the behavior of the physicians, in particular, to what extent physicians reconsider their prescription decisions based on warnings from Janusmed.

Keywords:

Drug Combinations, Drug Interactions, Prescriptions, Clinical Decision Support Systems, Medical Informatics

Introduction

Drug prescribing is becoming more complex due to the increasing number of drugs available as well as more information about their usage, effectiveness, and side effects. At the same time, patient safety is a major public concern emphasizing the importance of prescribing the right drug to the right patient in the right dose. Adverse drug reactions are a significant source of illness and also a cause of death [1], and drug-drug interactions represent a common cause of adverse drug reactions [2,3]. Thus, concise and up-to-date information about possible drug-drug interactions available at the point of drug prescription is vital for physicians [4, 5, 6, 7].

Janusmed Integrated (toolbar) is a clinical decision support system (CDSS) that includes several medication-related knowledge databases (modules) such as Janusmed Interactions, Janusmed Risk Profile, Janusmed Drugs and Birth Defects, Janusmed Drugs and Breast Feeding, Janusmed Drugs and Renal Function, and Janusmed Sex and Gender [5, 6]. The Janusmed system is designed especially for integration into electronic health record systems. It is available in most of the EHR systems in healthcare organizations in the Stockholm County Council in Sweden.

The Janusmed Interactions module warns for and provides information about (pharmacokinetic) drug-drug interactions based on 1) information in the drug-drug interaction knowledge base and 2) information in the electronic health record (EHR) system, such as the patient's present list of prescribed drugs. It will automatically alert for drug-to-drug interactions when physicians prescribe new drugs. The interactions are classified from A to D based on their clinical relevance and corresponding color codes have been developed to aid end users. A (green) is a minor interaction with no clinical relevance, B (white) means that the clinical outcome of the interaction is uncertain and/or may vary, C (yellow) is a clinically relevant interaction that can be handled, for example, by dose adjustments, and D (red) is a clinically relevant interaction that should be avoided [4].

Janusmed is widely available and used, but it is still unknown how the system is used and to what extent it actually affects the clinical practice, for example, if prescribing physicians do change their behavior due to warnings for drug-drug interactions. Filling this knowledge gap would support and guide the future development of Janusmed.

In order to address these issues, an evaluation project is being carried out between the Stockholm County Council responsible for developing the system and Stockholm University. The goal of the evaluation project is to determine how Janusmed is used in the clinical practice. More specifically, the project addresses three main research questions:

- How does the utilization of Janusmed Interactions change over time?
- How is the usage of Janusmed Interactions influenced by patient demography, especially age and gender?
- To what extent do physicians recognize and act on the drug-drug interaction warnings provided by Janusmed?

Method

A data survey was carried out to answer the three research questions. First, we operationalized the concepts used in the three research questions in order to understand which data to collect and analyze. Second, we identified data sources that contained relevant data. The first data source was the Janusmed log. In the log, a record is created every time a physician checks a patient's list of drugs, changes the list by adding another drug to it, or does a follow-up on a warning for a drug-drug interaction to obtain additional information from Janusmed. The second source was the Janusmed drug-drug interaction database that was used for reconstructing drug-drug interactions not included in the Janusmed Integrated log. The third source was Health Bank that contains all EHR records from the EHR systems in one of the major hospitals in Stockholm, Karolinska University Hospital. We needed this third source to determine whether the physician prescribes another drug instead of one of those causing the drug-drug interaction.

A major task in the project has been to address interoperability challenges. In particular, the information in the Janusmed log is anonymized in the sense that it does not identify any individual. Therefore, there is no way to directly map Janusmed log records to EHR records in Health Bank. Instead, we had to rely on combinations of patient attributes (for example, age and gender) and information on physician consultations (for example, place and time) to arrive at a mapping between the systems. Another

interoperability issue concerned different formats and identifiers for drugs.

The data analysis was primarily carried out by means of descriptive statistics. This allowed us to investigate the possibility to apply process mining techniques in the future. Process mining is a new paradigm of data science that offers AI techniques that support decision-making in process management. The techniques are based on analyzing event logs, that is, logs of the activities that users have carried out during a process [8, 9].

Results and Conclusions

The data is still being analyzed but an initial result is that the utilization of Janusmed Integrated (toolbar) has increased over time. To investigate the utilization, we measured the number of follow-ups the physicians did to obtain additional information. This measurement was based on all care units using Janusmed in the Stockholm County Council for a time period of 15 months. We are presently analyzing the usage patterns of the different modules in Janusmed, in particular, Janusmed Interactions.

We are also currently investigating whether the usage of Janusmed depends on demographic factors, in particular, the age and gender of the patients. We are also investigating how the warnings affect the behavior of the physicians, depending on the class of clinical relevance. In particular, to what extent physicians carry out follow-ups and reconsider their prescription decisions.

In order to better understand how the physicians are using Janusmed, a qualitative investigation is planned using observations at the point of prescription, followed by semi-structured interviews with the prescribing physicians.

Acknowledgements

The project is funded by the Stockholm County Council, Sweden.

References

[1] Lazarou J, Pomeranz BH, Corey PN (1998) Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA 279(15):1200–1205

[2] Juurlink DN, Mamdani M, Kopp A, Laupacis A, Redelmeier DA (2003) Drug-drug interactions among elderly patients hospitalized for drug toxicity. JAMA 289(13):1652–1658

[3] Bjerrum L, Andersen M, Petersen G, Kragstrup J (2003) Exposure to potential drug interactions in primary health care. Scand J Prim Health Care 21(3):153–158

[4] Y. Bottiger, K. Laine, M.L. Andersson, T. Korhonen, B. Molin, M.L. Ovesjo, T. Tirkkonen, A. Rane, L.L. Gustafsson, B.Eiermann, SFINX-a drug-drug interaction database designed for clinical decision support systems, Eur. J. Clin. Pharmacol.65 (6) (2009) 627–633.

[5] B. Sjöborg, T. Bäckström, L.B. Arvidsson, E.Andersén-Karlsson, L.B. Blomberg, B. Eiermann, et al., Design and implementation of a point-of-care computerized system for drug therapy in Stockholm metropolitan health region – bridging the gap between knowledge and practice, Int. J. Med. Inf. 76 (7) (2007) 497–506.

[6] M. Eliasson, P. Bastholm, P. Forsberg, K. Henriksson, L.Jacobson, A. Nilsson, et al., Janus computerised prescribing system provides pharmacological knowledge at point of care– design, development and proof of concept, Eur. J. Clin.Pharmacol. 62 (2006) 251–258.

[7] Nabovati, E. et al., 2017. Information Technology-Based Interventions to Improve Drug-Drug Interaction Outcomes: A Systematic Review on Features and Effects. Journal of medical systems, 41(1), p.12.

[8] Van der Aalst, Wil MP. "Process discovery: An introduction." Process Mining. Springer, Berlin, Heidelberg, 2011. 125-156.

[9] Rojas, E., Munoz-Gama, J., Sepúlveda, M., & Capurro, D. (2016). Process mining in healthcare: A literature review. Journal of biomedical informatics, 61, 224-236.

Address for correspondence

Paul Johannesson, pajo@dsv.su.se
Department of Computer and Systems Sciences
Stockholm University
SE-106 91 Stockholm, Sweden
https://processminingjanusmed.blogs.dsv.su.se/

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190505

Augmenting Medical Device Evaluation Using a Reusable Unique Device Identifier Interoperability Solution Based on the OHDSI Common Data Model

Guoqian Jiang[a], Yue Yu[a], Paul R. Kingsbury[a], Nilay Shah[a]

[a] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

Abstract

The objective of the study is to augment safety and effectiveness evaluation of medical devices through building a reusable unique device identifier (UDI) interoperability solution. We propose a framework for building a UDI research database for medical device evaluation using the OHDSI common data model (CDM). As a pilot study, we design, develop and evaluate a UDI vocabulary, which would enable tackling challenges of data islands and standardization for medical device evaluation.

Keywords:

Equipment Safety, Reference Standards, Controlled Vocabulary, Electronic Health Records.

Introduction

Reliable and timely data and information on the performance and safety of medical devices such as stents and artificial joints are critical to the success of postmarket surveillance of medical devices. In recent years, there are increasing needs for capturing Real World Evidence (RWE) from claims and electronic health records (EHR) data for evaluations of medical device safety and effectiveness [1]. In particular, the Food and Drug Administration (FDA) and medical device evaluation community have envisioned a system that can not only promote patient safety through earlier detection of safety signals, but also generate, synthesize and analyze evidence on real-world performance and patient outcomes [2].

However, valuable data for evaluating device exists largely as data islands with only limited connectivity. It has been well recognized that a standard is key for documentation and linking of medical device identification information to diverse data soures. Fortunately, FDA initiated the regulation of the Unique Device Identifier (UDI) implementation and established a Global Unique Device Identifier Database (GUDID) for making uniquely identify medical device possible [3]. By September 24, 2018, all Class III and Class II devices are required to bear a permanent UDI. Note that such a classification is risk based. Class I includes device with the lowest risk and Class II includes those with the great risk. Meanwhile, a number of demonstration projects have been conducted to demonstrate the feasibility of informatics technology on building a medical device evaluation system and to identify keys to success and challenges to achieving targeted goals [4]. These projects provided proof of concept that UDIs can be used as the index key to combine device and clinical data in a database useful for device evaluation. However, the community lacks of an open-source solution for community-based medical device evaluation.

The Observational Health Data Sciences and Informatics (OHDSI) Common Data Model (CDM) has been increasingly used to build a large-scale international data network in support of observational studies [5]. In this study, we propose a reusable UDI interoperability solution based on the OHDSI CDM that tackles the challenges of data islands and standardization. We aim to build a prototype device evaluation system with the capability to uniquely identify medical devices, implementing standardized clinical vocabularies and dictionaries, and linking diverse, strategically complementary data sources.

Methods

OHDSI CDM: The OHDSI CDM 5.0.1 version contains 39 database tables in 6 categories: standardized clinical data, standardized health system data, standardized health economics, standardized metadata, standardized vocabularies and standardized derived elements [5]. Of them, 10 tables are designed for handling standardized vocabularies, including tables describing CONCEPT, CONCEPT_RELATIONSHIP, CONCEPT_CLASS, DOMAIN, VOCABULARY. We installed an OHDSI virtual machine (VM) that is conformant to the OHDSI CDM 5.0.1 version. The VM contains the full OHDSI technology stack and is loaded with both standard vocabularies and sample data.

FDA GUDID: The database contains key device identification information submitted to the FDA about medical devices that have UDI [3]. As of September 1, 2018, a total of 1,765,653 UDIs have been recorded, comprising 45,011 UDIs for Class III devices, 1,456,524 UDIs for Class II devices and 264,128 UDIs for Class I devices. In this study, we used the UDI information for Class III devices for prototyping the UDI vocabulary.

System Architecture: We propose a framework for building a UDI research database for medical device evaluation using the OHDSI CDM (**Figure 1**). The framewrok comprises three modules: 1) UDI Vocabulary; 2) ETL for UDI-Enabled Device Data; and 3) Analytic methods for medical device evaluation.

Prototype Implementation: In this study, we prototyped a UDI vocabulary using the UDI information for Class III devices.

First, we downloaded the Class III device dataset from the GUDID website. The dataset contains 45,011 records for Class III device. The fields of the dataset are "DEVICE_ID|DEVICE_ID_ISSUING_AGENCY|BRAND_NAME|COMPANY_NAME|VERSION_MODEL_NUMBER|MRI_SAFETY_STATUS|LABELED_CONTAINS_NRL|GMDN_TERMS."

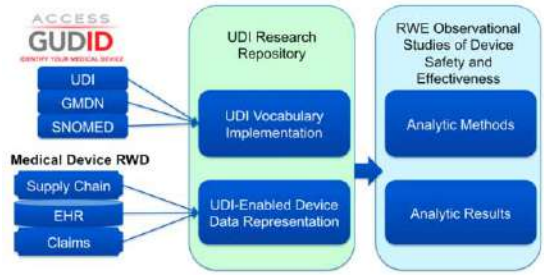


Figure 1. System architecure of our proposed framework. RWD: Real-World Data; RWE: Real-World Evidence.

Second, we retrieved the SNOMED CT information associated with each single device using the GUDID web service API called Device SNOMED. A device identifier and a UMLS single-use ticket are required to receive a response.

Third, we used the feature of custom structures of HAPI FHIR API [6] and generated Java-based data model for the OHDSI Standardized Vocabulary CDM which comprises 10 tables. With this advanced feature, vocabulary content can be readily represented in the FHIR-compliant XML or JSON format.

Fourth, we parsed the Class III device dataset and generated the concept_id from each single device. And then we populated the CONCEPT, CONCEPT_RELATIONSHIP, CONCEPT_CLASS, DOMAIN and VOCABULARY tables. The key design points here include 1) the concept_name is formed as a combination of BRAND_NAME, COMPANY_NAME, and VERSION_MODEL_NUMBER; 2) We tentatively set the status of STANDARD_CONCEPT as non-standard for each single device is linked to a SNOMED CT concept code which by default is used as standard concept in the OHDSI CDM. 3) The Global Medical Device Nomenclature (GMDN) term information is recorded in the CONCEPT_RELATIONSHIP table. The concept_id is also generated for each GMDN term. An intermidiary FHIR Bundle XML file is generated, which can be easily exported into standard OHDSI vocabulary release CSV files.

Preliminary Evaluation: We loaded the UDI vocabulary release files into a OHDSI VM and used its ATLAS tool to explore the utility of the vocabulary. We compared the string search vs. concept-based search.

Results

A total of 51,597 records were generated in the CONCEPT table, in which 45,011 records are for Class III devices and the rest are for the GMDN terms and a few of relationship concept ids. A total of 17,052 records were generated in the CONCEP_RELATIONSHIP table, in which half of records are for the relationships to SNOMED CT and the other half for the relationships to GMDN terms.

We successfully loaded the UDI vocabulary into an OHDSI VM platform. We used the ATLAS concept search tool for the string search of "Thermocool" and obtained 76 records. When we used the SNOMED CT concept code for "Cardiac mapping/radio-frequency ablation catheter, single use", we obtained 250 records (**Figure 2**). These results indicated that concept-based search may be more reliable if users would like to get comprehensive query results.

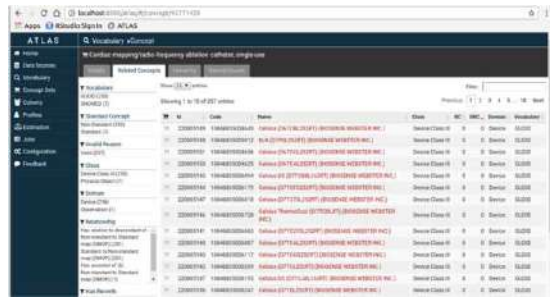


Figure 2. An ATLAS screenshot for concept-based search.

Discussion and Conclusions

In this study, we proposed a framework for building a OHDSI CDM-based UDI research database and prototyped a UDI vocabulary for Class III devices as a pilot study. Such a standard UDI vocabulary will serve as semantic foundation to enable both individual-level and class-level device data analyses and eventually augment safety and effectiveness evaluation of medical devices. We believe that the approach can be readily generalized for the devices in other classes. There are a number of design points that would need a community-based consensus. We plan to initiate a Working Group in the OHDSI community for the consensus development. In addition, more rigorous evaluation of the UDI vocabulary will be conducted in the near future after loading the real-world device data into the UDI research database driven by specific use cases.

Acknowledgements

This study is supported in part by NIH grant U01 HG009450.

References

[1] J. Corrigan-Curay, L. Sacks, J. Woodcock. Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness. JAMA. 2018;320(9):867-868.

[2] M. W. Krucoff, A. Sedrakyan, S. T. Normand. Bridging Unmet Medical Device Ecosystem Needs With Strategically Coordinated Registries Networks. JAMA. 2015;314(16):1691-1692.

[3] FDA GUDID URL: https://accessgudid.nlm.nih.gov/.

[4] J.P. Drozda, J. Roach , T. Forsyth, P. Helmering, B. Dummitt, J.E. Tcheng. Constructing the informatics and information technology foundations of a medical device evaluation system: a report from the FDA unique device identifier demonstration. J Am Med Inform Assoc. 2018 Feb 1;25(2):111-120.

[5] OHDSI CDM: https://github.com/OHDSI/CommonDataModel/wiki.

[6] HAPI FHIR API URL: http://hapifhir.io/.

Address for correspondence

Guoqian Jiang, MD, PhD, Mayo Clinic – jiang.guoqian@mayo.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190506

Validation of a Customized Algorithm for the Detection of Diabetic Retinopathy from Single-Field Fundus Photographs in a Tertiary Eye Care Hospital

Sheila John[a], Sangeetha Srinivasan[b], Rajiv Raman[c], Keerthi Ram[d], Mohanasankar Sivaprakasam[d]

[a]*Teleophthalmology, Sankara Nethralaya, Chennai, India*
[b]*Vision Research Foundation, Chennai, India*
[c]*Shri Bhagwan Mahavir Vitreoretinal Services, Sankara Nethralaya, Chennai, India*
[d]*Healthcare Technology Innovation Center, Indian Institute of Technology, Chennai, India*

Abstract

The study was done to validate the real time efficacy of a customised algorithm in detecting diabetic retinopathy (DR) among diabetic patients being examined at the vitreo retinal outpatient department (VR OPD) of a tertiary care hospital, Diabetic Retinopathy algorithm showed sensitivity of 79% and specificity of 57% which is an acceptable methodology to diagnose diabetic retinopathy and avoid unnecessary referrals.

Keywords:

Diabetic Retinopathy, Algorithm, Photograph

Introduction

Diabetic retinopathy (DR) is a microvascular complication of diabetes and is a potentially blinding disease, especially in its advanced stages[1]. In India, one of the several challenges faced in screening is inadequate man power, with only one ophthalmologist for every 107,000 people [2]. This scenario warranted an urgent need for alternative methods of screening. However, screening, timely referral, and intervention can lessen the likelihood of blindness and can reduce the health care costs[3].

Several studies have reported the effectiveness of computer-assisted screening tool for DR detection [4-8]. HTIC developed and reported an algorithm for the detection of presence of DR[9]. We reported that the algorithm can identify the presence or absence of DR in subjects with diabetes under mydriatic and non mydriatic conditions using retinal photographs.

Methods

Our DR image screening algorithm [9] includes modules for analyzing retinal images to identify normal retinal structures such as the optic disk, macular region, segmentation of blood vessels, detection of three classes of pathologic lesions in the retina: red lesions (such as blot and flame haemorrhages), bright lesions (hard exudates, circinates, cotton-wool spots, ischemic zones), and microaneurysms. Each of the modules derives parameters from the retinal image using image analysis techniques such as object detection and segmentation, and computes parameter values corresponding to the different features. The composite low level information of the locality, morphology, appearance and extent of the disease signs is considered and solved for the clinical decision (DR vs no-DR) using a multi-parametric machine learning based classification. The gradeability of the image, as well as presence of severe and proliferative DR signs are added as high-level information of gradeability and advanced late stage disease, and are combined to frame the final decision on DR referability, giving a value in the range of 0 to 1, with 1 indicating higher confidence score on the decision of 'DR-present'. The optimal operating point for the referability decision is chosen by doing a ROC analysis with multiple publicly available datasets, and a large retrospective dataset.

We validated the algorithm prospectively for real time use in an outpatient department of Sankara Nethralaya, a tertiary care hospital and compared the performance of the algorithm with that of vitreo-retinal surgeon (reference standard) and optometrist under non-mydriatic condition. This study was approved by Institutional Review Board of Vision Research Foundation, Chennai, India. Patients provided written informed consent. The study was conducted from Jan 2015 to May 2017. Patients with type 2 diabetes aged ≥35 years were included. Those with miotic pupil, nystagmus, history of other intraocular surgeries except cataract surgery and injections for diabetic macular edema were excluded.

848 eyes of 485 patients underwent 45-degree fundus photography (Forus3nethra Classic Non Mydriatic Fundus Camera, Bangalore, India) in both eyes after dark adaptation. The photo graders received a CD-ROM containing de-identified images in JPEG format. Information on the patients age, sex and duration of diabetes was shared with the reader and were masked to the rest of the demographic characteristics. The images were run through the automated system and were also graded by human graders (optometrist) and by an experienced vitreoretinal surgeon (gold standard). The patient was also examined by any one vitreoretinal surgeon in the outpatient department with the indirect ophthalmoscope who provided advise treatment/disposal of the patient, independent of the result of the algorithm.

Results

The effectiveness of the CAD to detect diabetic retinopathy lesions was examined in comparison to that by the human graders, optometrist and ophthalmologist ('reference standard'). The mean age of patients was 58.2±7.5 years and 68% were men. The mean duration of diabetes was 13.1±7.9 years (median=13 years, range=0.5 – 34 years) and the mean fasting blood sugar was 128±45mg% (median = 120mg%, range=70-250mg%). DR Diagnosis and gradeability were assessed as two separate tools for retinal photographs.

The algorithm successfully graded 634 out of 848 possible images (75%) and diagnosed presence of DR in 583 images. Compared to the ophthalmologist (Reference standard), 214 (25%) images were ungradable by the algorithm. Overall, the ophthalmologist found only 9 images to be ungradable compared to 214 images by the algorithm. There was only slight agreement in terms of image gradeability between the

ophthalmologist and algorithm, Kappa= 0.016 (95% CI = -0.013 – 0.044).

The sensitivity of the algorithm to detect DR for gradable images was found to be 79% and specificity was found to be 57% in detecting DR compared to ophthalmologist with an area under the receiver operator curve (AUC) was 0.67 (95%CI=0.63 to 0.71). With respect to vision threatening DR (VTDR), there were no significant differences in the image gradeability of the algorithm based on the VTDR status of the eye as graded by the ophthalmologist (P=0.32) by chi square test.

Compared to the optometrist, 25% images were ungradable by the algorithm. Overall, the optometrist found only 11 images to be ungradable compared to 214 images by the algorithm. There was slight or no agreement in terms of image gradeability between the ophthalmologist/ optometrist and algorithm, Kappa= 0.002 (95% CI = -0.024 – 0.028)

Conclusion

We found that the computer assisted algorithm developed by HTIC showed 78 - 79% sensitivity and 55 - 57% specificity in detecting DR. The algorithm appears to be highly sensitive at the expense of lower specificity for DR detection .. The strength of our study is that the algorithm was tested under physiological dilation and therefore more closely applicable to real-life situations. The main areas that require enhancement is the reduction in ungradable images and specifically improve agreement of gradeability and DR status with human graders.

References

[1] Resnikoff S, Pascolini D, Etya'ale D, et al. Global data on visual impairment in the year 2002. *Bull World Health Organ* 2004 82:844-851.
[2] Kumar R. Ophthalmic manpower in India—Need for a serious review. *Int Ophthalmol* 1993; 17:269–275.
[3] Rachapelle S, Legood R, Alavi Y, et al. The cost-utility of telemedicine to screen for diabetic retinopathy in India. *Ophthalmology* 2013; 120:566-573.
[4] Sil Kar S, Maity S. Automatic detection of retinal lesions for screening of diabetic retinopathy. *IEEE Trans Biomed Eng* 2018; 65:608-618.
[5] Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology* 2017; 124:962-969.
[6] Saleh GM, Wawrzynski J, Caputo S, et al. An automated detection system for microaneurysms that is effective across different racial groups. *J Ophthalmol* 2016; 2016:4176547. doi: 10.1155/2016/4176547. Epub 2016 Dec 15.
[7] Abràmoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci* 2016; 57:5200-5206.
[8] Gulshan V, Peng L, Coram M, Stumpe MC et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016; 316:2402-2410.
[9] John S, Ram K, Sivaprakasam M, et al. Assessment of Computer-Assisted Screening Technology for Diabetic Retinopathy Screening in India - Preliminary Results and Recommendations from a Pilot Study. *Stud Health Technol Inform* 2016; 231:74-81.

Address for Correspondence:

Dr. Sheila John
Department of Teleophthalmology
Sankara Nethralaya,
No 41 (old no 18), College Road,
Chennai - 600006.
Email: sheilajohn24@gmail.com
Phone: +91 9841081712

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190507

Development of Integrated Data and Prediction System Platform for the Localized Prostate Cancer

Sun Jung Lee[a,b], Sung Hye Yu[a], Yejin Kim[c], Jun Hyuk Hong[d], Choung-Soo Kim[d], Seong Il Seo[e], Chang Wook Jeong[f], Seok-Soo Byun[g], Byung Ha Chung[h], Ji Youl Lee[h], In Young Choi [a,b]

[a]Department of Medical Informatics, College of Medicine, Catholic University of Korea, Seoul, Republic of Korea,
[b]Department of Biomedicine & Health Sciences, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea
[c]Department of Creative IT Engineering, POSTECH, Korea
[d] Department of Urology, University of Ulsan College of Medicine, Seoul, Republic of Korea
[e]Department of Urology, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[f] Department of Urology, Seoul National University College of Medicine, Seoul, Republic of Korea
[g]Department of Urology, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
[h]Department of Urology, Yonsei University College of Medicine, Seoul, Republic of Korea

Abstract

In this study, we built a multi-center integrated database platform of localized prostate cancer and developed biochemical recurrence (BCR) prediction system with Gradient Boosted Regression model using Korean Prostate Cancer Registry (KPCR) database. This platform will facilitate clinical management of patients with prostate cancer, and it will also help develop appropriate treatment of prostate cancer.

Keywords:

Prostatic Neoplasms, Biochemical Recurrence Prediction, Database Management Systems

Introduction

Prostate cancer is the fifth most common cancer in men in South Korea, and the incidence is increasing [1]. Prostate cancer is a high incidence of disease and it is important to prevent and manage recurrence after curative treatment. By managing the integrated data of prostate cancer and predicting biochemical recurrence (BCR), progress and pattern of prostate cancer is possible to analyze and may help to choose a treatment in the future. Currently, there are prostate cancer data portals for scientists and clinical doctors, but most of the studies were based on Americans and Europeans data. The cancer incidence rates are different among the countries [2], and there is also racial difference both in cancer incidence and survival rate [3]. Therefore, in order to analyze and predict the prostate cancer data of Koreans, Korean Prostate Cancer Database System is required. The purpose of this study is to construct a multi-center integrated database platform using the Korean Prostate Cancer Registry (KPCR) data and develop a biochemical recurrence (BCR) prediction system to facilitate clinical management of prostate cancer.

Methods

The KPCR Database has collected 7,394 prostate cancer data of Korean from six domestic EMR-based medical institutions. We implemented a system platform to visualize and manage this data. We implemented a web-based platform using Java language and Tomcat Server. In order to predict BCR, we extracted 5,119 analytic data from the KPCR database, generated train dataset for BCR prediction and trained the dataset. We used 15 statistical analysis models to predict BCR including decision tree, logistic regression, and neural networks.

Results

This platform has been implemented features of integrated data management and BCR prediction. The AUC (Area Under the Curve) of BCR prediction is shown in Table 1. 3 years post-surgery, the highest AUC of BCR prediction is achieved by Gradient Boosted Regression, and the value is 0.8419. 5 years after surgery, the highest AUC is achieved by Ridged regression with a value of 0.8071. Finally, we implemented a BCR prediction system using a Gradient Boosted Regression model which reaches the highest average of AUC.

The 3-year and 5-year BCR prediction system with Gradient Boosted Regression model was implemented using the Python language, and we used the GradientBoostingClassifier of the sklearn module in Python. In order to integrate the Web-based platform implemented in Java with the BCR prediction system implemented in Python, we used the Django framework, a Python web framework.

The variables used for BCR prediction include age, initial PSA, clinical stage, pathology Gleason score, pathology stage, surgical margin, perineural invasion, seminal vesicle, extracapsular extension, and lymphovascular invasion. We used the feature_importances_method of GradientBoostingClassifier to output the variable importance, and the result is shown in Table 2.

The result shows that age, initial PSA, and pathology Gleason score are important variables in predicting BCR with gradient boosting regression model.

We developed biochemical recurrence (BCR) prediction system using data of Korean prostate cancer patients. In the input screen, enter the value of age, initial PSA, clinical stage, pathology Gleason score, pathology stage, surgical margin,

perineural invasion, seminal vesicle, extracapsular extension, and lymphovascular invasion. The input value uses the predict_proba method of GradientBoostingClassifier to determine the probability of occurrence of BCR and the probability of non-occurrence of BCR after 3-year and 5-year of surgery. In the result screen, BCR prediction is implemented by outputting the probability of BCR occurrence after 3-year and 5-year of prostate cancer surgery using Gradient Boosting Regression model.

Table 1 – BCR Prediction AUC

Analysis	Method	3-year	5-year
Trees	Decision Tree	0.8186	0.7561
	Random Forest(ntrees=20)	0.8320	0.7956
	Random Forest(ntrees=50)	0.8349	0.8047
	Random Forest(ntrees=80)	0.8362	0.8050
Logistic Regression	Ridged Regression(L2)	0.8288	0.8071
	Lasso(L1)	0.8319	0.7993
	Gradient Boosted	0.8419	0.8031
Neural Networks	1hidden, dropout=0.3 at input	0.7939	0.7895
	1hidden dropout=0.1 at input	0.8027	0.7977
	1hidden dropout=0.1 at input, hidden	0.7984	0.7923
	2hidden, dropout=0.3 at input	0.7978	0.7941
	2hidden dropout=0.1 at input	0.7988	0.7967
	2hidden dropout=0.1 at input, hidden	0.8016	0.7943
Survival Regression	Cox PH	0.7944	0.7816
	Random survival forest	0.7645	0.6920

Table 2 – Variable Importance

Variable	3-year	5-year
Age	0.1601	0.1870
Initial PSA	0.2813	0.2284
Clinical T stage	0.1357	0.1343
Pathology Gleason Score Sum	0.1484	0.1546
Pathology T stage	0.0656	0.0445
Surgical margin	0.0321	0.0525
Perineral invasion	0.0380	0.0483
Seminal vesicle	0.0470	0.0247
Extracapsular extension	0.0055	0.0083
Lymphovascular invasion	0.0454	0.0360

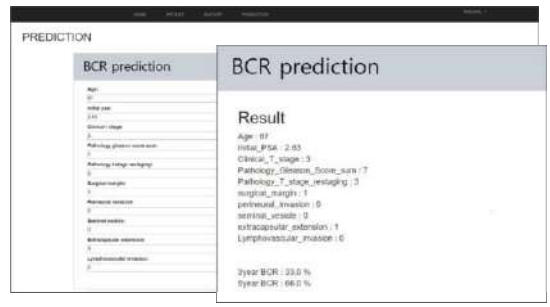


Figure1 – BCR Prediction System

Conclusions

In this study, we designed a management system platform for Korean prostate cancer patients. The platform was implemented on web-based to facilitate data access, management and BCR prediction. It will further facilitate clinical management of patients with prostate cancer and it will also help develop appropriate treatment of prostate cancer in future.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIP,2016R1A2B4015922)

References

[1] 2016 cancer statistics. National Cancer Information Center. [accessed on November 12, 2018]. https://www.cancer.go.kr/lay1/S1T639C641/contents.do.

[2] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, and A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J Clin 68 (2018), 394-424.

[3] R.M. Hoffman, F.D. Gilliland, J.W. Eley, L.C. Harlan, R.A. Stephenson, J.L. Stanford, P.C. Albertson, A.S. Hamilton, W.C. Hunt, and A.L. Potosky, Racial and ethnic differences in advanced-stage prostate cancer: the Prostate Cancer Outcomes Study, J Natl Cancer Inst 93 (2001), 388-395.

Address for correspondence

In Young Choi
222 Banpo-daero, Seocho-gu, Seoul 06591, Korea
Tel: +82-2-2258-7648, E-mail: iychoi06@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190508

Towards Structured Data Quality Assessment in the German Medical Informatics Initiative: Initial Approach in the MII Demonstrator Study

Gaetan Kamdje-Wabo[a], Tobias Gradinger[a], Matthias Löbe[b], Robert Lodahl[c], Susanne Andrea Seuchter[d], Ulrich Sax[c], Thomas Ganslandt[a]

[a]*Department of Biomedical Informatics of the Heinrich-Lanz-Center, Mannheim University Medicine, Ruprecht-Karls-University Heidelberg, Mannheim, Germany,*
[b]*Institute of Medical Informatics, Statistics and Epidemiology, University Leipzig, Leipzig, Germany,*
[c]*Department of Medical Informatics, University Medical Center Goettingen, Goettingen, Germany,*
[d]*Medical Center for Information and Communication Technology, University Hospital Erlangen, Erlangen, Germany*

Abstract

The Demonstrator study aims to analyse comorbidities and rare diseases among patients from German university hospitals within the German Medical Informatics Initiative. This work aimed to design and determine the feasibility of a model to assess the quality of the claims data used in the study. Several data quality issues were identified affecting small amounts of cases in one of the participating sites. As a next step an extension to all participating sites is planned.

Keywords:

Data Accuracy; Electronic Health Records; Medical Informatics

Introduction

The German Medical Informatics Initiative (MII) is a long-term nationwide program established to improve research opportunities and patient care through Information Technology (IT) solutions, as well as to facilitate the exchange and use of data across German University Hospitals [1]. It involves four consortia implementing Data Integration Centers (DICs) in every participating university hospitals (MII sites) [1]. The Demonstrator Study was initiated with the intention of a continuous cross-consortia cooperation throughout the project duration. The study carries out a retrospective, descriptive and geo-regionalized exploration of comorbidities and rare diseases of patients in MII sites. The study leverages the standardized "§21" data format mandated by the German Diagnosis Related Group-based billing regulations [2], which enables multicentric analyses even before interoperable data structures and interfaces were fully developed and deployed across the MII sites. The §21-conformant datasets provide a narrow set of data elements for inpatient cases including demographics, diagnoses, procedures and further administrative data. The Demonstrator Study makes use of the Integrated Data Repository Toolkit (IDRT) to extract, transform and load the source §21-conformant data into local i2b2 data repositories (Informatics for Integrating Biology and the Bedside), including a pseudonymization step (see Figure 1). The Demonstrator Study carries out analyses on these repositories to calculate Comorbidity scores and to analyse rare diseases from ICD10 diagnosis data, under consideration of patient ages, encounter duration, hospital admission and discharge reasons, patient zip codes areas and the treating university hospital.

The use of routine clinical data (specifically: claims data) in the Demonstrator Study raises questions concerning the fitness for purpose of the data for the planned analyses. In this regard, the combination of data completeness, data conformance and data plausibility aspects of the data quality (DQ) may considerably determine the validity and veracity of analysis results [3].

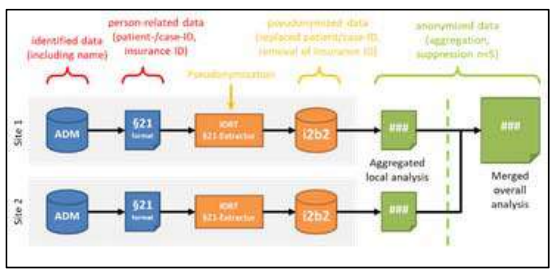


Figure 1 - Data processing pipeline of the MII Demonstrator study

Even though syntactic and coding aspects of the §21 source dataset quality is already tightly defined by billing regulations and controlled yearly by an independent institute in order to ensure their adequacy for billing purposes [2], aspects of data completeness and conformance taking account the aspects of the Demonstrator Study should be additionally addressed, as well as plausibility aspects of the variables in the datasets. The goal of the project was to evaluate the feasibility of constructing and implementing a DQ assessment framework based on a pretest of aggregated §21-conformant data from just one MII site taking part in the Demonstrator Study, before a large scale rollout to all the MII sites is conducted.

Methods

Quality ontology and analysis methods

According to the harmonized terminology of Kahn et al [3], the three quality dimensions completeness, conformance, and plausibility were choosen based on their relevance for the initial DQ approach. Operationalized quality indicators provided by the DQ Assessment guidelines from the German Technology and Method Platform for Networked Medical Research [4] were assigned to each dimensions, and quantitatively assessed using the verification method [3], through seven items formulated in consideration of longitudinal and cross-sectional aspects of the Demonstrator Study (see Table 1).

Implementation of the data quality queries

At the time of the submission, a pre-test was carried out based on the billing data of one of the participating MII sites based on an i2b2 database with data already loaded. We implemented

Table 1- Initial DQA-Framework

DQA-Dimensions	DQA-Indicators	Measurement Items
Completeness	Missing values for data elements (ID-1013)	**1-** Frequency of missing values in data elements
Conformance	Invalid values for qualitative data elements (ID-1021)	**2-** Frequency of invalid values contained in qualitative data elements
	Invalid values for quantitative data elements (ID-1024)	**3-** Frequency of invalid values contained in quantitative data elements
Plausibility	Freedom of Contradiction (ID-1003)	**4-** Men with malignant neoplasms of the female genital organs (C51-C58) as a hospital diagnosis **5-** Women with malignant neoplasms of the male genital organs (C60-C63) as a hospital diagnosis **6-** Men with in-patient delivery (birth) as admission reason **7-** Patients with a difference between the maximum and minimum age within the study period greater than 3 years (study duration)

SQL queries against the i2b2 database resulting in aggregated results. Data use within the project was approved by the ethics commissions of all involved MII sites lead by the Mannheim University Medicine Ethics Board (2018-864R-MA).

Results

Cohort description

The dataset consisted of 152,806 inpatient cases assigned to 94,570 patients. 25,238 patients (26.7%) had more than one case assigned to them (ranging from 1 to 66 cases assigned to a single patient during the observation period: 2015-2017).

Data quality analysis

The DQA results are displayed in

Table 2.

Table 2 - DQA results

DQA-Dimensions	Items	Results
Completeness	1	2532 missing values (1.6%) in the variable “Postal code” 37 missing values (0.02%) in the variable “ICD codes of the principal hospital diagnosis”
Conformance	2	178 invalid codes (0.12%) in the variable “Reason for the hospital admission”
	3	Range of patient ages at the hospital admission: 0-116
Plausibility	4	1 case
	5	5 cases
	6	0 cases
	7	2 patients

Conclusion

The results showed that a subset of 1.6% of cases did not contain a postal code, which, however, can be traced back to foreign patients or homeless patients for which zip codes are either unavailable or are excluded from documentation within the definition of the §21 dataset. In addition a small subset of cases did not contain a diagnosis (0.02%) or contained truncated or invalid codes not listed in the ICD10-GM classification (0.12%). The plausibility analysis revealed only a very small amount (single-digit) of inconsistent cases (see

Table 2). These first results address the feasibility of constructing and implementing a structured DQ analysis model focusing on quality aspects of data completeness, data conformance and data plausibility of the datasets used in the Demonstrator study. In the next steps we plan to implement a visualization of the DQ analysis using the R statistics software. The resulting scripts will be deployed to all MII sites participating in the MII Demonstrator study. This will allow us to carry out and compare DQ analysis based on the datasets from each of the participating MII locations. Further investigation will be perfomed to identify the reasons behind the formation of the problematic cases and to decide about how to handle the given cases before the datasets are used for statistical analysis.

Acknowledgements

This work was funded in part by the German Federal Ministry of Education and Research within the Medical Informatics Initiative, Grant IDs 01ZZ1801A (MIRACUM), 01ZZ1609A (SMITH), 01ZZ1802A (HiGHmed).

References

[1] S. Gehring, and R. Eulenfeld, German Medical Informatics Initiative: Unlocking Data for Research and Health Care. *Methods of information in medicine* 57 (2018), e46-e49.

[2] Institute for Reimbursement System in Hospital. *Data according to §21 KHEntgG - Version 2018 for the data year 2017*, http://www.webcitation.org/74AR3gjC0, Accessed on March 22, 2019.

[3] M.G. Kahn, T.J. Callahan, J. Barnard, A.E. Bauck, J. Brown, B.N. Davidson, H. Estiri, C. Goerg, E. Holve, and S.G. Johnson, A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. eGEMs 4 (2016).

[4] M. Nonnemacher, D. Nasseh, and J. Stausberg, Data quality in medical research, *Medical Scientific publishing company* (2014).

Address for correspondence

Gaetan Kamdje Wabo,

Department of Biomedical Informatics of the Heinrich-Lanz-Center. Medical Faculty Mannheim of the University of Heidelberg,

Theodor-Kutzer-Ufer 1-3,

68167 Mannheim, Germany.

Email: gaetankamdje.wabo@medma.uni-heidelberg.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190509

An Adversorial Approach to Enable Re-Use of Machine Learning Models and Collaborative Research Efforts Using Synthetic Unstructured Free-Text Medical Data

Suranga N. Kasthurirathne[a,b], Gregory Dexter[a], Shaun J. Grannis[a,c]

[a] *Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, Indiana, USA*
[b] *Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, Indiana, USA*
[c] *School of Medicine, Indiana University, Indianapolis, IN, USA*

Abstract

We leverage Generative Adversarial Networks (GAN) to produce synthetic free-text medical data with low re-identification risk, and apply these to replicate machine learning solutions. We trained GAN models to generate free-text cancer pathology reports. Decision models were trained using synthetic datasets reported performance metrics that were statistically similar to models trained using original test data. Our results further the use of GANs to generate synthetic data for collaborative research and re-use of machine learning models.

***Keywords*:**

Neural Networks (Computer); Machine Learning; Dataset

Introduction

Large scale adoption of Health Information Systems (HIS), together with the rapid evolution of Artificial Intelligence (AI) and various analytical and machine learning toolkits have led to the widespread development of machine learning solutions to address various healthcare challenges using patient data. However, legislation on sharing of Patient Health Identifiers (PHI) restricts researchers from (a) re-using machine learning solutions across larger audiences, (b) fostering inter-organizational collaboration addressing various healthcare challenges, and (c) building generalized machine learning models targeting larger, diverse populations.

Current efforts to enable better data sharing focus on de-identification efforts, where PHI is scrubbed from patient data. However, de-identified free-text data may be vulnerable to re-identification based on clinical data elements. In contrast, alternate approaches to create synthetic data that mimic clinical patterns in data present considerably lower re-identification risk [1]. We leverage recent advances in Generative Adversarial Networks (GAN) [2] to produce synthetic unstructured free-text medical data, and assess (a) possibility of using these datasets to replicate machine learning results generated using original patient data, and (b) levels of re-identification risk posed by these synthetic datasets.

Methods

We leveraged a convenience sample of 7,000 free-text pathology reports on potential cancer cases from the Indiana Network for Patient Care (INPC) [3], a statewide Health Information Exchange (HIE) to build decision models capable of identifying positive cancer cases for public health reporting. Positive and negative report sets were extracted, and used to train SeqGAN [2] models of varying epoch sizes. We selected optimal GAN models for positive and negative cancer report sets by comparing synthetic data generated by these models with original test data using Bilingual Evaluation Understudy (BLEU) scores. We created vectors by counting presence of each stemmed feature in positive and negative contexts across each report in the report set. Next, we developed decision models to predict cancer cases using the Random Forest classification algorithm [4] and the top 5, 10, 20 and 50 features selected from the original test and synthetic feature sets using the Kullback-Leibler divergence (information gain) method [5].

We compared the performance of these models using sensitivity, specificity, F1-measure (harmonic mean between precision and recall) and area under the ROC curve values (AUC), together with their 95% confidence intervals. We assed re-identification risk for presence disclosure [6], where attackers in possession of a set of patient records can determine if any of them were used to train GAN models by comparing these records against the synthetic patient dataset using Hamming scores, a measure of variation between two binary strings.

Results

The 7,000 free-text cancer cases consisted of 1,950 (27.86%) positive reports and 5,050 (72.14%) negative reports [7]. Comparison of BLEU scores identified models trained for 70 epochs as the optimal synthetic data generation model for both positive and negative cancer reports. We extracted the top 50 features from each of the original and synthetic datasets using information gain scores (*Table 1*). Feature selection identified a 36% overlap between the top 50 features extracted from each dataset (*Table 2*).

Figure 1 presents variance of information gain scores across each of the top 50 feature sets. Decision models trained using the top 5, 10, 20 and 50 features extracted from the synthetic and original datasets reported performance metrics that were significantly high (sensitivity: 77-92%, specificity: 95.7-99.8%, F1-measure: 91-97%, AUC: 90-99%). Further, there was no statistically significant difference between many performance metrics reported by models trained using original or synthetic datasets. Presence disclosure tests performed using Hamming score comparisons indicated relatively low probability for re-identification across positive and negative synthetic reports.

Table 1 - List of top 20 stemmed features selected from the original and synthetic datasets using information gain scores.

Rank	Original dataset	Synthetic dataset
1	Tumor	consult
2	Carcinoma	slide
3	Invasion	node
4	Slide	lymph
5	Cell	malign
6	Metastat	grade
7	Lymph	right
8	Node	pathologist
9	Return	submit
10	adenocarcinoma	prostat
11	Margin	collect
12	involve	carcinoma
13	Consult	section
14	differenti	receiv
15	mass	left
16	cassett	specimen
17	phone	posterior
18	left	surgic
19	grade	identifi
20	microscop	tumor

Table 2 - Intersection of top 5, 10, 20 and 50 features selected from the original and synthetic datasets using information gain scores.

Feature subset size	# features overlap	List of features present in both datasets
5	1 (20%)	slide
10	3 (30%)	slide, lymph, node
20	8 (40%)	slide, consult, node, grade, tumor, lymph, left, carcinoma
50	18 (36%)	note, section, prostat, cassett, slide, malign, consult, node, grade, tumor, right, margin, phone, submit, case, lymph, carcinoma, left

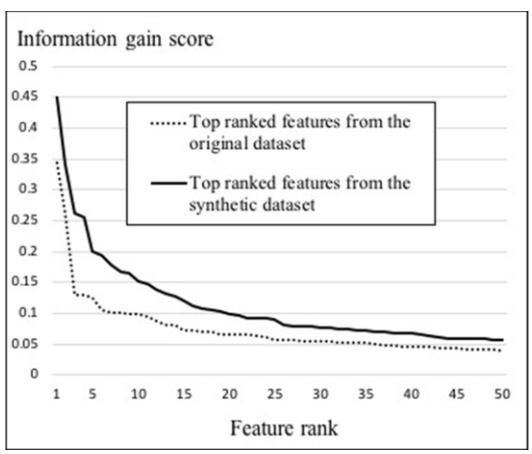


Figure 1 - Variance of information gain scores reported by the top 50 original and synthetic features.

Conclusions

Our results indicate that GAN methodologies can generate synthetic free-text medical data with limited re-identification risk, and that synthetic datasets can be used to develop machine learning models with statistically similar performance metrics to decision models trained using original test data. As such, they are of considerable importance for enabling cross-institutional collaboration and broader dissemination of machine learning models. Adoption of GAN models alone does not result in de-identified data. However, synthetic data generation reduces re-identification risk by creating new patient records with similar, but different data. It also removes any 1-to-1 mapping between test and synthetic reports. Our results demonstrate that synthetic datasets poise a significantly lower chance of re-identification based on clinical information. However, synthetic data produced by these efforts must undergo rigorous de-identification of PHI elements before they can be distributed for public use. Future research includes use of GAN models to create truly de-identified synthetic free-text data that does not require additional de-identification, and expansion of our work across other more challenging healthcare datasets.

We propose the following hypothetical scenario to demonstrate how our approach could be applied in a real-life setting. An organization that possesses rich free-text data sources, but lacks adequate machine learning expertise can leverage our approach to create synthetic data. They de-identify and share the synthetic data with experts who use it to build machine learning models. Once optimal models have been identified, they can be implemented across the original dataset with compatible performance measures.

References

[1] E. Choi, S. Biswal, B. Malin, J. Duke, W.F. Stewart, and J. Sun, Generating multi-label discrete patient records using generative adversarial networks, *arXiv preprint arXiv:1703.06490* (2017).

[2] L. Yu, W. Zhang, J. Wang, and Y. Yu, SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient, in: *AAAI*, 2017, pp. 2852-2858.

[3] C.J. McDonald, J.M. Overhage, M. Barnes, G. Schadow, L. Blevins, P.R. Dexter, B. Mamlin, and I.M. Committee, The Indiana network for patient care: a working local health information infrastructure, *Health affairs* **24** (2005), 1214-1220.

[4] L. Breiman, Random forests, *Machine learning* **45** (2001), 5-32.

[5] D. Polani, Kullback-leibler divergence, *Encyclopedia of Systems Biology* (2013), 1087-1088.

[6] M.E. Nergiz and C. Clifton, δ-presence without complete world knowledge, *IEEE Transactions on Knowledge and Data Engineering* **22** (2010), 868-883.

[7] S.N. Kasthurirathne, B.E. Dixon, J. Gichoya, H. Xu, Y. Xia, B. Mamlin, and S.J. Grannis, Toward better public health reporting using existing off the shelf approaches: A comparison of alternative cancer detection approaches using plaintext medical data and non-dictionary based feature selection, *Journal of biomedical informatics* **60** (2016), 145-152.

Address for correspondence

Suranga N. Kasthurirathne, PhD; E-mail: snkasthu@iu.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190510

AutoScribe: Extracting Clinically Pertinent Information from Patient-Clinician Dialogues

Faiza Khan Khattak[a,b*], Serena Jeblee[a,b*], Noah Crampton[c], Muhammad Mamdani[c], Frank Rudzicz[a,b,c,d]

[a] *Department of Computer Science, University of Toronto, Toronto, Ontario, Canada (*equal contribution)*
[b] *Vector Institute for Artificial Intelligence, Toronto, Ontario, Canada*
[c] *Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, Ontario, Canada*
[d] *Surgical Safety Technologies Inc, Toronto, Ontario, Canada*

Abstract

We present AutoScribe, a system for automatically extracting pertinent medical information from dialogues between clinicians and patients. AutoScribe parses the dialogue and extracts entities such as medications and symptoms, using context to predict which entities are relevant, and automatically generates a patient note and primary diagnosis.

Keywords:

Medical Records, Machine Learning, Medical Informatics

Introduction

Currently, clinicians spend up to 50% of their time entering information from patient interviews into electronic medical records (EMRs) [1]. This reliance on slow, laborious, and inconsistent data entry results in wide variability in the quality of EMR data [2], which presents a challenge to clinical data analytics [3]. Recent machine learning (ML) algorithms, such as recurrent neural networks and word embeddings [4], have been applied to tasks such as disease and mortality prediction from EMR data [5,6]. This suggests that a significant portion of clinical data entry can be automated by analyzing patient-clinician dialogues.

Here, we optimize an ML model, *AutoScribe,* to classify dialogue phrases from patient interviews as contextually pertinent to clinical documentation, which is the foundational step to generating EMR data from the analysis of patient-clinician dialogues. We extract medically relevant entities such as signs, symptoms, diagnoses, therapies, and referrals through natural language processing. Unlike systems which primarily use lexicon-based term matching, our system also uses linguistic context and time information.

Data

The data consists of 800 audio patient-clinician dialogues and their transcripts, purchased from Verilogue Inc[1], including primary diagnosis codes. The most frequent are *ADHD*, *COPD*, *depression*, and *influenza.*

We developed a new annotation tool and are doubly annotating all dyads for relevant medical entities. Of the 30 dialogues that have been completed, the annotations have .53 agreement (Krippendorff's alpha [8]) and .80 partial match F_1 score. We also have 302 dialogues with annotations from one physician at present. We present a synthetic patient-clinician dialogue in Figure 1 with the output of our system compared to human annotation.

Methods and Results

The AutoScribe system currently consists of several modules. The cumulative output of these models constitutes the initial AutoScribe system. We evaluate each component of the system using F_1 measure, considering tags that overlap with the human annotation as correct. For entity tagging, we also calculate inter-annotator agreement between the physicians and the automatic pipeline using Krippendorff's alpha [8]. All but the utterance type classification model are evaluated on 302 conversations.

Utterance type classification

Each utterance in the dialogue is automatically labeled as a *question, statement, positive answer, negative answer, backchannel* or *excluded.* We use a two-layer bidirectional gated recurrent unit (GRU) neural network [12], implemented in PyTorch. Each word is represented as a 200-dimensional vector using the freely available Wikipedia-PubMed word embedding model[2]. We evaluate the utterance type classifier on 20 conversations, annotated independently by 2 annotators with inter-annotator agreement of .77 (Cohen's kappa). For training, we use two external, publicly available datasets: the Switchboard corpus [10], and the AMI corpus[3]. Our model achieves .71 F_1 score on Verilogue data.

Time expression identification

Phrases in the dialogue that reference absolute and relative times and dates are automatically tagged and converted to standardized values using HeidelTime [11], a freely available temporal tagger. For example, in a document dated Jan 1, 2018, the phrase *tomorrow* would be normalized to *2018-01-02*.

Medical entity identification

AutoScribe currently identifies the following medical concepts: anatomical locations, signs and symptoms, diagnoses, medications, referrals, investigations and therapies, and reasons for visit. The identification uses lexicon look-up using terms from BioPortal[4], Consumer Health Vocabulary (CHV)[5], SNOMED-CT[6], and RxNorm[7], and achieves an average F_1 score of .63 and .55 Krippendorff's alpha. Entity identification is currently limited to the terms present in our reference lists, which are large but cannot cover all possible expressions of relevant entities. There may be many valid variations of these entities that we hope to be able to identify in the future.

1 http://www.verilogue.com
2 http://bio.nlplab.org/
3 http://groups.inf.ed.ac.uk/ami/corpus/
4 https://bioportal.bioontology.org/ontologies
5 http://consumerhealthvocab.chpc.utah.edu/CHVwiki/
6 http://www.snomed.org/
7 https://www.nlm.nih.gov/research/umls/rxnorm/

DR: *How's the [numbness in your toes]*$_{\text{Sign/Symptom}}$ */[toes]*$_{\text{Anatomical Location}}$ *?* ***PT:*** *The same. I'm used to it by now.* ***DR:*** *Okay, that's good. Let's keep you on the [same dose of Metformin]*$_{\text{Medication}}$ *[for now]*$_{\text{TIMEX3}}$ *then we'll check your [a1c]*$_{\text{Investigation/Therapy}}$ *again [in three months]*$_{\text{TIMEX3}}$ *, and then I'll [see you back here after that]*$_{\text{Disposition plan}}$.	***DR:*** *How's the numbness in your [toes]*$_{\text{Anatomical Location}}$ *?* ***PT:*** *The same. I'm used to it by [now]*$_{\text{TIMEX3}}$ *.* ***DR:*** *Okay, that's good. Let's keep you on the same [dose]*$_{\text{Medication}}$ *of [Metformin]*$_{\text{Medication}}$ *for [now]*$_{\text{TIMEX3}}$ *then we'll check your a1c again in [three months]*$_{\text{TIMEX3}}$ *, and then I'll see you back here after that.*

*Figure 1 – Example dialogue - **(a)** Human annotation. **(b)** Automatic annotation. In both **1a** and **1b**, highlight indicates the annotated entities; darker highlights indicate overlap between human and automatic annotations. Subscripts indicate the entity type (TIMEX3 indicates time phrases).*

Attribute classification

Once the entities have been identified, the system should determine which are actually pertinent to the diagnosis. For instance, a physician or patient might mention a medication that they have never actually taken, so the system should not record that medication as part of the patient's history. Currently, we classify two attributes: *modality* and *pertinence*. The modality indicates whether the event actually occurred (actual, negative, possible), and pertinence indicates the condition to which the entity is medically relevant (i.e., *ADHD*, *COPD*, *depression*, *influenza*, other). The attribute classifier is a support vector machine (SVM) trained with stochastic gradient descent [9].

Each medical entity is represented as the average word embedding, concatenated with the word embeddings for the previous and next 5 words. We also include the speaker code of the utterance in which the entity appears. The system achieves .77 F_1 score for modality classification, and .62 for pertinence. Pertinence classification currently performs worse than modality, perhaps because it requires more global information.

Primary diagnosis classification

We classify the primary diagnosis on each patient-clinician conversation to be used for billing codes. We train and test the models on a 5-fold cross validation of the 800 dyads. We apply tf-idf on the cleaned text of each dyad and use logistic regression, SVMs, and random forest models. The F_1 scores of classification are calculated based on the human-assigned labels available in the transcription of the conversation's 'primary diagnosis' field. Diagnosis classification currently handles 6 classes only and does not account for conditions other than the primary diagnosis that may be discussed in the conversation.

F_1 scores (Linear SVM**):** *Influenza .93*±.04, *ADHD .83*±*.05*, *COPD .68*±*.14, Osteoporosis .78*±.04, *Type II diabetes .76*±.07*, Depression .71*±*.08,* and *Other .76*±*.05.*

Discussion & Conclusion

We have presented a novel approach to clinician-patient dialogue parsing, whose outputs are oriented toward pragmatic linguistic features, and the needs of clinicians. Specifically, we have developed machine learning models based on recurrent neural networks that extract medical linguistic entities and their time-based contextual partners, as well as primary diagnoses from dialogue. Future directions include extracting other key contextual entities within clinical dialogues that are pertinent to clinical documentation, such as quantity, quality, and severity words and phrases, as well as accounting for similar medical terms and spelling variations. Training will be expanded to include more entities, more conversations, more diagnoses, and multiple diagnoses per conversation.

Acknowledgments

Research Ethics Board approval from St Michael's Hospital (REB # 18-082) and the University of Toronto (REB# 00036367).

References

[1] Sinsky C. et al. Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties. *Ann Intern Med.* 2016 Dec 6; 165 (11): 753-760. doi: 10.7326/M16-0961

[2] Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. *Br Med J*. 2003; 326(7398): 1070–2.

[3] Roth CP, et al. The challenge of measuring quality of care from the electronic health record. *American Journal of Medical Quality* 2009; 24(5): 385–94. http://dx.doi.org/10.1177/1062860609336627. PMid:19482968.

[4] Mikolov T, Sutskever I, Chen K, Corrado G, and Dean J. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2* (NIPS'13) 2013; 3111-3119.

[5] Rajkomar A, Oren E, Chen, K, Dai AM, Hajaj N, Liu PJ, … Dean J. (2018). Scalable and accurate deep learning for electronic health records. NPJ Digital Medicine (2018): 1–10. https://doi.org/10.1038/s41746-018-0029-1

[6] Mullenbach J, Wiegreffe S, Duke J, Sun J, and Eisenstein J. Explainable prediction of medical codes from clinical text. In *Proceedings of NAACL-HLT* 2018; 1101–1111.

[7] Liu PJ. Learning to write notes in electronic health records. *ArXiv preprint 1808.02622v1*, 2018.

[8] Krippendorff K. Content Analysis: An Introduction to Its Methodology, Chapter 11. Sage, Beverly Hills, CA, 2004.

[9] Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, and Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2014): 1724–1734.

[10] Calhoun S, Carletta J, Brenier J, Mayo N, Jurafsky D, Steedman M, and Beaver D. The NXT-format Switchboard Corpus: A Rich Resource for Investigating the Syntax, Semantics, Pragmatics and Prosody of Dialogue. *Language Resources and Evaluation* 2010; 44(4), 387-419, doi:10.1007/s10579-010-9120-1.

[11] Strötgen J and Gertz M. Multilingual and cross-domain temporal tagging. *Language Resources and Evaluation* 2013; 47(2):269–298.

[12] Pedregosa F, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 2011; 12: 2825-2830.

Address for correspondence

Faiza Khan Khattak. Email: faizakk@cs.toronto.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190511

Development of a Common Data Model Facilitating Clinical Decision-Making and Analyses

Eizen Kimura[a], Hideo Suzuki[b]

[a] *National Institute of Public Health, Wako, Saitama, Japan,*
[b] *MoDeL Corp., Yokohama, Kanagawa, Japan*

Abstract

Currently, the Common Data Model (CDM) for primary use (as distinct from models designed for secondary use) poorly supports clinical decision-making and medical process analyses. We designed a CDM featuring a search flow that identifies facts after defining the clinical process, and we make the data definition language of the CDM employed freely available as open source.

Keywords:

Data warehousing, clinical decision support systems, data mining

Introduction

To allow medical records to be used in clinical research, Common Data Models (CDMs) have been developed by several groups; the models include Informatics for Integrating Biology and the Bedside (i2b2) [1], Observational Health Data Sciences and Informatics (OHDSI) [2], and the Cancer Biomedical Informatics Grid (caGRID) [3]. The i2b2 allows the research community to find interesting patients by reviewing electronic records while preserving patient privacy. The OHDSI features an international network of real-world data facilitating collaborative observational studies. These CDMs feature tables with relatively few columns and shallow hierarchical star schemas; they embrace an Entity-Attribute-Value (EAV) philosophy allowing rapid exploratory interrogation. While CDMs for secondary use are widespread, few CDMs for primary use supporting clinical decision-making and analyses are available. CDMs for primary use should essentially be the good source of business intelligence tools, must be readable by clinicians, and allow facile interrogation.

It is first necessary to minimize the complexity of interrogation (thus enhancing usability) by ensuring that the database is appropriately denormalized. Usually, the normalization of relational databases is considered desirable. However, as the number of tables increases, querying becomes difficult. In addition, the CDM must feature extensive analyses axes, allowing analysis of various medical processes.

Arden Syntax etc. has been developed as Domain Specific Language (DSL) [4] for medical knowledge supporting clinical decision-making. However, the "curly brace problem" is in play; no standard data retrieval method is yet available [5]. Queries depend on the database structure and the terminology used by the system; it is difficult to port a given logic to another clinical decision-making system [6]. In addition, readability declines if system-specific descriptions become admixed, rendering it increasingly difficult for medical professionals who are not familiar with computers to read the query logic [7]. Although FHIR [8] uses a standard query method to explore routine information, and openCDS [9] attempts to standardize both the data and the API of the clinical support system, neither is easy to use and the data can be employed only to aid in clinical decision-making. Therefore, a CDM that can retrieve medical information after receipt of a simple query is required. A few previous CDMs sought to attain this goal; here, we present a CDM for primary use that is easy to use.

Methods

HL7 messages are associated with prescriptions, injections, radiographic procedures, and specimen examinations. During DWH construction, many ordered HL7 message tables must be created to allow full exploration of the clinical processes used. The number of tables increases as the clinical process of interest becomes more complex, negatively affecting query structure. We built a database that aggregates clinical processes, identifies medical procedures requiring analyses, and then outputs facts. As analyses of clinical decision-making and the associated financial implications require answers to many time-sensitive questions, we optimized the table index to facilitate time-series retrieval. To facilitate analyses of the business processes of an organization, we code names, places, organizations, and teams in a standardized manner; allowing searching based on organizational role and function. To optimize performance, we create specialized tables and indices for specific analysis without creating SQL views; we intentionally allow data duplication and schema denormalization. During the Extract-Transform-Load (ETL) process, we standardize the notation, units, and ranges; searching is consistent across different electronic medical record systems. In addition, traceability is ensured to prevent tampering.

Results

The CDM is currently managed by the Semantic Data Model (SDM) consortium [10]; all specifications are free to all consortium members. We have published the Data Definition Language (DDL) for the relational database systems SQLite, MySQL, PostgreSQL, Microsoft SQL Server, Oracle, and DB2, and the source code of the DDL generator at GitHub [11]. As of October 2018, SDM ver 1.09 features 91 tables, 2,833 unique columns, and 40 common columns allowing construction of dimensions.

Acts (including orders) are collected in the SDM_ORDER table. The SDM_CALENDAR table lists all events during hospitalization, with the dates of major events (including initial hospitalization, discharge, and surgery) being linked, in a simple manner, to specific events. The SDM_INDEX table lists all events and is key when analyzing medical processes. The SDM_KEYS table recodes the editing histories of medical records, ensuring authenticity (the table is an audit trail). Whenever the text of a medical record changes, a hash value is calculated and added to the SDM_KEYS table with 50 attributes such as document related attributes, creators, and modification dates. The SDM_REPORT table manages the links among reports allowing medical staff to view all relevant reports contemporaneously. Thus, tables associated with clinical processes are centralized and factual tables are placed at the periphery; the structure of the database is a shallow star schema (Fig. 1).

A single prescription/injection creates one record row in a table. However, common laboratory data are collectively stored in a single row. Since the results of the laboratory exams often evaluate multiple items at the same time, consideration is given so that a plurality of sample inspection results can be taken out collectively with one query.

The codes use standard terminologies such as the JAHIS Protocol for Clinical Laboratory Data Communication (equivalent to LOINC), the Standard Master for Pharmaceutical Products (equivalent to RxNorm), and the ICD10-based Standard Disease Code Master for the ETL process.

All tables feature 40 common columns storing data on actors, wards, dates, and times, facilitating multi-dimensional analyses. As of 2018, the SDM consortium cooperates with the electronic medical record vendors Canon, Fujitsu, Hitachi, IBM, NAIS, NEC, SSI, and Softmax.

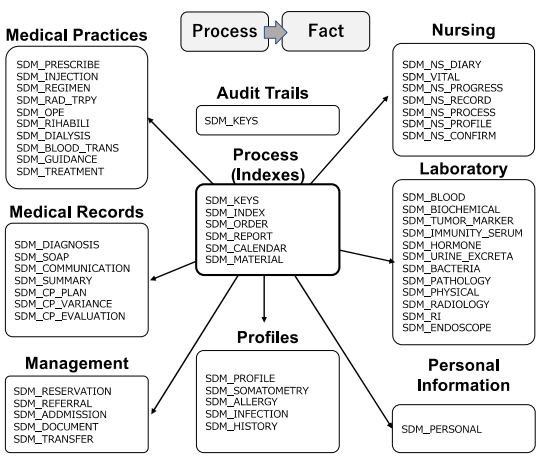


Fig1-The Overview of the Semantic Data Model ables

Discussion

Our CDM includes only information required in medical practice, thus not the items of standard information models such as the FHIR. As the data structure differs for each vendor of electronic medical records, we developed an independent ETL process featuring a trade-off between developmental cost and completion of all SDM tables. This study has limitations that it has no quantitative evaluation of the proposed model yet such like the time for building ETL, the concept coverage rate compared with another CDMs, the effectiveness of denormalization on the degree of contribution to simplifying the queries on CDM. We will harmonize the standard medical information models with the SDM tables. Thus, data conforming to the standard model for medical information is output by the electronic medical record system, the need for further ETL development is reduced, and the fulfillment of SDM items will increase.

Conclusions

In the future, we will seek to conform to the FHIR which is increasingly the accepted model of standard information.

References

[1] S.N. Murphy, G. Weber, M. Mendis, V. Gainer, H.C. Chueh, S. Churchill, and I. Kohane, Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2), Journal of the American Medical Informatics Association 17 (2010), 124-130.

[2] G. Hripcsak, J.D. Duke, N.H. Shah, C.G. Reich, V. Huser, M.J. Schuemie, M.A. Suchard, R.W. Park, I.C.K. Wong, and P.R. Rijnbeek, Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers, Studies in health technology and informatics 216 (2015), 574.

[3] S. Oster, S. Langella, S. Hastings, D. Ervin, R. Madduri, T. Kurc, F. Siebenlist, P. Covitz, K. Shanbhag, and I. Foster, caGrid 1.0: A grid enterprise architecture for cancer research, in: AMIA Annual Symposium Proceedings, American Medical Informatics Association, 2007, p. 573.

[4] A. Van Deursen, P. Klint, and J. Visser, Domain-specific languages: An annotated bibliography, ACM Sigplan Notices 35 (2000), 26-36.

[5] H.C. Karadimas, C. Chailloleau, F. Hemery, J. Simonnet, and E. Lepage, Arden/J: an architecture for MLM execution on the Java platform, Journal of the American Medical Informatics Association 9 (2002), 359-368.

[6] T.A. Pryor and G. Hripcsak, Sharing MLM's: an experiment between Columbia-Presbyterian and LDS Hospital, in: Proceedings of the Annual Symposium on Computer Application in Medical Care, American Medical Informatics Association, 1993, p. 399.

[7] J. Choi, Y.A. Lussier, and E.A. Mendonça, Adapting current Arden Syntax knowledge for an object-oriented event monitor, in: AMIA Annual Symposium Proceedings, American Medical Informatics Association, 2003, p. 814.

[8] D. Bender and K. Sartipi, HL7 FHIR: An Agile and RESTful approach to healthcare information exchange, in: Computer-Based Medical Systems (CBMS), 2013 IEEE 26th International Symposium on, IEEE, 2013, pp. 326-331.

[9] K. Kawamoto, OpenCDS, in, http://www.opencds.org/, 2010.

[10] SDM Consortium, in, https://sdm-c.org/schedule.html, 2014.

[11] E. Kimura, DDL Generator and DDL for Semantic Data Modeling v.1.09, in, https://github.com/kandalva/SDM_CDM, 2018.

Address for Correspondence

Eizen Kimura, kimura.e.aa@niph.go.jp , ekimura@m.ehime-u.ac.jp
2-3-6 Minami, Wako-shi, Saitama 351-0197 Japan

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190512

Scientific Challenge in eHealth: MAPPATHON, a Metadata Mapping Challenge

Ann-Kristin Kock-Schoppenhauer[a], Philipp Bruland[b], Dennis Kadioglu[c], Dominik Brammen[d], Hannes Ulrich[a], Kerstin Kulbe [e], Petra Duhm-Harbeck[a], Josef Ingenerf[a,e]

[a] IT Center for Clinical Research, Lübeck, University of Lübeck, Germany
[b] Institute of Medical Informatics, University of Münster, Germany
[c] Medical Informatics Group, University Hospital Frankfurt, Frankfurt am Main, Germany
[d] Department of Anesthesiology, University of Magdeburg, Germany
[e] Institute of Medical Informatics, University of Lübeck, Germany

Abstract

Scientific challenges based on benchmark data enable the comparison and evaluation of different algorithms and take place regularly in scientific disciplines like medical image processing, text mining or genetics. The idea of a challenge is rarely applied within the eHealth community. Mappathon is a metadata mapping challenge that asks for methods to find corresponding data elements within similar datasets and to correlate data elements among each other.

Keywords:
Metadata, Telemedicine, Benchmarking

Introduction

To enable secondary use of routinely collected medical data, it is necessary to create a general understanding of given information. As a common practice, this understanding is achieved through metadata about data elements and the interconnections between data elements. Metadata can be stored in so-called metadata repositories (MDR). The functionalities of such an MDR include pure storage, administration and other specific metadata functionalities like matching and mapping. Mapping means transforming instance data using mapping rules. The technical term matching is used for searching corresponding, semantically similar data elements at the schema level. Matching and mapping rules are difficult to determine and often require manual effort. Therefore, there is a great need for advanced data analysis techniques promoting the definition of matchings and mappings, where matching refers to the discovery of related or equivalent metadata and mapping to the relationship between data elements such as conversion rules. We organized a scientific challenge enabling competition between approaches for determining automatically related data elements [1].

Methods

The challenge was organized in two phases: The training phase, where datasets have been published and a restful service for the evaluation of the results have been implemented, and the test phase, where the algorithms have been compared. Within the training phase, participants were invited to download the training dataset consisting of metadata about eCRFs and the included data elements as well as a gold standard mapping of related data elements created by medical experts. This allowed the validation and optimization of algorithms and methods. Any automatic method that predicts the valid mapping was asked for. There was no restriction on new, innovative or unpublished methods and no limitations on including external information like terminologies. Participants were explicitly invited to use of coding systems and terminology servers. The Mappathon Challenge provided two tasks, which were independently evaluated:

1. The first task, multilabel classification, is the problem of classifying source data elements to one or more target data elements.
2. The second task, multiclass classification, is the problem of classifying source data elements in specific relation to (multiple) target data elements.

Data

We provided no instance level data, but only metadata sets in the German language. Dealing with German metadata is one of the main reasons for this challenge because most existing and freely available datasets (e.g., MIMIC II) are in English. At the same time, most available lexical or text mining tools do not fully support the analysis of German medical texts.

All training datasets are at least partly annotated with UMLS CUIs, but there were no annotations in the test dataset. For the challenge, datasets of routine documentation and clinical research are provided by the Portal of Medical Data Models [2]. Curated training datasets have been made available for download in different formats like FHIR questionnaires or CDISC ODM. All eCRFs refer to the use of case Emergency Department Visit. We provided seven training datasets consisting of 950 data elements and three test datasets with 579 data elements.

	relatedto	equivalent	equal	wider	narrower	undefined
relatedto	1,00	0,30	0,30	0,30	0,30	-1,00
equivalent	0,40	1,00	0,80	0,60	0,60	-1,00
equal	0,40	0,80	1,00	0,40	0,40	-1,00
wider	0,50	0,60	0,40	2,00	0,30	-1,00
narrower	0,50	0,60	0,40	0,30	2,00	-1,00

Figure 1 -- Evaluation matrix to calculated the Mappathon Score (left side: defined mapping, right side: gold standard mapping. Row names correspond to the equivalence classes referred to FHIR ConceptMaps.

Equivalence Classes

The Evaluation Classification of the Mappathon is based on the FHIR v3.0.1 ConceptMap [3], which uses the ConceptMapEquivalence value set. The number of classes has been significantly reduced and the classes are being utilized on data element level and not on the intended concept level.

Mapping

The validation was carried out on the basis of elaborately developed manual mappings (ground truth) in cooperation with clinical partners, see Fig. 2. The mappings are directional - from the source to the destination. Nevertheless, in many cases, reverse mappings are valid as well. Not all data sets were compared to each other, but suitable preselection took place. However, the manual mapping resulted in an overall of 104,207 possible mappings to consider in the training sets and 110,936 in the test sets. The quantitative output of valid mappings is less. The experts could identify 417 actual mappings in the training data and 419 in the test datasets.

Validator

The mapping results could be evaluated using the Online Mappathon Validator. The Mappathon challenge provided a RESTful interface for uploading mapping results under https://validate.mappathon.de and an additional client available on the ITCR GitHub repository [4]. The Mappathon Validator calculates the Zero-one Classification Loss and the Mappathon Score according to the published evaluation matrix, see Fig. 1.

Results

Six teams have registered to the challenge. During the on-site workshop held at the 63rd Annual Meeting of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS) e. V. in Osnabrück, a set of test cases was released of which participants were asked to apply their

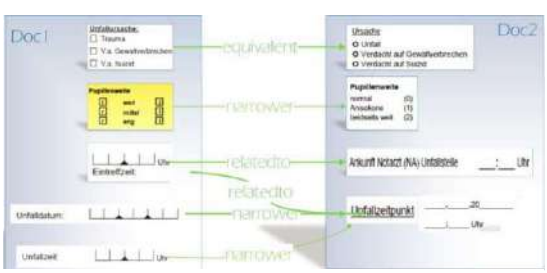


Figure 2 – Defined mapping according to the equivalence classes between several data elements of two datasets.

algorithm on and upload their mapping results. In the workshop, the teams finally had the opportunity to present and discuss their respective results and their methods in particular. Based on the evaluation matrix, which was also published at the beginning, the organizer selected the team whose results most closely corresponded to the gold standard. Both metrics have been calculated for each test dataset. The results are shown in Tab. 1. In addition, the workshop participants had the opportunity to vote on which of the shown solutions they considered to be the most innovative and original. The award for the best mapping went to the team "MDRCupid" consisting of Noemi Deppenwiese and Hannes Ulrich from the University of Lübeck and the prize for the most innovative solution to the team "Marvelous Mappers" consisting of Michael Storck, Philipp Neuhaus and Stefan Hegselmann from the University of Münster.

Table 1– Overvie of best Mappathon results in 2018.

Datase	Zero-One-	M-
TEST1	0,18	-8,7
TEST2	0,16	2,4
TEST3	0,0	-1,0

Discussion

Various datasets and the related heterogeneity make it nearly impossible to compare different approaches in a fair way. By providing a high-quality dataset publicly as well as pre-defined evaluation rules, this challenge aims to overcome these limitations and to create a common framework for a comprehensible and adequate comparison of results. Before a repetition of the challenge Mappathon, further detailed rules about the aims of the mapping (concept versus data element level) should be developed.

Conclusions

The Mappathon and the contributed solutions are the first steps towards to enable routine data for the secondary use in clinical research. We are planning another Mappathon and are already looking forward to furthering input, interesting solutions, and exciting discussions.

Acknowledgements

We would like to thank the GMDS for making the workshop at the annual conference possible as well as all the professional societies and data stewards to provide us with excellent data sets.

References

[1] A.-K. Kock-Schoppenhauer, Mappathon: A Metadata Mapping Challenge, *Mappathon: A Metadata Mapping Challenge*. https://www.mappathon.de/ (accessed November 19, 2018).

[2] M. Dugas, P. Neuhaus, A. Meidt, J. Doods, M. Storck, P. Bruland, and J. Varghese, Portal of medical data models: information infrastructure for medical research and healthcare, *Database*. **2016** (2016).

[3] ConceptMap - FHIR v3.0.1. https://www.hl7.org/fhir/conceptmap.html (accessed November 16, 2018).

[4] IT Center for Clinical Research-Universität zu Lübeck, MappathonValidatorClient, (2018). https://github.com/itcr-uni-luebeck/MappathonValidator_Client (accessed November 19, 2018).

Address for correspondence

Ann-Kristin Kock-Schoppenhauer
IT Center for Clinical Research, Lübeck
Universität zu Lübeck
Ratzeburger Allee 160, 23562 Lübeck
Telephone: +49 (0) 451 - 500 52702

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190513

Original Laboratory Test Code Mapping System Using Test Result Data on Electronic Health Record

Naoto Kume[a], Kenji Suzuki[b], Shinji Kobayashi[a], Hiroyuki Yoshihara[a], Kenji Araki[c]

[a] Department of Electronic Health Record, Graduate School of Medicine, Kyoto University, Kyoto City, Kyoto, Japan,
[b] SRL, Inc., Hachioji-shi, Tokyo, Japan,
[c] Faculty of Medicine, University of Miyazaki Hospital, Miyazaki City, Miyazaki, Japan

Abstract

Laboratory tests results have potential secondary usage. Each healthcare facility has a laboratory test code. Hence, test code mapping is required to support laboratory technicians. An automatic code mapping can reduce the burden of manual mapping during data preparation. The authors developed a semi-automatic mapping support system that uses the newest test results generated in the electronic health record.

Keywords:
Clinical Laboratory Information Systems, Electronic Health Records

Introduction

Recently, researchers applied various data analysis methods and artificial intelligence algorithms to process and analyze medical data. However, the lack of massive and properly structured can affect the performance of computer-aided data analysis. Several countries implemented Electronic Health Records (EHR) [1,2]. Healthcare facilities might implement various laboratory test coding systems where the same laboratory test might have different codes numbers. The EHR vendors have to maintain and map the laboratory test to each facility coding system that uses their EHR instead of delegating this task to the healthcare facility itself to ensure the correctness of the mapping between the EHR codes and the local coding system. Mapping the local codes that different healthcare facilities use can be valuable for secondary usage and data exchange. Hence automating the mapping of laboratory tests codes can reduce the manual efforts, and each healthcare facility can oversee and manitain the coding to ensure the correctness of the mapping. This study proposes a laboratory codes mapping support system that can reduce the manual mapping process.

Conventionally, the authors defined a common laboratory test code standard due to the lack of pervasiveness of LOINC and JLAC10 in many hospitals [3]. The original code has three segments. The first segment is an accounting system code, which the Ministry of Health and Welfare in Japan defines. The accounting system code reflects the clinical aspect of the test from the physician perspective rather than the technician perspective, which provide over precise classification. The second segment, subcode1, represents the classification of specimens and items. The third segment represents subcode2, the laboratory test characteristics such as standard and unit.

There are two significant factors for the efficiency of the mapping. The first is retrieving the newest facility code, that facility manager prepares and sends it to the EHR manager. Hence, manual mapping time depends on the facility manager. The second factor is different management of mapped facility code. The selection of the expired code, the mapped code, and the new code are the main tasks to generate the specific code that the managers assign to the original unique code.

Therefore, the mapping system has two functions. The system can automatically identify the different unregistered code in the original unique code master. The system can manage all the facility code, including active, rejected, and expired code, to track the mapping history.

Methods

This study applies two steps to create the continuous maintenance of the original unique code master. The first step classifies the facility code using the accounting code in the code master that the facility manager provides. Hence, the maintener can focus on the two subcodes mapping. In the second step, an extraction function in the EHR detects all test results items in the data received from all facilities by daily batch processing. The daily test result items include the newest 100 values. The proposed mapping support system identifies the new items of facilities using the retrieved list.

Consequently, a maintener can proceed with the mapping task using the following three steps;

1. The initial test code mapping: The facility code master is imported to the mapping support system. The mapping system proposes the unique code, which is the candidate of the mapping to the facility codes.
2. Difference detection: The newest test result items list are extracted from the EHR, then the different codes are registered for mapping task waitlist when the test item list are imported into the proposed system.
3. The maintener workflow: After insurance organizations announce a new test item, the maintener waits until the item appears on the EHR test result list. Afterwards, the EHR test result listis imported to the proposed system so that the maintener can start the mapping the subcode without any other preparation.

The proposed system described above is a continuous subcode mapping as a stational process after the initial master code mapping.

Definition of Laboratory test code master

Table 1 shows the definition of the original unique code structure. Accounting code, two subcodes and item name are defined.

Table 1 – Examples of the original unique code definition

Accounting code	Subcode1	Subcode2	Item name
160010010	000	00	HbA1c (NGSP)
160010010	000	01	HbA1c (JDS)
160022510	000	00	AST/GOT
160022610	000	00	ALT/GPT
…	…	…	…

Table 2 shows the definition of the EHR test result item list. OID indicates the hospital that reports the test result. Labtest OID indicates the laboratory that runs the test.

Table 2 – The EHR test result item list definiton

No	Item	Example
1	Organization ID	1.2.840.114319.5.1000.1.50.1
2	Labtest OID	1.2.840.114319.5.1000.1.50.1
3	Facility test code	68001
4	Item name	WBC
5	Reference value	3500—9200
6	Unit	x10^3/muL
7	Report date	2017/1/5 15:48
8	Value	5.8
9	Registered date	2017/1/5 23:59
10	New/Exist flag	New

Mapping Support System

Figure 1 illustrates the structure of the EHR, Electronic Medical Record (EMR) and the proposed system. Daily submitted test results from EMR are accumulated as a list in the EHR. The maintener imports the list into the proposed system.

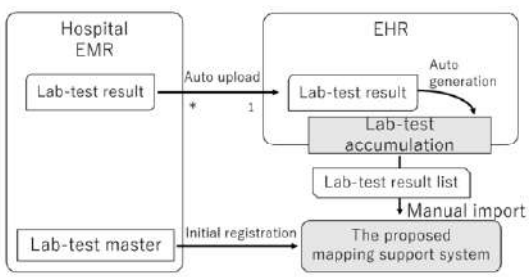


Figure 1- The mapping support system and the relation with EHR and EMR

Results

We compared the manual mapping workflow to the proposed system workflow. We described the manual mapping workflow in seven steps; 1. Comment code omission, 2. Sort by the accounting code to compare the code, 3. List the accounting code classification of the facility code, 4. Subcode mapping, 5. Retreive new master data from the facility manager, 6. Omit flag setup for the mapped code on the newest facility master, 7. Process step 2 to step 4 for the new code.

The proposed system workflow has four steps; 1. Comment code omission, 2. Format the facility code master for import, 3. Import EHR data and facility code master, 4. subcode mapping.

System operation performed all processes on the proposed system workflow. The proposed system reduced the additional wait time such as step 5 on the manual mapping workflow. The maintener only needed to import several files to the system before starting the actual mapping task. Hence, the proposed system drastically reduced mapping preparation tasks.

Conclusions

This study proposes a mapping support system for laboratory test code. The method leverages the EHR to retrieve the newest test item that is not in the original facility code master. The system reduced the burden of manually mapping the target code as well as parallelizing the workflow with several mainteners according to the active, rejected, and expired code management.

Acknowledgements

This research is supported by "ICT infrastructure establishment for clinical and medical research; Sustainable Massive Health and Clinical Data Repository for Secondary Use" from Japan Agency for Medical Research and Development, AMED (17lk1010001h0003).

References

[1] N. Kume, K. Suzuki, S. Kobayashi, K. Araki, and H. Yoshihara, Development of Unified Lab Test Result Master for Multiple Facilities, *Medinfo 2015: Ehealth-Enabled Health* **216** (2015), 1050-1050.

[2] F. Pecoraro, O. Tamburis, and F.L. Ricci, The Use of a Mathematical Model in the Adoption of an Integrated EHR System, *Medinfo 2013: Proceedings of the 14th World Congress on Medical and Health Informatics, Pts 1 and 2* **192** (2013), 1069-1069.

[3] T. Takemura, K. Araki, K. Arita, T. Suzuki, K. Okamoto, N. Kume, T. Kuroda, A. Takada, and H. Yoshihara, Development of Fundamental Infrastructure for Nationwide EHR in Japan, *Journal of Medical Systems* **36** (2012), 2213-2218.

Address for correspondence

Naoto KUME
kume@kuhp.kyoto-u.ac.jp
+81-75-874-7030.
KRP9-506
Awatacho91, Chudoji
Kyoto-city, Kyoto
Japan

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190514

An Ontology for Assessing Health Information Needed During Pregnancy

Joo Yun Lee[a]

[a] *College of Nursing, Gachon University, Incheon, Korea*

Abstract

In this study we developed an ontology for accessing online health information related to pregnancy. Social media data and the categories in the literature on pregnancy information were used to collect terms for identifying class and class hierarchy. The developed ontology included 241 classes and 788 synonyms, with six superclasses. This ontology can be used to provide appropriate information based on a needs assessment.

Keywords:

Consumer health informatics, pregnant women, social networking,

Introduction

Patients and consumers actively use online social media, including online communities, to obtain health-related information and to share experiences and opinions. According to the Pew Internet Report [1], the health information shared most frequently online was information on food, drug use, and pregnancy-related health. Although pregnancy is a normal life-cycle process, and not a disease, many pregnant women seek information during pregnancy in order to relieve anxiety and to help with decisions related to pregnancy [2].

Despite the high information needs, the reliability of online information tends to be underestimated compared with information provided by professional caregivers. In addition, pregnant women were found to judge the reliability of online information by their own standards [3].

Detailed information on the kind of health information a pregnant woman requires is necessary in order to provide reliable, high-quality information. Studies that have assessed the information needs of pregnant women have mainly used questionnaires or focus group interviews. As a result, most of the information needed involves broad topic levels, such as fetal growth, symptoms of pregnancy or its complications, and nutrition during pregnancy. An analysis of social media data should provide additional details on the diverse topics pregnant women want to learn about.

Recently, an analysis of social media data based on ontology has been used for effective data collection and extraction [4].

Therefore, in this study we developed an ontology for health information needs during pregnancy to be used for collecting and analyzing social data.

Methods

This study consisted of six steps [4, 5]. First, we determined the scope of the ontology. We collect termsed for identifying classes and the class hierarchy: , social media data written by pregnant women in online communities were used as a source of terms for class identification, and categories in the pregnancy literature, including textbooks and guidelines, were used as the source of class hierarchy. In Step 2, we enumerated terms for the ontology. Over the full month of July 2017, we used a web crawler to collect 11,290 pregnancy-related text posts from online communities running in the largest two web portal sites in Korea. Approximately 90 search keywords containing term information were used (e.g., 24 weeks pregnant or first trimester pregnancy). We used text mining to extract pregnancy-related terms. These collection and text mining processes were carried out in collaboration with a leading Korean telecommunications company. We then identified the classes and synonyms for the classes using collected terms. In Step 4, we defined the class hierarchy based on the categories in the pregnancy literature. In Step 5, additional terms were assessed as a second data collection from same online communities using a crawler from August 1, 2017, to July 31, 2018. Newly identified terms were added as classes or synonyms based on the developed ontology. Finally, in Step 6, the adequacy of the class hierarchy of the developed ontology was validated by interviewing three domain experts.

Results

In total, 578 terms were extracted from 11,290 pregnancy-related text posts using a text-mining technique. Of these, 212 terms were identified as classes and another 366 terms were identified as synonyms. Based on the categories in the pregnancy literature, class relationships and hierarchies were defined. The classes had three- or four-level hierarchies, with six superclasses: “pregnancy stages”, “fetal health”, “maternal health”, “medical care”, “support”, and “preparing for birth (supplies)” (See Figure 1). The “maternal health” superclass comprised physiology, psychology, and lifestyle. The “lifestyle” subclass included classes such as “diet”, “activity”, “beauty” and “travelling”. The “support” superclass was classified into “family”, “work”, and “community”. The “Work” subclass included classes such as “maternity leave”, “flextime” and “working mom”. The second round of data collection added 29 classes and 422 synonyms. For example, a “fine dust levels” class was added to “maternal health” and an “mobile application” class was added to “support”. Information related closely to health was included in four classes: “pregnancy stages”, “fetal health”, “maternal health”, and “medical care”. After reviewing the ontology, domain experts recommended that the “pregnancy stages” class be placed in a crossed relationship with the other classes. In this way, one can analyze the change in health information needs according to pregnancy stage.

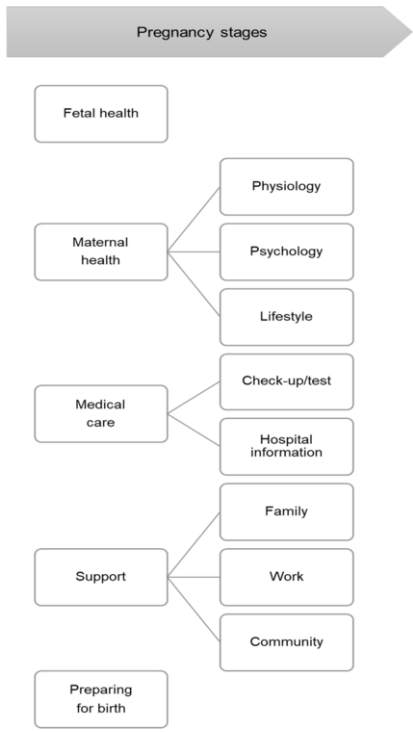


Figure 1. Superclasses in the Developed Ontology

Conclusions

The ontology developed in this study can be used to assess the detailed information needed by pregnant women. Based on the results from this needs assessment, health professionals can determine the reliable information to provide to pregnant women.

Acknowledgments

This work has been supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2018R1C1B5085259).

References

[1] Pew Research Center, Health Topics in,Pew internet, 2011.
[2] P. Sayakhot and M. Carolan-Olah, Internet use by pregnant women seeking pregnancy-related information: A systematic review, BMC Pregnancy Childbirth 16 (2016), 65.
[3] J. Huberty, D. Dinkel, M.W. Beets, and J. Coleman, Describing the use of the internet for health, physical activity, and nutrition information in pregnant women, Matern Child Health J 17 (2013), 1363-1372.
[4] H. Jung, H.-A. Park, and T.-M. Song, Ontology-based approach to social data sentiment analysis: Detection of adolescent depression signals, J Med Internet Res 19 (2017).
[5] N.F. Noy and D.L. McGuinness, Ontology development 101: A guide to creating your first ontology, in Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, Stanford, CA, 2001.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190515

Machine Learning Approaches for Extracting Stage from Pathology Reports in Prostate Cancer

Raphael Lenain[a], Martin G. Seneviratne[a], Selen Bozkurt[a,b], Douglas W. Blayney[c], James D. Brooks[d], Tina Hernandez-Boussard[a,b]

[a] *Department of Medicine, Biomedical Informatics, Stanford University, Stanford, CA, USA*
[b] *Department of Biomedical Data Science, Stanford University, Stanford, CA, USA*
[c] *Department of Medicine, Division of Medical Oncology, Stanford University, Stanford, CA, USA*
[d] *Department of Urology, Stanford University, Stanford, CA, USA*

Abstract

Clinical and pathological stage are defining parameters in oncology, which direct a patient's treatment options and prognosis. Pathology reports contain a wealth of staging information that is not stored in structured form in most electronic health records (EHRs). Therefore, we evaluated three supervised machine learning methods (Support Vector Machine, Decision Trees, Gradient Boosting) to classify free-text pathology reports for prostate cancer into T, N and M stage groups.

Keywords:

Prostate Cancer, Neoplasm Staging, Natural Language Processing

Introduction

Prostate cancer is the commonest non-cutaneous malignancies in men, with over 260,000 new cases annually in the United States [1]. The staging of these newly diagnosed cancer patients is one of the most important factors in determining treatment options and predicting patient survival [3]. Free-text pathology reports contain a wealth of staging information that is not captured in structured form in most electronic health records (EHRs). The ability to automatically extract stage from pathology reports would facilitate the creation of research cohorts from the EHR (e.g. pragmatic trials), provide a framework for quality assurance over time (e.g. assess bone scan adherence), and assist with harmonizing data across sites (e.g. evaulate population-level trends).

Natural language processing (NLP) has emerged as a promising tool for extracting stage from clinical texts. There have been various attempts to apply NLP to automatically extract stage from progress clnical notes and pathology reports across a range of tumor types including lung, breast, colorectal and prostate [4-9]. The majority of these studies have used a rule-based approach, relying on regular expressions associated with stage descriptions or smart text forms. However, rule-based approaches often have limited generalizability between tumor types and across institutions. Therefore, in this study, we aimed to evaluate the performance of different machine learning approaches for extracting staging information from pathology reports in prostate cancer using a more generalizable machine learning approach. This may help to inform the strategy of automated stage extraction from unstructured clinical text.

Methods

The Stanford prostate cancer research database was used for analysis, which is described in detail elsewhere [10]. We identified a cohort of prostate cancer subjects with at least one pathology report. This study was made possible due to linkage of the EHR with an institutional cancer registry, which contained ground-truth stage labels manually abstracted from the clinical notes. Stage annotations were defined at the time of diagnosis using the T, N, M classification (i.e. each document had a separate T, N and M annotation).

We included only reports within one year of the diagnosis date. As we are a teritary cancer center, one year post-diagnosis was used to ensure patients on active surveillance seeking secondary opinions were included. In the case where multiple reports appeared within one year of diagnosis, we treated each report as a separate training sample. In an effort to simplify the classification task, stage labels from the cancer registry were clustered intro groups under the guidance of clinical advisors (e.g. 7 separate T stage labels were grouped into 3). The cohort contained only tumors of T stage 2 and above, as lower-stage tumors were not biopsied.

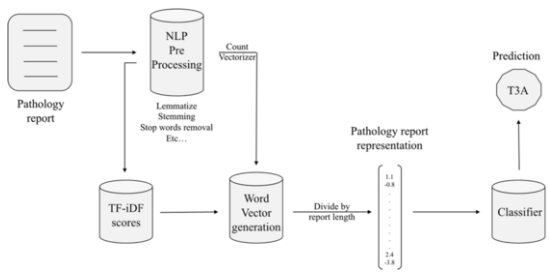


Figure 1 - Architecture of the NLP pipeline for classifying pathology reports into T, N, M stage categories

The pipeline was built with Python (version 3.6) using the Natural Language Toolkit (NLTK) for preprocessing, and scikit-learn for feature extraction and classification. Each report was put through a pre-processing pipeline consisting of stemming, lemmatizing, stop-word and punctuation removal. Subsequently, term frequency-inverse document frequency (TF-iDF) scores were generated for each term-document pair [11].

A bag-of-words representation for each document was generated, with word weighting by TF-iDF scores. A vocabulary was constructed using the entire document corpus. This vocabulary was used to generate document-level word vectors. Neural embeddings were not used because of the limited size of the corpus, and the fact that pre-trained embeddings such as GloVe (GLobal Vectors for Word Representation) were not well suited to the vocabulary of pathology reports. The NLP pipeline is illustrated in Figure 1.

For each of T, N and M stage classifications, we used the document-level vector representations to train a classifier against the ground-truth pooled stage labels from the cancer registry. We used an 80/10/10% split for training/validation/test sets. The following classifiers were trialed: support vector machines (SVM), random forest (RF), and extreme gradient boosting (XGB). With F1-score as our target metric, we used random hyperparameter search to tune our classifiers.

Results

This study cohort included 4,470 prostate cancer subjects with at least one pathology report, yielding a total of 13,595 unique reports. Table 1 shows the results of each classifier for the T, N, M classification tasks. The optimal F1-score achieved was 0.80 on pooled T stage (3 labels), 0.71 on unpooled T stage (7 labels), 0.98 on N stage and 0.99 on M stage.

Table 1 - Evaluation Results

Classifier	Model	Precision	Recall	F-Score
T	SVM	0.77	0.79	0.77
(3 labels)	Decision Trees	0.76	0.75	0.76
	Gradient Boosting	0.80	0.81	0.80
T	SVM	0.61	0.64	0.61
(7 labels)	Decision Trees	0.63	0.62	0.62
	Gradient Boosting	0.71	0.71	0.71
N	SVM	0.98	0.98	0.97
(2 labels)	Gradient Boosting	0.99	0.98	0.98
M	SVM	0. 99	0. 99	0. 99
(2 labels)	Gradient Boosting	0.99	0. 99	0. 99

This study is limited in analyzing pathology reports from a single institution, albeit one with a very diverse clinician and patient population over an extended timeframe. Further work is warranted to apply the pipeline to pathology reports from other sites in order to validate the putative generalizability of this machine learning approach relative to rule-based methods. In addition, the classification tasks were affected by the class imbalances in the dataset, especially between prostate M0 and M1, and breast M0 and M1. We have also made assumptions that a pathology report within one year of diagnosis date reflects the stage at the time of diagnosis - it is conceivable that the stage listed by the registry is not accurate at the time of the report.

Conclusions

Our NLP pipeline is able to efficiently classify pathology reports into T, N, and M stage categories, with strongest performance for N and M stage. This may be a more scalable method than rule-based systems for extracting staging data from unstructured text, which has implications for auto-populating registries and identifying observational research cohorts from EHRs.

Acknowledgements

Research reported in this publication was supported by the Stanford AstraZeneca Research grant and National Cancer Institute of the National Institutes of Health under Award Number R01CA183962. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

[1] R.L. Siegel, K.D. Miller, and A. Jemal, Cancer statistics, 2018, CA Cancer J Clin 68(1) (2018) 7-30.

[2] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2015, CA Cancer J Clin 65(1) (2015) 5-29.

[3] M.K. Buyyounouski, P.L. Choyke, J.K. McKenney, O. Sartor, H.M. Sandler, M.B. Amin, M.W. Kattan, and D.W. Lin, Prostate cancer - major changes in the American Joint Committee on Cancer eighth edition cancer staging manual, CA Cancer J Clin 67(3) (2017) 245-253.

[4] T.L. Evans, P.E. Gabriel, and L.N. Shulman, Cancer Staging in Electronic Health Records: Strategies to Improve Documentation of These Critical Data, J Oncol Pract 12(2) (2016) 137-9.

[5] I. McCowan, D. Moore, and M.J. Fry, Classification of cancer stage from free-text histology reports, Conf Proc IEEE Eng Med Biol Soc 1 (2006) 5153-6.

[6] I.A. McCowan, D.C. Moore, A.N. Nguyen, R.V. Bowman, B.E. Clarke, E.E. Duhig, and M.J. Fry, Collection of cancer stage data by classifying free-text medical reports, JAMIA 14(6) (2007) 736-45.

[7] A. Nguyen, D. Moore, I. McCowan, and M.J. Courage, Multi-class classification of cancer stages from free-text histology reports using support vector machines, Conf Proc IEEE Eng Med Biol Soc 2007 (2007) 5140-3.

[8] A.N. Nguyen, M.J. Lawley, D.P. Hansen, R.V. Bowman, B.E. Clarke, E.E. Duhig, and S. Colquist, Symbolic rule-based classification of lung cancer stages from free-text pathology reports, JAMIA 17(4) (2010) 440-5.

[9] J.L. Warner, M.A. Levy, and M.N. Neuss, ReCAP: Feasibility and Accuracy of Extracting Cancer Stage Information From Narrative Electronic Health Record Data, J Oncol Pract 12(2) (2016) 157-8; e169-7.

[10] M.G. Seneviratne, T. Seto, D.W. Blayney, J.D. Brooks, and T. Hernandez-Boussard, Architecture and Implementation of a Clinical Research Data Warehouse for Prostate Cancer, EGEMS (Wash DC) 6(1) (2018) 13.

[11] J. Ramos, Using tf-idf to determine word relevance in document queries, Proceedings of the first instructional conference on machine learning, 2003.

Address for Correspondence

Selen Bozkurt

selenb@stanford.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190516

Named Entity Recognition in Chinese Electronic Medical Records Based on the Model of Bidirectional Long Short-Term Memory with a Conditional Random Field Layer

Luqi Li [a], Li Hou [a]

[a] *Institute of Medical Information and Library, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China*

Abstract

Named entity recognition in electronic medical records is of great significance to the construction of medical knowledge maps. This paper proposes a model of bidirectional Long Short-Term Memory with a conditional random field layer(BiLSTM-CRF). In terms of simultaneously identifying 5 types of clinical entities from CCKS2018 Chinese EHRs corpus, the BiLSTM-CRF model finally achieved better performance than the baseline CRF model (F-score of 84.23% vs 82.49%).

Keywords:
Data Mining; Electronic Health Records; Natural Language Processing

Introduction

Electronic medical record (EMRs) contain diagnoses, examinations, medications and other important clinical evidence that are helpful to support clinical decision-making and health management. However, a large number of unstructured texts hinder the key information extraction and deep utilization of EHRs. Therefore, it is urgent to explore auto-information extraction methods to transform the natural language in EHRs into structured data that is easy for computers to understand and use.

Named entity recognition (NER) has received much attention recently as an important step for text mining, which aims to automatically detect the target concepts from given texts. In previous works, clinical named entity recognition systems ofen used rule-based NER approaches to match strings by rule templates and dictionaries. They could get good results and high efficiency in the recognition of small sample data, but relied too much on domain knowledge by experts and professional dictionaries. However, state-of-the-art CRF-based NER methods depend on effective feature engineering, i.e. the design of effective features using various NLP tools and knowledge resources, which is still a labor-intensive and skill-dependent task. Recently, deep learning has become prevalent in the machine learning research community. This method can automatically learn and predict the features of texts by using neural network models based on word embedding. In particular, the model of bidirectional Long ShortTerm Memory with a conditional random field layer (BiLSTM-CRF) has been widely used in NER tasks and has achieved excellent performance in various fields. Zeng et al. made outstanding achievements in the drug name recognition tasks of DDIExtrction2011 (DDI2011) and DDIExtrction2013 (DDI2013) by proposing a method combining bidirectional long short-term memory and CRF (BiLSTM-CRF) to explore word and character characteristics [1]. Therefore, in this paper, we propose the BiLSTM-CRF model to improve NER performance in Chinese EHRs, while the BiLSTM layer is used to solve the problem of artificial feature extraction and the CRF layer is used to capture sentence-level tag information.

Methods

In this section, embedded features used in our neural network model are introduced. Then, we employ a CRF model as the baseline for performance comparison. Last, the BiLSTM-CRF model is presented.

Recently, distributed feature representation or word embedding has been widely used in the field of NLP, especially for deep learning methods. It can capture both the semantic and syntactic information of words from a large unlabeled corpus, and has attracted considerable attention from many researchers [2]. Compared with the bag-of-words (BOW) representation, word embedding is low-dimensional and dense. In recent years, word2vec and GloVe have been widely used in the field of NLP. In order to avoid clinical word segmentation interference caused by common word segmentation tools, we employ character embedding as a basic feature. To achieve high-quality character embedding, we collected a total of 2,605 clinical texts from the China Conference on Knowledge Graph and Semantic Computing (CCKS) 2017 challenge. Then these 2,605 texts and another 1,000 texts from the CCKS 2018 NER task were used to train 100-dimensional character embedding by the word2vec tool as pre-trained character embedding.

CRF is a probabilistic undirected graphical model which can take into account the joint probability distribution of the output sequence of labels. The CRF-based model has been widely used for sequence labeling tasks like NER. In this study, we use the CRF++ package for the implementation of CRF model and fix the content window size at 5, then build 18 Unigram templates to extract context characters. As an additional feature, part of speech (POS) tags are generated by the Jieba segmentation system [3].

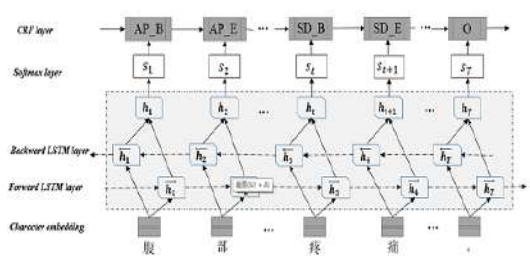


Figure 1–Structure of The BiLSTM-CRF Model

The architecture of a BiLSTM-CRF model is illustrated in Figure 1 [4]. In this paper, each tagged entity in the corpus was encoded in a BIESO (Begin, Internal, End, Single, Other) tagging scheme. We divide the EMR corpus into sentences by periods. The model will predict a label corresponding to each of the input tokens in the sentence. The first layer of the model is the embedding layer. Through the pre-trained character embedding table lookup operation, the input sentence is represented as a sequence of vectors $[X]_1^T = (X_1, X_2, X_3, \dots, X_T)$, where T is the length of the sentence. Next, the vectors are input to the BiLSTM layer. The output $[h]_1^T$ is obtained by connecting the output of the forward LSTM $h_f = < h_{f1}, h_{f2}, h_{f3}, \dots, h_{fT} >$ and the output of the backward LSTM $h_b =< h_{b1}, h_{b2}, h_{b3}, \dots, h_{bT} >$,the meaning of h_t is the deep characteristic representation of X_t. Then, through the softmax function, the decision probability of h_t mapping to the annotated result is expressed as $P_{ht,yt} = softmax(h_t)$. Finally, instead of modeling tagging decisions independently, the CRF layer is added to decode the best of all possible tag paths. In the CRF layer, the current input is predicted by the past input and the state to which the input belongs; the score by moving from state i to state j is represented by the probability transfer matrix $T_{i,j}$; $[O]_1^n$ is defined as an output state sequence; therefore the entire network and the output state sequence have a transition score of

$$\mathrm{S}([X]_1^T, [O]_1^T, \theta, \mathrm{T}) = \sum_{t-1}^{T}(P_{ht,yt} + T_{t-1,t}) \qquad (1)$$

In order to optimize the parameters of the model, the loss function is defined $£(\theta, \mathrm{T})$, which regularizes all possible state sequences $[j]_1^T$ with log likelihood to avoid overfitting:

$$£(\theta, \mathrm{T}) = \mathrm{S}([X]_1^T, [O]_1^T, \theta, \mathrm{T}) - \log \sum_{\forall [j]_1^T} e^{\mathrm{S}([X]_1^T, [j]_1^T, \theta, \mathrm{T})} \quad (2)$$

During the training phase, the objective of the model is to maximize the log-probability of the correct tag sequence. This can be computed using dynamic programming. In the prediction phase, the Viterbi algorithm always is used to discover the state sequence with the highest transition score as the prediction result.

The dataset used for performance evaluation in this study is derived from the NER task in CCKS2018. This study focuses on the identification of three major types of entities-"Symptom", "Drug", and "Surgery". Considering the "Symptom" type of entities are more structured, they are further classified into three categories: "Anatomical Position" (e.g., "***腹***痛 ***abdominal*** pain"), "Symptomatic Description" (e.g.,"腹***痛*** abdominal ***pain***") and "Independent Symptom"(e.g.,"***咳嗽 cough***"). The training set involves 600 annotated clinical texts and another 400 are served as test data. Precision, Recall and F-score are used to evaluate the performance.

Results

The distribution of entities in the training set and the test set is shown in Table 1.

Table 1–Distribution of Entities Among the Training Set and the Test Set

Data set	AP	SD	IS	Drug	Surgery
Training set	52%	14%	20%	7%	7%
Test set	63%	9%	13%	8%	7%

AP: *Anatomical Position;* ***SD:*** *Symptomatic Description;* ***IS****: Independent Symptom*

Figure 2 shows the recognition performances of the BiLSTM-CRF model and the CRF model for five different entities. The BiLSTM-CRF model with character embedding achieves better F-score performance (84.23%) than the CRF model with POS feature (82.49%). The BiLSTM-CRF model is superior in four types of clinical entity recognition but slightly inferior in "Symptomatic Description" because of the small training set and the complexity of Chinese NERs. In view of the unsatisfactory performance of "Drug," caused by a large number of new words and acronyms, we will consider using drug dictionaries and other features (such as chunk and stroke) to improve the ability of the model to discover new words. At the same time, tagging inconsistency of the document-level was found. Furthermore, the BiLSTM-CRF model combined with attention mechanism solved the problem of label inconsistency.

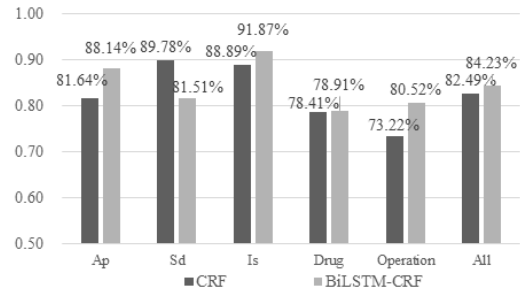


Figure 2–Comparision of F-scores Between Two Models Among 5 Entity Types

Conclusions

We have developed an automatic method of NER in Chinese EMRs and achieved a good overall recognition performance. This BiLSTM-CRF model with character embedding lays a good foundation for future work.

Acknowledgements

This work was supported by the National Engineering Laboratory for Internet Medical Systems and Applications (No.NELIMSA2018P02);National Social Science Foundation (No.14BTQ032). The authors would like to thank the CCKS 2017 and CCKS 2018 CNER challenge organizers for providing the training and test corpora.

References

[1] D. Zeng, C. Sun, L. Lin, and B. Liu, LSTM-CRF for drug-named entity recognition, *Entropy* **19** (2017), 283.
[2] T. Mikolov, K. Chen, G. S. Corrado, and J. Dean. Efficient estimation of word representations in vector space, *Proceedings of Workshop at ICLR* (2013).
[3] jieba [homepage on the Internet]. Python Software Foundation; 2018 [updated 2018; accessed on Nov 5, 2018]. Available from: https://pypi.python.org/pypi/jieba/.
[4] Z. Huang, W. Xu, and K. Yu, Bidirectional LSTM-CRF models for sequence tagging. *CoRR* (2015).

Address for correspondence

Corresponding author : Li Hou . No. 3 Yabao road, Chaoyang district, Beijing, China. Hou.li@imicams.ac.cn

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190517

Exploring Hidden In-Hospital Fall Clusters from Incident Reports Using Text Analytics

Jiaxing Liu[a], Zoie Shui-Yee Wong[b], Kwok-Leung Tsui[a], Hing-Yu So[c], Angela Kwok[c]

[a] School of Data Science, City University of Hong Kong, Hong Kong, China,
[b] Graduate School of Public Health, St. Luke's International University, Tokyo, Japan,
[c]New Territories East Cluster (NTEC), Quality and Safety Division Prince of Wales Hospital, New Territories, Hong Kong, China

Abstract

Retrospective analysing of fall incident reports can uncover hidden information, identify potential risk factors, and improve healthcare quality. This study explores potential fall incident clusters using word embeddings and hierarchical clustering. Fall incident reports from 7 local hospitals in Hong Kong were catalogued into 5 potential clusters with significantly different fall severity, gender, reporting department, and keywords. This study demonstrates the feasibility of using text clustering methods on real-world fall incident reports mining.

Keywords:
Unsupervised machine learning, natural language processing, patient safety

Introduction

In-hospital patient fall is one of the most commonly found incidents in hospitals. Falls can cause patients to suffer, prolong the length of stays and increase hospital resources. It is essential to identify fall risk factors through post-fall analysis and learn lessons from these incidents to provide better clinical care. And it would be more efficient if this can be achieved by automatic text mining reviews. Existing studies mainly target effectively mining the narrative text of patient safety reports that contain fall incidents. Some of the studies successfully identified fall incident reports from other types of patient safety incident reports using text classification or text clustering methods [1-3]. However, the nature, cause and mechanism of in-hospital falls are not studied specifically using text mining. We envision the potential to use recent advances in natural language processing and modelling techniques for automatic fall incident report classification, retrieval, and analysis.

Methods

We retrospectively analysed a cohort of in-hospital fall incidents from AIRS system in Hong Kong from January 2011 to September 2014. The workflow of the text analysis is shown in Figure 1. Narrative text including a brief description and immediate action taken was retrieved and preprocessed. Next, the word2vec method is used to map words to vector representations of words that capture the syntactic and semantic information. Word2vec has two basic architectures: Continuous Bag of Words (CBOW) and Skip-gram. The CBOW architecture was chosen because of its fast computation time. CBOW targets to predict the current word based on the context. Our strategy is to train the word vectors on the whole corpus of the fall incident report narratives under CBOW architecture using negative sampling. Other hyperparameters were selected based on the optimal choice set from Google news [4]. Then we take an average of word embeddings of all the words appearing in the given document as the document's representation. Finally, we applied Ward's agglomerative hierarchical clustering algorithm with Euclidean distance to cluster the reports. The optimal number of clusters was determined by majority vote of indices in R package NbClust [5].

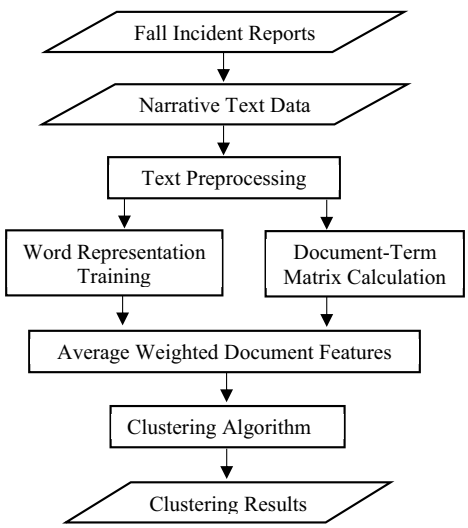


Figure 1– Text clustering procedure

Results

To evaluate our cluster results statistically, we compared the characteristics of the incidents in each cluster and the result is shown in Table 1. The patient sex, reporting departments, and severity levels in different groups are significantly different under the 0.05 significance level. Furthermore, we selected the top 15 TF-IDF words in each cluster's corpus as the characteristic words to represent the content. These keywords are presented in Table 2. From the top TF-IDF terms, it is more likely to reveal the action taken and type of injuries from the plain description text. Based on the current clustering outcome, we cannot make a conclusive catalogue over other pre-specific fall types frameworks. This may be due to the fact that the nature and mechanism of fall are still hidden under our current level of text analysis.

Table 1– Characteristic Words in Different Clusters

Cluster 1	wife, bridg(e), carri(ed), houseoffic(er), lobe, rehabilit(ation), lacer(ated), befor(e), accid(dentally), bone, ent(erance), similar
Cluster 2	Furnitur(e), LT(OT), resid(ent), educ(ate), circumstanti(al), ground, appar(ant), temp, nil, arm, elbow, madam, hrs, befor(e), ca, qud, toe, miss
Cluster 3	Hemorrhag(e), educ(ate), dress(ed), lacer(ated), CA, neurosurger(y), event, haemorrhag(e), shown, gc(s), hematoma, NS, shift, multipl(e), ST, befor(e), film, remind, effect, gauz(e), pleural, stitch
Cluster 4	Radiograp(h), cot, wait, therapist, nil, DMO, ISS, read, geriatr(ic), sustain, rehab, appar(ant), intend, stretcher, final, mang(ement), tri, bus, HDU, lift, occup(y)
Cluster 5	row, april, leg, rib, transfus(ion), bedpan, wait, film, necessar(y), physiotherapi(st), husband, Rt(right), wrist, current, react, sever(e)

Conclusion

This work contributes to quantitative research using text mining methods on the fall incident reports analysis and demonstrates a way to partition a large pool of incident reports based on the semantic use and identify characteristic groups automatically for the retrospective study. We plan to further explore improvement by pre-training word representation on other large open source fall incident related text sources, using other unsupervised machine learning methods, and trying different parts of text datasets as our corpus. The reports of resulting groups will be further reviewed by clinical experts to confirm the meaningful patterns in the group of falls.

Acknowledgements

This work was supported by Japan Society for the Promotion of Science KAKENHI Grant number 18H03336 and the Research Grants Council Theme-Based Research Scheme T32-102/14N.

References

[1] K. Fujita, M. Akiyama, K. Park, E. Yamaguchi, H. Furukawa, I. Sakata, and Y. Kajikawa, Detecting effective categories of medical incident reports for patient safety management, in: Technology Management for Emerging Technologies (PICMET), 2012 Proceedings of PICMET'12:, IEEE, 2012, pp. 3073-3082.

[2] Y. Wang, E. Coiera, W. Runciman, and F. Magrabi, Using multiclass classification to automate the identification of patient safety incident reports by type and severity, BMC medical informatics and decision making 17 (2017), 84.

[3] S.-i. Toyabe, Detecting inpatient falls by using natural language processing of electronic medical records, BMC health services research 12 (2012), 448.

[4] T. Mikolov, K. Chen, G. Corrado, and J. Dean, Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781 (2013).

[5] M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs, and M.M. Charrad, Package 'NbClust', Journal of Statistical Software 61 (2014), 1-36.

Address for correspondence

Zoie Shui-Yee Wong, Ph.D.
Graduate School of Public Health,
St. Luke's International University, Tokyo, Japan
Email: zoiesywong@gmail.com

Table 2– Statistical Comparison between Clusters on Characteristics of the In-Hospital Falls.

	Cluster 1 (N=85) n (%)	**Cluster 2 (N=795) n (%)**	**Cluster 3 (N=409) n (%)**	**Cluster 4 (N=335) n (%)**	**Cluster 5 (N=59) n (%)**	**P values[b]**
Patient age, y	68.5±13.8	66.7±19.3	67.4±19.4	65.3±22.0	68.9±19.4	0.529[a]
Sex						
Male	82 (96.5)	436 (54.8)	218 (53.3)	184 (54.9)	0 (0)	<0.001
Female	3 (3.53)	359 (45.2)	191 (46.7)	151 (45.1)	59 (100)	
Reporting Department						
Medicine	47 (55.3)	186 (23.4)	255 (62.3)	32 (9.6)	21 (35.3)	<0.001
Orthopedics & Traumatology	8 (9.4)	115 (14.5)	42 (10.3)	14 (4.2)	10 (16.9)	<0.001
Accident & Emergency	2 (2.4)	115 (14.5)	12 (2.9)	56 (16.7)	4 (6.8)	<0.001
Geriatrics	3 (3.5)	48 (6.0)	7 (1.7)	91 (27.2)	1 (1.7)	<0.001
Psychiatry	2 (2.4)	63 (7.9)	14 (3.4)	40 (11.9)	0 (0)	<0.001
Clinical Oncology	5 (5.9)	70 (8.8)	24 (5.9)	12 (3.6)	3 (5.1)	0.021
Surgery	5 (5.9)	56 (7.0)	25 (6.1)	11 (3.3)	9 (15.3)	0.007
Hospice/Palliative	0 (0)	30 (3.8)	2 (0.5)	24 (7.2)	0 (0)	<0.001
Paed and Adolescent Medicine	0 (0)	18 (2.3)	10 (2.4)	7 (2.1)	0 (0)	0.484
Rehabilitation	2 (2.4)	31 (3.9)	4 (1.0)	4 (1.2)	3 (5.1)	0.009
Obstetrics	0 (0)	14 (1.8)	1 (0.2)	5 (1.5)	2 (3.4)	0.090
Others	11 (12.9)	49 (6.2)	13 (3.2)	39 (11.6)	6 (10.2)	<0.001
Severity Index groups						
0 – 1(No injury sustained)	28 (32.9)	411 (51.7)	92 (22.5)	140 (41.8)	25 (42.4)	<0.001
2 – 3 (Minor injury occurred)	53 (62.4)	348 (43.8)	285 (69.7)	170 (50.7)	26 (44.1)	<0.001
4 – 6 (Significant injury or death)	4 (4.7)	36 (4.5)	32 (7.8)	25 (7.5)	8 (13.6)	0.015

Notes: [a]ANOVA (analysis of variance). [b]Chi-square test. The significant level is 0.05.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190518

Design of Metadata Services for Clinical Data Interoperability in Germany

Matthias Löbe[a], Oya Beyan[b,c], Sebastian Stäubert[a], Frank Meineke[a], Danny Ammon[d], Alfred Winter[a], Stefan Decker[b,c], Markus Löffler[a] and Toralf Kirsten[e,f]

[a] *Institute for Medical Informatics, Statistics and Epidemiology (IMISE), Leipzig University, Leipzig, Saxony, Germany*
[b] *Fraunhofer FIT, Sankt Augustin, North Rhine-Westphalia, Germany*
[c] *Informatik 5, RWTH Aachen University, Aachen, North Rhine-Westphalia, Germany*
[d] *Jena University Hospital, Jena, Germany*
[e] *University of Applied Sciences Mittweida, Mittweida, Saxony, Germany*
[f] *Leipzig Research Centre for Civilization Diseases - LIFE, Leipzig University, Leipzig, Saxony, Germany*

Abstract

Secondary use of electronic health record (EHR) data requires a detailed description of metadata, especially when data collection and data re-use are organizationally and technically far apart. This paper describes the concept of the SMITH consortium that includes conventions, processes, and tools for describing and managing metadata using common standards for semantic interoperability. It deals in particular with the chain of processing steps of data from existing information systems and provides an overview of the planned use of metadata, medical terminologies, and semantic services in the consortium.

Keywords:

Metadata, Information Systems, Health Information Interoperability

Introduction

The re-use of clinical data for research purposes has been an active field of medical informatics for a long time [1-2], but recent developments such as Fast Healthcare Interoperability Resources (FHIR) and machine learning enable new levels of interoperability between sites and more sophisticated ways of phenotyping. In 2018, Germany launched a major project called the Medical Informatics Initiative to make such data available for research [3]. To this end, all university hospitals have joined one of the four consortia that want to meet this goal with different technical approaches. The main objective of the initiative is to make patient data from clinical care available for biomedical research in new organizational units to be set up called Data Integration Center (DIC) in every participating site. One of the consortia is *SMITH* [4].

Methods

In the area of data exchange, SMITH focuses on the HL7 Clinical Document Architecture (CDA) standard for clinical documents and FHIR for individual medical data points as well as appropriate profiles from Integrating the Healthcare Enterprise (IHE). Information systems supporting these standards are referred to as type A data sources (see Figure 1). However, many existing systems in hospitals cannot directly support these standards and their data need to be transformed. Type B sources export data in standard formats such as HL7, DICOM etc., but need data transformations for a unified data exchange. Finally, type C sources are proprietary such as data provided by CSV files or arbitrary SQL databases. Metadata Services that can semantically describe captured medical data are therefore essential for data curation, analysis, and sharing.

Results

The SMITH Consortium will provide Metadata Services to promote secondary use of clinical and research data not only between the DICs but also with other data providers. First, barriers between “data consumers” and “data providers” will be lowered by promoting user-friendly services to discover, interpret, and provide access to data. Second, consortium members will be provided with means to improve metadata quality and reusability of shared data. In line with the FAIR (Findable, Accessible, Interoperable, Reusable) data principles [5] developed for sharing research data, the services provided will guide DIC and external data providers to generate metadata, annotate the shared data, define access rights and protocols, integrate heterogeneous data models, and eventually support discoverability and reusability of data. Following subsections will outline required infrastructure and processes to comply with the FAIR principles, including metadata capturing and management, quality-driven metadata improvements through curation on every site in the consortium as well as metadata harmonization (e.g., naming, coding scheme) across different sites.

Metadata Capturing and Management

A Metadata Repository (MDR) is a system that describes the syntax and semantics of each data element representing the clinical facts (data points). In SMITH, we will establish a MDR at each site, i.e., in Aachen, Jena, and Leipzig. Each local MDR will manage metadata according to the clinical information systems available at each site. Additionally, a consortia-level MDR will be established to facilitate central metadata interoperability. The consortia-level MDR will provide persistent identifier (PI) services for identifiable data and metadata, as well as cross-site metadata harmonization and query services (API). Figure 1 illustrates the set-up and metadata-related components relevant to each site and across all the sites, and their relationship to the planned national metadata services. A major step before metadata are uniformly managed by a dedicated MDR is to capture relevant metadata and import them into the MDR. While some clinical information systems store metadata inherently so that they can

be extracted automatically, transformed, and then loaded to the MDR, in other cases only a portion or none of the necessary metadata are electronically available. The proposed connectors will read available metadata wherever possible. For all other cases, standardized manual metadata capturing processes will be established. The Raw Metadata component stores all the metadata as submitted by the Metadata Capturing Service within the DIC. In order to enable querying and subscription services, the metadata need to be mapped and aligned, after which they are stored in the Aligned Metadata component taking provenance aspects into account. Currently, querying and subscription services use this aligned metadata storage. In the Consortium-Level MDR, this step is replaced by a Metadata Harmonization component, which ensures that metadata are compatible across sites.

Metadata Curation

The SMITH metadata curation and harmonization services will define a process to harmonize common metadata elements from each site by creating descriptive metadata at both document and data element levels including semantic, technical, and provenance metadata. Varying coding schemes and vocabularies will be semantically annotated, mapped, and harmonized using the tools and algorithms developed. A quality management process will be set up to ensure better metadata quality at the metadata entry stage by applying terminology management and semantic data validity checks.

Terminology and Ontology Service

We will set up terminology services to provide detailed representation of conceptual entities. This includes browsing using terminologies, displaying concepts, facetted searching in terminologies, and exporting. We will import standard terminologies for coded medical data such as ICD-10-GM (diagnoses), OPS (procedures, modified version of ICPM), LOINC (laboratory), ATC/DDD (pharmaceutical agents); UCUM (measurement units); terminologies and ontologies used for text mining and phenotyping such as MeSH, UMLS, or HPO; and also local vocabularies used in a DIC. The use of SNOMED-CT is being prepared as this terminology would cover many topics (e.g., pathogen/microbiology), but its use depends on acquisition of a license. Furthermore, new vocabularies and concepts can be created and mapped to terminologies to facilitate use of core data sets. The terminology services will be accessible through a graphical user interface as well as via a ReST API.

Conclusions

We expect that the concept outlined here will be powerful enough to map the heterogeneous health information technology landscape and processes in university hospitals in Germany, along with the integration of other partners such as non-university hospitals, general practitioners, data from epidemiological studies, health insurance companies, and patient-reported outcomes.

Acknowledgements

The research was funded by the German Ministry of Education and Research (reference numbers: 01ZZ1803A, 01ZZ1803C, 01ZZ1803K) and German Research Foundation (reference number: WI 1605/10-1).

References

[1] S. Rea *et al.,* "Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: the SHARPn project," (eng), *Journal of biomedical informatics*, vol. 45, no. 4, pp. 763–771, 2012.

[2] N. G. Weiskopf, G. Hripcsak, S. Swaminathan, and C. Weng, "Defining and measuring completeness of electronic health records for secondary use," (eng), *Journal of biomedical informatics*, vol. 46, no. 5, pp. 830–836, 2013.

[3] S. C. Semler, F. Wissing, and R. Heyder, German Medical Informatics Initiative, *Methods of information in medicine* 57 (2018), no. S 01, e50-e56.

[4] A. Winter *et al.,* "Smart Medical Information Technology for Healthcare (SMITH)," (eng), *Methods of information in medicine*, vol. 57, no. S 01, e92-e105, 2018.

[5] M. D. Wilkinson *et al.,* "The FAIR Guiding Principles for scientific data management and stewardship," (eng), *Scientific data*, vol. 3, p. 160018, 2016.

Address for correspondence

Matthias Löbe, matthias.loebe@imise.uni-leipzig.de, IMISE
Universität Leipzig, Härtelstraße 16-18, 04107 Leipzig, Germany

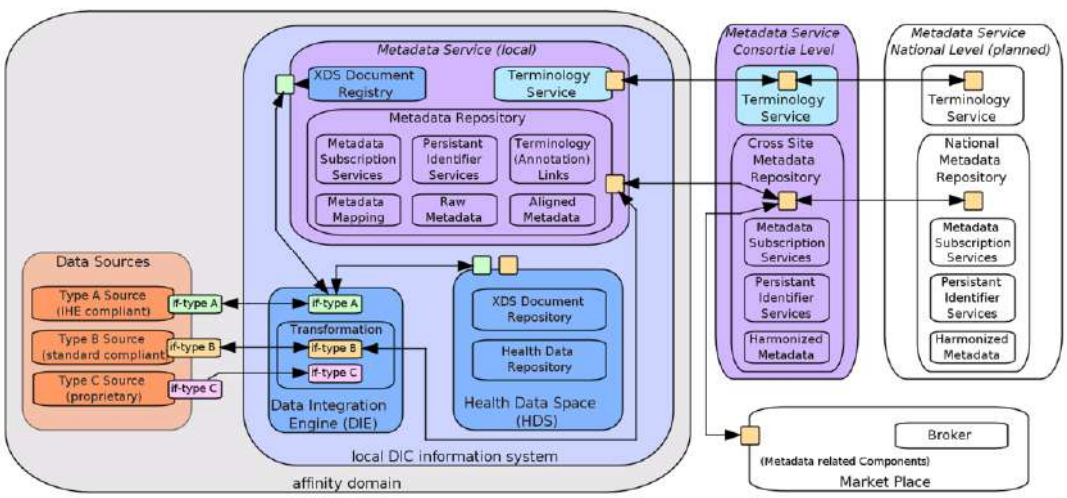


Figure 1– High-level System Architecture of Metadata-related Services at different Levels

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190519

Characterizing the Scope of Exposome Research Through Topic Modeling and Ontology Analysis

Guillermo Lopez-Campos[a,b], Philip Kiossoglou[b], Ann Borda[b], Christopher Hawthorne[a], Kathleen Gray[b,c], Karin Verspoor[b,c]

[a] *Wellcome-Wolfson Institute for Experimental Medicine, Queen's University of Belfast, Belfast, Northern Ireland, United Kingdom,*
[b] *Health and Biomedical Informatics Centre, The University of Melbourne, Parkville, Victoria, Australia,*
[c] *School of Computing and Information Systems, The University of Melbourne, Parkville, Victoria, Australia*

Abstract

Exposomics is a field of research which is receiving growing attention. In this work, we characterize the exposome research landscape and update our previous study of formal knowledge representation approaches to this field. We applied a deductive analysis using the National Center for Biomedical Ontology Recommender for comparability of the results generated from a literature dataset and newly available ontologies with our previously published work. We highlight the changes in ontology recommendations.

Keywords:

Biomedical Ontologies, Data Mining, Environmental Exposure, Publications

Introduction

Environmental factors and exposures remain the major risk factors for most diseases and certainly for those that are a major burden for the health system. The concept of the *exposome* was introduced in 2005 when C.P. Wild coined the term, defining it as an exposure-oriented analogue of the genome [1]. The concept has since gained significant attention, particularly with the recent establishment of several large research collaborations and because of the demonstrated relevance to the new paradigm of "precision medicine" [2].

The increasing corpus of literature focused on the exposome provides opportunities to characterize the breadth of the field. As an interdisciplinary and evolving field, exposome research covers a broad range of topics and methodologies that could be related to different areas such as public health and biomarker detection.

Formal knowledge representation is an important tool for organization of information in many different areas and disciplines. The exposome poses complex challenges for biomedical informatics and requires the adaptation and development of suitable methods and tools [3]. Knowledge representation methods, such as ontologies, have also significantly supported "omics" research; examples of well-adopted informatics tools include the Gene Ontology [4] and the Human Phenotype Ontology, both of which provide formal vocabularies and represent relationships among concepts within these vocabularies.

In an associated work on which we base this study [5], we proposed a methodology to characterize this evolving exposome landscape.

The deductive analysis approach of the prior work used the National Center for Biomedical Ontology (NCBO) Recommender [6], which allowed for contextualization of current formal knowledge representation of the exposome. However, only a small portion of the overall recommended ontologies could be deemed as broadly useful in the field of exposome research at that point. Importantly, the results showed that none of the frequently recommended ontologies had major associations with the science of environment or exposures, suggesting that there was a need to bridge interdisciplinary research gaps to build knowledge in this field.

This study seeks to re-apply our previously developed deductive analysis to analyze the evolution of knowledge representation tools covering exposome research, as well as to ensure comparability with new publication results, indicated by the growth of the corpus of exposome literature, which has more than doubled in the last two years (2017-2018).

Methods

We analyzed the recommended ontologies using an expanded dataset of papers up to October 2018. The new dataset used a similar search strategy to our previous work [5]. Briefly, we considered papers published between January 2005 and October 2018 with the term "exposom*" in the title, abstract or keyword fields.

The abstracts from the retrieved documents were subsequently submitted for ontology recommendation using the NCBO Ontology Recommender tool API. Default parameters were used to ensure the comparability of the results generated using the literature dataset up to 2018 and the new ontologies, with those previously published.

For topic modeling, we used the R package "text2vec" (v0.5.1) on the subset of documents that contain the word exposome in their abstract.

Results

The results from the ontology analyses using NCBO's Ontology Recommender for the search 2005-2018 resulted in the identification of 171 ontologies for 2018 and 165 for 2016. Of those just 20 ontologies were present in more than 50% of the documents (50 ORF – ontology recommended fraction). The comparison of these ontologies with the results previously published shows that there is a significant overlap among the ontologies recommended for the three datasets in Table 1.

To identify the number of topics we used the R library "LDATuning" determining the presence of 25 topics, based on a balance among different criteria.

Table 1 – List of Ontologies Recommended with ORF >50 in 2016 and in 2018.

Shared 2016 & 2018		Found in 2018	Found in 2016
NCIT	EFO	IOBC	SNMI
SNOMEDCT	CRISP	CHEAR	RH-MESH
MESH	HL7	GO-EXT	
LOINC	NDFRT	MEDDRA	
HUPSON	EDAM	NBO	
ONTOAD	GO		
NIFSTD	RCD		
MEDLINEPLUS			

Discussion

Deductive analyses using the knowledge representation tools led to the identification of 171 recommended ontologies for the 2018 dataset of documents, in comparison with 165 ontologies identified in 2016. The differences between the 2016 and 2018 recommendations are slightly reduced after applying the ORF >50 threshold.

However, there are five differences in the ontologies that fulfill the requirement of being recommended for at least 50% of the documents that appear as significant in the 2018 analyses. Among these five newly identified ontologies it is important to note that three of them, NBO (Neurobehaviour Ontology), MEDDRA (Medical Dictionary for Regulatory Activities Terminology) and GO-EXT (Gene Ontology Extension), were present in the 2016 results but did not reach the significance threshold of ORF >50 at that time. In the most recent analysis, these three have jumped to be recommended for around 60% of the documents. This change was more relevant for NBO which experienced a nearly 50% increase in its recommendations (from 41% to 62% of the documents). In terms of the recommendation ranking, MEDDRA did not change its average recommendation position (a corrected average of 20.5) whereas in the case of GO-EXT it improved its ranking in three positions, from an average corrected rank of 20.7 to 18.7. Similarly NBO's rank improved from 22.3 to 19.5.

These changes in the recommended ontologies can be explained by the update of existing technologies and by an expansion of the exposome research in areas already covered by the ontologies that led them to be deemed as significant.

The other two newly recommended ontologies CHEAR (Children's Health Exposure Analysis Resource) [7] and IOBC (Interlinking Ontology for Biological Concepts) are recently developed and cover important aspects related to the exposome. CHEAR has been recommended for 64% of the documents. This resource was developed for the analysis of environmental exposures in an "exposome context" and therefore it could be considered as an "authentic" exposome ontology. IOBC is a recently developed ontology containing both biological concepts and earth and environmental sciences concepts.

Conclusions

Exposomics research and literature have doubled over the last two years. This increment reflects the increasing relevance of the field. Analysis of the recommended ontologies that could potentially be used to annotate or model exposome research still shows that broad and general tools such as National Cancer Institute Thesaurus (NCIT), Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT), Medical Subject Headings (MESH), and Logical Observation Identifiers Names and Codes (LOINC) would cover aspects of all the documents. However, in the last couple of years, CHEAR and IOBC have emerged as two new and promising tools to describe the knowledge around the exposome more formally.

Acknowledgements

C. Hawthorne is funded by a Northern Ireland Department for the Economy (DfE) postgraduate studentship Award.

References

[1] C.P. Wild, Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology, *Cancer Epidemiol Biomarkers Prev* **14** (2005), 1847-1850.

[2] National Research Council (U.S.). Committee on A Framework for Developing a New Taxonomy of Disease., Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease, *National Academies Press*, Washington, D.C., 2011.

[3] F. Martin Sanchez, K. Gray, R. Bellazzi, and G. Lopez-Campos, Exposome informatics: considerations for the design of future biomedical research information systems, *J Am Med Inform Assoc* **21** (2014), 386-390.

[4] S. Carbon, H. Dietze, S.E. Lewis, C. J. Mungall, M.C. Munoz-Torres, and S. Basu, The Gene Ontology Consortium. Expansion of the gene ontology knowledgebase and resources: The gene ontology consortium, *Nucleic Acids Res* **45** (2017), D331-D338.

[5] P. Kiossoglou, A. Borda, K. Gray, F. Martin-Sanchez, K. Verspoor, and G. Lopez-Campos, Characterising the scope of exposome research: A generalisable approach, MEDINFO 2017: Precision Healthcare through Informatics, *Stud Health Technol Inform* **245** (2017), 457-461.

[6] C. Jonquet, M.A. Musen, and N.H. Shah, Building a biomedical ontology recommender web service, J Biomed Semantics 1 Suppl 1 (2010), S1.

[7] D.M. Balshaw, G.W. Collman, K.A. Gray, and C.L. Thompson, The Children's Health Exposure Analysis Resource: enabling research into the environmental influences on children's health outcomes, Curr Opin Pediatr 29 (2017), 385-389.

Address for correspondence

Dr. Guillermo Lopez Campos.
Email: G.LopezCampos@qub.ac.uk

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190520

Using Enriched Samples for Semi-Automated Vocabulary Expansion to Identify Rare Events in Clinical Text: Sexual Orientation as a Use Case

Kristine E Lynch[a,b], Patrick Alba[a,b], Benjamin Viernes[a,b], Scott L DuVall[a,b]

[a]*VA Informatics and Computing Infrastructure, Department of Veterans Affairs, Salt Lake City, UT, USA,*
[b]*Department of Internal Medicine, Division of Epidemiology, University of Utah, Salt Lake City, UT, USA*

Abstract

We demonstrate the utility of concept lexicon expansion and evaluation using enriched samples of patients and documents with sexual orientation as a use case for rare event detection in electronic medical records. Using this approach, we found 7 additional words and 21 misspellings beyond our initial set of five seed words. We can use the expanded vocabulary to further develop a full natural language processing system to identify instances where sexual orientation is documented.

Keywords:

Natural Language Processing, Electronic Health Records

Introduction

Finding instances of rare events in medical notes with natural language processing (NLP) can be challenging. When the goal of a study is to identify all patients who have experienced an event, a concept lexicon with contextual diversity and widespread coverage is sine qua non. Survey research often deliberately oversamples racial, ethnic and other minority groups to attain adequate numbers to make meaningful comparisons [1]. In a similar fashion, sampling from document groups that, based on a priori knowledge are likely to have a higher proportion of mentions than other groups, can help discover and evaluate terms that may have otherwise been overlooked. We explore sexual orientation documentation in the Department of Veterans Affairs' (VA) electronic health record (EHR) as a motivating example.

Sexual orientation documentation is expected to be a rare occurrence in VA. Based on estimates of active military (2.8% of all military personnel are homosexual [2]) and estimates of Veteran disclosure rates (30% [3] to 80% [4]), we expect to discover 71,400 to 190,400 of the 8,500,000 total Veterans with an outpatient visit in VA in the last 18 years to have at least one mention of lesbian, gay, or bisexual (LGB) status in the EHR. Documentation of sexual orientation is not representative of all LGB Veterans nor all who have disclosed to a provider. Rather, documentation is evidence that information about sexual orientation was exchanged either through patient or provider initiated communication and the provider felt it relevant enough to the patient's health to record it in the EHR. Failure to assess and document sexual orientation status of patients can contribute to the perpetuation of health inequalities that LGB Veterans may face compared to their heterosexual counterparts. Understanding the extent to which documentation occurs is therefore important, but unlike age and race, sexual orientation is not formally incorporated into the EHR of most healthcare systems as a structured demographic field. Consequentially, the only place to find documentation of sexual orientation is in providers' medical notes.

A common way that NLP is used in EHR research is to supplement variables for a study population that has been formerly defined using structured data, such as administrative codes. Correspondingly, only a small subset of potentially relevant documents is required to be processed. Administrative codes do not exist for sexual orientation as they do for gender identity disorder, for example, which can be used to identify a subset of transgender patients and aid in the search of transgender words or phrases [5], so an NLP assisted approach was needed to build our vocabulary. In this study, we describe our strategy for developing a representative sexual orientation lexicon among unstructured text of clinical notes of the VA EHR. This work forms the bedrock for a future, validated NLP system to be used in health disparity research.

Methods

Data Source

The VA is the largest integrated healthcare delivery network in the United States, capturing healthcare information from patients across all medical specialities with broad geographical coverage. The VA's Corporate Data Warehouse (CDW) is a nationwide repository with historical data dating back to October 1, 1999, including patient encounters, pharmacy, lab/chemistry, microbiology, vital sign, and radiology related data. For the present study, we evaluated clinical notes from the CDW from 2000-2017 for patients enrolled in VA after 2000 with a record of at least one outpatient visit (~8.5 million Veterans). This study was approved by the University of Utah's Institutional Review Board.

Vocabulary Expansion

Our LGB vocabulary started with an initial set of five seed terms (lesbian, gay, homosexual, LGBT, and bisexual). We used a combination of four approaches to expand our word set. First, we manually reviewed LGBT health literature and added any words thought to be relevant. Second, we used a full text index built in the Microsoft SQL Server instance storing clinical notes to find terms within an edit distance of 3 from any of the five initial terms that had at least 100 instances. Third, we conducted an analysis of bigrams and trigrams that surrounded the original 5 terms to determine frequently occurring terms in the same immediate context [6]. A search was done using these terms to see if any other similar terms existed in the same context. Lastly, we used a lexicon discovery tool [7] implemented with an Ipython Jupyter Notebook to automatically find semantically and lexically similar words from a corpus to words from the initial term set. The tool uses a combination of word embeddings, a mathematical representation of the context in which a string is used and edit distance. The settings for the tool were set to a cosine similarity threshold of 0.75 and a maximum edit distance threshold of 3.

Terms at or above the threshold were reviewed for inclusion in our vocabulary.

We applied the lexicon discovery tool to a random set of documents and from patients hypothesized to have a high volume of sexual orientation documentation relative to other groups (i.e., enriched samples). We evaluated documents from patients diagnosed with Human Immunodeficiency Virus (HIV,) according to International Classification of Disease diagnostic codes. It is estimated that 68% of newly diagnosed HIV patients are homosexuals.

From each of the two cohorts, 1,000 random documents containing each of the five initial terms were processed (5,000 documents per cohort). We manually reviewed the output from the four different approaches for additional potentially relevant terms. We processed every document from all patients in our sample and the entire HIV cohort separately. We used cumulative binomial probabilities $\sum_{k=0}^{n} \binom{n}{k} x^k p^{n-k}$, to determine the probability of getting x or more patients without documentation given a hypothesized background rate of no documentation. In this equation n = number of HIV patients, k = number of patients without documentation, and p = estimated background documentation rate. In other words, we compared the proportion of observed patients without keywords to the proportion expected in to determine if additional keywords were potentially needed. Expected background rate in this context refers to the number of patients that would have sexual orientation documented had the rates of sexual orientation and disclosure been equal to documented prevalence. For example, if we assumed that 68% of HIV patients are homosexual, 60% disclose and 90% that disclose are documented we expect to see ≥36% of the HIV cohort with ≥1 keyword (64% without a keyword). Vocabulary expansion is the first of multiple steps needed to develop and deploy an accurate NLP system. As we formally incorporate exclusionary rules that consider surrounding context the number of relative instances of LGB will decrease. Assuming an accurate background rate, observing more patients without documentation than expected, may indicate that our vocabulary is inadequate and additional work is needed.

Results

Our initial set of five keywords was found among 441,870 of the 8.5 million patients across 1,100,737 documents. The literature review added 6 terms (His husband, Her wife, His boyfriend, Her girlfriend, Queer, and Same-sex). The full text index search identified 8 misspellings and 1 additional term. We assessed five bigrams/trigrams ("sexuality:", "sexual preference:", "orientation:", "identifies as", "came out as") and no additional terms were found. The lexicon discovery tool identified 11 misspellings and 5 new terms, of which one was also found in the literature review (same-sex) and the full text index search (lesbianism). Our final vocabulary consisted of 28 distinct words including misspellings. A sample list of the most common terms is presented in Table 1.

We found 28,437 patients with a diagnosis of HIV. Before including additional terms found through our four approaches we found 11,550 patients (126,559 documents) of the HIV cohort with at least one instance of any of the five initial keywords. After re-processing using our expanded list of new and misspelled terms, the number of HIV patients with a keyword increased to 12,051 patients (42%) across 146,689 documents. Assuming 64% Veterans diagnosed with HIV will not have LGB documentation, the probability of getting 58% of HIV patients without documentation in a sample of 28,437 is 0.98, suggesting that our vocabulary is an adequate starting point.

After all documents from the 8.5 million Veterans were passed through the entire expanded vocabulary, we identified an additional 19,543 (+4.4%) and 49,716 (+4.5%) potentially relevant patients and documents, respectively, beyond what was found with the initial five term set.

Conclusions

We demonstrate that using enriched cohorts is advantageous for two reasons. First, the approach resulted in an expanded vocabulary that identifies additional patients and documents that may be relevant but otherwise been overlooked. Second, enriched cohorts can lend insight into whether the vocabulary will capture all instances of the event by comparing observed rates to expected rates. We will use the expanded vocabulary to further develop a validated NLP system to identify instances where sexual orientation is documented. We demonstrated our approach using sexual orientation documentation, but it can be tailored to any scenario in which instances are expected to be rare and precise structured data is unavailable to aid the construction of a comprehensive concept lexicon.

Table 1. Keywords by document and patient count

Rank	Keyword	Document count	Patient count
1	Gay	708,341	292,726
2	Bisexual	270,866	167,589
3	Homosexual	240,585	139,970
4	Lesbian	151,151	87,939
5	LGBT	58,734	15,274
6	Same-sex	37,637	12,622
7	LGB	26,789	12,476
8	Queer	8,584	3,992
9	Homosexual-lesbian	3,668	8,014
10	Homosexual-gay	684	759

Acknowledgements

This work was supported using resources and facilities at VINCI, VA HSR RES 13-457.

References

[1] R. Vaughan, Oversampling in health surveys: Why, when, and how?, *Am J Public Health* **10** (2017), 1214-1215.

[2] G. Gates, *Gay men and lesbians in the military: Estimates from Census 2000,* (2004, September 28)

[3] T.L. Simpson, K.F. Balsam, B.N. Cochran, K. Lehavot, and S.D. Gold, Veterans administration health care utilization among sexual minority veterans, *Psychol Serv* **10** (2013), 223-232.

[4] K.M. Mattocks, J.C. Sullivan, C. Bertrand, R.L. Kinney, M.D. Sherman, et al, Perceived stigma, discrimination, and disclosure of sexual orientation among a sample of lesbian veterans receiving care in the Department of Veterans Affairs, *LGBT Health* **2** (2015), 47-53.

[5] R. Roblin, J. Barzilay, D. Tolsma, B. Robinson, L. Schild, et al, A novel method for estimating transgender status using electronic medical records, *Ann Epidemiol* **26** (2016), 198-203.

[6] P.R. Alba, O.V. Patterson, B. Viernes, D. Denhalter, N. Baily, et al, *The super annotator: A method of semi-automated rare event identification forl large clinical data sets,* Poster session presented at AMIA Annual Symposium, Chicago, IL, 2015.

[7] A.B. Chapman, P.R. Alba, B.T. Bucher, S.L. DuVall, and O.V. Patterson, A user-friendly interface for concept dictionary expansion using word embeddings and SNOMED-CT. Poster session presented at AMIA Annual Symposium, San Francisco, CA, 2018.

Address for Correspondence

Kristine E Lynch, MS, PhD. E-mail: Kristine.Lynch@hsc.utah.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190521

A Cross-Lingual Effort Towards Managing English-Chinese Cancer Education Resources

Hetong Ma[a], Jiansong Ren[b], Xuwen Wang[a], An Fang[a], Jiao Li[a], Qing Qian[a]

[a] *Institute of Medical Information/Library, Chinese Academy of Medical Sciences/Peking Union Medical College, Beijing, China*
[b] *National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences/Peking Union Medical College, Beijing, China*

Abstract

As translated education resource plays an important role in healthcare providers' training and medical knowledge dissemination, we proposed a method to manage cross-lingual education resources with the goal of facilitating the medical education and physician training. We created an English-Chinese cancer knowledge base including bilingual description on cancer diagnosis, prevention, screening, treatments, etc. We developed a workflow to create the bilingual corpus, and applied it to six cancer monographs in PDQ (Physician Data Query).

Keywords:

Cancer education resource, Cross language, Parallel corpus

Introduction

Multi-lingual resources promote knowledge dissemination and communication. Many efforts have been made for the construction from expert translation, machine translation, crowd-sourcing (e.g., Wikipedia [1]), etc. In this field, literatures are still mainly recorded in English, however, health professionals often need to learn those written in native languages. Local languages are most common with least difficulty, which will naturally lead to the highly requirement of processing and transformation among various languages. Since manual translation will not meet the requirements of mass data, automatic machine translation will be an irresistible trend.

Based on multi-languages similarities and ambiguities, an acceptable translation is supposed to be both correct with analyzing and generating, and to be humanlike where a great amount of processing and understanding are completed [2]. Still, medical field can be much different for its own professional knowledge, which may lead to higher-level difficulty, and may require a better quality of translation where high-quality parallel corpus are intensely needed. A good parallel corpus has high demand for content, either in specific realm or aligned pattern.

In this paper, we constructed an English-Chinese cancer parallel corpus using Physician Data Query (PDQ). Up to now, no English-Chinese parallel corpus of cancer subject in the same type has been established yet, which is quite substantial for data collection, analysis, and dissemination. We used the Physician Data Query (PDQ) as the original corpus, which was NCI's comprehensive source of cancer information, for its authoritativeness, specialty and leading role in cancer information. For better reuse and redistribution of PDQ information, NCI encourages the translation in different language. As Chinese translated version of PDQ was rigorously translated by experienced clinicians and experts in both Chinese and English in a recognized cancer hospital, responsible for cancer patients, the dataset was greatly precious for its golden standard. The format of original data is XML, which is open accessible. PDQ English version (PDQ_{EN}) describes information ranged from cancer, drug, genetic terms, and other related knowledge [3] with its own presence entrance, so is the Chinese version (PDQ_{CH}) [4]. Since the corpora was translated at different time, in this paper, The PDQ_{EN} and PDQ_{CH} we use were from 2013 edited in XML.

We selected education monographs related to six cancers as our data set, including liver cancer, colorectal cancer, breast cancer, lung cancer, stomach cancer and esophageal cancer. Inside these cancers, there were various themes including prevention, screening, diagnosis and treatment. In this corpus, there were 22 pairs of file, each containing different content from others, either in cancer type or in cancer theme. Every file in PDQ was stored in XML format structure, which had many elements embedded in other elements. The total numbers of English sentences, Chinese sentences, English words, Chinese words are respectively 22091, 12365, 291095, and 170688.

Despite PDQ_{CH} came from manual translation of PDQ_{EN}, layers, elements and even the whole structure were not always the same. Some of translation were lost, some paragraphs had not yet been translated, and some elements had been deleted from PDQ_{CH}, which led to a major difficulty when trying to make the corpus parallel. Nesting elements are quite common in original corpus, while there was only one element possessing ID that can help with the alignment and location of two corpus. The other great difficulty was that appearance times of the same element can be quite different, thus, we could not make sure which two paragraphs or sentences should be aligned.

Methods

Following the framework shown in Figure 1, the construction contained three stages: cancer resource parsing, data processing and parallel corpus release. In the first stage, we examined the similarities and differences of the two PDQ corpora in structure features and possible alignment anchors. Next come to the

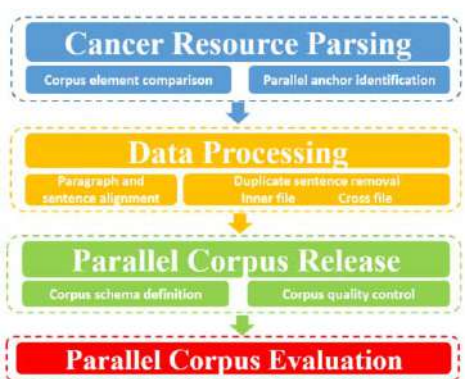


Figure 1 – The Workflow of Parallel Corpus Construction.

second stage where extraction of content in elements that occur in both corpora were completed, which brought a large amount of paragraphs. It was followed by the alignment strategies setting, alignment of paragraphs and sentences, and duplicate sentence removal. Then it enters into the third stage where we defined the corpus schema, controlled the corpus quality and evaluated the corpus.

We set a rule that text should be classified into words, phrases, and sentences. Words were mostly professional terms that could help us just like a dictionary. Phrases were not same with sentences, they had smaller numbers of words and may not expressed an integrated meaning. The rule could be summarized through length, those with 1-2 words should be sorted into 'Word', those with 3-7 words should be sorted to 'Phrase', and those over 7 words should be sorted to 'Sentence'.

We used the sentence alignment ratio as a metric to evaluate the parallel corpus quality. To ensure the quality of our corpus, we mannualy assessed the corpus sentence by sentence. The criteria could be categorized as: correct, incorrect and partially correct, where partially correct means one sentence was partially aligned to the counterpart but still there was some part unaligned. The evaluation formula is as follows, where α, β, and γ respectively represent the numbers of correct, partially correct, and incorrect alignments:

$$\text{Ratio} = 100\% * \frac{\alpha + 0.5 * \beta + 0 * \gamma}{\alpha + \beta + \gamma}$$

Results

We constructed the parallel corpus where PDQ_{EN} and PDQ_{CH} were separated apart but combined with the same ID. It formed a 23-pair parallel corpus focused on medical cancer information with words, phrases and sentences. Finally, we constructed a parallel corpus of 7188 pairs consisting of 514 words, 1236 phrases and 5438 sentences involving information of six cancers and 23 detailed groups including the newly-created group. The whole corpus contained 7026 one-to-one corpus pairs, 160 one-to-many corpus pairs and two many-to-many corpus pairs. The statistics was shown in Figure 2. After checking, the number of partially correct alignments is 47 and the number of incorrect alignments is 23, where we could get the alignment ratio 99.35%.

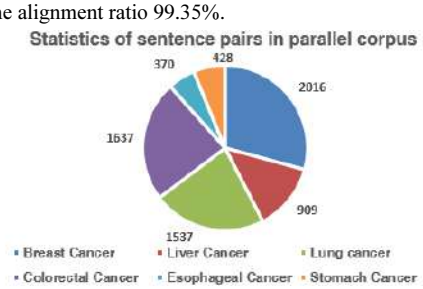


Figure 2 – Statistics of Sentence Pairs in English-Chinese PDQ Parallel Corpus.

Discussion

This corpus was constructed to provide the original prerequisite resouce data for real medical needs not limited to as follows:

Decision Making

Staff like medical researchers and doctors need an updated, authorized, and reliable medical knowledge base on cancer to efficiently assist them with decision making and problem solving.

Knowledge Dissemination

More people and patients are eager to learn professional knowledge about cancer, a knowledge base with cancer theme can satisfy these needs, and can help public education.

Cross-lingual Scenario

Parallel corpus, as a prerequisite resource, can help with machine reading, machine translation and machine understanding, which exactly matches the crucial requirement of those using different languages such as Chinese.

Conclusion

To help medical domain machine reading, understanding, and generating for possible smart question and answering systems, we have created a cancer focused multilingual parallel corpus of 7,188 pairs using the peculiar XML corpus, containing 5,438 pairs of English-Chinese sentences, 1,236 pairs of phrases and 514 pairs of words involving information of six cancers and 23 detailed groups through steps of corpus research, data processing, and corpus construction, which can be clarified into comparison between elements, analysis on parallel anchor, element alignment, sentence splitting, sentence alignment, design and analysis on data processing steps, design of parallel corpus construction strategy, and the final evaluation for the whole corpus.

In summary, the corpus we constructed is a deep-level knowledge Chinese-English corpus on medical domain, specifically, cancer information. General corpus is mostly sentence-level, we decided to make it a hybrid. To our knowledge, there is no similar or comparable resource like the corpus we presented in this paper. Our future work will keep on deeper research of better applications including machine translation on cancer knowledge, automatic information retrieval to achieve automatic update, automatic question and answer system based on this corpus.

Acknowledgements

The authors would like to thank the National Cancer Center of China and the US National Cancer Institute (NCI) for providing the bilingual cancer information. This study was supported by Chinese Academy of Medical Sciences CAMS Innovation Fund for Medical Sciences (Grant No. 2017-I2M-B&R-10), The National Key Research and Development Program of China (Grant No. 2016YFC0901901), the National Population and Health Scientific Data Sharing Program of China, and the Chinese Academy of Medical Sciences (Grant No. 2017PT63010, 2018PT33024).

References

[1] https://www.wikipedia.org/ [accessed March 29, 2019]
[2] J. Hirschberg and C.D. Manning, Advances in natural language processing, *Science* **349** (2015), 261-266.
[3] https://www.cancer.gov/publications/pdq [accessed March 29, 2019]
[4] http://www.phoc.org.cn/ailment/ [accessed March 29, 2019]

Address for Correspondence
Prof. Qing Qian is the corresponding author.
Tel: +86(10)52328701
Email: qian.qing@imicams.ac.cn
Address: No.3 Yabao Road, Chaoyang District, Beijing, China

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190522

eHOP Clinical Data Warehouse: From a Prototype to the Creation of an Inter-Regional Clinical Data Centers Network

Julia Madec, Guillaume Bouzillé, Christine Riou, Pascal Van Hille, Christian Merour, Marie-Lisen Artigny, Denis Delamarre, Veronique Raimbert, Pierre Lemordant , Marc Cuggia

Univ Rennes, CHU Rennes, Inserm, LTSI – UMR 1099, F-35000 Rennes, France

Abstract

Creation of networks such as clinical data centers within the hospital enables efficient exploitation of clinical data from a local to an inter-regional scope. This work present the structuration of the French Western Clinical Data Center Network (FWCDCN) conducted between 2016 and 2018. As of November 2018, FWCDCD is compounded with 7 institutions. CDW of the combinded Clinical Data Centers (CDC) contains the data of over 4 million patients followed since 2000.

Keywords:
Clinical Data Warehouse, Health Information System, Health Big Data

Introduction

Among strategic data sets, those produced at the hospital during patient care are already bringing new knowledge and technological innovations, from whom society and patients will benefit [1].

A CDC is an organization dedicated to exploit health bigdata for the hospital. The main IT infrastructure of a CDC relies on a cutting edge Clinical Data Warehouse technology called eHOP, developed by the academic hospital of Rennes.

The deployment of a such system in hospital institutions supposes to set up 3 main works: technical, functional and organizational.

Methods

Technical work

eHOP is an indoor development of Rennes University Hospital, using up-to-date technologies and compatible to the open-source platforms such as i2b2 and SHRINE

The system encompasses:

- Libraries : R
- Database management system : Oracle & MySQL
- Web servers used : Apache PHP (application server) and Java (database server)
- ETL : ENOVACOM Suite V2
- Interoperability standards : HL7, HPRIM, PN13, RSS
- Reference terminologies and mapping managed from UMLS

Functional work

Data flows have been chosen by each institution among all the data sources provided by the hospital (biology, drugs…). Data flows regarding the patient master index and hospital structures were mandatory to enable any clinical data flows .

Data integration procedures were the following :

- Real-time data integration : batches of dataset
- Takeover of the existing system

Every center has set up the request circuit with user and profile management. eHOP is only accessible on the center internal network from a web browser. Access is secured using the HTTPS protocol and traceability of requests and studies is ensured.

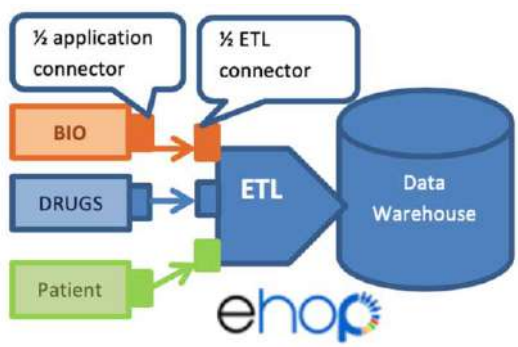


Figure 1 - eHOP data integration process

Organizational work

This work contents :

- CNIL (the French supervisory authority) formalities
- Reglementary procedure : patient information, patient opposition, user charters…
- Human resources assignation for project
- Creation of the corresponding structures :
- An administrative structure to set up strategic orientations of the CDC

- A regulatory structure to validate the access and exploitation policy of the CDW
- An operational structure for data processing. It undertakes to respect the following principles:
- Neutrality: No particular interest in the exploitation of data
- Confidentiality - Security: Exclusive and secure access to all (sensitive) data entrusted to it.
- Transparency: Carries out processing operations only if they are subject to agreement with the organizations and actors providing the data.
- Multidisciplinary medical (Methodological, IT and statistical), regulatory (Data Protection) and technical expertise.

eHOP CDW projects typologies

- **Non-interventional data research**: ex : epidemiological studies
- **Vigilance studies:** ex: therapeutic follow-up in real life
- **Support to health professionals**: ex: assistance in signal interpretation
- **Evaluation of practices**: ex: analysis of health trajectories

Results

The FWCDCN is now compounded with the following institutions : Angers University Hospital, Brest Regional University Hospital, West Cancerology Institute, Nantes University Hospital, Poitiers University Hospital, Rennes University Hospital, Tours University Hospital.

These institutions represent almost 4 million hospitalizations and consultations per year, with a total of 13 000 beds and a budget of almost 4 m€.

Table 1– eHOP documents by flows (mn) with historical depth-example of Rennes University Hospital (as of November 14, 2018)

Type of data flow	2010	2015	2018
medical surveys	-	1.1	3.5
PMSI	2.7	6.3	7.1
emergency report	0.3	0.7	1.8
medical report & doc	0.6	2.6	6.1
biology	0.4	4.9	8.3
anapath	0.1	0.2	0.3
drugs	-	1.2	22.1
Total	4.1	17	49.2

Conclusions

The FWCDCN structuring network makes it possible to share both an organizational model and good practices for the exploitation of data, and innovative tools designed to accelerate research. It creates the conditions to integrate and securely exploit other data sources such as registers epidemiological, administrative database data, and connected objects.

The main objective for the network for 2019 will be to position itself as a local node for the French national Health Data Hub.

Table 2– aggregated content of FWCDCN (as of November 14, 2018)

Historical Depth	Number of patients (mn)	Number of documents (mn)	Number of structured data (mn)
2000	0.1	-	-
2006	0.7	3	16
2009	1.4	12	109
2015	3.1	57	610
2018	4.3	102	1067

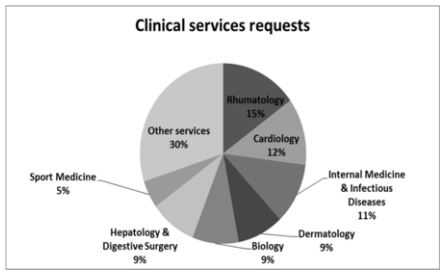


Figure 2– Clinical services requesting by number of studies- Rennes University Hospital CDC (From January 2014 to July 2017)

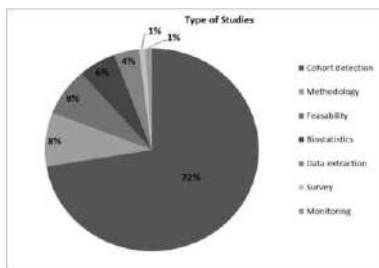


Figure 3– Type of studies at Rennes University Hospital CDC (From January 2014 to July 2017)

Acknowledgements

The authors would like to thank all the members of the French West Clinical Data Centers network for the conduction of the organizational work inside their institution during the deployment of eHOP clinical data warehouse.

References

[1] Andreu-Perez, C. C. Poon, R. D. Merrifield, S. T. Wong, et G.-Z. Yang, « Big data for health », IEEE J Biomed Health Inf., vol. 19, no 4, p. 1193–1208, 2015.

Address for correspondence

France - julia.madec@chu-rennes.fr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190523

Medical Equipment Replacement Prioritisation: A Comparison Between Linear and Fuzzy System Models

Norbert Maggi[a], Antonella Adornetto[a], Stefano Scillieri[a], Ezio Nicholas Bruno Urbina[b], Carmelina Ruggiero[a], Mauro Giacomini[a]

[a] Department of Informatics, Bioengineering, Robotics and Systems engineering, DIBRIS, University of Genoa, Italy
[b] S.C. RUP - Nuovo Galliera, EO Ospedali Galliera, Genoa, Italy;

Abstract

In hospital management, health technology assessment techniques are being increasingly developed. This paper presents a comparison of the results obtained using two models for replacement priority value calculation applied to the Galliera hospital in Genoa (Italy). One the models was developed at the Galliera Hospital along the lines of the model by Fennigkoh and addresses four primary replacement issues: equipment service and support, equipment function, cost benefits and clinical efficacy, by a "yes-no" scheme. This model is compared with a model based on fuzzy logic. The comparison between the two models shows a conservative behaviour by the Galliera model, according to which 77.4% of the analysed instrumentation is maintained, whereas the classification by the fuzzy model allows for a better discrimination among the devices.

Keywords:
Fuzzy Logic, Health technology assessment model

Introduction

In the last few years, very fast health technology changes have taken place, but they have not been paralleled by progress in management. Medical equipment requires very time-consuming and costly maintenance, which makes it crucial to introduce innovative technology management strategies focused on appropriateness, efficiency and cost effectiveness. Many health technology assessment (HTA) procedure have been proposed, often identifying mathematical models which require mostly subjective parameters. Many hospitals are carrying out research on models that objectively express obsolescence.

A linear multiparametric model for medical equipment replacement has been proposed by Fennigkoh [3]. This model can be adapted to each specific hospital by changes in the parameters and has often been used. A model based on this approach has been developed for the Busto Arsizio hospital (Northern Italy) [2] and has subsequently been adapted to meet the needs of the Galliera hospital in Genoa (Italy). Subsequently, a model based on fuzzy logic has been developed. A comparison of the two models is presented.

Methods

The model based on the Fennigkoh approach follows the one which has been adopted by Caimmi et al. [2] for the Busto Arsizio hospital, but it is simpler. In the first place some difficulties deriving from the retrieval of data such as maintenance cost for individual devices have been avoided. Secondly, the model by Fennigkoh requires choices as to which parameters should be regarded as relevant for decisions about substitutions. In this respect, some of these choices may not be generally valid and may not be the most appropriate in different contexts. Moreover, obtaining subjective opinion of medical users about the equipment in use and/or to be purchased is difficult. Therefore, the procedure for calculating the replacement priority value (RPV) has been divided into two phases. In the first phase (RPV1) only objective parameters are considered – and only if they exceed a threshold. Subjective data are considered in the second phase (RPV2). This allows to reduce the overall workload and the dependence on subjective factors. RPV2 is calculated only if RPV1 is above a predetermined threshold.

The following parameters have been taken into account.

- Age (x_1) – It is calculated by the following

$$x_1 = \begin{cases} 0, & if \ \dfrac{\text{current year} - \text{year of purchase}}{\text{functional age}} < 0 \\ 1, & otherwise \end{cases} \quad (1)$$

identifying if a device is over its functional age. The functional age is indicated by the manufacturer of the equipment and may also vary among devices belonging to the same group.
- Downtime (x_2) – A threshold value of 6 days was set up as indicated by the hospital. x_2 takes the value of 0 if it is below the threshold and 1 if it is above the threshold.
- Equipment function (x_3) – Same as in BA model (1 to 4).
- Manufacturer support, maintenance service, and availability of parts (x_4) – As in Fennigkoh's model, $x_4 = 0$ if parts, consumables, maintenance service, or manufacturer support are available or adequate (availability is guaranteed by law for 10 years after purchase); otherwise $x_4 = 1$.

RVP1 has been calculated as linear combination of these parameters according to the following

$$RPV_1 = 9(x_1 + x_2) + 7.5x_3 + 25x_4 \quad (2)$$

Ranking of the result is the same as in BA Model. Specifically: RPV1 < 40 → good conditions, no need to proceed, 40 ≤ RPV1 ≤ 60 → critical device, enter the second step, RPV1 > 60 → very critical device, replacement suggested as soon as possible. If RPV1 is between 40 and 60, it enters the second assessment phase, as in the BA model.

Fuzzy logic-based systems have been developed in hospital management only recently even though for analogue qualitative reasoning approaches has already been applied many years ago in bioengineering [1; 4; 5; 7; 8], where the problem of uncertain values of variables has similar effects. The limited availability of financial resources and of qualified personnel are among the

causes of improper and incorrect management of biomedical instrumentation. This has inspired some work [9] and the set up a fuzzy model for the classification of biomedical instrumentation according to the risk level [10]. Moreover, an inferential fuzzy model has been proposed [6].

Biomedical equipment can be classified by type of device (life support, therapeutics, diagnostics and others). Moreover, all instruments in the same class of equipment may not have the same functional age. In this respect, different fuzzy models have been produced which differ as relates to two features: rules (which allow a stricter assessment for life support or therapeutic equipment than for diagnostic or other devices) and age-related membership function. The variables that have been chosen are the same as for the GA model.

The model that has been developed is a fuzzy *mamdani* model in which preamble and conclusions are linguistic concepts (model with set of inferential rules). The model has the four variables (corresponding to the variables used in RPV1) and provides an output that suggests whether a device should be replaced. Specifically, the variable can give four suggestions: maintain, maintain over the functional age, re-evaluate (second phase), replace. The Fuzzy model presented here aims to replace the first phase of the GA model. The model has been implemented using the MATLAB Fuzzy Logic Toolbox.

For each input and output variable, the corresponding membership functions have been created. The membership functions define how each input (or output) value is mapped between 0 and 1. For each input, the degree of membership in fuzzy sets has been calculated by the application of IF-THEN type rules which have been set up on the basis of the suggestions by the Galliera Hospital. The fuzzified inputs have been applied to the antecedents: since the rules used have more than one antecedent, the AND intersection operator has been used to obtain the result of the antecedents. This operator provides the minimum degree of belonging among those present. The result of the antecedents has been applied to the function of membership of the consequent one according to the Clipping method, which simply cuts the consequent at the level of membership of the antecedent. The aggregation has allowed the unification of the consequent of all rules. The fuzzy set for the output variable has been obtained from this process. The last step is defuzzification, transforming the aggregates fuzzy set into RPV1.

Results and discussion

This work focuses on the analysis of six categories of instrumentation: electro-controlled bed for intensive care (BED), patient monitors (MON), biological refrigerators (BIR), surgical lights (SUL), ultrasound machines (USM), armchairs for therapy and blood samples (ARC). Both the GA model and the fuzzy model have been tested on the same equipment classes. Although most of the instrumentation has been analysed, for each class there is a number of biomedical equipment "not found" or for which it has not been possible to find the data necessary to calculate the RPV. The reasons for this are a number of factors, including problems in the outsourced service of management and monitoring of the technological park, difficulties in obtaining direct feedback from suppliers and/or technicians and equipment ageing.

The two models have been defined on the basis of different logics, but they do not differ greatly as relates to identification of the instrument RPV classes. The GA model has a tendency towards a more conservative behaviour, about 78.2% of the analysed instrumentation is maintained, 22% pass to the second phase of evaluation and suggests replacement only for 0.6%. The Fuzzy model suggests maintenance of the devices analysed for 70.8%. It also further specifies that 58.3% of the instruments are certainly to be maintained and that 12.6% of the instruments at this time are not a source of concern, but that their age or functional state could cause a significant increase in RPV in the future. This model also identifies the need for re-evaluation for 25.1% of the instrumentation analysed and 4% of the total recommends replacement without further evaluation.

The two largest differences between the models are found for ultrasound machines (USMs) and patient monitors (MON). Specifically, it is suggested to maintain 83.3% of the total USMs analysed according to the GA model and 58.3% according to the Fuzzy model, while for MON, the GA model identifies 36.7% of instruments to be maintained, 57.6% to be revalued and no instruments to be replaced immediately. The Fuzzy model identifies 15.1% of MON to be replaced.

At this initial stage, both models have been validated on a limited number of equipment classes. This is the basis of a much broader analysis. The Galliera Hospital showed interest in the Fuzzy model, whose results have been regarded as more satisfactory than those obtained by the GA model. Exploiting the uncertainty elements, as allowed by the fuzzy model, could be an advantage for the evaluation of the future availability of spare parts for each specific instrument.

In this respect, the Galliera Hospital intends to extend the application of the fuzzy model to the its entire technology park.

References

[1] S. Badaloni, B. Di Camillo, and F. Sambo, Qualitative reasoning for biological network inference from systematic perturbation experiments, *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* **9** (2012), 1482-1491.

[2] C. Caimmi, A. Bocchieri, A. Buizza, and G. Genduso, A new model for the replacement priority value of medical equipment, in: *5th Mediterranean Conference on Medical and Biomedical Engineering and Computing 2001: MEDICON 2001, 12-15 June 2001, Pula, Croatia*, 2001, pp. 183-186.

[3] L. Fennigkoh, A medical equipment replacement model, *J Clin Eng* **17** (1992), 43-47.

[4] S. Gaglio, M. Giacomini, C. Nicolini, and C. Ruggiero, A qualitative approach to cell growth modeling and simulation for cancer chemotherapy, *Biomedical Engineering, IEEE Transactions on* **38** (1991), 386-389.

[5] P. Linko, Uncertainties, fuzzy reasoning, and expert systems in bioengineering, *Annals of the New York Academy of Sciences* **542** (1988), 83-101.

[6] G. Mummolo, L. Ranieri, V. Bevilacqua, P. Galli, F. Menolascina, and G. Padovano, A fuzzy approach for medical equipment replacement planning, in: *Proceedings of the 3rd International Conference on Maintenance and Facility Management*, 2007, pp. 229-235.

[7] C. Ruggiero, M. Giacomini, and S. Gaglio, A qualitative model of the dynamics of blood glucose and its hormonal control, *Computer methods and programs in biomedicine* **40** (1993), 117-130.

[8] C. Ruggiero, M. Giacomini, O. Varnier, and S. Gaglio, A qualitative process theory-based model of the HIV-1 virus-cell interaction, *Computer methods and programs in biomedicine* **43** (1994), 255-259.

[9] S. Taghipour, D. Banjevic, and A.K.S. Jardine, Prioritization of medical equipment for maintenance decisions, *Journal of the Operational Research Society* **62** (2010), 1666-1687.

[10] B. Tawfik, B.K. Ouda, and Y.M. Abd El Samad, A Fuzzy Logic Model for Medical Equipment Risk Classification, *Journal of Clinical Engineering* **38** (2013), 185-190.

Address for correspondence

Mauro Giacomini, via all'Opera Pia, 13 – 16145 Genova, Italy

Email: mauro.giacomini@dibris.unige.it Tel +39 010 35 56546

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190524

Recovery Medication from Free Text to a Structured Form

Humberto Fernán Mandirola Brieux[a], Santiago Orozco[a], Marisa Lanfranconi[a], Jorge De All[a], Nahin Chedresse[b], Sebastián Guillen[c], Alejandro Deporte[c]

[a]Sanatorio Otamendi, Buenos Aires, Argentina,
[b]Hospital Naval Puerto Belgrano (HNPB), Argentina,
[c]BIOCOM Sistemas, Argentina

Abstract

We studied methods to convert medical prescriptions in free text to a structured form for pharmacy insructions and planning nursing activities in hospitalized patients. We compared Natural Language Processing (NLP) with Parsing Process (PP), both for the Spanish language. We studied 87,750 and processed 65,000 prescriptions and recovered 62% and 65% with NLP and PP to a structured format respectively. The difference between the methods is significant ($p < 0.001$) and further work is needed to determine if combining them will have higher performance.

Keywords:

Natural Language Processing, Prescriptions, Medical Order Entry Systems

Introduction

Physicians usually prescribe medications in free text, however, computer systems need this information in a structured way to organize the data so that the pharmacy can deliver the medications and nurses can organize their tasks [1]. Computerized Provider Order Entry (CPOE) can reduce medication errors; but, its benefits are only achieved when data is entered in a structured format and entries are properly coded [2].

To process the information by computer, the system extracts the data from many fields. The system analyzes the following information: generic drug name, dose, strength, posology, route of administration, care and recommendations that are required to comply with the medication.[3].

FHIR standard (Fast Healthcare Interoperability Resources) has resources to manage pharmaceutical orders but it needs this information to be structured. Thus it is fundamental to perform the data extraction in different fields to match the FHIR interoperability standard [4].

Automated dispensing cabinets for pharmacies use HL7 and the information to construct the messages needed in a structured way [5] [6] .

A prescription is an order for medication which is dispensed to a patient. The medical prescription is not enough for the pharmacy to deliver the medication of the day to different departments. It requires prior processing so that the pharmacist knows the daily amount of each drug to deliver [7]. The pharmacist verifies the legality, safety and appropriateness of the prescription order, checks the patient medication record before dispensing the prescription (when such records are kept in the pharmacy), ensures that the quantities of medication are dispensed accurately and decides whether the medication should be handed to the patient, with appropriate counselling, by a pharmacist [8].

Even if doctors are becoming more accustomed to entering data in a structured way using CPOE, there are prescriptions in free text form that need to be processed.

Our question is: “How good can the medical prescription extraction process be when using a free text input for Spanish language?”

We studied Natural Language Processing (NLP) and Parsing Process (PP) as two possible methods to process unstructured text.

Methods

The objective of this work is to demonstrate that it is possible to recover information about medications in a structured format. Passing the narrative medical prescription through NLP is a helpful tool for doctors to better transform it in structured format for hospitalized patients.

We build the NLP routine for this study in C# .net using the Stanford CoreNLP technology tools.

The PP routine was developed by our programmers. Since the data we have uses a fixed pattern, we use the regular expression technique to extract and control its elements.

A retrospective,,observational, cross-sectional, descriptive study was designed. The sample consisted of 87,750 medical prescriptions collected consecutively during March 1th, 2017 and March 1th, 2018. The medical prescriptions were made by the institution's staff and independent external physicians who assist patients within the institution.

We used the following criteria for the study inclusion:

- Adults pharmacological prescriptions
- Singular generic drugs (for example, Losacor®", that has only one drug (Losartan) but not "Losacor D®" that has two drugs Losartan and Hydrochlorothiazide)
- Commercial names where used (for example, "Losacor®" is acceptable for Losartan)
- One prescription (100 characters per line or field) and one drug per line
- Spelling mistakes and different ways of writing each prescription elements that we analyzed

We used the following criteria for the study exclusion:

- Non-pharmacological prescriptions such as prescription for nutrition, vital signs, studies or any other type of indication that is not pharmacological.
- Pediatric medication

- Parenteral hydration plans
- Medications with more than one drug
- Multiple line medications

We included 65,000 and excluded 22,750 medical prescriptions collected consecutively with this process.The medical prescriptions that met the inclusion criteria were processed by the two algorithms.

The average age of the study population of the samples was 67 years old, with the maximum age of 96 and the minimum age of 22 years. The percentage of males participating in the study was 45.3% and the female participation was 54.7%.

In this study we tested two algorithms:

1. The PP method
2. The NLP algorithm

We considered it a success when the algorithm could complete the five fields and the pharmacist confirmed that it was correct.

Results

The results were the following: we were able to recover 62% and 65% with the NLP and the PP method respectively to a structured format that could be processed by computer systems. Please see tables 1 and 2.

The difference between the two methods was statistically significant when calculated using the Z-test with $p < 0.001$.

Table 1 – Results in Absolute Values

	NLP	PP
Correct	40,300	42,250
Incorrect	24,700	22,750

Table 2 – Results in Percentage

	NLP	PP
Correct	62%	65%
Incorrect	38%	35%

Discussion

The PP routine used the regular expression technique to extract data (i.e., narrative text or speech data) in the context of a specific task. Our results have shown that both algorithms are capable of detecting more than 60% of the medical prescriptions made with free text. However, there is a significant difference between the two methods.

The NLP method has the possibility to learn and improve over time, while the PP cannot improve by itself. It requires more expressions to be loaded by manual programming.

Conclusions

It remains to be seen if these methods can achieve higher performance in future. It is possible to make the system learn from medical prescriptions through a knowledge thesaurus. We will work to incorporate these challenges in future work.

Acknowledgements

We thank Dr. Fernán Gonzalez Bernaldo de Quirós , Dr. Daniel Luna, Dra Paula Otero & Dr. Carlos Otero from Hospital Italiano de Buenos Aires for their asistance and corrections of this work.

References

[1] E.S. Chen, P.D. Stetson, Y.A. Lussier, M. Markatou, G. Hripcsak, and C. Friedman, Detection of practice pattern trends through Natural Language Processing of clinical narratives and biomedical literature, *AMIA Annu Symp Proc*. (2007) 120–124.

[2] L. Zhou, L.M. Mahoney, A. Shakurova, F. Goss, F.Y. Chang, D.W. Bates, and R.A. Rocha, How Many Medication Orders are Entered through Free-text in EHRs? - A Study on Hypoglycemic Agents, *AMIA Annu Symp Proc*. **2012** (2012) 1079–1088.

[3] Ö. Uzuner, I. Solti, and E. Cadag, Extracting medication information from clinical text, *J Am Med Inform Assoc*. **17** (2010) 514–518. doi:10.1136/jamia.2010.003947.

[4] Extension: rateGoal - FHIR v3.0.1, (n.d.). https://www.hl7.org/fhir/extension-pharmacy-core-rategoal.html (accessed November 17, 2018).

[5] Automation Anywhere - Largest Enterprises Have Made Their Choice for RPA, (n.d.). https://www.automationanywhere.com/lp/largest-enterprises-have-made-their-choice?utm_source=google&utm_medium=cpc&utm_campaign=rpabroad_ROW&keyword=Robotic%20Process%20Automation&match=b&network=g&place=&gclid=CjwKCAiAlb_fBRBHEiwAzMeEdn7b6HadbZih5i83jnJhLb1KbcqaR8K5JzQ2EWdB-njKFYFOV1PPiBoC970QAvD_BwE (accessed November 17, 2018).

[6] N.F. Carolina PhD, PharmDDirector, Pharmacy Practice LaboratoriesWingate University School of PharmacyWingate, North, Safety of Automated Dispensing Systems, (n.d.). https://www.uspharmacist.com/article/safety-of-automated-dispensing-systems (accessed November 17, 2018).

[7] H. Liu, Q. Burkhart, and D.S. Bell, Evaluation of the NCPDP Structured and Codified Sig Format for e-prescriptions, *Journal of the American Medical Informatics Association : JAMIA*. **18** (2011) 645. doi:10.1136/amiajnl-2010-000034.

[8] The Role of the Pharmacist in the Health Care System: PART I: THE ROLE OF THE PHARMACIST IN THE HEALTH CARE SYSTEM: 4. The scope of pharmacy and the functions of pharmacists: 4.2 Community pharmacy, (n.d.). http://apps.who.int/medicinedocs/es/d/Jh2995e/1.6.2.html (accessed November 13, 2018).

Address for correspondence

Humberto F Mandirola Brieux. Email hmandirola@gmail.com; Mobile Phone +54911-44174879.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190525

Disseminating Research Findings: The Crowdhealth Paradigm

Andriana Magdalinou[a], John Mantas[a], Parisis Gallos [a], Lydia Montandon[b]

[a]European Federation for Medical Informatics, Lausanne, Switzerland,
[b]ATOS, Spain

Abstract

The CrowdHEALTH project intends to integrate high volumes of health-related heterogeneous data from multiple sources with the aim of supporting policy making decisions. The European Federation of Medical Informatics supports the development of an effective Communication and Collaboration Plan.A dissemination strategy has been applied for this purpose considering appropriate messages, target audience, tools, and channels to achieve the highest impact and the first results of social media dissemination are presented here.

Keywords:
Social Media, Policy Making, Intention

Introduction

The CrowdHEALTH project intends to integrate high volumes of health-related heterogeneous data from multiple sources with the aim of supporting policy-making decisions. The CrowdHEALTH is delivering a secure ICT platform that seamlessly integrates Big Data technologies, providing Data as a Service (DaaS) and a Data Analysis Toolkit. The transition from patient health records towards the Holistic Health Records (HHRs) and Social HHR is also proposed.

The European Federation of Medical Informatics (EFMI) supports the development of an effective Communication and Collaboration Plan identifying the messages and the tools and channels in disseminating the project and its outcomes to the target audience based on the McGuire approach. The main objective of the strategy envisioned is to engage and inform the target audience about the CrowdHEALTH project and its outcomes [1]. It is important that a large number of audience is addressed through appropriate means in order to be engaged with the progress of the project and raise its awareness about the project and the deliverable outcomes. Each applied tool has different strengths and weaknesses in reaching the audience and therefore by using more than one tool, they complement one another to produce a strong dissemination plan [2]. Social media can play a significant role in all phases of the research lifecycle, from identifying research opportunities to disseminating findings. Scientific journals, conferences, and edited books remain the core traditional means of disseminating research, but social media has become an important channel for disseminating research [3].

In the CrowdHEALTH project a cyclical process was developed. In the cyclical models, individual components are linked, and the process is depicted as interactive and ongoing. This is the case with Graham et al's knowledge to action model where aspects of the research, context, knowledge transfer intervention, and evaluation lead to the identification of new opportunities [4]. Researchers need to adopt a theory-informed approach to their research dissemination [5].

Methods

Dissemination strategy is defined through a cyclical process. The McGuire Persuasive communication approach was implemented for the development of an effective Collaboration and Communication plan, figuring out the dissemination objective, the message to be communicated, the tools and channels of communication, and the target audience. The statistics of the CrowdHEALTH Facebook page are monitored to give first evidence of audience engagement in social media and the published material.

Results

Defining a Dissemination Strategy Through a Cyclic Process

Defining the CrowdHEALTH dissemination strategy is an ongoing cyclic process as shown in Figure1. Past or published ongoing dissemination efforts are reviewed, the available resources are identified, the momentum and opportunities are considered and the four dimensions of the dissemination strategy consisting of the objective of the dissemination, the key messages to be communicated, the appropriate dissemination tools and channels, and the target audience are defined (Figure2). KPIs are monitored in order to evaluate dissemination efforts.

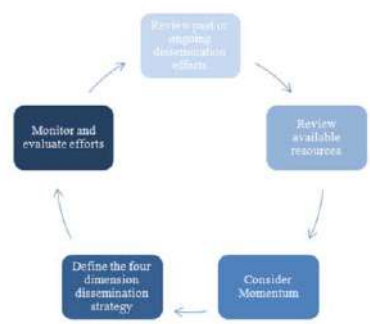


Figure 1 – The Cyclical Process of a Dissemination Strategy

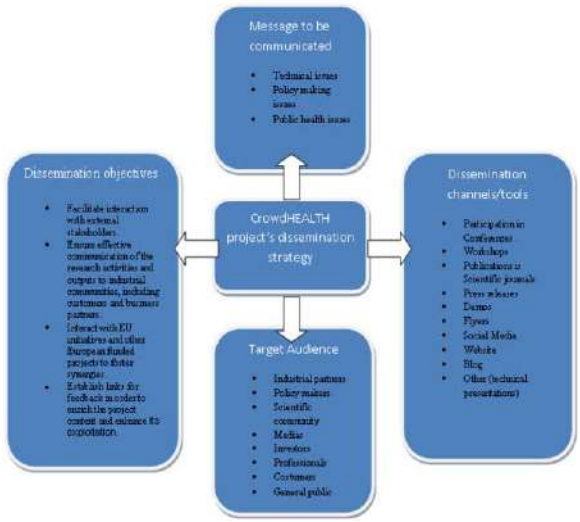


Figure 2 – The Four Dimensions of the Dissemination Strategy

Monitoring Efforts of Social Media Dissemination

Data regarding the people reached, the post clicks, and reactions (shares, likes, comments) are collected from 28 December 2017 to 24 November 2018 from the CrowdHEALTH Facebook page as illustrated in Figure 3.

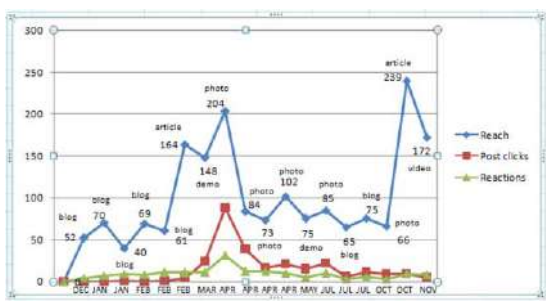


Figure 3 – The Impact of Blog Posts and Other Published Material on the CrowdHEALTH Facebook Page

Discussion

The CrowdHEALTH project implemented a cyclical process for defining the dissemination strategy consisting of five steps. Taking into consideration this framework , the four pillars of the dissemination strategy were described, defining the dissemination objective, the message to be communicated, the tools and channels of communication, and the target audience. Early evidence from the CrowdHEALTH social page on Facebook suggests that audience is more engaged to articles, demos, videos, and photos from events and conferences and less on blog posts.

Conclusions

The implemented framework identified a dissemination strategy to better communicate project's findings. Optimal communication and dissemination would also facilitate interaction with the target groups. It is important that a large number of audience be addressed through appropriate means in order to be engaged with the progress of the project and raise their awareness about the project and its deliverable outcomes. Social media can play a significant role in all phases of the research lifecycle indluding the dissemination of research findings. Scientific journals, conferences, and edited books remain the core traditional means of disseminating research, although social media has become an important channel for disseminating research. Each applied tool has different strengths and weaknesses in reaching the audience and therefore by using more than one tool, they complement one another. First evidence of the CrowdHEALTH project suggests that press releases connected with social media pages of the project can engage audience in social media. Visual materials such as demos, videos, pictures, and photos can highly engage audience in social media.

Acknowledgements

The CrowdHEALTH project is co-funded by the Horizon 2020 Programme of the European Commission Grant Agreement number: 727560 – Collective wisdom driving public health policies.

References

[1] World Health Organization, *Disseminating the research findings,* WHO Document Production Services, Geneva, Switzerland.

[2] K. G. Shaver, Principles of Social Psychology, Third Edition, *Psychology Press* (2015), 203-205

[3] D. Nicholas and I. Rowlands, Social media use in the research workflow, *Information Services & Use* (2011), 61-63

[4] V. Ward,A. House,and S. Hamer , Developing a framework for transferring knowledge into action: a thematic analysis of the literature, *J Health Serv Res Policy (*2009) 156–164

[5] P. M. Wilson, M. Petticrew, M.W. Calnan, and I. Nazareth, Disseminating research findings: what should researchers do? A systematic scoping review of conceptual frameworks, *Implement Sci.* ,(2010) 91

Address for correspondence

Andriana Magdalinou, RN, MSc
Email address : alfami@nurs.uoa.gr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190526

The Opportunities and Challenges of Pragmatic Randomized Trials Using a Specialized Software: CloudTrials Project

Santiago Márquez Fosser [a], Andrey Kamozin [b], Emily Gibson McDonald [b], Robyn Tamblyn [a,c], Todd Campbell Lee[a,b,c]

[a] Clinical and Health Informatics Research Group, McGill University, Montréal, Canada
[b] Department of Medicine, McGill University, Montréal, Canada
[c] Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montréal, Canada

Abstract

Pragmatic randomized trials are essential to improve knowledge of real world clinical practice. Pragmatic trials design reflect routine clinical care, which have advantages to initiate and conduct, compared to classical clinical trials. Trials help to: engage bedside clinicians, increase efficiency of patient recruitment and follow up, minimize loss to follow up and include technological patient reported outcomes. The objective is to describe the opportunities and challenges designing a specialized software to administrate pragmatic randomized trials.

Keywords:
Research Design, Patient Selection, Software Design

Introduction

A randomized controlled trial (RCT) is considered the best method to evaluate therapeutic efficacy. In a RCT, one or more groups of patients receive different treatments in order to compare the results across the different groups with minimal bias. Ideally, RCTs should apply interventions that resemble actual clinical practice and enroll representative samples of clinicians and patients. However, they are generally very expensive, time consuming, and are complex to carry out [1]; consequently, they are not accessible to most physicians and patients.

The objective of these trials (also known as explanatory trials) is to test the superiority, equivalence, or non-inferiority of the new treatment being tested [2]. Clinical trials also commonly require the collection of highly detailed patient data. This necessitates onsite staff training regarding data collection and management, which presents a burden to study sites, interferes with usual care, and presents a barrier which often excludes research-native practitioners and sites from participation [3-4].

A different type of trial, known as a pragmatic trial, aims to identify the best choice between clinical decisions which are usually made in real practice or which answer important questions about whether a practice is beneficial. Pragmatic trials need to closely align all study-related interventions with routine clinical practice so as to minimize interference with routine care and enhance the generalizability of the results to the real-word setting [5].

Our objective was to create an application to serve as a clinical trials platform which overcomes many of the barriers and methodological challenges to designing, administering, conducting, monitoring, and analyzing pragmatic clinical trials. This platform was designed to be independent from the EMR to allow participation in many centers independent of their underlying system and to include, from the beginning, a patient voice in terms of patient reported outcomes. The software seeks to do all of this while maintaining the highest methodological rigor for the study and minimizing the administrative and human resource costs of participation.

Methods

This article uses a methodology that assesses and describes obstacles in the conduct of pragmatic randomized trials as well as avenues for future developments of software-based solutions.

Setting

The McGill Clinical and Health Informatics (MCHI) is a research group in the Faculty of Medicine whose research activities focus on the development and evaluation of health software and the analysis of data generated by these systems.

Challenge

When interference with usual care or the administrative and/or data collection burden for physicians is too high, many physicians will not participate. By contrast, if the data collection is incomplete or not standardized, the entire methodology of the study can be called into question. A strategy that aims to collect information through an EHR has the advantage of not interfering with routine clinical care, however, it may not have the data quality needed [6].

Results

CloudTrials components

Establishing common structures across different clinical trials allows for the design of a scalable software which could facilitate multiple trials.

User administration

The development of methods of user registration and validation was essential. A user signs up by filling in their basic professional information with a system verification of their email account, and, in this case, a manual verification by an administrator who confirms identity with provincial provider databases and again via direct communication and the

completion of any legal contracts and agreements necessary in that jurisdiction.

Patient enrollment

Each doctor is responsible for the recruitment and enrollment of their patients in the system. For that a 5-step wizard was designed, with the minimum requirements to incorporate a patient in the study: *Patient profile, Eligibility by Clinical Information, Verbal Consent. Follow Up, Summary.*

Randomization

At the end of patient enrollment, the system assigns the patient to an intervention group using a pre-prepared lookup table which has been previously produced for the study in question using proper statistical methods for the question being investigated.

Patient list

To allow for management of their enrolled patients and to identify tasks remaining to be performed, a list of patients had to be developed.

Follow up

Following a pre-designed schedule which is customized for each trial, this module allows for documentation of follow up visits and tests during the period of observation.

Data monitoring

An administrator view is necessary for the supervision of any RCT. This user is required to oversee the study progress to apply corrections early on when the study is deviating from the planned course.

Privacy and security

The biggest concern while developing the application was the safety of patients' identities. Outside of important security features employed at the server levels, several additional measures were put in place:

Ethical and regulatory challenges

While somewhat outside of the scope of a document discussing the software development of a clinical trials platform, it is relevant to consider the challenges of performing crowd sourced clinical trials. Namely, such studies need to be mindfully designed such that they allow clinicians to enroll patients ethically within their own practice drawing upon the pre-existing informed consent process and professional code of conduct.

Discussion

This article describes the issues and challenges we faced in the design of our software which allows for the administration of PRTs. The incorporation or creation of a module for the generation of new structured clinical trials by end users is a key step in our future to allow for maximal reuse of existing code prior to launching a new trial. We are excited to see our first two trials, involving hundreds of patients, enrolled in 2019.

Conclusions

This article describes the design of a software which allows for the administration of PRTs.

Acknowledgements

Drs. Lee and McDonald receive research salary support from the Fonds de Recherche Québec, Santé and the CloudTrials project has benefited from financial support from the Clinical Practice Assessment Unit of the McGill University Health Centre and the material support of the McGill Clinical Health Informatics Research Group.

References

[1] T.-P. van Staa, L. Dyson, G. McCann, S. Padmanabhan, R. Belatri, B. Goldacre, J. Cassell, M. Pirmohamed, D. Torgerson, S. Ronaldson, J. Adamson, A. Taweel, B. Delaney, S. Mahmood, S. Baracaia, T. Round, R. Fox, T. Hunter, M. Gulliford, and L. Smeeth, The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials, *Health Technol. Assess* **18** (2014), doi:10.3310/hta18430.

[2] D. Schwartz, and J. Lellouch, Explanatory and Pragmatic Attitudes in Therapeutical Trials, *J. Clin. Epidemiol* **62** (2009), 499–505.

[3] J. McCambridge, J. Witton, and D.R. Elbourne, Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects, *J. Clin. Epidemiol* **67** (2014), 267–277.

[4] K. Martin, B. Bégaud, P. Latry, G. Miremont-Salamé, A. Fourrier, and N. Moore, Differences between clinical trials and postmarketing use, *Br. J. Clin. Pharmacol.* **57** (2003), 86–92.

[5] M.G.P. Zuidgeest, I. Goetz, R.H.H. Groenwold, E. Irving, G.J.M.W. van Thiel, D.E. Grobbee, and GetReal Work Package 3, Series: Pragmatic trials and real-world evidence: Paper 1. Introduction, *J. Clin. Epidemiol* **88** (2017), 7–13.

[6] A.-K. Meinecke, P. Welsing, G. Kafatos, D. Burke, S. Trelle, M. Kubin, G. Nachbaur, M. Egger, M. Zuidgeest, and work package 3 of the GetReal consortium, Series: Pragmatic trials and real world evidence: Paper 8. Data collection and management, *J. Clin. Epidemiol* **91** (2017), 13–22.

Address for correspondence

Dr. Márquez Fosser Santiago. Postdoctoral Fellow. McGill Clinical Health Informatics Research Group.

email: santiago.marquezfosser@mcgill.ca

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190527

Understanding Urgency in Radiology Reporting: Identifying Associations Between Clinical Findings in Radiology Reports and Their Prompt Communication to Referring Physicians

Xing Meng[a†], Michael V. Heinz[b†], Craig H. Ganoe[c], Ryan T. Sieberg[d], Yvonne Y. Cheung[d], Saeed Hassanpour[a,c,e]

[a] *Department of Computer Science, Dartmouth College, Hanover, New Hampshire, USA*
[b] *Department of Psychiatry, Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire, USA*
[c] *Department of Biomedical Data Science, Dartmouth College, Hanover, New Hampshire, USA*
[d] *Department of Radiology, Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire, USA*
[e] *Department of Epidemiology, Dartmouth College, Hanover, New Hampshire, USA*
[†] *These authors contributed equally to this work*

Abstract

In this study, we aim to develop an automatic pipeline to identify clinical findings in the unstructured text of radiology reports that necessitate communications between radiologists and referring physicians. Our approach identified 20 distinct clinical concepts and highlighted statistically significant concepts with strong associations to cases that require prompt communication.

Keywords:

Natural language processing, communication, radiology.

Introduction

It is well understood in clinical practice that lapses in communication, either due to delays or a lack of communication altogether, increase the likelihood of adverse patient outcomes. In radiology, high case volumes and the expectation of timely clinical interpretation present challenges in identifying and communicating cases that require urgent management on the part of the referring physician. In this study, we developed a pipeline for the identification of critical clinical concepts that are the most likely to appear in patient cases requiring prompt communication between a radiologist and a referring physician.

Related Work

There are many examples of previous uses of automatic information extraction from medical records, particularly with radiology report text analysis [1]. MetaMap has been developed and utilized to map radiology reports and other clinical note texts to concepts in the Unified Medical Language System (UMLS) Metathesaurus [2]. Apache clinical Text Analysis and Knowledge Extraction System (cTAKES) is a system to analyze clinical notes and annotate UMLS concepts in free text [3], which is utilized in this project and will be further discussed in the following sections. Of note, this study is built on our previous work to use machine learning to identify urgent cases for prompt communication in radiology reporting [4].

Methods

We developed a natural language processing (NLP) pipeline to extract clinical findings from a corpus of free-text radiology reports at our institution that required prompt communication between radiologists and referring physicians. We then compared these clinical findings to the extracted findings from regular radiology reports. We used a statistical test of significance to identify critical findings in these reports that are associated with prompt communication. The overview of our pipeline is shown in Figure 1.

Dataset

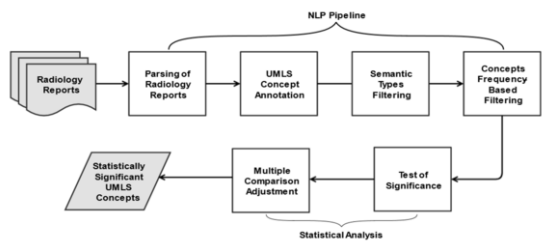


Figure 1– An Overview of Our Framework to Identify Statistically Significant Clinical Findings Requiring Prompt Communication with Referring Physicians Based on Radiology Reports

We took a retrospective approach in this research, starting with 1,460 radiology reports extracted from electronic medical records within the years of 2010 – 2018 from Dartmouth-Hitchcock Medical Center (DHMC). These reports came from a mixture of three imaging modalities: (1) computed tomography (CT); (2) magnetic resonance imaging (MRI); and (3) X-ray. Two radiologists manually annotated each of these reports, adding a binary label, either "this report requires prompt communication to referring physician" or "this report does not require prompt communication." In order to minimize the errors, reports that received the same label from both radiologists were included in the analysis and those that did not were excluded. Our final data set included 1,389 radiology reports. 1,057 of these were interpreted by both radiologists as requiring urgent communication to a referring physician, while 332 were interpreted as not requiring urgent communication.

NLP Pipeline

Parsing of Radiology Reports

We focused only on analyzing the text that appears in the "impression" sections of radiology reports. This is because the impression section contains information about diagnosis and follow-up recommendations.

UMLS Concept Annotation

We applied Apache clinical Text Analysis and Knowledge Extraction System (cTAKES), which identifies nouns and noun phrases in clinical text that correspond to UMLS concepts. Using cTAKES, we identified the concepts in radiology reports that are classified in 5 high level classes based on a subset of UMLS semantic types: (1) disorders/diseases, (2) signs/symptoms, (3) procedures, (4) anatomy, and (5) drugs. cTAKES also provides preferred names from UMLS for extracted concepts to facilitate a standard representation. In

addition, this pipeline incorporates NegEx [5] to identify the concepts that are used in a negative context in the radiology report. We mapped all the medical terms and phrases that appeared in the impression section to UMLS concepts, excluding those that were negated.

Semantic Types Filtering

The identified UMLS concepts were aggregated according to the five high level semantic type classes. Using the full list of concepts identified by cTAKES, we found that the aggregated concepts helped to narrow down a list of potential concepts associated with the need for prompt communication with referring physicians. After consultation with the domain-expert clinicians involved in our study, we focused solely on UMLS concepts belonging to the diseases/disorders class in our analysis from the five aforementioned UMLS classes returned by cTAKES.

Concepts Identified by Frequency-Based Filtering

The identified UMLS concepts found in the reports were also filtered by frequency, including those that were only found in 2% or greater of reports. This frequency-based filtering reduced the noise in our analysis and directed our focus to reasonably common findings in radiology reports that encompass the majority of cases requiring prompt communication. The concept frequencies were further analyzed in the following statistical analysis steps to differentiate reports requiring prompt communication from those not requiring prompt communication.

Statistical Analysis

We used Fisher's exact test to find statistically significant associations between identified UMLS concepts and the radiology reports that require prompt communication. The concepts with a P-value < 0.05 are considered strongly associated with necessary prompt communication with referring physicians.

We addressed the multiple comparisons problem in evaluating the significance of extracted UMLS concepts. The false discovery rate provides a complementary measure to positive predictive value, which indicates the probability of a positive test result being accurate. In order to identify truly significant UMLS concepts while still maintaining a low false positive rate, we adjusted the P-values calculated by Fisher's exact test. To adjust the P-values, we utilized the Benjamini-Hochberg method, which is a powerful method to address the multiple comparisons problem through false discovery rate.

Results

The described NLP pipeline identified 580 UMLS concepts. Table 1 shows the extracted concepts with adjusted P-values less than 0.5 from radiology reports that are associated with reports requiring prompt communication. The most commonly identified UMLS concepts in radiology reports that required prompt communication with referring physicians are shown in Table 2.

Table 1– Frequency of Selected UMLS Concepts in Radiology Reports that Require Prompt Communication
P = Occurrences with Prompt Communication
N = Occurrences without Prompt Communication

UMLS Concept	*Semantic Type*	*P*	*N*	*Odds Ratio*	*P-value*	*Adjusted P-value*
Pneumothorax	Disease or Syndrome	33	0	∞	2.25E-04	4.19E-02
Malignant Neoplasms	Neoplastic Process	73	3	8.05	2.66E-06	7.72E-04
Fracture	Injury or Poisoning	127	7	6.27	3.34E-09	1.94E-06
Nodule	Acquired Abnormality	107	13	2.73	2.89E-04	4.19E-02

Conclusions

We developed an automatic pipeline to identify clinical concepts that are statistically significant for radiology reports requiring prompt communication with referring physicians. However, our imbalanced dataset containing urgent and non-urgent cases for prompt communication is a potential limitation of this study. We plan to address this limitation by extending our dataset in our future work.

Table 2– Most Frequency UMLS Concepts in Impression Sections of Radiology Reports that Require Prompt Communication

P = Positive for Occurrences，N = Negative for Occurrences

UMLS Concept	*Semantic Type*	*P*	*N*
Fracture	Injury or Poisoning	127	7
Nodule	Acquired Abnormality	107	13
Malignant Neoplasms	Neoplastic Process	73	3
Communicable Diseases	Disease or Syndrome	62	6
Neoplasms	Neoplastic Process	53	8
Pleural effusion disorder	Disease or Syndrome	50	5
Traumatic injury	Injury or Poisoning	49	5
Cyst	Disease or Syndrome	43	6
Disorder of skeletal system	Disease or Syndrome	42	16
Protrusion	Anatomical Abnormality	39	15
Intervertebral Disc Degeneration	Disease or Syndrome	39	11
Pneumothorax	Disease or Syndrome	33	0
Abscess	Disease or Syndrome	29	2
Pneumonia	Disease or Syndrome	28	3
Squamous intraepithelial lesion	Neoplastic Process	26	13
Laceration	Injury or Poisoning	23	14
Arthropathy	Disease or Syndrome	22	4

Acknowledgements

The authors would like to thank William Keller, MD, Lamar Moss, and Sophie Montgomery for their help in this project.

References

[1] S. Hassanpour, and C.P. Langlotz, Unsupervised topic modeling in a large free text radiology report repository, J. Digit Imaging, 29, 1 (2016) 59-62.

[2] A.R. Aronson, Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program, Proc. AMIA Symp. (2001) 17.

[3] G.K. Savova, J.J. Masanz, P.V. Ogren, J. Zheng, S. Sohn, K.C. Kipper-Schuler, and C.G. Chute, Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, component evaluation and applications, J. Am. Med. Informatics Assoc. 17 (2010) 507–513.

[4] X. Meng, C.H. Ganoe, R.T. Sieberg, Y.Y. Cheung, and S. Hassanpour, Assisting radiologists with reporting urgent findings to referring physicians: A machine learning approach to identify cases for prompt communication, J. Biomed. Inform. (2019) In press.

[5] W.W. Chapman, W. Bridewell, P. Hanbury, G.F. Cooper, and B.G. Buchanan, A simple algorithm for identifying negated findings and diseases in discharge summaries, J. Biomed. Inform. 34 (2001) 301–310.

Address for correspondence
Saeed Hassanpour, PhD
One Medical Center Drive, HB 7261, Lebanon, NH 03756, USA
Fax: (603) 650-1966 Email: Saeed.Hassanpour@dartmouth.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190528

Enabling West African Herbal-Based Traditional Medicine Digitizing: The WATRIMed Knowledge Graph

Borlli Michel Jonas Somé[a], Georgeta Bordea[b], Frantz Thiessard[b], Gayo Diallo[b]

[a] Ecole Supérieure d'Informatique, Université Nazi Boni, Bobo Dioulasso, Burkina Faso
[b]Team ERIAS, Bordeaux Population Health INSERM 1219, Univ. Bordeaux, F-33000, Bordeaux, France

Abstract

The purpose of this study is to describe the design and development of the first release of the West African Herbal based Traditional Medicine Knowledge Graph (WATRIMed). It is a resource containing Traditional Medicine (TM) related entities and linked with publicly available knowledge bases in order to facilitate bringing West African TM into the digital world. The core model comprises currently 556 concepts including 143 identified West African medicinal plants and 108 recipes used by tradi-practitioners to treat 110 diseases and symptoms which are commonly encountered in this part of the world.

Keywords:
African Traditional Medicine, Knowledge Bases; Phytotherapy

Introduction

For many people in Africa, Traditional Medicine (TM) is either the first line of treatment or is used as a last resort when all the available possibilities in the conventional medicine are exploited. Although its affordability, it comes with various issues, in particular due to the oral transmission of knowledge and lack of digitized resources that could contributed to making gathered experiences sustainable.

The high usage of TM is often driven by the inaccessibility, unaffordability or unavailability of conventional health care services and medicines in socioeconomic settings that are characterised by a high rate of poverty and a lack of suitable and affordable conventional medicine services and drugs, like in West Africa. That underserved and mostly illiterate, rural people account for the majority of the population, is an additional barrier to access healthcare. In response to the growing recognition of the potential of TM, the supra-national West African Health Organisation (WAHO)[1] has given priority to TM in 2007, with the objective of supporting the institutionalization of African Traditional Medicine (ATM) in member countries' health systems, followed up by WAHO's 2016-2020 Strategic Plan. Within this plan, an important action item is the standardisation of descriptions of herbal and TM. Together with the lack of computable TM data, it is difficult to take benefit from them for primary and secondary use cases including patient follow-up and public health statistics or phytovigilance about available herbal medications. An important step was the launch of the first edition of the West African pharmacopeia in 2013, with inputs from ATM experts coming from different member states [1]. This step is a good asset towards the West African TM standardisation.

In this study, we aim at relying on this endeavor in order to design and develop the first release of the West African Herbal Traditional Medicine (WATRIMed) Knowledge Graph (KG), with the objective of helping to preserve TM knowledge and for bringing West Africa TM to the digital world using a state-of-the-art, flexible and shareable knowledge representation approach. Further, helping to establish bridges with conventional medicine, similarly to previous attempts of digitizing Chinese TM [2] and more general African TM [3,4].

Methods

The West African Herbal Pharmacopeia (aka WAHO herbal pharmacopeia) gathers information on medicinal plants used in West Africa, building on a first African Pharmacopeia including 105 plants created in 1985, followed by a book on medicinal plant analysis in 1986 [1]. It describes every plant by the following features: a summary description of the plant, its ethno-medicinal usage, related clinical information and safety, its chemical constitution, contraindications, the regions where the plant grows, a photograph, information on biological and pharmacological activity, and possible dosages and mode of administration.

The workflow to build the WATRIMed KG comprises three main components:

1. Designing and feeding the TM relational database (TradiMed) from the WAHO Herbal Pharmacopeia resources;
2. Designing the Herbal-based TM (HTM) Ontology, and establish links between TM and conventional medicine entities. This is particularly relevant for supporting phytovigilance activities and taking benefit from the advances in terms of pharmacovigilance and drug usage assessment in conventional medicine;
3. Mapping TradiMed and HTM Ontology and linking them to the external publicly available Knowledge Bases (KB) relevant to the domain.

We identified a set of publicly available KBs, which allow to enrich the core information of West African TM and to widen its scope while opening up the perspective of wide-scale integration: Therefore, the following KBs are considered:

[1] http://www.wahooas.org/index3.php?lang=en

DBpedia[2] for plants and diseases; STITCH[3] and PubChem for chemical compounds; IPNI for plants names and bibliographic references; GeoNames for information about countries and regions; Wikidata and Yago for local dialects and vernacular names of plants and recipes.

The OpenRefine tool[4] is used to query these external KBs and to perform the mapping with the core WATRIMed model.

Results

The TradiMed Database

The database comprises of 25 relational tables which total 3544 tuples. It is hosted in a PostGreSQL server and the schema is available online[5]. Table 1 indicates the statistics about the main entities of TradiMed. As can be seen, 143 plants are identified and documented from the WA countries members.

The Herbal based Traditional Medicine Knowledge Model

The Herbal-based TM knowledge model, referred as HTM comprises 556 main Concepts and 75 Properties. The latter are subdivided into 18 Datatype Properties and 57 Object Properties. The main component is *MedicinalPlant*. It is linked with the *ChemicalCompound* entity by the object property *HasChemicalComponent*. A *MedicinalPlant* has a *BotanicalName* (which is a Datatype Property). The *PlantInRecipe* entity illustrates the n-ary relationship pattern. A *PlantInRecipe* defines the plant part components which constitute a given Recipe. Thus, two property restrictions are used to link *PlantInRecipe* respectively to Recipe and Plant-Part. For example, let's take as an illustration the *1.Rx* recipe of TradiMed, which is a treatment for Malaria. For 100g dosage, it is composed of 40g of the root of the *Cryotolepsis sanguinolenta*, 20g of *Moringa oleifera leaves*, 20g of *Cymbopogon citratus leaves* and 20g of *Khaya senegalensis stem barks*.

Table 1. Current Statistics of the Main Entities of the TradiMed Database

Component	Size (#tuples)
West African Plants	143
Countries	16
Therapeutic Indications	110
Contraindications	148
Local Dialects	122
Traditional Medicine Recipes	108
Chemical Compounds	179
Plant Parts	34

The WATRIMed Knowledge Graph

The following entities of the HTM Ontology have been linked to external resources identified among the publicly available external KBs: *MedicinalPlant*, *TheurapeuticIndication*, *ContraIndication*, *ChemicalComponent* and *Dialect*.

There are 143 *MedicinalPlant* respectively linked to 143 DBpedia entities and 143 IPNI resources. With setting up these external links we were able to enrich the description of the plants, because the information provided by the two KBs is complementary. IPNI references the bibliographic information about the first scientific publication that references a given plant; while DBpedia provides, among others, information including species, fat and fibre content of a given plant. For *TherapeuticIndication*, about 40% of them are linked to DBpedia entities (42 out of 110). However, only 6 out of 110 could be linked to some Yago entity.

Eighteen *ContraIndication* entities have been linked to Yago (12%). All the *ChemicalComponent* entities have been linked to external resources by fetching URLs from STITCH and PubChem. We ensured that any *ChemicalCompound* is consistent and correctly labelled.

For *Dialect*, we have identified 13 out of 122 links with Yago entities and 46% (56 out of 122) links with Wikidata.

Conclusion

We have briefly described the West African Herbal-based TM KG, which is made available to the community at www.watrimed.org together with a SPARQL endpoint. It could therefore be processed both by human and machines. It comprises 556 Concepts and 75 Properties. It is further mapped to a set of external KBs including DBpedia, PubChem and GeoNames. It has been built from the core component of the WAHO's Herbal Pharmacopeia resource and linked to publicly available KBs about plants, diseases and drugs. It is ongoing work which describes 143 plants and 108 traditional recipes identified as treatments of common diseases in West Africa. Future work includes further validation of WATRIMed with ATM experts and enriching the current graph with resources from additional external KBs.

Acknowledgements

Dr. Borlli Michel Jonas Somé has been supported by a mobility funding to Univ. Bordeaux by Nazi Boni University of Burkina Faso. We would like to thank WAHO (Dr Koffi Busia) for sharing their resources. We also thank Pr. A. Baldé CRVPM (Guinea) and Mr. D. Ba Hôpital Keur Massar (Senegal) for sharing their expertise about ATM.

References

[1] WAHO Herbal pharmacopoeia for Economic Community of West African States. Bobo Dioulasso, ISBN: 978-9988-1-8015-7, 2013, KS Printkraft Ghana, Ltd.

[2] Z. Wu, T. Yu, H. Chen, X. Jiang, C. Zhou, Y. Zhang, Y. Mao, Y. Feng, M. Cui, and A. Yin, Semantic Web Development for Traditional Chinese Medicine, in: Proc. Twenty-Third AAAI Conf. Artif. Intell. AAAI 2008 Chic. Ill. USA July 13-17 2008, 2008: pp. 1757–1762.

[3] G. Lô, V. De Boer, S. Schlobach, and G. Diallo, Linking African Traditional Medicine Knowledge, in: Semantic Web Appl. Tools Healthc. Life Sci. SWAT4LS, Rome, Italy, 2017.

[4] B. Kamsu-Foguem, G. Diallo, and C. Foguem, Conceptual graph-based knowledge representation for supporting reasoning in African traditional medicine, *Eng. Appl. Artif. Intell.* **26** (2013) 1348–1365.

Address for correspondence

Corresponding author : Gayo Diallo, Gayo.Diallo@u-bordeaux.fr.

[2] https://wiki.dbpedia.org
[3] http://stitch.embl.de/
[4] http://openrefine.org/
[5] http://www.watrimed.org

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190529

Unsupervised Phrase-Level Query Rewriting for Assisting Search in Clinical Free Text

Hans Moen[a], Laura Peltonen[b], Henry Suhonen[b], Mikko Koivumäki[b], Tapio Salakoski[a], Sanna Salanterä[b]

[a] *Department of Future Technologies, University of Turku, Turku, Finland,*
[b] *Department of Nursing Science, University of Turku, Turku, Finland*

Abstract

We report on the pilot evaluation of an experimental query-based search functionality that enables phrase-level query rewriting in an unsupervised way. It is intended for supporting search in clinical text. Qualitative evaluation is done by three clinicans using a prototype search tool. They report that they find the tested search functionality to be beneficial for making query-based searching in clinical text more efficient.

Keywords:

Natural Language Processing, Information Storage and Retrieval, Electronic Health Records.

Introduction

Query-based searching for information in clinical free text, such as information about a specific health-related phenomena or concept, can be challenging as the same concept can be described in many different ways. An important job for a search engine is to try to bridge the gap between user queries and how associated phrases of similar meaning (semantics) are documented in the targeted text. The aim of the presented work is to test an experimental search functionality that aims to find, suggest and highlight phrases which have similar meaning and length as queries provided by the user. Intended users of a search engine with this functionality are clinicans who are looking for previously documented information in a patient's electronic health record (EHR). One example could be a clinician who is interested in reading what, if any, information concerning *confusion symptoms* has been documented by searching with queries similar to how he/she would typically document such symptoms.

Unlike traditional information retrieval that focuses on retrieving relevant documents, paragraphs or sentences, we are focusing on phrase-level matching and highlighting. Labeled training data, e.g. in the form of search history logs (see e.g. [1]) for enabling supervised query rewriting and expansion techniques, is typically not readily available from EHR systems. Thus, our approach relies primarily on un-labeled free text for training. We are not aware of related works that have proposed suitable, unsupervised approaches/methods for enabling the search functionality that we are aiming for.

The two main challenges in the identified scenario are: 1) Individual words may have several synonyms, near synonyms, and/or closely related words which refers to the same or similar underlying concept (e.g., "oxygen" vs. "SaO2" (oxygen saturation)); 2) When using multiple words, more complex concepts may be expressed and there are typically a greater number of ways to describe a single concept with variations in the choice of words, compositionality and length (e.g., "DM II" vs. "type 2 diabetes mellitus," see also Table 1).

Our approach/method is based on unsupervised machine learning and uses primarily three components: A statistical language model (KenLM toolkit [2]); an off-the-shelf search engine (Apache Solr search platform); and a semantic model of word n-gram vectors (or embeddings) trained with the Word2Vec toolkit [3] – which learns semantics from word co-occurrence statistics in a large text corpora in an unsupervised way. Such models of distributional semantics have been shown to capture word synonymy and relatedness. Similar to Zhao et al. [4], we combine co-occurrence statistics from word n-grams of two different sizes: unigrams (single words) and bigrams (word pairs) when training a single semantic model. One motivation for doing this is that we want to be able to map from unigrams to semantically similar bigrams and vice versa. Another motivation is that Zhao et al. [4] found this to produce improved word representations compared to only using unigram co-occurrence statistics.

We hypothesize that the proposed search functionality (described in the Methods section) can be beneficial to clinicians for saving time and effort when seeking for information of interest in clinical free text. Evaluation of the search functionality is enabled and tested through a prototype query-based search tool/interface. We report on a pilot evaluation performed by three domain experts. The evaluation provides qualitative feedback on how the search functionality performs, and highlights strengths and weaknesses as seen by the evaluators.

Methods

The data set we use is a relatively large corpus of clinical text consisting of physician notes and nursing shift notes from patients admitted to a Finnish hospital. It consists of 136 million tokens (1.5 million unique tokens). For training the semantic model, we first preprocess a version of the corpus with lowercasing, tokenization and stemming (Snowball stemmer for Finnish). We decided to use stemming primarily to reduce the vocabulary size. Next, the text is converted into uni- and bi-grams. As an example, the sentence "a nice flower" becomes: "a a_nice nice nice_flower flower". We train the semantic model with Word2Vec using the SkipGram architecture and hierarchical softmax. We use a dimensionality of 300 and a window size of 6. To train the statistical langauge, we use a version of the corpus only containing stemmed and lowercased unigrams.

Briefly explained, the way we generate rewritten candidate suggestions for a given query is as follows. First, we create two query vectors by splitting the query into uni- and bi-grams before summing the associated vectors (normalized) from the semantic model. If, let us say, the query is: "this is a query", we create two query vectors as follows:

$\text{qvec}_{\text{unigram}} = \overrightarrow{this} + \vec{is} + \vec{a} + \overrightarrow{query}$
$\text{qvec}_{\text{bigram}} = \overrightarrow{this\_is} + \overrightarrow{is\_a} + \overrightarrow{a\_query}$

Next, these are both used to extract the semantically most similar unigrams and bigrams using the semantic model and cosine similarty (sim) as vector distance measure. In addition to forming individual candidate suggestions, we use them to generate multi-word phrase candidates. For bigrams, this is done by iteratively combining bigrams that have one overlapping word in order to create longer phrases (e.g., "is_a" and "a_solution" is combined to "is_a_solution"). To reduce the number of nonsensical phrases being generated, we use the statistical language model to iteratively assess whether or not a potential phrase candidate is likely to exist in the corpus. If no additional n-grams can be added to any of the phrase candidates, the process stops. Another stopping criteria we introduce is a max length relative to the length of the query. Next, we again use the semantic model, this time to create phrase vectors ($p_i\text{vec}$) for each phrase candidate in the same way as we did with the query (i.e., $p_i\text{vec}_{\text{unigram}}$ and $p_i\text{vec}_{\text{bigram}}$). Then, we calculate a similarity score between the query and each of the phrases before sorting them according to their similarity score. As similarity function between a query (q) and a phrase candidate (p_i) we use:

$$\text{sim}(q, p_i) = \text{avg}(\text{sim}(\text{qvec}_{\text{unigram}}, p_i\text{vec}_{\text{unigram}}), \text{sim}(\text{qvec}_{\text{unigram}}, p_i\text{vec}_{\text{bigram}}), \text{sim}(\text{qvec}_{\text{bigram}}, p_i\text{vec}_{\text{unigram}}), \text{sim}(\text{qvec}_{\text{unigram}}, p_i\text{vec}_{\text{bigram}}))$$

As a final step, we use the search engine (Apache Solr search platform) where we have indexed the clinical notes in the care episodes to search in. We use a filter that enables matching stemmed queries with their inflected forms in the indexed text. Finally, the search engine is used to find and highlight the top *n* phrase candidates that actually occur in the targeted care episode. Candidates not found are discarded.

Prototype Search tool and Experimental Setup

We asked three domain experts with a background as a hospital nurse to test the search functionality through a prototype query-based search tool/interface. The purpose was to have them evaluate how well the (rewrite) suggestions by the tool helps them to better find the information that they are searching for (if it exist at all), compared to only the exact matches of the query. When searching with a query, the tool highlights in the clinical notes/documents exact matches (if any) and additional rewritten phrase suggestions found with unique colors. Each unique match/suggestion are also listed in a separate table showing their similarity score to the query and occurrence count in the targeted care episode. The evaluators were given a set of patient phenomena to search for, including *state of mind*, *smoking status*, *secretion* and *activity level*. They were also encouraged to search for other phenomena of interest. We also encouraged them to use multi-word queries when searching. 20 different care episodes (i.e. all physician and nursing notes from 20 patients' hospital stays) were indexed and searchable (one at a time). The evaluators were given a set of questions to consider and comment on while they were doing the testing. These were questions about whether or not the tool made it easier and faster to find the information they were looking for (compared to only relying on the exact matches), weaknesses, strengths, problems and suggestions for improvements. Finally, their answers and comments formed the basis of a joint feedback meeting.

Results

All three evaluators reported that the suggestions provided by the tool makes it easier and faster to search for and find information in clinical text. Since the way clinicans document various phenomena can vary greatly, the rewrite suggestions help to identify potential words and phrases that are semantically related to the search query but written with different words. Even though some of the suggestions by the tool were not relevant, they reported that it was easy to simply ignore these. As problems, they reported that it can be difficult to know that something is not present in a care episode even if it is not found by the tool (false negatives). Also, they noticed that the use of stemming was not optimal as different inflections of the same word were not always connected and found. Examples of rewrite suggestions provided by the tool for a few queries can be seen in Table 1.

Table 1– Example of Query Rewrite Suggestions Found by the Tool. Translated from Finnish to English.

Query	Rewrite suggestions
"runs daily"	"exercise a lot"
"speaks funny things"	"speaks nonsense" "speaks delirious" "speaks by himself"
"patient sees small green men"	"sees a lot of things" "sees things that are not there" "sees illusions"

Conclusions

We present our ongoing work toward enabling unsupervised phrase-level query rewriting to support searching in clinical text. The evaluators report that they find the prototype search tool, with the underlying query rewrite functionality, to be beneficial for making searching and finding information in clinical text faster and easier compared to only exact query matching. As future work we plan to perform a quantitative evaluation. We also aim to test using this search functionality to supporting manual annotation of clinical text. Additional plans include looking into alternatives to stemming, the inclusion of trigrams in the semantic model(s), other similarity measures, query segmentation and algorithm optimization.

Acknowledgements

The research was supported by the Academy of Finland.

References

[1] Y. He, J. Tang, H. Ouyang, C. Kang, D. Yin, and Y. Chang. Learning to rewrite queries. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 2016, 1443–1452.

[2] K. Heafield. KenLM: Faster and smaller language model queries. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, 2011, 187–197.

[3] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26*, 2013, 3111–3119.

[4] Z. Zhao, T. Liu, S. Li, B. Li, and X. Du. Ngram2vec: Learning improved word representations from ngram co-occurrence statistics. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, 244–253.

Address for Correspondence

Corresponding author: Hans Moen. E-mail: hans.moen@utu.fi.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190530

Automatic Mapping Between Brazilian Portuguese Clinical Terms and International Classification for Nursing Practice

Lucas Brehm Ronnau[a], Fernanda Broering Gomes Torres[a], Lucas Emanuel Silva e Oliveira[a], Denilsen Carvalho Gomes[a], Marcia Regina Cubas[a], Claudia Moro[a]

[a] Health Technology Program, Pontifícia Universidade Católica do Paraná, Brazil

Abstract

This study describes MappICNP, an automatic method for mapping between Brazilian Portuguese clinical narratives in free text and International Classification for Nursing Practice (ICNP) concepts. It's composed of six natural language processing rules, related to terms comparison. A set of 2,638 terms extracted from hospitals nursing notes was mapped. MappICNP helps to map 1,607 terms, 113 less than a manual approach. The results demostrate its advantages in minimizing the time spent and reducing the scope of analysis through candidate terms of ICNP.

Keywords:
Standardized Nursing Terminology; Natural Language Processing; Health Information Interoperability.

Introduction

Most of Electronic Health Record (EHR)'s data is represented as free text. It's necessary to map natural language text to a standard terminology, i.e., terminological mapping, to improve their reusability[1; 2]. International Classification for Nursing Practice (ICNP) is one of the clinical standard terminologies which allows nominating and documenting the elements of nursing practice. It is used for nursing diagnosis, outcomes, and, interventions standardization [3]. It's developed by International Council of Nurses, and has a translation to Brazilian Portuguese (pt-br). The 2017 version of the ICNP contains 4,326 concepts and 2,401 of them are primitive.

There are previous studies focusing on cross-mapping from clinical narrative terms to ICNP [4]. However, no existing studies or tools to automatically map terms in pt-br to ICNP concepts were identified. The objective of this study was to propose an automatic method, MappICNP, to support mapping between pt-br terms in free text and ICNP concepts.

Method

MappICNP method was developed by applying natural language processing (NLP) techniques and focuses on ICNP primitive concepts. It was structured in two phases.

The first phase performs normalization of terms in nurse notes and ICNP (2.401 primitive) concepts: accentuation, special characters and stopwords removal; and lowercasing. In the second phase, six NLP rules were created to perform terms comparison between an input term (nurse note term) and ICNP concepts. Except for the first rule, all input and ICNP terms are modified to cover orthographic variants possibilities. For each rule, the comparison between terms is carried out through Levenshtein's distance-editing algorithm.

Terms Comparison is the first rule. Each input term is compared to all ICNP terms until 100% of similarity is met. If the similarity is above 90%, the ICNP term is added to a list of candidate terms that represented the input term. Even if a candidate term is identified, the input term is compared to the next term of ICNP, until the end of ICNP list. So, in the end this rule is possible to have an ICNP term with 100% of similarity with the input term or a list of candidate terms. When this rule achieved 100% of similarity, the other rules are not executed. Otherwise, it goes to the second rule, trying to find more candidate terms.

In the second to sixth rules, all similar ICNP terms (from 90% to 100%) are registered in a list of candidate terms. These all five rules are executed even many candidate terms were met before. Stemming is applied as the second rule. Input terms are transformed into its orthographic root. The third rule applies a pt-br lemmatizer that transforms each input term and ICNP term into their motto, before comparing them. The fourth rule uses an orthographic synonyms dictionary to find a set of different terms that may represent the input term. Then, these synonyms are compared to ICNP terms applying the previous rules. The fifth rule aims to find an ICNP term with a semantic meaning encompassed by the input term. If this occurs, the concept is added to the list of candidate terms. Terms are separated into words and using the first 3 rules is checked if the ICNP concepts content all the words of the input term. The goal of Rule 6 is to find an ICNP concept with a semantic meaning that encompasses the input term, which are separated into words, and applied the first 3 rules to verify if all the words of the ICNP concept are present in the input term, adding the concept to the list of candidate terms.

A set of 2,638 most frequent terms extracted from 148,299 pt-br nursing notes created as free text generated in a period of 2 years, from a university hospital in south of Brazil. This set was also used to a manual mapping describes by Cubas et. al [5].

Results

MappICNP enable to identify ICNP candidates' concepts for the 1,607 input terms (≈62%). From these, 556 (≈ 21% of input terms) were associated with a single ICNP concept by rule 1 (100% of similarity). The rest of the terms (1,051) were mapped by rules 2 to 6, generating a set of candidate concepts. Figure 1 shows examples of all six rules results.

RULES	ICNP Example
1. Direct Comparison	tronco -> tronco|10020180
2. Stemming	responsivo -> responsividade|10017091
3. Lemmatization	dores -> dor|10013950
4. Orthographic Synonyms	higienizar-> limpar|10004444
5. Specific term	acidose -> acidose metabolica|10032010
6. Comprehensive term	acesso intravenoso -> acesso|10000340

Figure 1– MappICNP rules results examples.

Discussion

The results applying MappICNP are similar to manual mapping described by Cubas et al. [5], MappICNP mapped 1607 input terms and in Cubas et al. were identified 1720 (4% of difference). The difference of 113 terms between the results can be explained by the fact that in manual approach 443 terms were found in the definition of another term, a strategy that could be considered as future work. Where found 556 terms identical to the ICNP concepts, in the manual mapping 655 were considered identical. Rule 1 presented the best results, as it does not make changes in the structure of the terms, improving the accuracy of the mapping. It may introduce some errors when similarity is below 100%. Rules 2 and 3 results need analysis, as often they do not represent the correct semantic meaning of the term. The fourth rule can approximate the semantic equivalence of the mapped terms or indicate semantically distinct terms. It may be improved by using a synonyms dictionary specific to clinical or health domain. The fifth and sixth rules do not find equivalent concepts for the input terms, their goal, however, is to find terms with near meaning.

Conclusions

The method proposed in this study helps to map the terms of nursing notes to a standardized terminology; minimizing the time spent searching for identical terms and reducing the scope of analysis through candidate terms.

Table 1-- Mapping rules results

Terms covered	Candidate Terms	Rule	Similarity
568	556	1	100%
	13	1	>90% e <100%
122	129	2	100%
	10	2	>90% e <100%
140	144	3	100%
	4	3	>90% e <100%
598	593	4	100%
	14	4	>90% e <100%
540	1867	5	100%
	59	5	>90% e <100%
90	109	6	100%
	10	6	>90% e <100%

Acknowledgements

Pontifícia Universidade Católica do Paraná and "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001" partially financed this study.

References

[1] J. Dziadek, A. Henriksson, and M. Duneld, Improving terminology mapping in clinical text with context-sensitive spelling correction, *Informatics for Health: Connected Citizen-Led Wellness and Population Health* **235** (2017), 241.

[2] I.O.f. Standardization, ISO/TR 12300:2014(en) Health informatics — Principles of mapping between terminological systems, in, Geneva: International Organization for Standardization, 2016.

[3] ICN., International classification for nursing practice, in, ICN, 2002.

[4] T.Y. Kim, Automating lexical cross-mapping of ICNP to SNOMED CT, *Informatics for Health and Social Care* **41** (2016), 64-77.

[5] M.R. Cubas, L.E. Pleis, D. Carvalho Gomes, E.C. Rodrigues Da Costa, V.D. Peluci, A. Paula, M.A. Hochuli Shmeil, G. Carvalho, and C. Maris, Mapping and definition of terms used by nurses in a hospital specialized in emergency and trauma care, *Revista de Enfermagem Referência* **4** (2017).

Address for correspondence

Claudia Moro email: c.moro@pucpr.br

Table 2– Examples mapping results.

Portuguese Term	Term	CIPE Term	ICNP Term	ICNP Code	Rule	perc
abdome	abdomen	Abdome	Abdomen	10000023	1	100%
salivar	salivary	Salivação	Salivation	10017460	3	100%
		glândula salivar	Salivary gland	10017456	5	100%
alimentar-se	feed his/her self	Alimentar	Feed	10007786	2	100%
		Alimento	Food	10008089	2	100%
		capacidade para alimentar-se	Ability to feed His/Her Self	10000166	5	100%
audição	impaired hearing	Audição	Hearing	10008814	6	100%
prejudicada		Prejudicado	Impaired	10012938	6	90%
bactéria	bacterium	Microrganismo	Microorganism	10012014	4	100%
hemático	hematic				0	0%
fragilidade	fragility	Fraqueza	Weakness	10024897	4	100%
morder	to bite	Dente	Tooth	10019830	4	100%

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190531

Mining Pharmaceutical Product Data Related to Payment Pattern from the CMS Open Payments Data: A Case Study in Thoracic Surgery

Xu Na, Haihong Guo, Sizhu Wu, Jiao Li

Institute of Medical Information / Library, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

Abstract

This study used descriptive statistical analyses to investigate the payment characteristics and to discuss the regularity of highest paying industries. Payments by 4.70% of highest paying industries (N=446) accounted for 85% of the total (US $72,458,304) in 2014-2016. A tiny minority of highest paying industries control the majority of payments. Large payments from these industries are highly associated with few specific products. Furthermore, payment patterns among the industries include concentration and diversification.

Keywords:
Health expenditures, drug industry, medical informatics.

Introduction

Industry transactions to surgeons have become transparent and publicly accessible under the Open Payments program (OPP), a national financial disclosure program in healthcare systems in the U.S. OPP has been viewed as the most potent antidote for potential conflicts of interest engendered by financial interactions between industry and the health care delivery enterprise. Therefore, speaking with authority on the regularity from these transactions recorded in the OPP database becomes important in the open data science and medical informatics studies. Several studies have been conducted for analyzing the open payment. Piller found a pattern of after-the-fact compensation by industry to those advising the US government on drug approvals [1]. Na *et al.* discovered the massive transfer from industries to thoracic surgeons has a strong "apical dominance" and excludability [2].

In the case of thoracic surgery, we aimed to investigate the product associated payment pattern and discuss the regularity of industries who transfer large payments to surgeons.

Methods

We accessed the general payments in 2014-2016 focusing on the non-research and non-ownership payments, which are publicly accessible. We limited physician specialty to thoracic surgery (cardiothoracic vascular surgery) and excluded payments for *current or prospective ownership* or *investment interest* or *valued $0*. The final cohort included 197,592 payments transferred by 446 industries to 4552 surgeons in three-year period.

All analyses were performed with R 3.4.1. The pareto distribution was used to show the distribution of payments transferred by pharmaceutical industries in three years. Descriptive statistics were calculated to analyze the distribution of the highest annual payment associated with a specific product made by the tiny minority of highest paying industries.

Results

Distribution of payments transferred by pharmaceutical industries

Payments transferred by 4.70% (n=21) of highest paying pharmaceutical industries accounted for 85% of the total payments (72,458,304) in thoracic surgery in 2014-2016. The cumulative percentage of payments transferred by 10% of highest paying industries was 94.44%. Specifically, the highest paying industry transferred cumulative payments totaling $20662039, accounting for 28.52% of the total. 14 industries transferred cumulative payments more than one million in 3 years (see Figure 1). The result illustrated that distribution of payments shows a great discrepancy in pharmaceutical industries.

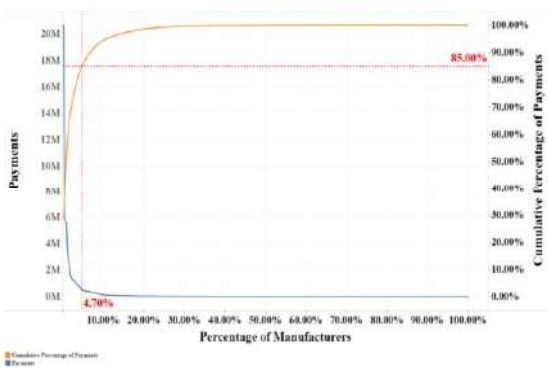


Figure 1– Pareto Distribution of Payments Transferred by Pharmaceutical Industries in 2014-2016

Distribution of the highest annual payment associated with specific product by the 21 highest paying industries

The percentage of the highest paying product transferred by 15, 15 and 13 manufacturers were more than 50% of total, among which 9,9 and 7 were approximately 90% in the three-year period, respectively. Furthermore, the highest annual payment transferred by 16 industries were associated with certain product in each of 3 years; meanwhile, 6 transferred more than 80% of the total (see Figure 2). The result reflected that payments transferred by these highest paying industries are specifically targeted to certain products. The payments pattern vary greatly in different industries.

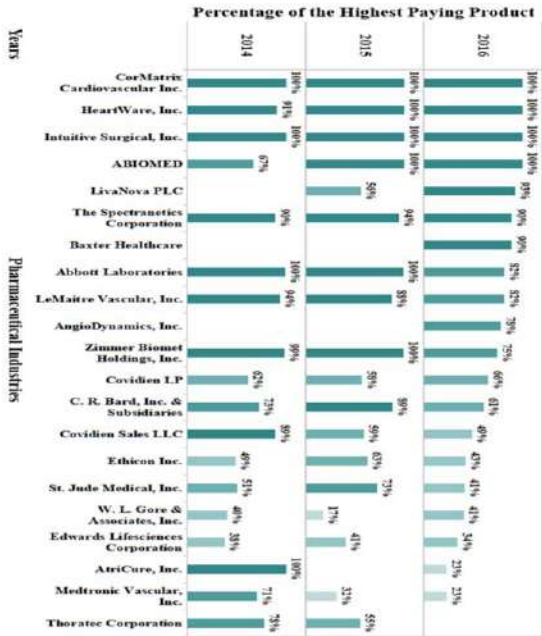


Figure 2– Proportion of the Highest Annual Payment Made by Industries Associated with Specific Product in 2014-2016

Discussion

The tiny minority of highest paying industries controlling the vast majority of payment

Our study indicated that there was significant difference in the distribution of payment made by pharmaceutical industries, with only 4.70% of highest paying industries controlling as high as 85.00% of the total payment. The pareto distribution is a power-law probability distribution that is originally applied to describing the distribution of wealth in a society, fitting the trend that a large portion of wealth is held by a small fraction of the population. This distribution gave the inequality of payment a very comprehensible and visible display that the tiny minority of industries determine the vast majority of payment in our research. In addition, we found that 3.14% of total industries transferred cumulative payments more than one million in 3 years, among which the highest paying industry accounted for approximately 30% of the total. The highly skew distribution of payments remain signification among the critical minority. Therefore, we focused on analyzing the payments pattern made by the tiny minority of highest paying industries.

Characteristic of payments pattern among the tiny minority of highest paying industries

Our research showed that the most majority of annual payment transferred by the tiny minority of highest paying industries were highly associated with their specific products. Nearly half of these highest paying industries transferred more than 50% of the total targeting to certain product in each of 3 years. Moreover, 28.57% of the industries transferred as high as four-fifths of the total. It indicated that these industries are strongly inclined to centralize the annual payment to one specific product. Conversely, others are more likely to pay for diversification. The implication is that there exists a great discrepancy in the payments patterns among the tiny minority of industries.

Modern portfolio theory assumes that investors are risk averse, meaning that when given two portfolios that offer the same expected return, investors will prefer the less risky one to lower the uncertainty. For this reason, an investor will only take on increased risk in return for higher expected returns. Industries will evaluate the trade-off differently in order to make the decision of benefit maximization, based on their individual risk aversion characteristics.

For the centralized paying industries, concentration means they may receive the maximum expected returns by using minimum payment in the certain period. In the meantime, these industries have to accept the high-risk that come with the high returns. Therefore, industries transferred almost all of the annual payments targeted to the potentially most valuable product after the analysis and assessment of potential risks and benefits. For the diversified paying industries, diversification may allow for the same expected returns with reduced risk. Due to the unpredictability of the market, industries can decrease the exposure to individual product risk by paying a diversified portfolio of the most valuable products.

Conclusions

There exists a great discrepancy in the distribution of payments transferred by pharmaceutical industries to healthcare professionals. The tiny minority of highest paying industries controlling the vast majority of payment. The largest majority of payments from these top-paying industries are mainly associated with few specific products. Furthermore, payments pattern among the tiny minority of industries include concentration and diversification. In future, we will attempt to investigate the driving factors of the concentration of certain pharmaceutical products via integrating more open datasets such as clinical trials and FDA approval documents.

Acknowledgements

This work was supported by the Chinese Academy of Medical Sciences (2017PT63010, 2018PT33024), the National Population and Health Scientific Data Sharing Program of China(NCMI-MCE02-201806), the National Key Projects of Precision Medicine Research Program (2016YFC0901602), and the Key Laboratory of Knowledge Technology for Medical Integrative Publishing.

References

[1] Piller C. Hidden conflicts? *Science* 2018 Jul 06; 361(6397): 16-20. doi: 10.1126/science.361.6397.16.

[2] Xu Na, Haihong Guo, Yu Zhang, Liu Shen, Sizhu Wu, Jiao Li. Mining Open Payment Data: Analysis of Industry Payments to Thoracic Surgeons From 2014-2016. *J Med Internet Res* 2018;20(11): e11655. doi:10.2196/ jmir.11655.

Address for correspondence
NAME: Jiao Li, PhD, EMAIL: li.jiao@imicams.ac.cn
ADDRESS: No.3 Yabao Road, Chaoyang District, Beijing, China

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190532

Artificial Intelligence in Diabetic Retinopathy: Insights from a Meta-Analysis of Deep Learning

Tahmina Nasrin Poly[a,b], Md. Mohaimenul Islam[a,b], Hsuan Chia Yang[b], Phung-Anh Nguyen[b], Chieh Chen Wu[a,b], Yu-Chuan (Jack) Li[a, b]

[a] *Graduate Institute of Biomedical Informatics, Taipei Medical University, Taipei, Taiwan,*
[b] *International Center for Health Information and Technology (ICHIT), Taipei, Taiwan*

Abstract

The demand for AI to improve patients outcome has been increased; we, therefore, aim to establish the diagnostic values of AI in diabetic retinopathy by pooling the published studies of deep learning on this subject. A total of eight studies included which evaluated deep learning in a total of 706,922 retinal images. The overall pooled area under receiver operating curve (AUROC) was 98.93% (95%CI:98.37%-99.49%). However, the overall pooled sensitivity and specificity for detecting referable diabetic retinopathy (RDR) was 74% (95% CI: 73%-74%), and 95% (95% CI: 95%-95%). The findings of this study show that deep learning had high sensitivity and specificity for identifying diabetic retinopathy.

Keywords:

artificial intelligence, deep learning, diabetic retinopathy

Introduction

Artificial intelligence (AI) is a machine intelligence that is meant to mimic human cognitive functions [1]. A deep learning system (DLS) is an advanced form of artificial intelligence allow an algorithm to program itself by learning from a large set of data. With rapid advances in technology, AI has recently been applied to many areas in healthcare due to paramount significant to enable early disease detection, and diagnosis decision. AI has already been created an opportunity to improve productivity in several areas of the economy, whereas in the healthcare need AI to improve efficiency, affordability, and accessibility [2].

The prevalence of diabetic retinopathy (DR) has been increasing. It is estimated that the number will be approximately 600 million by 2040 with one-third expected to have DR [2]. Retinopathy remains a leading cause of blindness globally. Therefore, timely screening for DR has accepted strategy for blindness prevention and proper treatment. However, there are several challenges to delivery of care: (1) clinical diagnosis is highly subjective; (2) lack of human assessors; and (3) long-term financial sustainability. Automated grading of DR indeed has potential benefits of increasing efficiency, reproducibility, and large-scale screening program. AI system would help to improve the quality of care by providing early diagnosis and treatment. Deep learning has had a profound impact on detecting DR from retinal photographs. We, therefore, evaluated the performance of deep learning in detecting DR by pooling the published literature.

Methods

This systematic review and meta-analysis were conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). A review written protocol was not drafted.

We performed a comprehensive search of electronic databases, such as PubMed, Embase, Web of Science, Scopus, Google Scholar, Cochrane, and IEEE between 2000 and 2018. These databases provide a useful source of relevant and potential information as the subject of this study encompassed with two contexts including artificial intelligence and diabetic retinopathy. We used various keywords to search for most potential relevant studies. Keywords, synonyms of keywords, specific characters, and the Boolean operators are used in the search strategies as follows: “deep learning”, OR “artificial intelligence”, OR “automated computerized system”, OR “DL”, OR “CBIA”, AND “diabetic retinopathy”.

We carefully screened all the studies in the initial search and checked their relevant titles and abstracts. Two authors (MMI, TNP) examined these studies independently. To be included, all the studies had to fulfill the following criteria: (1) articles published in English and be peer-reviewed, (2) articles that provided an outcome diabetic retinopathy, (3) articles that provided information regarding evaluation measurements like receiver operating curve (ROC), accuracy, sensitivity, specificity, (4) articles that provided total number of DR patients (5) provided clear definition of DR, (6) articles that clearly described deep learning models and full information of dataset, (7) articles that provided explicit overview of DR image information. We excluded studies if they published in the form of editorials, letter to editor, review. All disagreement between two authors for selecting potential studies were then resolved by the chief investigator (YC, L).

Data extraction was conducted by the same two authors who used a predefined, standardized protocol, and data collection instrument. However, data were entered into Review Manager software (RevMan-5) and checked for accuracy. All information regarding extracted data were (a) title (b) method: study design, data source, number of image, number of study centers, and location, study setting, date of data analysis, process of data analysis, (c) participants: number of patients with diabetes, patients’ age, patients’ gender, age range, percentage of gender, diagnostic criteria, (d) outcome: primary and secondary outcome. We used the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool-2 to assess the quality of the included studies in terms of risk of bias and concerns regarding applicability over four domains.

We calculated the sensitivity, specificity, and ROC value with 95% confidence interval (CI) for detecting diabetic retinopathy. A ROC curve is plotted the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The values of the ROC curve 0.9-1, 0.8-0.9, 0.7-0.8, 0.6-0.7, 0.5-0.6 are defined as excellent, good, fair, poor and fail, respectively. The values of ROC, sensitivity, and specificity were measured with 95% CI in the final analysis. Every analysis was carried out in the random effect model. However, We used MetaDiSc (version 1.4) for the pooled estimate of AUROC, sensitivity, specificity, and diagnostic odds ratio.

Results

Through the initial searches of the electronic bibliographic databases, we identified 125 original articles. One author (HC, -Y) carried out additional searches of trial registers, commercial websites, conference proceedings, and reference lists of included studies; we identified one further study for inclusion. Therefore, the total number of records identified was 135. We then removed 87 duplicate studies. Furthermore, two authors (MMI, TNP) assessed 48 unique abstracts and eliminated obviously irrelevant studies from the titles and abstract alone. Of which, 12 articles went through full-text evaluation. Finally, 8 unique articles were eligible for the final meta-analysis [3-10].

Eight studies measured the performance of deep learning algorithms to detect referable diabetic retinopathy correctly and accurately from the retinal fundus photograph. The pooled AUROC of the deep learning model for diabetic retinopathy was 98.93% (95%CI:98.37%-99.49%), sensitivity was 74% (95%CI: 73%-74%), and specificity was 95% (95%CI:95%-95%). However, the diagnostic odds ratio was 250.83 (95%CI: 77.72-809.49). Three studies evaluated the performance of the deep learning algorithms to detect vision threating diabetic retinopathy. The overall pooled AUROC was 99.46% (95%CI: 99.27%-99.65%). However, sensitivity and specificity was 98% (95%CI: 97%-99%), 92% (95%CI: 92%-92%) respectively. In addition, positive likelihood ratio, negative likelihood ratio, and the diagnostic odds ratio was 25.08(95%CI: 11.36-55.39), 0.02(95%CI: 0.00-0.43), and 1316.41(95%CI:110.91-1562.29), respectively.

Conclusion

This systematic review and meta-analysis demonstrated that deep learning algorithms are highly accurate for the detection of referable diabetic retinopathy, with 74% sensitivity and 95% specificity. These results may change the way diabetic retinopathy is being diagnosed and would be helpful to implement AI in low resource areas. However, one of the key findings is that a deep learning system can detect DR with comparable or better performance than experts. The deep learning algorithm outperformed not only the most expert but also other traditional systems used previously. Overall, our study demonstrated that deep learning model has an acceptable degree of sensitivity and specificity for detecting diabetic retinopathy. The excellent likelihood and diagnostic odds ratio suggest that deep learning may be useful in the real-world clinical setting. However, deep learning algorithms would diagnose diabetic retinopathy better than human experts. In fact, it has potential applications in disease detection, monitoring, and prognosis in patients at risk of diabetic retinopathy.

References

[1] Erickson BJ, Korfiatis P, Akkus Z, Kline TL: Machine learning for medical imaging. Radiographics 2017;37:505-515.

[2] Helmchen LA, Lehmann HP, Abràmoff MD: Automated detection of retinal disease. The American journal of managed care 2014;20:eSP48-52.

[3] Gulshan, V., et al., Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. Jama, 2016. **316**(22): p. 2402-2410.

[4] Ting, D.S.W., et al., Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. Jama, 2017. **318**(22): p. 2211-2223.

[5] Wan, S., Y. Liang, and Y. Zhang, Deep convolutional neural networks for diabetic retinopathy detection by image classification. Computers & Electrical Engineering, 2018. **72**: p. 274-282.

[6] Mansour, R.F., Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy. Biomedical Engineering Letters, 2018. **8**(1): p. 41-57.

[7] Lin, G.-M., et al., Transforming Retinal Photographs to Entropy Images in Deep Learning to Improve Automated Detection for Diabetic Retinopathy. Journal of ophthalmology, 2018. **2018**.

[8] García, G., et al. Detection of Diabetic Retinopathy Based on a Convolutional Neural Network Using Retinal Fundus Images. in International Conference on Artificial Neural Networks. 2017. Springer.

[9] Gargeya, R. and T. Leng, Automated identification of diabetic retinopathy using deep learning. Ophthalmology, 2017. **124**(7): p. 962-969.

[10] Abràmoff, M.D., et al., Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Investigative ophthalmology & visual science, 2016. **57**(13): p. 5200-5206.

Address for correspondence:

Yu-Chuan (Jack) Li,

Email: jack@tmu.edu.tw; jaak88@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190533

Do You Need Embeddings Trained on a Massive Specialized Corpus for Your Clinical Natural Language Processing Task?

Antoine Neuraz[a,b,c], Vincent Looten[b,d], Bastien Rance[b,d], Nicolas Daniel[d], Nicolas Garcelon[a,e], Leonardo Campillos Llanos[c], Anita Burgun[a,b,d], Sophie Rosset[c]

[a] *Institut National de la Santé et de la Recherche Médicale (INSERM), Centre de Recherche des Cordeliers, UMR 1138 Equipe 22, Paris Descartes, Sorbonne Paris Cité University, Paris, France*
[b] *Department of Medical Informatics, Necker-Enfants Malades Hospital, Assistance Publique des Hôpitaux de Paris (AP-HP)*
[c] *LIMSI, CNRS, Université Paris Saclay*
[d] *Hôpital Européen Georges Pompidou, AP-HP, Université Paris Descartes, Sorbonne Paris Cité, Paris, France*
[e] *Institut Imagine, Paris Descartes Université Paris Descartes-Sorbonne Paris Cité, Paris, France*

Abstract

We explore the impact of data source on word representations for different NLP tasks in the clinical domain in French (natural language understanding and text classification). We compared word embeddings (Fasttext) and language models (ELMo), learned either on the general domain (Wikipedia) or on specialized data (electronic health records, EHR). The best results were obtained with ELMo representations learned on EHR data for one of the two tasks(+7% and +8% of gain in F1-score).

Keywords:
Natural language processing, electronic health records.

Introduction

State of the art methods in natural language processing (NLP) usually include learning a statistical representation of the vocabulary such as word embedding (e.g., word2vec, Glove, Fasttext [1]) or language models (e.g. ELMo [2]). One of the advantages of these approaches is that they are unsupervised: they do not require the learning datasets to be manually annotated with a ground truth. However, in the biomedical domain, the vocabulary is very specific. Would the representations learned from general domain corpus be usable for NLP tasks in the biomedical domain? Wang *et al.* [3] showed that learning embeddings using clinical notes from EHRs increased the performances. Nonetheless, it may be challenging to have access to the large number of documents requested (*e.g.,* more than 100k patients in the study of Wang *et al.*) or even have access to embedding matrices learned on such corpus, due to privacy issues.

This work explores the impact of the method and data source of embeddings on different tasks. We compared the performances of two types of word representations: Fasttext and ELMo. We also compared two different learning sets in French: Wikipedia or a set of 1M clinical notes from our local EHR. We evaluated the performances of these approaches on two different NLP tasks in French: natural language understanding and text classification.

Methods

Embeddings

We compared two types of embeddings: 1) continuous skip-gram model with sub-word information (*i.e.,* each word is represented as a bag of character n-grams), as implemented in fastText[1] and 2) embeddings from language models (ELMo) where the vectors are learned from the internal states of a deep bidirectional language model as described in [2]. As a baseline, we use a continuous skip-gram model learned only on the training set (no external dataset). We also compared the performances of these methods when learned on either a general domain dataset (French Wikipedia, Wiki) or a specialized dataset (1m clinical notes from a French AP-HP childrens hospital in Paris, EHR).

Tasks

Natural language understanding in a virtual assistant (VA task) This task aims to provide natural language understanding (NLU) in a virtual assistant for clinicians to explore biological tests results information in the patient's record in natural language (e.g., Donne moi le dernier résultat de créatinine 'Give me the last result of creatinin', Comment a évolué l'hémoglobine depuis 2 ans 'How has the hemoglobin evolved for the last 2 years'). The set of queries that a physician may have about characteristics and results of a patient is broad and diverse. Therefore, enabling queries in natural language may help access information more efficiently.

Given that no public dataset is available for this task in French, we generated a training set of 16,000 questions from templates, terminologies and paraphrases as described in a previous work [4]. This dataset contains 144k words (mean length of a question is 9). We also generated a development set of 4,000 questions for the tuning of the models. For the evaluation, we collected from physicians in our hospital, a set of 178 questions that they would like to ask in such a system.

For this study, we focused on the slot filling task (VA-sequence) consisting of a sequence labeling task aiming at

identifying whithin the question the lab test mention (e.g., *créatinine* 'creatinin') and the date-related information (e.g., *22/04/2012*, *depuis 4 semaines* 'for 4 weeks'). In the training set, the number of distinct lab mentions was 336 with a length ranging from 1 to 11 tokens and a median length of 2.

Text classification of absconding behaviour in nurses' notes (ABS-task) Absconding, where a patient leaves the **hospital without informing staff, is a rare but serious adverse event that can affect the safety of both the patient and people outside the hospital. Identifying hospital encounters with absconding patients can be challenging in most hospital information systems, given that those events are not systematically collected in a structured format.**

We collected nurses' notes from the clinical data warehouse of the European Hospital Georges Pompidou, an AP-HP hospital located in Paris. We collected a set of 2,605 notes that were manually classified by two experts by whether it mentions an actual absconding or not.

Models

As we aim to evaluate the impact of embeddings methods and data sources, we fixed the model used for the tasks and only let vary the embeddings.

We evaluated 5 different configurations of embeddings: a continuous skip-gram model of 300 dimensions learned on the training set; a fasttext embedding of 300 dimensions learned on the Wiki; a fasttext embedding of 300 dimensions learned on EHR; a ELMo embedding of 1024 dimensions learned on Wiki; and a ELMo embedding of 1024 dimensions learned on EHR.

Sequence Labeling task (VA-sequence) For the sequence labeling task (VA-sequence), we used a recurrent neural network (RNN) based on long short term memory units (LSTM). The architecture is based on a 2 layers bidirectional LSTM of 256 units.

CNN for Text Classification (ABS-task) For the classification tasks we used a convolutional neural network (CNN). After the embedding layer, the model contains a 1 dimensional convolutional layer of 250 units, kernel size of 3, ReLU activation, followed by a max pooling layer and a dense fully connected layer with softmax or sigmoid activation depending on the number of classes to predict.

We used a weighted F1-score to evaluate the results of the different models with 5 fold cross-validation.

Results

Sequence labeling task (VA-sequence) All the results for this task are presented in Table 1. The best results are obtained with ELMo learned on EHR with a F1-score of 0.76 (95%CI [0.74-77]) compared to Fasttext on EHR (0.67, 95%CI [0.61-0.73]). Interestingly, we show similar results between ELMo on Wiki (0.69, 95%CI [0.67-0.70]) and Fasttext on EHR (0.67, 95%CI [0.61-0.73]).

Text classification On this task, all the models give similar results with a low variation around the baseline (F1-score of 0.89 (95%CI [0.88-0.90])) as shown in Table 2.

Table 1 – Results for the NLU task for the virtual assistant: sequence labeling (Bi-LSTM)

Method	F1-score [95%CI]
Baseline (only training set)	0.62 [0.61-0.64]
Fasttext on Wiki	0.69 [0.67-0.70]
Fasttext on EHR	0.67 [0.61-0.73]
ELMo on Wiki	0.69 [0.67-0.70]
ELMo on EHR	**0.76 [0.74-0.77]**

Table 2 – Results for the classification of absconding in nurses' notes (CNN)

Method	F1-score [95%CI]
Baseline (only training set)	0.88 [0.87-0.89]
Fasttext on Wiki	0.85 [0.84-0.86]
Fasttext on EHR	0.87 [0.86-0.88]
ELMo on Wiki	**0.89 [0.88-0.90]**
ELMo on EHR	0.84 [0.83-0.85]

Conclusions

Depending on the task, embeddings learned on large corpora can have a significant impact on NLP tasks in the biomedical domain in French. Moreover, learning these embeddings on clinical notes will increase the performances compared to general domain. As it may not be feasible to access a large corpus of clinical notes, it is still profitable to use advanced methods such as ELMo learned on general domain and obtain reasonable results. When the task does not rely on a large specialized vocabulary, the impact of external embeddings might be reduced.

References

[1] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, Enriching word vectors with subword information, *ArXiv Preprint ArXiv:1607.04606*. (2016). https://arxiv.org/abs/1607.04606 (accessed February 14, 2017).

[2] M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, Deep Contextualized Word Representations, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), Association for Computational Linguistics, New Orleans, Louisiana, 2018: pp. 2227–2237. doi:10.18653/v1/N18-1202.

[3] Y. Wang, S. Liu, N. Afzal, M. Rastegar-Mojarad, L. Wang, F. Shen, P. Kingsbury, and H. Liu, A comparison of word embeddings for the biomedical natural language processing, *Journal of Biomedical Informatics*. **87** (2018) 12–20. doi:10.1016/j.jbi.2018.09.008.

[4] A. Neuraz, A. Burgun, L.C. Llanos, and S. Rosset, Natural language understanding for task oriented dialog in the biomedical domain in a low resources context, *Machine Learning for Health (ML4H) Workshop at NeurIPS 2018*. (2018).

Address for correspondence
Antoine Neuraz, antoine.neuraz@aphp.fr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190534

Rapid-Cycle Implementation of a Multi-Organization Registry for Heart Failure with Preserved Ejection Fraction Using Health Information Exchange Standards

Ambarish Pandey[a], James MacNamara[a], Satyam Sarma[a,b], Ferdinand Velasco[b], Vaishnavi Kannan[a], John Willard[b], Cheryl Skinner[b], Tony Keller[b], Mujeeb Basit[a], Benjamin Levine[a,b], Duwayne Willett[a]

[a] Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas, U.S.A.,
[b] Texas Health Resources, Arlington, Texas, U.S.A.

Abstract

Constructing multi-site specialty registries typically proves time-consuming. Electronic health record (EHR) data collected during clinical care affords a pragmatic approach to accelerating registry implementation. Heart failure with preserved ejection fraction (HFpEF) is an increasingly common and morbid condition. Building a multi-site registry for HFpEF proved feasible using EHR data coded in standard terminologies (SNOMED CT, LOINC) and shared via Health Information Exchanges.

Keywords:

Registries; Health Information Exchange; Heart Failure, Diastolic

Introduction

After the major investment in migration to electronic health records in the U.S. [1], how can we leverage now-adopted EHR terminology standards for practical clinical benefit? Identifying and following patients with a specialized and/or rare condition across multiple centers/sites can help in better understanding the natural history of the condition, developing optimized diagnostic and management algorithms, and improving care. However, setting up multi-site specialty registries typically proves laborious and time-consuming. Agile development of EHR-based specialty registries leveraging data collected during clinical care shortens the development cycle time[2]. Additionally, shared definitions of conditions across multiple sites are now practical using standard terminologies for exchange of EHR data, such as SNOMED CT (Systematized Nomenclature of Medicine-Clinical Terms). SNOMED CT enables more focused condition definitions than possible with ICD (International Classification of Diseases) alone[3].

Heart failure with preserved ejection fraction (HFpEF, also termed "diastolic heart failure") is common, increasing in prevalence, and associated with a high burden of morbidity, mortality, and healthcare system expense[4]. Several cardioprotective pharmacotherapies have failed to improve clinical outcomes in patients with HFpEF largely due to the underlying heterogeneity in the disease pathophysiology, lack of uniform diagnostic criteria, and existence of clinically diverse phenotypes of the disease. Thus, there is an unmet need to better characterize patients with clinical HFpEF and develop targeted therapeutic strategies for their management. However, this information is not collected in standard cardiovascular registries. Accordingly, our institutions—an urban academic medical center, a large community hospital network, and an affiliated network of primary care physicians on disparate EHRs—wanted to create a pragmatic shared EHR-based registry of HFpEF patients for population health, quality improvement, and clinical research. Given that standard terminologies enable semantic interoperability between certified EHRs, we wanted to validate that leveraging those standards, specifically SNOMED CT and LOINC (Logical Observation Identifiers Names and Codes), would be helpful.

Methods

Software

The authors' institutions each separately employ Epic (Verona, WI, USA) as their EHR. Community physicians in our network are on a variety of EHRs (Allscripts, eClinicalWorks, NextGen, others). EHR diagnoses are recorded with clinical terms pre-mapped to SNOMED CT. Continuity of Care Documents (CCDs) conforming to U.S. standards were exchanged among EHRs via several health information exchanges (HIEs), both public and private: Carequality, eHealth Exchange, Epic's Care Everywhere, and our network's HIE (dbMotion, Pittsburgh, PA, USA). Diagnoses received via HIE are stored as "external problems." In 2017, UT Southwestern exchanged 7.3 million patient documents with other institutions' EHRs via HIEs.

Procedures

The HFpEF registry was constructed during two-week iterations, as previously described[2]. Both a broad (sensitive) and a narrower (specific) inclusion rule were constructed. The full registry was based on the broad rule, with the narrower definition applied as a report filter. Both inclusion rules were articulated initially during conference calls with cardiologists, captured as textual Boolean logic statements, then created within the EHR. Following medical record review of a sample of included patients by an expert clinician, the inclusion rules were further refined. Diagnosis groupers (value sets) for both HFpEF and co-morbid conditions were constructed using SNOMED CT concept hierarchies combined with boolean logic, and clinically vetted[2,3]. Existing SNOMED CT groupers for co-morbid conditions were re-used. LOINC codes were identified using the LOINC search browser.

Results

The broad (case-finding) registry inclusion rule employed five SNOMED CT condition-defining groupers (value sets) as inclusion or exclusion criteria. Two LOINC value sets were used, for left ventricular ejection fraction (LVEF) and estimated glomeular filtration rate (eGFR). The more specific inclusion rule added six SNOMED CT groupers defining exclusionary conditions (Table 1). An additional fifteen co-morbid conditions (e.g. diabetes, atrial fibrillation, etc.) employ SNOMED CT condition definitions (not shown).

Table 1 – Conditions Evaluated in Registry Inclusion Rules, Defined with SNOMED CT Concept Hierarchies

Role	Condition	Broad	Narrow
Inclusion	HFpEF-Diastolic Heart Failure	Yes	Yes
Inclusion	Congestive Heart Failure (with LVEF >=50%)	Yes	Yes
Exclusion	Systolic Heart Failure	Yes	Yes
Exclusion	CKD-5 (chronic kidney disease – stage 5)	Yes	Yes
Exclusion	On Renal Dialysis	Yes	Yes
Exclusion	CKD-4	No	Yes
Exclusion	Amyloidosis	No	Yes
Exclusion	Carcinoid Heart Disease	No	Yes
Exclusion	Cardiac Sarcoidosis	No	Yes
Exclusion	Hemochromatosis	No	Yes
Exclusion	Mitral Valve Replacement	No	Yes

SNOMED CT-defined conditions and LOINC-defined results were combined by boolean logic to form the registry inclusion rule, following the decision tree vetted with cardiologists (Figure 1).

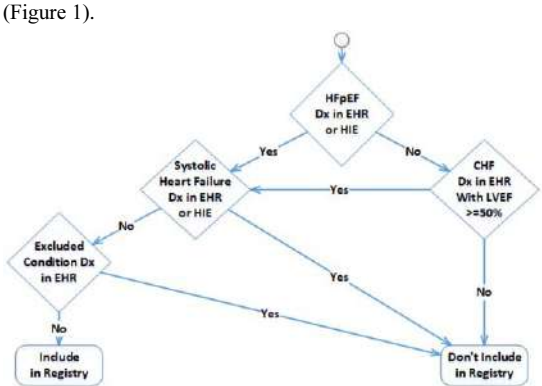


Figure 1 – Decision Tree for HFpEF Registry Inclusion Rule (Dx: diagnosis; CHF: congestive heart failure)

Registry Patient Characteristics

3,758 living patients are currently in the broad registry. 83% (3,111) also meet the more specific "primary" HFpEF inclusion rule. The population was primarily elderly, with the largest 10-year age band 70-79 (Figure 2). Female preponderance was seen in all age bands (overall 64% female, 36% male).

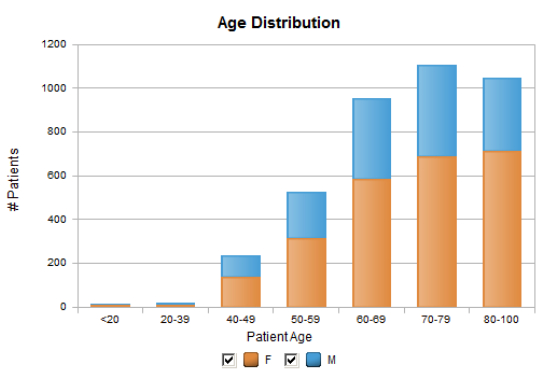


Figure 2 – Age-Sex Distribution of HFpEF Registry Patients

40% of registry patients had HFpEF on their external problem list data received via a CCD. Most of these (36% of registry patients) were included on the broad registry solely based on external data received via an HIE (Figure 3).

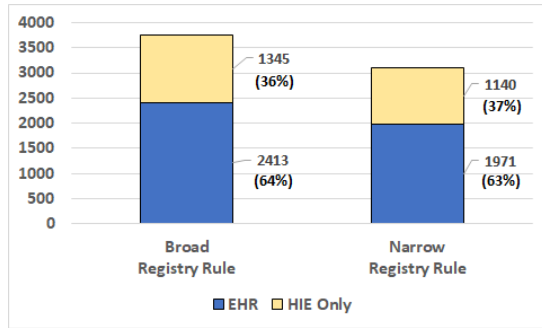


Figure 3 – Source of Information Placing Patient on Registry

Conclusions

Some specialized conditions such as heart failure with preserved ejection fraction (HFpEF) require complex inclusion rules evaluating both diagnosis and test result data. Even so, construction of a pragmatic multi-site EHR-based registry for HFpEF proved feasible using EHR data coded in standard terminologies. 36% of registry members were identified solely by clinical data received via an HIE. Sharing SNOMED CT rule-based value sets for computable clinical phenotype definitions, e.g. via the U.S. National Library of Medicine's Value Set Authority Center, would further expedite construction of such registries across multiple organizations using disparate EHRs. For specialized and/or rare conditions, using SNOMED CT-encoded EHR data has particular potential to help advance collaborative population health initiatives and to conduct multi-site pragmatic clinical research.

Acknowledgements

Funding from Southwestern Health Resources (Arlington, TX).

References

[1] D. Blumenthal and M. Tavenner, The "Meaningful Use" regulation for electronic health records, *N Engl J Med* **363** (2010), 501-504.

[2] V. Kannan, J. Fish, J. Mutz, A. Carrington, K. Lai, L.Davis, J. Youngblood, M. Rauschuber, K. Flores, E. Sara, D. Bhat, and D. Willett, Rapid development of specialty population registries and quality measures from electronic health record data, *Methods Inf Med* **56** (2017), e74-e83..

[3] D.Willett, V. Kannan, L. Chu, J. Buchanan, F. Velasco, J. Clark, A. Ortuzar, J. Fish, J. Youngblood, D. Bhat, and M. Basit, SNOMED CT concept hierarchies for sharing definitions of clinical conditions using electronic health record data, *Appl Clin Inform* (2018), 678-693.

[4] A. Pandey, J. Berry, M. Drazner, J. Fang, W. Tang, and J. Grodin, Body mass index, natriuretic peptides, and risk of adverse outcomes in patients with heart failure and preserved ejection fraction: analysis from the TOPCAT trial, *J Am Heart Assoc* **7** (2018), e009664.

Address for correspondence

Duwayne Willett, MD, MS

Email: Duwayne.Willett@UTSouthwestern.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190535

Effect of Governance Functionality for Data Standardization Management of the Medical Information Database Network Project

Jinsang Park[a], Takanori Yamashita[a], Atsushi Takada[a], Chinatsu Nojiri[a], Rieko Izukura[a], Yasunobu Nohara[a], Taeko Hotta[b], Dongchon Kang[b], Naoki Nakashima[a]

[a] *Medical Information Center, Kyushu University Hospital, Fukuoka, Japan*
[b] *Department of Clinical Chemistry and Laboratory Medicine, Kyushu University Hospital, Fukuoka, Japan*

Abstract

Data standardiztion an important aspect to ensure data quality for utilizing large-scale, medical information databases such as the Medical Information Database Network (MID-NET) Project in Japan. We established a governance center to assess the consistency of standard codes across MID-NET-cooperating medical institutions. Moreover, we developed a real-time validation tool and determined its effect in improving data quality in medical institutions by providing a central feedback on the detected differences in standard disease-name codes.

Keywords:

Medical Information Database; Governance; Standardization

Introduction

The Medical Information Database Network (MID-NET) project (previously known as the "Japanese Sentinel Project") is establishing new "real-world data" (RWD) from multiple medical institutions in Japan. This project was established by the Ministry of Health, Labor, and Welfare, Japan, as a scientific approach to opt safety measures for adverse drug reactions to pharmaceuticals. The full-scale operation of the project was inititaed in FY2018 by the Pharmaceuticals and Medical Devices Agency (PMDA). The MID-NET database comprises RWD collected from approximately 4 million patients and consists of 10 medical institutions, including 23 hospitals in Japan. The project aimed to promote effective safety measures for drugs through pharmacoepidemiological methods using RWD. One of the most important factors for appropriately conducting pharmacoepidemiological research using medical information from multiple medical institutions is to systematically evaluate the quality of the data provided. However, the MID-NET Validation Project observed that, prioritized the management of medical institutions local codes associated to medical services, resulting in delays in clinical coding or omissions in standardized data [1]. As these factors are temporary, and continuously and unexpectedly occur at each institution, integrated management is extremely difficult and results in poor data quality. Therefore, our research group received support from the Japan Agency for Medical Research and Development (AMED) to establish an experimental "Governance Center" at the Kyushu University Hospital for the maintanence and management of the standard codes of the MID-NET project-cooperating medical institutions. Moreover, we developed a "real-time validation tool (Version 1)" for the management of standard codes and introduced this tool in two MID-NET project-cooperating medical institutions. Thus, the present study aimed to evaluate the standardization of codes across cooperating medical institutions and analyze the effect of this real-time validation tool in data standardization.

Methods

In this project, medical information from cooperating medical institutions was stored in the MID-NET integrated data source system from the Hospital Information System (HIS) via the Standardized Structured Medical Information eXchange Version 2 (SS-MIX2) [2]. This integrated data source system is used to network and comprises standardized database systems retrieved from electronic health records of cooperating institutions for the analysis and evaluation of adverse drug reactions. The MID-NET integrated data source system includes 11 types of currently available standard codes, such as the International Classification of Diseases 10th Revision (ICD10), Pharmaceutical Standard Code (HOT, YJ) and Laboratory Test Standard Code based (JLAC10) on SS-MIX2 standard storage. The table that was used while saving to a MID-NET integrated data source system was referred to as the "mapping table." The consistency of the standard codes from each cooperating medical institution that as extracted from the database of the Governance Center of the Kyushu University Hospital was evaluated using the mapping table (Fig. 1).

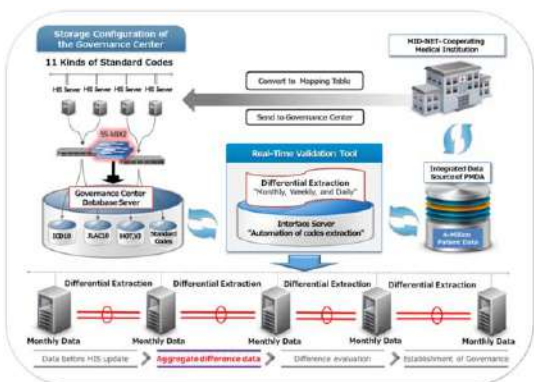


Figure 1–Configuration of the storage and real-time validation tool of the Governance Center

We evaluated the impact of the initial introduction of the real-time validation tool in two medical institutions for identifying differences in the standard disease-name codes from June 2018 to September 2018. The real-time validation tool automatically provided differential information on a monthly, weekly, and daily basis for 11 types of standard codes. Here, differential implies the addition of or change in the standard disease-management number and the standard disease name that matched with

the standard disease-name code in ICD10 as per the local code unique to that particular medical institution. We used the table for standard disease-name codes (v4.95 6/1 revised version) developed by the Medical Information System Development Center (MEDIS) as the key table (Fig. 2) to evaluate the accuracy of the difference detected by the real-time validation tool. Moreover, in an effort to determine the factors affecting the difference, we conducted an interview survey on the registration status of the standard disease names in the Electronic Health Records Management Department of medical institution A.

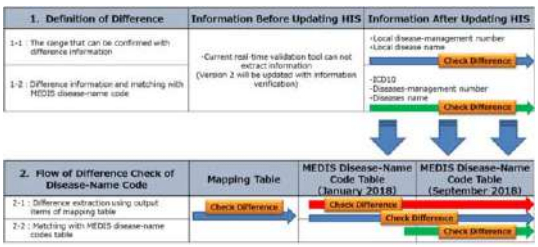


Figure 2–Flow of the procedure to detect differences in standard disease-name codes

Results

In total, 83 differences were detected by the real-time validation tool in medical institution A. Of these 83 differences, medical institution A showed a variation in the matching rate of the standard disease-name codes corresponding to 19 records in terms of consistency with the MEDIS key table. In medical institution B, 68 differences detected, which were consistent in the MEDIS key table (Table 1). Moreover, we confirmed three issues that were identified by an interview survey that assessed the flow of the registration of disease-name codes in the medical record management department.

1. 1. Medical records management department requested the registered, changed, and deleted disease-name codes from the medical information management department and were reflected in HIS.
2. 2. When registering the disease-name code in HIS, the codes were not evaluated for consistency with those in the MEDIS key table.
3. 3. When registering a disease-name code in HIS, the medical information department registered the medical institution unique disease-name code to the HIS when no suitable corresponding code was available in the MEDIS key table.

Table 1–Results of differential aggregation using MEDIS standard disease-name code

Time of Extraction	Medical Institution A (n=83)		Medical Institution B (n=68)	
Month	**Match**	**Mismatch**	**Match**	**Mismatch**
2018/6	0	4	20	0
2018/7	0	5	48	0
2018/8	0	5	0	0
2018/9	64	5	0	0
Total	**64**	**19**	**68**	**0**

Conclusions

The results of the present study indicate that the consistency between standard disease-name codes in the mapping table differed across medical institutions. At the clinical site, as only the institution unique code can be used, it is necessary to have a system that simultaneously, automatically, and correctly assigns corresponding standard codes. Moreover, using the real-time validation tool, we were able to visualize situations where large differences were noted in standard disease-name codes; these corresponded to differences detected by additions and changes at the clinical site. This may be attributed to the update of the MEDIS key table, which is conducted twice a year. Thus, it is presumed that the consistency may be affected following the update. Conversely, medical institution B showed a high consistency between and accuracy of disease-name codes regardless of the timing of the update of the MEDIS key table. Thus, these results indicate the following three reasons for the differences obseved in disease-name codes. First, while registering disease-name codes in HIS, if no code corresponded to those in the MEDIS key table, there was a high tendency to register a disease-name code unique to the medical institution. Second, when codes were stored in the mapping table in the SS-MIX2-standardized storage device, there was no real-time validation with the MEDIS key table. Third, coding errors and mistakes of the medical records management department were not evaluated when registering disease-name code in HIS. This study shows that consistency in standard disease-name codes by governance was effective in improving the data quality management system by detecting and unifying the mapping situation of standard disease-name codes in the MID-NET project. Taken together, our findings reinforce the importance of real-time management of standardized data and show that it is possible and feasible to improve data quality management by integrating and managing standard codes using a central governance method. While data inconsistencies were observed in the initial stage, the data quality dramatically improved following collaborative efforts between hospitals and PMDA. The MID-NET project is thus successfully progressing for full-scale implementation in FY2018.

Acknowledgments

This research was supported by AMED (Japan Agency for Medical Research and Development) under Grant Number (17mk0101064h0002) and (17mk0101075h0002). The authors thank the Ministry of Health, Labor, and Welfare, PMDA, and the MID-NET-cooperative medical institutions for their cooperation in the research.

References

[1] N. Nakashima, Pharmaceutical Regulatory Harmonization and Evaluation Research Project "Practical analysis method and education on drug epidemiology research for benefits and risk assessment of pharmaceuticals using MID-NET, The Japan Agency for Medical Research and Development (AMED) Outsourced Research and Development Results Report (15mk0101014h0102), 2016.

[2] M. Kimura, K. Nakayasu, Y. Ohshima, N. Fujita, N. Nakashima, H Jozaki, T. Numana, T. Shimizu, M. Shimomura, F. Sasaki, T. Fujiki, T. Nakashima, K. Toyoda, H. Hoshi, T. Sakusabe, Y. Naito, K. Kawaguchi, H. Watanabe, S. Tani, SS-MIX: a ministry project to promote standardized healthcare information exchange, Methods. *Inf. Med* 50 (2011), 131-139.

Address for correspondence

Jinsang Park, Ph.D, Medical Information Center, Kyushu University Hospital, 3-1-1 Maidashi Higashi-ku Fukuoka 812-8582,Japan; park21@med.kyushu-u.ac.jp

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190536

An Information Retrieval Approach to ICD-10 Classification

Hee Park, José Castaño, Pilar Ávila, David Pérez, Hernán Berinsky, Laura Gambarte, Daniel Luna, Carlos Otero

Departamento de Informática en Salud, Hospital Italiano de Buenos Aires, Argentina

Abstract

ICD-10 (International Classification of Diseases 10th revision) is a classification code for diseases, signs and symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or diseases. This paper describes an automatic information retrieval approach to map free-text disease descriptions to ICD-10 codes. We use the Hospital Italiano de Buenos Aires (HIBA) terminology data mapped to ICD-10 codes as indexed data to find an appropriate ICD-10 code using search engine similarity metrics.

Keywords:

Clinical Coding, International Classification of Diseases, Natural Language Processing

Introduction

The International Classification of Diseases code [1] is a very useful tool in a wide range of medical processes including statistical analysis of morbidity rate and mortality rate, medical reimbursement and resource allocation [2]. The registry by medical practitioners needs additional training time and manual encoding is not a simple task because not every health professional records in the same way [3].

Automated coding of ICD-10 is not an easy string matching work due to patient context's dependency [4]. There are two approaches to automated coding of medical texts to ICD-10: dictionary projection and supervised machine learning methods or the combination of both [3,4,5,7,8,9,10]. Dictionary projection methods use string matching techniques: n-grams, edit-distance, synonyms and abbreviations expansion, stop-word elimination, frequency of use, spell checkers and stemming. Machine learning methods include Support Vector Machine (SVM), Naive Bayes, Maximum Entropy Modeling (MaxEnt) and Gradient Boosting (XGBoost) [6].

The Hospital Italiano de Buenos Aires (HIBA) has a Spanish interface vocabulary where each term is mapped to SNOMED CT as its reference vocabulary. It has 207,000 post-coordinated concepts in its terminology system. This system provides services to the Hospital Italiano healthcare facilities as well as other healthcare institutions in Argentina, Uruguay, and Chile.

A major benefit of the local interface vocabulary is its size and coverage. Two previous works at HIBA are relevant to the work presented here. First, string similarity approaches [6] are used to map new description terms to existing terms in the interface vocabulary. Second, a terminology search-engine server was implemented [12]. We used a similar information retrieval approach to map a description term to the corresponding ICD-10 code. We constructed an ICD-10 search-engine that given a description term as a query, it finds the most similar existing term with an ICD-10 code and reports the corresponding ICD-10 code. We used the HIBA terminology data mapped to ICD-10 and the Spanish 2014 version of ICD-10 data. We also used HIBA terms frequency, stopwords, synonyms, collocations, abbreviations and frequent typographical errors term expansion. The ICD-10 data has 14,359 codes but HIBA mapping has 343,072 distinct terms that match only 8,537 codes (59.45%). We also added all the description terms corresponding to ICD-10 codes. We discovered a performance of 66.35% of recall @1, as an automated process [11,12]. We also evaluated recall @3, @5, @10. It should be noted that we obtained more than 80% with recall @5. These values are relevant for interactive approaches, where a health-care professional uses keywords or description terms to find the appropriate ICD-10 code for a given situation.

Methods

In this study, we constructed several indices using Elasticsearch to evaluate several alternative configurations. We used a dataset consisting of 6,787 distinct terms provided by an external institution that has been encoded by HIBA terminologists. Each electronic health record term can be codified to one or more ICD-10 codes, for example: "insuficiencia respiratoria con requerimiento alto de oxígeno" ("respiratory failure with high oxygen requirement") codes are J969 and R068.

Index Construction

The following parameters were used in each index construction with Elasticsearch:

1. Each data-set was indexed using Elasticsearch search engine using both n-gram and word analyzers.
2. For each concept in the HIBA terminology, we used its frequency to rank relevant concepts for a given mapping.
3. Mapping from HIBA terminology to ICD-10 concepts is a many-to-many mapping. We implemented these strategies using concept frequency:
   a) Maximum frequency of a HIBA concept mapped to an ICD-10 code.
   b) Minimum frequency of a HIBA concept.
   c) Average frequency of all HIBA concepts mapped to an ICD-10 code.
4. If a description term is mapped to more than one ICD-10 codes, if there is a more general term it is preferred (e.g. Z88 vs. Z889). At evaluation time, more general matches were considered correct.

5. Synonyms, collocations, abbreviations and frequent typographical errors term replacement. [6].

To construct the index, we evaluated 3 alternative data-sets:

I. The Spanish 2014 version of ICD-10 data.
II. The HIBA terminology.
III. The combination of both.

Results

We used several alternative configurations to construct the search-engine index. We used n-grams combined with word analyzers (1) and the term expansion (5) as the baseline approach. We used first only the Spanish 2014 version of ICD-10 data as the dictionary corpus (I), HIBA terminology frequency, and a general response as a valid match. This configuration resulted in 41.12 % recall @1.

Using the HIBA terminology as the dictionary corpus (II), we obtained a recall @1 of 62.84% improving the performance, without considering HIBA concept weights in ICD-10 codes. Using the two-dictionary corpus (III) provides a similar result, recall @1 62.9%.

The best result, recall @1 of 66.35, was obtained using minimum frequency strategy (3.b) stopwords and considering general codes as valid compared to a more specific one.

To be sure that the results returned by the search engine have better quality, a threshold can be used. We are considering a valid match only when the score is greater than the threshold. In other words, the search engine will not return a result under the confidence value. Using the score threshold of 6,000 we have a good balance, providing a better precision without losing too much recall. In Figure 1, we show the results of recall at more than one result for interactive choose.

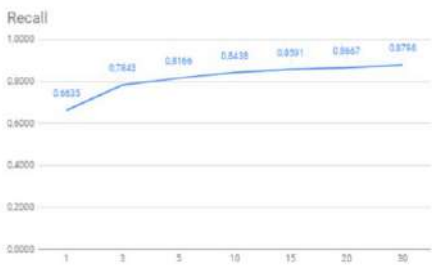


Figure 1. Recall @1, @3, @5, @10, @15, @20, @30

Discussion

We can see that the HIBA terminology to ICD-10 code mapping terms corpus, provides the best performance. The HIBA terminology covers only 59.45% of the ICD-10 codes but the richness in synonyms and frequent typographical errors enable a good result. We are using a combination of HIBA mapping to ICD-10 and the ICD-10 to complete all the ICD-10 codes but this addition doesn't provide a significant improvement.

The selection of the strategy (3.b) above, which uses the minimal frequency produces a better result than using the maximal, or the average frequency, which is somewhat surprising.

In the context of a health-care professional recording using an EHR, the system can provide the possibility of choosing the appropriate ICD-10 code from the list that returns the search engine, and as we have shown, recall @5 and @10 are very good, in the mid-80s.

Conclusions

In this work we showed several alternatives to construct an index for automatic ICD-10 term codification. We obtained a very reasonable performance. Some limitations of the current approach may be related to the differences on the corpus of the HIBA terminology data compared to the external institution data. Our approach is a dictionary projection method. A hybrid approach using supervised machine learning techniques may improve the performance.

References

[1] W.H. Organization, International statistical classification of diseases and health related problems (The) ICD-10, World Health Organization, 2004.
[2] O'malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM., Measuring diagnoses: ICD code accuracy, *Health services research*. **40(5p2) (**2005), 1620-39.
[3] Arifoğlu, D., Deniz, O., Aleçakır, K., & Yöndem, M., CodeMagic: semi-automatic assignment of ICD-10-AM codes to patient records, *Information Sciences and Systems* (2014), 259-268. Springer, Cham.
[4] Nitsuwat, S., & Paoin, W., Development of ICD-10-TM ontology for a semi-automated morbidity coding system in Thailand, *Methods of information in medicine* **51(06)** (2012), 519-528.
[5] Zweigenbaum, P., & Lavergne, T. (2016). Hybrid methods for ICD-10 coding of death certificates. In Proceedings of the Seventh International Workshop on Health Text Mining and Information Analysis (pp. 96-105).
[6] Castaño, J., Gambarte, M. L., Park, H. J., Williams, M. D. P. A., Perez, D., Campos, F., & Zanetti, S. (2016). A machine learning approach to clinical terms normalization. In Proceedings of the 15th Workshop on Biomedical Natural Language Processing (pp. 1-11).
[7] Mujtaba, G., Shuib, L., Raj, R. G., Rajandram, R., Shaikh, K., & Al-Garadi, M. A., Automatic ICD-10 multi-class classification of cause of death from plaintext autopsy reports through expert-driven feature selection, *PloS one* ***12*****(2)** (2017).
[8] Boytcheva, S. (2011). Automatic matching of ICD-10 codes to diagnoses in discharge letters. In Proceedings of the Second Workshop on Biomedical Natural Language Processing (pp. 11-18).
[9] Chen, Y., Lu, H., & Li, L., Automatic ICD-10 coding algorithm using an improved longest common subsequence based on semantic similarity, *PloS one* **12(3)** (2017)
[10] Jia, Z., Qin, W., Duan, H., Lv, X., & Li, H, A Hybrid Method for ICD-10 Auto-Coding of Chinese Diagnoses, *Studies in health technology and informatics*, **245** (2017) 427-431.
[11] Manning CD, Raghavan P, Schütze H. Introduction to Information Retrieval. New York, NY, USA: Cambridge
[12] Castaño, J., Park, H. J., Williams, M. D. P. A., Perez, D., Gambarte, M. L. Luna,D., (2018). Evaluating the Performance of a Terminology Search Engine Using Historical Data, *Medinfo* 2019.

Address for correspondence

Hee Joon Park
Ingeniería de Software
Hospital Italiano de Buenos Aires
hee.park@hospitalitaliano.org.ar

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190537

An Improvised Classification Model for Predicting Delirium

Sai Pavan Kumar Veeranki[a], Dieter Hayn[a], Stefanie Jauk[b,d], Franz Quehenberger[d], Diether Kramer[c], Werner Leodolter[c], Günter Schreier[a]

[a] *AIT Austrian Institute of Technology, Graz, Austria,*
[b] *CBmed, Graz, Austria,*
[c] *Steiermärkische Krankenanstaltengesellschaft m.b.H. (KAGes), Graz, Austria,*
[d] *Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria*

Abstract

With the vast increase of digital healthcare data, there is an opportunity to mine the data for understanding inherent health patterns. Although machine-learning techniques demonstrated their applications in healthcare to answer several questions, there is still room for improvement in every aspect. In this paper, we are demonstrating a method that improves the performance of a delirium prediction model using random forest in combination with logistic regression.

Keywords:
Algorithms, Delirium, Logistic Models

Introduction

With the increase in the rate of generation and collection of digital healthcare data, there is the need to mine the data for the extraction of knowledge. In the context of secondary use of health care data, the question arises if the large set of health care data can prevent some unwanted events during the health care process [1].

Predicting Delirium in Hospitalized Patients

Delirium is a neuro psychiatric syndrome which is very common in the elderly and is prevalent in the hospital setting. ~Up to 60% of elderly patients develop delirium during hospitalization and are often under diagnosed [2]. ~30-40% of such cases can be prevented with an intervention [2]. Delirium can cause further morbidities and may lead to death [2]. Thus, delirium is being recognized as a hospital quality measure [3]. Recent publications dealing with delirium prediction were based on delirium assessment tests and some selected features [2]. The idea was to identify patients who are at risk of developing delirium during their stay at the point of hospitalization using electronic medical records (EMR). Related work has been presented in previous publications, in which we have selected Random Forest (RF) to predict delirium as it is transparent to explain the decisions when compared to its counterparts [4].

Classification with Random Forest

Predicting delirium is a binary classification problem (either delirium or not). The present work was motivated to find ways to further improve the performance of classification, by combining RF with other algorithms.

RF is an ensemble of a set of decision trees, where each decision tree is a weak learner as only a subset of the data is used. Figure 1 shows a forest of n decision trees where the labeled nodes are the leaves where the decisions of the observations of the data took place. In prediction step, for an observation of the test set, data traverses through the tree and lands in only one of the leaves in each tree of the forest (red node). The corresponding decision of the selected leaf is considered for the prediction.

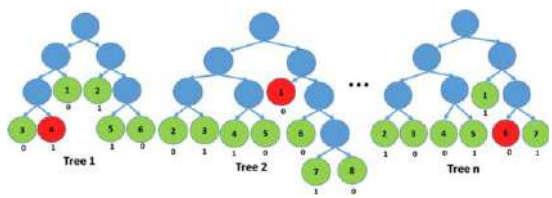


Figure 1– For prediction, each observation lands in one of the leaves (Decision leaf is red coloured node) in each tree of RF.

Objectives

The aim of this work was to present a method that would assign weights to each leaf (terminal node) in each tree of the RF, which would, subsequently improve the performance of the delirium prediction model and provide a transparent way of explaining the decisions.

Methods

Data, Sample Selection and Featureset

The data includes patients' demographical data, diagnoses, procedures, laboratory results, nursing assessment, etc. These data are partly based on international standards such as ICD - 10, ICPM, LOINC, etc. (see [4]). The data was provided by Steiermärkische Krankenanstaltengesellschaft m.b.H (KAGes).

After administering a set of criteria (see [4]), the delirium cohort consisted of 4,596 patients. 25,000 patients were selected randomly as control group in which 24,972 patients were considered for the analysis. After a systematic feature selection and engineering (see [4]), the final featureset consisted of 500 features with purely numerical values.

Algorithm

The algorithm described in this paper is based on an example published by Tim Head [5]. Initially, the given dataset was divided into three subsets with a ratio of 35%, 35%, and 30%. These datasets were used for training RF, Logistic Regression (LR), and evaluation, respectively. Figure 2 shows training set one, training set two and test set along with corresponding response variables y_{tr_1} , y_{tr_2} and y_{test} respectively (ground truth).

Step 1: Training the RF (turquoise): The training set one along with the corresponding response were used to train RF model. The leaves of individual weak learners of the RF model were transformed into a new feature matrix using One Hot Encoding [5].

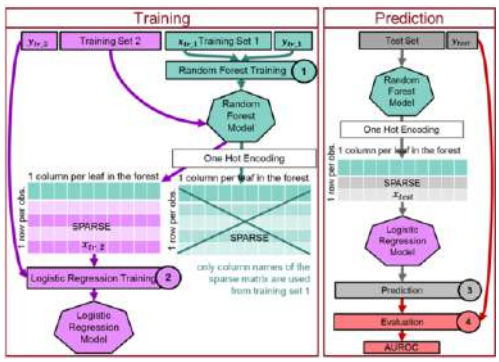


Figure 2– Overview of the algorithm, consisting of 1) training a random forest from training set one, two) deriving a new feature matrix, and 3) testing the final model with the test set.

Step 2: Training the LR model (purple): The training set two was applied to the RF model. Each observation in training set two leads to one leaf in each tree of the forest. The leaf indentifiers (IDs) of the corresponding decision leaves of each tree were fitted to the columns of the new feature matrix. The purple new feature matrix x_{tr_2} was considered as a training set along with a corresponding response variable $\{(x_{tr_2}, y_{tr_2})\}$ to train a LR, with regularisation. Step 3: Evaluation: the test set was fed into the RF model; the leaves of the decisions in the forest were fitted to the columns of the new feature matrix, which was used for prediction by the LR model. The predictions were then compared with the response variable y_{test} for evaluation. In this algorithm, we consider the IDs of decision leaves instead of the decision itself.

The length (number of columns) of the new feature matrix is denoted by $l = \sum_{i=1}^{n} N_i$, where, n is number of trees and N_i is number of leaves of the i^{th} tree. The coefficients (β_{ij}) that are derived from learning LR act as the weights of corresponding leaves in the trees of RF. The estimated outcome $\widehat{y_p}$ is derived from the formula given below.

$$\widehat{y_p} = \beta_0 + \sum_{i=1}^{n} \sum_{j=1}^{Ni} \beta_{ij} . x_{pij} \qquad (1)$$

Where $x_{pij} = 1$ when the decision for patient p lands in j^{th} node of the i^{th} tree, otherwise $x_{pij} = 0$. The classification will be obtained based on the selected threshold over the $\widehat{y_p}$.

Evaluation

We evaluated our approach by comparing the area under the receiver operating characteristic (AUROC) with that of LR, RF with majority voting and RF with weighted averaging. The results presented were obtained from built in functions of MATLAB for RF weighted average and LR. We have implemented proposed feature transformation method and RF Majority voting as described in evaluation in MATLAB. All RFs were trained with 1,000 trees, classification method with out of bag (OOB) prediction 'on' and without pruning and minimum leaf size selection.

Results

Figure 3 shows the AUROC of the four methods listed above. The AUROC of RF alone was 90.20%. With introduction of logistic regression, the performance increased by 1.17%.

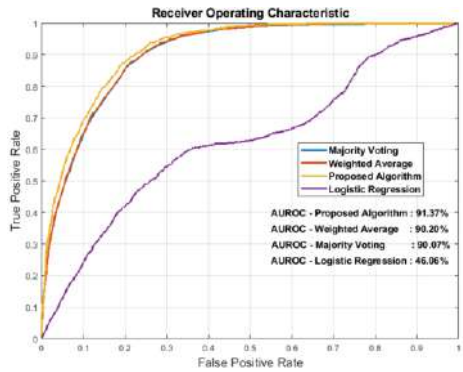


Figure 3 – AUROC of proposed method along with RF weighted average, majority voting and LR

Conclusions

The proposed method has outperformed the classical selected algorithms. However, there is a need for detailed investigation for real time health care applications. The merger of RF and LR has potential to find a reasonable way to explain the decisions.

Acknowledgements

This work has been carried out with K1 COMET Competence Centre CBmed, which is funded by Austrian Ministry of Transport, Innovation, and Technology, Ministry of Science, Research and Economy; Styrian Business Promotion Agency; and Vienna Business Agency. The COMET program is executed by FFG. KAGes and SAP provided resources, manpower and data as basis for research and innovation.

References

[1] S.M. Meystre, C. Lovis, T. Bürkle, G. Tognola, A. Budrionis, and C.U. Lehmann, Clinical Data Reuse or Secondary Use: Current Status and Potential Future Progress, *Yearb. Med. Inform.* **26** (2017) 38–52. doi:10.15265/IY-2017-007.

[2] A. Wassenaar et. al., Multinational development and validation of an early prediction model for delirium in ICU patients, *Intensive Care Med*. **41** (2015) 1048–56. doi:10.1007/s00134-015-3777-2.

[3] V. Douglas, and S. Rogers, Delirium and Hospital Quality, *Netw. Investig. Delirium Unifying Sci.* (2017). https://deliriumnetwork.org/delirium-and-hospital-quality/.

[4] D. Kramer, S. Veeranki, D. Hayn, F. Quehenberger, W. Leodolter, C. Jagsch, and G. Schreier, Development and Validation of a Multivariable Prediction Model for the Occurrence of Delirium in Hospitalized Gerontopsychiatry and Internal Medicine Patients, *Stud. Health Technol. Inform.* **236** (2017) 32–39.

[5] T. Head, Feature Transformations With Ensembles of Trees (2018). https://scikit-learn.org/stable/auto_examples/ensemble/plot_feature_transformation.html#sphx-glr-download-auto-examples-ensemble-plot-feature-transformation-py.

Address for correspondence

Sai Pavan Kumar Veeranki, AIT Austrian Institute of Technology, Reininghausstrasse 13, A-8020 Graz, Austria, E-Mail: sai.veeranki@ait.ac.at

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190538

Development of a Method for Extracting Structured Dose Information from Free-Text Electronic Prescriptions

Man Qing Liang[a,b], Vivek Gidla[c], Aman Verma[d], Daniala Weir[d,e], Robyn Tamblyn[d,e], David Buckeridge[d,e], Aude Motulsky[b,f]

[a] Faculty of Pharmacy, Université de Montréal, Montréal, Canada,
[b] Research Center, Centre hospitalier de l'Université de Montréal, Canada
[c] Faculty of Engineering, McGill University, Montréal, Canada
[d] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Canada
[e] Clinical and Health Informatics Research Group, Faculty of Medicine, McGill University, Montreal, Canada
[f] Department of Management, Evaluation and Health Policy, School of Public Health, Université de Montréal, Montréal, Canada

Abstract

In this study, we sought to develop an automatic parser tool for unstructured free-text electronic prescriptions, focusing specifically on defining the daily dose. We manually coded a set of electronic discharge prescriptions and established the most reliable rules to structure the medication data. A named-entity recognition (NER) parser tool was implemented, which was capable of identifying 90% of the doses and 86% of the frequencies from 255 dosage instructions.

Keywords:

Electronic Prescribing, Natural Language Processing, Prescription Drugs

Introduction

Electronic prescription technology is a digital health solution that seeks to improve the productivity of prescription filling, reduce the number of prescription errors in traditional prescription script writing, and optimize the appropriate use of drugs. Currently, in Quebec, electronic prescribing systems within electronic medical records normally allow prescribers to input either structured or free-text drug dosage instructions, the latter which precludes the systematic application of computer algorithms for further analysis (e.g., for pharmacoepidemiology studies or clinical decision tools). In order to maximize the potential of electronic prescription technologies, it is essential to have access to structured and standard drug dosage instructions. This study seeks to develop an automatic parser for unstructured electronic instructions.

Several studies have presented methods to extract medication data. Most notably in the literature, the third i2b2 Workshop on Natural Language Processing promoted research in medication data extraction from discharge summary [1]. For example, Xu et al. developed a natural language processing (NLP)-based parser to capture signature information (e.g., strength, route, frequency) and discussed the potential of NLP technologies to extract contextual information ("start", "stop", "increase") [2]. Based on previous designed models, we decided to build a parser adapted to Quebec's prescribing practice using real-life e-prescriptions.

In this study, we developed an NLP-based parser for structuring the dose and frequency data from free-text electronic discharge prescriptions in Quebec, and compared its performance with a baseline using only regular expression.

Methods

Data Sources, Data Analysis, and Semantic Categories

The hospital discharge data originated from a previous study on medication reconciliation at the MUHC [3]. Between October 2014 and November 2016, the study collected 131,336 hospital discharge prescriptions written in English by 202 physicians. We manually structured the dose and frequency data to establish the reference standard.

Opioids (oral form) were targeted given their high level of risk, variability in dose and frequency instructions, which are prone to parsing errors. ATC codes were used to identify and filter the opioid therapeutic class of drugs (ATC code starting with N02A), resulting in a total of 925 free-text opioid prescriptions.

Based on the available data for the input fields (Table 1), we established our desired output fields (Table 2) to determine the daily dose (daily drug exposure). For reference standard, the dose and frequency could either be fixed (one value) or variable (dose range per day with one minimum and one maximum value). To provide a better context of the daily drug exposure, we defined contextual tags (e.g., if the drug is "taken if needed", a "combination drug" or "variable dose over time"). The ultimate goal of this study is to define the daily dose, which could then be used to evaluate patient drug exposure on a daily basis and then applied to population-level data. We manually entered the dosage instruction in the appropriate fields according to our established reference standard, which was later compared with the results generated by the parser tool.

Implementation of a Natural Language Processing Tool

We compared two approaches to structuring the prescription data: regular expressions and an NLP-based approach. Although relatively simple to implement, regular expressions can only detect specific words or pattern of letters in a fairly rigid framework. Consequently, they could fail at any little divergence from the specific patterns defined. To produce a more robust parser, we used named-entity recognition (NER), an advanced natural language processing technique, for a

more generalized pattern recognition. NER is applied to detect different parts of a prescription such as the dosage or frequency. We implemented this method in SpaCy, a open-source NLP library that does NER [4]. The results were compared with the manually coded reference standard. Errors (e.g., not a perfect match) were then manually reviewed by the pharmacy intern (MQL), who analyzed their cause and provided feedback.

Table 1- Data source available for input for the parser tool

Input source	Examples
Generated by the prescriber	
Generic name	HYDROMORPHONE
Free text posology	1–2mg po q4-6h PRN for 2 weeks
DIN	885444
Generated through a look-up table with Vigilance (via DIN)	
Pharmaceutical form	TABLET
Strength	1MG
ATC code	N02AA03

Table 2- Desired output source for the parser tool

Example: 1–2mg po q4-6h PRN for 2 weeks	
Output field	**Desired output**
Dose information	
Dose range tag	1
Fixed dose	
Minimum dose	1
Maximum dose	2
Dose unit	mg
Frequency information	
Variable frequency	1
Fixed frequency	
Minimum frequency	4
Maximum frequency	6
Other qualifiers	
If needed (PRN)	1
Variable dose over time	0
Combination medication	0
Daily dose	
Minimum daily dose	4
Maximum daily dose	12

ResultsOverall, the NLP-based parser achieved a similar recognition success rate as the regular expression baseline. Both methods had similar frequency detection rate (91%). The dosage detection was better in the SpaCy-based parser (90%) than the baseline (86%) (Table 3). However, regular expressions had a better perfect match rate (84%) than NER-based approach (79%). The errors were analyzed to determine the cause. The main cause of errors was associated with combination medications (e.g., two or more active ingredients contained in a single pharmaceutical form), as the parsers had not been trained to detect those patterns yet. During parsing, the dose obtained was expressed in unit (e.g., number of tabs) instead of the dose per intake (in mg), which mislead the calculation of the daily dose. The daily dose was calculated according to the strength of each ingredient contained in the combined medication multiplied by the unit. Alternate doses in a single instruction also caused confusion during parsing. Lastly, the prescriber could prescribe either by product strength (e.g. tablets of 5mg) or dose (2.5 mg). When prescribed by product strength the parser had difficulty recognizing the difference between the dose per intake and the dose per tablet.

Table 3- Success rate for the extraction of unstructured dosage instructions

Match Type	Regular expressions (230 samples)	NLP-based parser (225 samples)
Fixed dose	86.5%	89.8%
Dose range	N/A	86.3%
Dose units	N/A	85.5%
Number of tabs	N/A	79.2%
Number of tabs range	N/A	81.2%
Strength	N/A	74.9%
Fixed frequency	91.7%	90.2%
Frequency range	N/A	80.8%
Taken if needed	N/A	94.1%
Variable dose over time	N/A	58.4%
Perfect match	83.9%	79.6%

Conclusions

In this study, we demonstrated that unstructured dosage instructions can be accurately extracted from hospital opioid prescriptions using standard NLP techniques. With promising results at an early development stage, it is reasonable to assume that the parser will have the potential to detect particular context cues and calculate the appropriate daily drug exposure. We can hope that future pharmacoepidemiology studies can benefit from this tool by extracting a variety of dosage instruction information, including the unstructured data without proper standardized fields. The ability to structure free-text opioid dosage instructions can be used to detect particular prescribing behaviors and resulting consequences, which are particularly relevant in the current context of opioid epidemic and misusage.

Acknowledgements

Man Qing Liang was funded by IVADO Undergraduate research scholarship. Aude Motulsky was funded by a scholarship from the Fonds de la recherche du Québec – Santé.

References

[1] Ö. Uzuner, I. Solti, and E. Cadag, Extracting medication information from clinical text, Journal of the American Medical Informatics Association 17 (2010), 514-518.

[2] H. Xu, S.P. Stenner, S. Doan, K.B. Johnson, L.R. Waitman, and J.C. Denny, MedEx: a medication information extraction system for clinical narratives, Journal of the American Medical Informatics Association 17 (2010), 19-24.

[3] R. Tamblyn, A.R. Huang, A.N. Meguerditchian, N.E. Winslade, C. Rochefort, A. Forster, T. Eguale, D. Buckeridge, A. Jacques, and K. Naicker, Using novel Canadian resources to improve medication reconciliation at discharge: study protocol for a randomized controlled trial, Trials 13 (2012), 150.

[4] M. Honnibal and M. Johnson, An improved non-monotonic transition system for dependency parsing, in: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015, pp. 1373-1378.

Address for correspondence

Aude Motulsky. Research Center, Centre hospitalier de l'Université de Montréal, 850 St-Denis, S01-130, Montréal, H2X 0A9. aude.motulsky@umontreal.ca; +1-514-343-5631

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190539

Prediction of Clinical Events in Hemodialysis Patients Using an Artificial Neural Network

Firdani Rianda Putra[a], Aldilas Achmad Nursetyo[a], Saurabh Singh Thakur[b], Ram Babu Roy[b], Shabbir Syed-Abdul[a, d*], Shwetambara Malwade [d], Yu-Chuan (Jack) Li[a,c, d]

[a] Graduate Institute of Biomedical Informatics, Taipei Medical University, Taipei, Taiwan
[b] Rajendra Mishra School of Engineering Entrepreneurship, Indian Institute of Technology Kharagpur, Kharagpur, India
[c] TMU Research Center of Cancer Translational Medicine, Taipei Medical University, Taiwan
[d] International Center for Health Information Technology, Taipei Medical University, Taiwan

Abstract

Advanced chronic kidney disease (CKD) requires routine renal replacement therapy (RRT) that involves hemodialysis (HD) which may cause increased risk of muscle spasms, cardiovascular events, and death. We used Artificial Neural Network (ANN) method to predict clinical events during the HD sessions. The vital signs, captured using a non-contact bed-sensor, and demographic information from the electronic medical records for 109 patients enrolled in the study was used. Weka Workbench software was used to train and validate the ANN model. The prediction model was built using a Multilayer perceptron (MLP) algorithm as part of the ANN with 10-fold cross-validation. The model showed mean precision and recall of 93.45% and AUC of 96.7%. Age was the most important variable for static feature and heart rate for dynamic feature. This model can be used to predict the risk of clinical events among HD patients and can support decision-making for healthcare professionals.

Keywords:
Renal dialysis, neural networks, electronic health records

Introduction

Taiwan reached peak incidence of chronic kidney disease (CKD) in 2014. CKD could advance into end stage renal disease (ESRD) which can be mentally and financially burdening. ESRD requires renal replacement therapy (RRT) to maintain quality of life and extend life expectancy among ESRD patients [1]. Hemodialysis is considered as treatment of choice for renal CKD and ESRD [2]. Despite its benefits, the procedure comes with risk of muscle spasms, cardiovascular events, and even death [3].

Artificial neural networks (ANN) have been known to provide state-of-the-art results for most of the classification tasks [4]. Despite its limitation to handle imbalanced data, recent oversampling method could overcome the problem correctly [5]. In this study, we aimed to predict clinical events during hemodialysis session using Multilayer Perceptron (MLP) algorithm.

Methods

In this study, the vital parameters were monitored and recorded for 109 patients during their HD session at Taipei Medical Hospital, Taipei, Taiwan. The data were collected using non-contact piezoelectric sensor which can be placed under the mattress. Total 3,237 HD sessions were observed in 23 weeks. During each session the data captured included: heart rates (HR), respiration rates (RR) and movement data (MD). In addition, the demographic details of patients such as gender, age, height, and weight, were also obtained. Following data collection, data cleaning and preparation procedures were conducted, using the synthetic minority over-sampling technique (SMOTE) [6] to handle unbalanced data between classes. Patients who reported emergency visit (ER), muscle spasm (MS), inpatient (IP), emergency visit and inpatient (ERIP) or sudden death (SD) during observation were classified in the class-event, and not reported clinical events as no-event. Weka Workbench software was used to train and validate the ANN model [7]. We performed 10-fold cross validation to evalute the model.

Results

The performance of the model developed in this study as shown in Table 1, showed an ROC of 0.96 (Figure 1) with False Positive Rate (FPR) and True Positive Rate (TPR).

Table 1 – Weighted Average Performance Model Result

	Precision	Recall	ROC
MLP	0.93	0.93	0.96

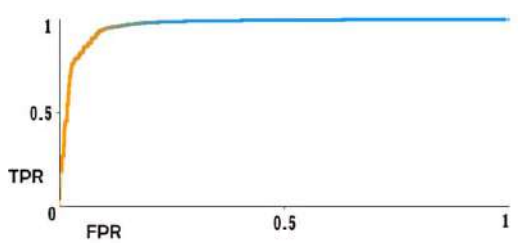


Figure 1 – ROC Curve
(TPR: True Positive Rate, FPR: False Positive Rate)

6 nodes in hidden layer MLP model showed top 5 significant variables (Table 2) as following: Heart Rate First Five Minutes (HR_FFM), Respiration Rate First Five Minutes (RR_FFM), Age, Respiration Rate Last Five Minutes (RR_LFM), and Heart Rate Last Five Minutes (HR_LFM) (Table 2).

Table 2 – Total weight variable in hidden layer MLP

No	Variable	Total Weight
1	HR_FFM	51.65743
2	RR_FFM	27.29772
3	Age	18.02295
4	RR_LFM	9.87849
5	HR_LFM	6.46442

Discussion

Our study indicated that ANN approach when applied to the classification task, performed with an ROC curve of 0.96 %. A study conducted by Barbieri et. al used AI to improve anaemia management during HD [8]. Our previous study [9] applied commonly known statistical learning methods such as k-Nearest Neighbour (kNN) and Support Vector Machine (SVM). In our study, MLP, as a feature selection algorithm, extracted HR_FFM as the most important variable for static feature and age for dynamic feature, which indicated that the model used the non-contact sensor to come up for clinical suggestions determine event or non-event patient. These suggestions could be sent to physicians to assist with their clinical decision making. Feedback about the suggestions (correct or wrong) would be collected, and as feedback into the AI system so that it can keep improving accuracy.

Conclusions

Our study demonstrated a new approach using non-contact sensor data and demographic data of HD patients, MLP being the proposed prediction model. MLP algorithm shows better performance than commonly known statistical learning methods. Our model could be used to predict the risk of clinical events among HD patients to support decision for healthcare professionals.

References

[1] M.H. Tsai, C.Y. Hsu, M.Y. Lin, M.F. Yen, H.H. Chen, Y.H. Chiu, and S.J. Hwang, Incidence, Prevalence, and Duration of Chronic Kidney Disease in Taiwan: Results from a Community-Based Screening Program of 106,094 Individuals, *Nephron* **140** (2018), 175-184.

[2] G. Villa, Z. Ricci, and C. Ronco, Renal Replacement Therapy, *Crit Care Clin* **31** (2015), 839-848.

[3] P. Ravani, S.C. Palmer, M.J. Oliver, R.R. Quinn, J.M. MacRae, D.J. Tai, N.I. Pannu, C. Thomas, B.R. Hemmelgarn, J.C. Craig, B. Manns, M. Tonelli, G.F. Strippoli, and M.T. James, Associations between hemodialysis access type and clinical outcomes: a systematic review, *J Am Soc Nephrol* **24** (2013), 465-473.

[4] N. Carlini and D. Wagner, Towards Evaluating the Robustness of Neural Networks, in: *2017 IEEE Symposium on Security and Privacy (SP)*, 2017, pp. 39-57.

[5] M. Buda, A. Maki, and M.A. Mazurowski, A systematic study of the class imbalance problem in convolutional neural networks, *Neural Netw* **106** (2018), 249-259.

[6] W.P. Kegelmeyer, L.O. Hall, K.W. Bowyer, and N.V. Chawla, SMOTE: Synthetic Minority Over-sampling Technique, *Journal of Artificial Intelligence Research* **16** (2002), 321-357.

[7] E. Frank, M.A. Hall, and I.H. Witten, *The WEKA Workbench*, 2016.

[8] C. Barbieri, M. Molina, P. Ponce, M. Tothova, I. Cattinelli, J. Ion Titapiccolo, F. Mari, C. Amato, F. Leipold, W. Wehmeyer, S. Stuard, A. Stopper, and B. Canaud, An international observational study suggests that artificial intelligence for clinical decision support optimizes anemia management in hemodialysis patients, *Kidney Int* **90** (2016), 422-429.

[9] S.S. Thakur, S.S. Abdul, H.-Y.S. Chiu, R.B. Roy, P.-Y. Huang, S. Malwade, A.A. Nursetyo, Y.-C.J. Li, S.S. Thakur, S.S. Abdul, H.-Y.S. Chiu, R.B. Roy, P.-Y. Huang, S. Malwade, A.A. Nursetyo, and Y.-C.J. Li, Artificial-Intelligence-Based Prediction of Clinical Events among Hemodialysis Patients Using Non-Contact Sensor Data, *Sensors* **18** (2018), 2833.

Address for correspondence

Shabbir Syed-Abdul

E-mail : drshabbir@tmu.edu.tw

Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, 15F., No. 172-1, Sec. 2, Keelung Rd., Da'an Dist., Taipei, Taiwan 10675.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190540

Design of a System of Systems (SoS) for the Interoperability of Non-Invasive Sensors for the Care of Older People

Carla Taramasco[a,b], Fabián Riquelme[a,b]

[a] *Escuela de Ingeniería Civil Informática, Universidad de Valparaíso, Chile,*
[b] *Centro de Investigación y Desarrollo en Ingeniería en Salud, Universidad de Valparaíso, Chile*

Abstract

We propose the design of a system of systems (SoS) called Seniors Guardian as an integrated environment for healthcare monitoring, in which several non-invasive sensors coexist to monitor falls, nocturia, perseverant behaviors, air quality (carbon monoxide), humidity, and temperature in real time. The modular architecture also allows the integration of new components. This research adheres to the use of Internet of Things to develop ambient assisted living in smart homes.

Keywords:

Frail elderly, Homes for the Aged, Environment, Internet

Introduction

A permanent risk in older adults is falls. Around one-third of adults older than 65 suffer a fall each year, and only half can get back on their own after they fall [1]. This can lead to fear, pressure ulcers, osteopenia, loss of muscle mass, dehydration, hypothermia, pneumonia, and even death [2]. The financial costs that countries pay to cover injuries caused by falls in adults over 65 are substantial and sometimes unsustainable over time [3]. Other risky event that may occur inside homes of the elderly is nocturia – the need to wake up at night many times to go to the bathroom. Nocturia is associated with sleep disorders that can lead to reduced concentration, cognitive decline, poor levels of energy, and overall, negative effects on the quality of life [4].

To deal with these issues, a patient-centered care model is recommended, where not only emergency cases are accepted, but also patients are actively committed to their care and therapeutic processes [5]. The advances in information technology (IT) and the reduction in hardware costs favor a preventive and proactive healthcare experience. Non-invasive monitoring of elderly allows for this experience and takes care of early symptoms of neurodegenerative diseases and the generation of treatments, plans, and programs geared towards the needs of this age group. In this context, *smart homes* quip homes with technology elements (such as sensors, microcontrollers) to monitor actions or events of the residents, as well as environmental conditions at the location, in order to detect emergencies early and reduce the times in making subsequent decisions [6,7].

There are several recent advances in low-cost devices to detect falls [7,8] and nocturia [9] in older people. It is known that people with nocturia are exposed to falls, due to the state of drowsiness at night [10]. However, there are no long-term measures to determine what percentage of falls in older adults are caused by nocturia. To solve this problem, the interoperability of monitoring systems would be necessary.

Methods

The goal is to develop a platform for continuous monitoring of the adult's daily activities to detect risk events both *inside* and *outside* of home. The key is to provide them with tools while safeguarding their autonomy and independence. The System of Systems (SoS) should contain software and hardware (sensor network) components. The monitoring systems considered are the following:

Actimetry (M1): To classify the daily activities of the elderly through infrared sensors of low resolution (for privacy). This will allow generating patterns of daily behavior. After the elderly activity profile is created, a probabilistic model based on Markov chains can determine abnormal behaviors. This module will be used to detect falls with an accuracy of 90%, by means of AI algorithms already in use [8].

Harmful gases detector (M2): To detect abnormal concentrations in environment and thus prevent accidents (such as explosions, inhalation of carbon monoxide) from sources of emission of gases (such as kitchen, stove, heater).

Temperature and humidity sensors (M3): To measure environment indicators that could negatively affect the health of the elderly.

Alert module (M4): To receive alerts generated by the elderly out of home through a 3G key chain with a panic button. Alerts are sent together with the elderly location to his/her contacts and emergencies.

Wandering analysis module (M5): To detect the wandering of the elderly. A mathematical model on the key chain of M4 monitor elderly trajectory and determines if she/he is lost.

With the data collected by the sensors, daily, monthly, and annual reports on the activities of the patient will be made. By unifying information from several patients, it will be possible to perform advanced analysis through data mining, to find non-trivial patterns: characteristics of the evolution of old age, disability, dependency, and more in order to guide treatments, plans, programs, and policies.

Results

We propose *Seniors Guardian*, an integrated environment for healthcare monitoring (Figure 1). The modularity approach allows the integration of new components. If one module stops

working, the others can continue to work. Data access will be given by two types of interfaces: a web application, for healthcare staff and decision makers that displays data about follow-up of the elderly, alerts issued, and descriptive statistics of their evolution; and a mobile application that receives the alerts and displays them in the cell phones of the contacts (user's family/caregiver) and the physician in charge.

The hardware drivers consist of a Local Processing Unit (LPU) to process the signals captured by the sensors, and a Remote Supervision Unir (RSU) to receive and manage the information sent from the LPU or the 3G key chain. The thermal sensor-actimetry is an AMG8833 (Figure 2) with 8x8 resolution pixels (viewing angle of 60°x60°). Information is processed by an ODROID-C1+ minicomputer connected to the sensors by an ATMEGA328P microcontroller (Figure 3). The key chain 3G uses a wireless button with 3G connection, 3.7V 800mAh battery, GPS and will allow calls. A DHT22 sensor detects humidity between 0-100% RH with 5% accuracy and temperature from -40°C to 125°C with 0.5°C of precision. It has a sampling rate of 0.5 [Hz] (it measures every 2 seconds). The MQ-9 sensor can detect concentrations of harmful gases from 100-10000ppm.

Figure 1 – The Proposed Solution

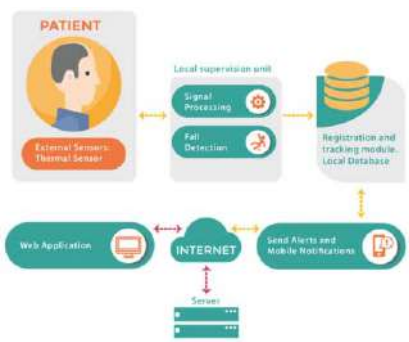


Figure 2 – Thermal Sensor Diagram

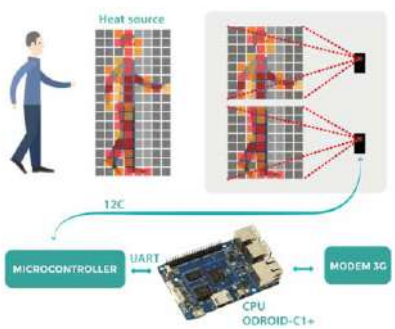


Figure 3 – Older Adult Detection and Sensor Coverage Area

Conclusions

Nowadays, the patient-centered care model is fundamental to improve the quality of life of the elderly people. In this work, we proposed a SoS that allows the interoperability of all these systems in a modular way, favoring the crossing of information to improve patient care, detect emerging risks more quickly, and respond more efficiently when a fall or some other problem occurs. Once the system is implemented, it will be validated by patients in a real-world environment.

Acknowledgements

This work was partially funded by CORFO – CENS (Grant #16CTTS-66390) and FONDEF IDeA I+D #ID18I10212. This work has the approval of ethics (code CEC181-19) granted by the accredited scientific ethics committee of the University of Valparaíso, Chile.

References

[1] A. Ambrose, G. Paul, and J. Hausdorff, Risk factors for falls among older adults: a review of the literature, *Maturitas* 75 (2013), 51-61.

[2] J. Fleming, and C. Brayne. Inability to get up after falling, subsequent time on floor, and summoning help: prospective cohort study in people over 90, *BMJ* 337 (2008).

[3] J. Davis, M. Robertson, M. Ashe, T. Liu-Ambrose, K. Khan, and C. Marra, International comparison of cost of falls in older adults living in the community: a systematic review, *Osteoporosis International* 21(2010), 1295-1306.

[4] B. Breyer, A. Shindel, B. Erickson, S. Blaschko, W. Steers, and R. Rosen, The association of depression, anxiety and nocturia: a systematic review, *The Journal of Urology* 190 (2013), 953-957.

[5] M. Facchini, Behavior changes in long-term treatment: Physician-patient relationship. *Medicina (B. Aires)* 64 (2004), 550-554.

[6] L. De Silva, C. Morikawa, and I. Petra, State of the art of Smart homes, *Engineering Applications of Artificial Intelligence* 25 (2012), 1313-1321.

[7] G. Demiris, M. Rantz, M. Aud, K. Marek, H. Tyrer, M. Skubic, A. Hussam, Older adults' attitudes towards and perceptions of 'smart home' technologies: a pilot study, *Med Inform Internet Med* 29 (2004), 87-94.

[8] F. Harow, N. Zerrouki, Y. Sun, and A. Houacine, Vision-based fall detection system for improving safety of elderly people, *IEEE Instru Meas Mag* 20 (2017), 49-55.

[9] C. Taramasco, T. Rodenas, F. Martínez, P. Fuentes, R. Munoz, R. Olivares, V. De Albuquerque, and J. Demongeot, A novel monitoring system for fall detection in older people, *IEEE Access* 6 (2018), 43563-43574.

[10] C. Taramasco, T. Rodenas, F. Martínez, P. Fuentes, R. Munoz, R. Olivares, V. De Albuquerque, and J. Demongeot, A novel low-cost sensor prototype for nocturia monitoring in older people, *IEEE Access* 6 (2018), 52500-52509.

[11] R. Stewart, M. Moore, F. May, R. Marks, and W. Hale, Nocturia: a risk factor for falls in the elderly, *J Am Geriatr Soc* 40 (1992), 1217-1220.

Address for correspondence

Corresponding Author: Fabián Riquelme (e-mail: fabian.riquelme@uv.cl;
address: General Cruz #222, 3th floor, Valparaíso, Chile)

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190541

Comparing Two Standardized Value Sets of Infectious Agents: Implications for Semantic Interoperability

Julian Sass[a], Moritz Lehne[a], Sylvia Thun[a, b, c]

[a] *Berlin Institute of Health (BIH),10178 Berlin, Germany,*
[b] *Charité – Universitätsmedizin Berlin, 10117 Berlin, Germany*
[c] *Hochschule Niederrhein – University of Applied Sciences, 47805 Krefeld, Germany*

Abstract

Infections are a global public health concern. For coordinated actions against infectious diseases, semantic interoperability between infection control systems is crucial. This requires a consistent use of standard terminologies such as SNOMED CT. Here, we compare two value sets of infectious agents annotated with SNOMED CT (WHONET 2018 vs. pathogens reported under the German Protection against Infection Act). Our comparison revealed several inconsistencies, highlighting the importance of the consistent and coordinated use of standard terminologies.

Keywords:

Infection control, SNOMED CT, Standardization

Introduction

Infections are among the leading causes of death worldwide [1]. As communicable diseases, infections can lead to dangerous outbreaks, claiming a high number of victims (recall, for example, the 2013-2016 Ebola epidemic in West Africa, which killed over 60% of the patients infected [2]). Multidrug-resistant organisms exacerbate the danger of infectious diseases and are especially challenging for hospitals, in which the total burden of hospital-acquired infections largely exceeds other communicable diseases [3]. Improving our understanding of infections and developing effective surveillance, treatment and prevention programs are therefore high-priority public health concerns, which can save lives and mitigate the risk of local and global epidemics.

Establishing effective infection control programs requires a common language—a standard terminology—for infectious organisms. Such a standard terminology can ensure that, once detected, infectious agents are documented in a systematic way so that different institutions or systems can easily exchange information. This semantic interoperability can enable large-scale analyses across different hospitals, institutions and countries leading to better prediction, early intervention and, ultimately, prevention of infectious diseases.

SNOMED CT (www.snomed.org) is a powerful terminology of medical concepts that provides a possible solution for developing standardized value sets of infectious agents. The top-level concept "Organism" that is part of SNOMED CT includes 34,405 subconcepts, among them a multitude of bacteria, viruses, prions, fungi, nematodes and others. This makes SNOMED CT a suitable terminology for defining comprehensive value sets of infectious agents.

The clinical concepts included in SNOMED CT have been used to generate standardized value sets for infectious organisms. For example, WHONET 2018 [4], a microbiology laboratory database software developed for the surveillance of infectious diseases, includes a list of 2,551 infectious agents annotated with SNOMED CT. Similarly, a German project annotated a list of infectious organisms with SNOMED CT to improve the communication between laboratories and public health authorities [5]. This list comprises the notifiable infectious organisms that have to be reported to the Robert Koch Institut (RKI) under the German Protection against Infection Act (Infektionsschutzgesetz, IfSG) [6].

However, to ensure semantic interoperability between different value sets (such as the two described above), standard terminologies have to be applied consistently. If different value sets use different codes for the same organism—even if they originate from the same terminology—interoperability is not eventually achieved.

This study aims to evaluate the consistency between two value sets that use SNOMED CT as a standard terminology for infectious agents. Using the value sets from WHONET 2018 and the RKI list of infectious agents annotated with SNOMED CT, we compare the agreement between concepts in the two value sets.

Table 1 – Example of infectious organisms reported under the German Protection against Infection Act (FSN: Fully specified name).

Organism	SNOMED CT identifier	SNOMED CT FSN
Adenovirus	74871001	Human adenovirus (organism)
Bacillus anthracis	21927003	Bacillus anthracis (organism)
Bordetella parapertussis	26183002	Bordetella parapertussis (organism)
Bordetella pertussis	5247005	Bordetella pertussis (organism)
Borrelia recurrentis	34726005	Borrelia recurrentis (organism)
Brucella sp.	26250004	Genus Brucella (organism)
Campylobacter	35408001	Genus Campylobacter (organism)
Chlamydia psittaci	14590003	Chlamydiophila psittaci (organism)

Methods

The starting point of this analysis is the list of microorganisms included in the WHONET 2018 software [4], which additionally contains a map to SNOMED CT identifiers for each of its organisms. A second list of 55 organisms is provided by the Robert Koch Institut (RKI) under the German Protection against Infection Act [6]. This set of notifiable pathogens was annotated with SNOMED CT as part of a project in the German federal state of North Rhine-Westphalia aimed at modernizing the reporting system between laboratories and health authorities [5]. Table 1 shows an excerpt of the organisms included in the RKI list. Initially, this analysis tested, how many of the 55 microorganisms from the RKI set were included in the WHONET set. In a second step, as illustrated in table 2, we compared if SNOMED CT identifiers for organisms in one set matched the identifiers of equivalent organisms in the other set.

Table 2 – Example comparison of agreement between organisms mapped to SNOMED CT from WHONET and RKI

WHO organism	SNOMED CT identifier	RKI organism	SNOMED CT identifier
Adenovirus	74871001	Adenovirus	74871001
Campylo-bacter sp.	116457002 (deprecated)	Campylo-bacter	35408001
Varicella zoster virus	80298008	Varicella-Zoster Virus	19551004

Results

Of the 55 pathogens listed in the RKI value set, 49 (89.1%) could also be identified in the WHONET value set. Of these, 32 (58.2%) were annotated with the same SNOMED CT identifier; 12 (21.8%) were annotated with different, sometimes more specific SNOMED CT identifiers; 2 (3.6%) were annotated with inactive identifiers; 3 (5.5%) did not have identifiers in the WHONET list (Figure 1).

Conclusions

The comparison between both sets of microorganisms annotated with SNOMED CT presented several inconsistencies. For example, semantically equal concepts did not have identical identifiers or matched just narrowly in subtype relationships. Note that with 2,551 organisms, the WHONET value set is much larger than the RKI value set and therefore can use a more fine-grained classification of organisms. This could explain why, for some organisms in the RKI value set, no direct matches were found in the WHONET value set. Furthermore, WHONET contains some deprecated codes, which by now have been replaced by active SNOMED CT concepts. This analysis highlights that, to achieve semantic interoperability between systems, coordinated efforts with regular updates, consistency checks and quality assurance are necessary. For terminology services it should be considerd to cluster subspecies of organisms, if reasonable, for example through SNOMED Expression Constraint Language (ECL). Algorithms for infection control systems may then be able to detect outbreaks where hospital information systems report organisms in varying granularity.

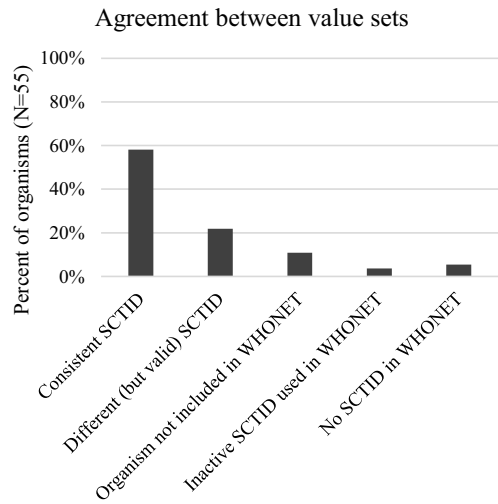


Figure 1 – Agreement between the infectious organisms reported under the German Protection against Infection Act compared with the WHONET 2018 value set (SCTID: SNOMED CT identifier).

References

[1] GBD 2016 Causes of Death Collaborators, Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016, *Lancet* **390** (2017), 1151-1210.

[2] T. Garske, A. Cori, A. Ariyarajah, I. M. Blake, I. Dorigatti, et al., Heterogeneities in the case fatality ratio in the West African Ebola outbreak 2013–2016, *Phil Trans R Soc B* **372** (2017), 20160308.

[3] A. Cassini, D. Plachouras, T. Eckmanns, M. Abu Sin, H.P. Blank, et al., Burden of Six Healthcare-Associated Infections on European Population Health: Estimating Incidence-Based Disability-Adjusted Life Years through a Population Prevalence-Based Modelling Study, *PLOS Med* **13** (2016), e1002150.

[4] WHONET 2018 (Build date: 2018-10-31), whonet.org.

[5] German federal state project funded by ZTG Zentrum für Telematik und Telemedizin GmbH: Einführung des überregionalen Echtbetriebs eines elektronischen Meldeverfahrens zur Übermittlung von meldepflichtigen Krankheiten und Erregern von Laboren an die zuständigen Gesundheitsämter [A cross-regional electronic system for reporting infectious diseases and organisms from laboratories to health authorities] (2014).

[6] German Protection against Infection Act: Infektionsschutzgesetz - IfSG.

Address for correspondence

Moritz Lehne, Berlin Institute of Health (BIH), Anna-Louisa-Karsch-Str. 2, 10178 Berlin, Germany, E-mail: moritz.lehne@bihealth.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190542

Dashboard and a Model of Predictive Analysis for Cerebrovascular Diseases in Primary Health Care

Jades Fernando Hammes[a], Grace T M Dal Sasso[b]

[a] *Bridge Laboratory, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil*
[b] *Health Sciences Center, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil*

Abstract

Cerebrovascular and hypertensive diseases are among the leading causes of death in the world. This study was developed in two stages: the development of a dashboard with specific reports of cerebrovascular diseases and applied a preliminary predictive analysis model for cerebrovascular diseases. The results demonstrate the abitity to predict a citizen's chance acquiring a cerebrovascular disease.

Keywords:
Data Mining, Primary Health Care, Cerebrovascular Diseases.

Introduction

The Ministry of Health (MS) is working on the guidelines on e-Health, inline with the principles of the Unified Health System (SUS) and the e-Government policies (e-Gov). The e-Health Strategy document for Brazil is targeting 2020 in order to increase the quality and the access for health care, as well as consolidate e-Health as a relevant State Policy for SUS [1].

Health organizations around the world are generating large amounts of data. Thus, organizations throughout the world face dilemmas due to the growth and volume of health data. Making decisions based on facts is not dependent on the amount of data you have. According to the SAS Institute [2], working with so much data can be scary. In this sense, success in decision making will depend on how quickly information is discovered and how this information are used to generate best practices in the organization. The SAS Institute also emphasizes that data visualization showing important metrics and indicators can help facilitate the understanding of health information via dashboards. Predictive analysis, data mining, machine learning and decision management will be key to assessing and predicting some behavior based of that data. Thus using it to transform data into proactive ideas with actions that are used in operational processes. This is a key interest of this study. The ability to store this data enhances both learning and the ability to develop knowledge that promotes improvements in health outcomes. In this way, it is fundamental to understanding the behavior of a disease and improving its clinical outcomes [3]. The objectives of this study were to develop a georeferenced map of cerebrovascular diseases in primary health care as support for decision-making to sensitive health conditions from integrated e-SUS AB data in Brazil; and to present a preliminary model predictive report of citizens at risk of cerebrovascular diseases for possible integration with e-SUS primary care in Brazil.

Methods

The study was divided into two stages. The first was the development of a dashboard with specific reports of cerebrovascular diseases in primary health care, presenting individual citizens who either had or have a cerebrovascular disease. Another report presented the indicators of each ICD10 in the cerebrovascular disease group, with the number of cases, grouped by sex, mean age and distribution per period. Also, in the dashboard, the cases of citizens who have already been affected by this disease were presented georeferenced on a map. The second stage was the development of a preliminary predictive analysis model, based on logistic regression, where an algorithm learns from historical data on characteristics and risk factors for cerebrovascular diseases with people who already have a group disease and predicts a citizen's chance to have, for example, a stroke.

The predictive model used in the study was a regression optimized by an optimization algorithm, Stochastic Descending Gradient (SGD) in which the stochastic downward gradient does not use all the data but a fraction of the data. For the tool developed, data was randomly generated from 170,567 citizens. These people served as reference for the algorithm to learn the patterns and relationships in which the citizen develops a cerebrovascular disease. The variables that served as analysis for this preliminary model were age, sex, race-color, obese, ex-obese, smoker and ex-smoker, hypertension, diabetes, pregnant women, alcohol and ex-alcohol, drugs and ex-drugs, among others. The results of this predictive analysis model are presented in the dashboard in the form of a report.

The tool has been developed according to software engineering tools, from prototyping, database modeling, migration process, tool construction and predictive analytics using free technologies.

The distribution of Cerebrovascular Disease functionality was divided into three tabs: a general Dashboard, a Heat Map and a Georeferenced Map. The georeferenced map shows the residences of people with cerebrovascular diseases identified by a point on the map (Figures 1).

To view or apply the preliminary model, the tool has a screen to show the chance of an individual getting a cerebrovascular disease in percent. The name of the functionality in the system is Risk of Cerebrovascular Diseases (Predictive) Analyze. (Figure 2)

Figure 1: Distribution of Cerebrovascular Diseases – Heat Map.

Figure 2: Risk of Cerebrovascular Diseases (Predictive Analysis)

Results

After structuring the database in PostgreSQL®, 5.3 the data and consolidated information is presented ina dashboard built with JAVA using Spring. Next, the data and information is analyzed on a map with geoprocessing technique from the Google® Geocoding API tool, and the database addresses were converted by geographic coordinates in a process called geocoding. Thus the citizens who have had or have cerebrovascular diseases for monitoring and management of health care are identified. Another important report was the presentation of the indicators of cerebrovascular diseases.

For the preliminary predictive analysis with citizens who have risks of developing cerebrovascular diseases to health professionals, an algorithm using Python v3.7.0 was applied. Initially, we exported the prediction table that was based off data from the e-SUS (system) basic care for cerebrovascular diseases. In the preparation to receive the source data, the input columns were defined, that is, the variables that the prediction algorithm would use to learn the pattern and to generate the probabilities. It was necessary to normalize age using Fuzzy Sets. For the execution of the algorithm, the prepared data were converted to X_train and X_test and executed using logistic regression optimized by an optimization algorithm, Stochastic Descending Gradient (SGD). It was possible to observe the correlation between the training and test variables to discover data failures, because if they differ greatly, it would be a problem for the algorithm, but in this case, they are similar. Heat maps of this correlation in algorithm model learning (Figure 3) showed the negative relationship between males and females, which validates that the data has no error in relation to these columns. On the other hand, cerebrovascular disease showed a positive correlation with hypertension, so it is expected that these two variables will similarly affect the outcome for both males and females.

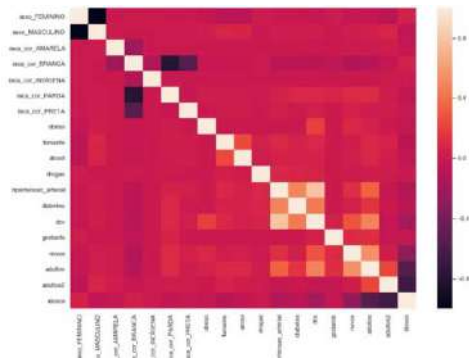


Figure 3: Heat map of correlation in learning

Finally, the trained algorithm was prepared to execute on the whole set of data and obtain the probability result. The report of Cerebral Vascular Diseases (Predictive Analysis) shows a list of all citizens of the base, after applying the preliminary predictive algorithm, ordered by the highest probability of acquiring the disease.

Conclusions

The development of the present study offered to the health professionals and health managers' geo-referenced demographic reports to visualize the health conditions in the territory, as well as to have follow-up reports of citizens with diseases and visualize indicators of each disease in your health region. The report of Risk of Cerebrovascular Diseases by the predictive analysis, although still preliminary, presents to the citizen the chance of the person being surprised by the disease according to its parameters of health risks. In addition, a risk assessment of cerebrovascular diseases will enable preventive actions to be performed before the clinical condition occurs, with the purpose of reducing costs of hospital admission, providing subsidies for health professionals to take decision-making and planning.

References

[1] BRASIL). Ministério da Saúde Estratégia e-Saúde para o Brasil. Brasília. 2014.

[2] SAS - Statistical Analysis System at North Carolina State University The Power to know. Data Mining From A to Z: How to Discover Insights and Drive Better Opportunities 2016.

[3] M.Prosperi, J. S. Min, J.Bian, and F. Modave, Big data hurdles in precision medicine and precision public health, BMC Medical Informatics and Decision Making(2018), 1-15.

[4] Sommerville, Software Engineering, 9th Edition, Adison – Welley, Boston, US, 2011.

Address for correspondence

Grace T M Dal Sasso. Campus Universitário s/n - 88040-900

Fpolis – SC. Brazil. Caixa Postal: 476

email:grace.sasso@ufsc.br or gracetmds@gmail.com

Phone: +55(48)999891409 or +55(48)37216495

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190543

Novel Analytics Framework for Universal Healthcare Insurance Claims Database

Jumpei Sato[a], Kazuo Goda[a], Masaru Kitsuregawa[a], Naoki Nakashima[b], Naohiro Mitsutake[c]

[a] *Institute of Industrial Science, The University of Tokyo, Meguro-ku, Tokyo, Japan,*
[b] *Medical Information Center, Kyushu University Hospital, Higashi-ku, Fukuoka, Japan,*
[c] *Institute for Health Economics and Policy, Minato-ku, Tokyo, Japan*

Abstract

Medical insurance claims are useful data to offer a big-picture view and insight of a nation-wide healthcare system. Yet, formal description of the logic to analyze the claims has not been established. So far, we proposed a description scheme of analytics logic over claims database. In this paper, we propose a novel analytics framework based on the description scheme. By showing a case study, we demonstrate the effectiveness of the framework.

Keywords:
Analytic Data, Insurance Claims Analysis, Logic.

Introduction

Japan's Ministry of Health, Labour and Welfare (MHLW) has continuously collected anonymized insurance claims data from all the insurers throughout Japan since April 2009 to build National Database of Health Insurance Claims and Specific Health Checkups of Japan (a.k.a. NDB) [1]. All medical services in Japan are basically covered under the universal health insurance system. The database potentially offers great opportunities to analyze all the medical services provided in the country. The database holds more than 14.8 billion claims as of March 2018 [1], which is one of the largest big data in the field of healthcare. To facilitate the intensive and extensive utilization of this database, we have formed a research team comprising researchers from both healthcare and informatics and built a home-grown analytics platform that accommodates 6-year claims data (provided by MHLW) to offer analytics services to healthcare researchers and practitioners [2].

So far, several analytics tools were developed to specific purposes [3]. Similar to big data analytics in other fields, most of the analytics on insurance claims database are basically *ad-hoc* and *wide-range*. Yet, a method to formally describe the analytics logic of insurance claims has not been established. Healthcare researchers and practitioners often have difficulties to discuss and share their analytics ideas since they have no common language or scheme to clearly describe the analytics logic. To solve this problem, we proposed a description scheme of analytics logic for claims data in our previous study [4]. In this paper, we present a novel analytics framework that we developed based on the description scheme. By presenting a typical case study that we observed on the prototype system, we demonstrate the effectiveness of the framework.

Methods

Scheme for describing analytics logics

So far, we have defined domain-specific data types and functional operator types that are typically utilized for insurance claims data processing. One can formally define an analytics logic as a functional operator or its combination by instantiating the defined data types and functional operator types. Suppose a typical analytics logic composed of three steps: (1) selecting claims from the database based on a certain condition, (2) performing a certain arithmetic calculation on the claim attributes, and (3) summarizing them in a spreadsheet. For such a specific logic, we have defined $f_{rez\_to\_tup}$ that inputs a claim set, applies selection, conversion and extraction operations and then outputs the results into a spreadsheet. In this way, we have finally defined six different functional operator types indicated in Figure 1, which are good enough for describing most of analytics logics that the healthcare reasearchers and practitioners need in our experience.

$$f_{rez\_to\_tup} : \mathbf{C} \xrightarrow{selection} \mathbf{C} \xrightarrow{conversion} \mathbf{C} \xrightarrow{extraction} \mathbf{R}$$
$$f_{rez\_to\_ep} : \mathbf{C} \xrightarrow{selection} \mathbf{C} \xrightarrow{conversion} \mathbf{C} \xrightarrow{sessionization} \mathbf{E}$$
$$f_{rezs\_to\_rez} : \mathbf{C} \times \mathbf{C} \xrightarrow{reduction} \mathbf{C}$$
$$f_{ep\_to\_tup} : \mathbf{E} \xrightarrow{selection} \mathbf{E} \xrightarrow{conversion} \mathbf{E} \xrightarrow{summarization} \mathbf{R}$$
$$f_{eps\_to\_ep} : \mathbf{E} \times \mathbf{E} \xrightarrow{reduction} \mathbf{E}$$
$$f_{tup\_to\_tup} : \mathbf{R} \xrightarrow{selection} \mathbf{R} \xrightarrow{sql-based-conversion} \mathbf{R}$$

Figure 1 – Typical fucntional operator types

Implementation of analytics framework

We have newly prototyped an analytics framework, which is capable of accepting an analytics logic that a user describes based on the above noted description scheme, compiling the logic into database queries, executing the queries and then answering an execution result to the user. The framework offers a visual user interface (UI) to help the user to easily and intuitively describe an analytics logic and interprete the execution result (Figure 2).

Case Study

We have demonstrated dozens of analytics on the framework. In this paper, we present one representative case (nation-wide medical expenditure analysis).

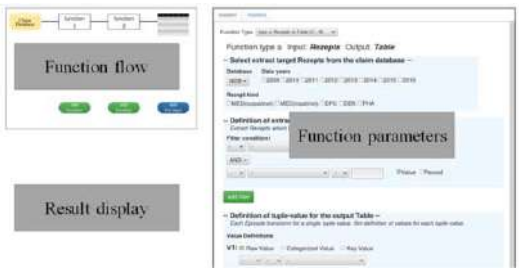


Figure 2 – User interface

Results

Figure 3 denotes an analytics logic defined for the study. Insurance claims are to be categorized by their medical service types: fee-for-service (MED_out-patinet, MED_in-patinet), diagnosis procedure combination (DPC), dental clinic (DEN) and pharmacy (PHA). The analytics logic is composed of two functional operator types: $f_{rez\_to\_tup}$ and $f_{tup\_to\_tup}$. (1) $f_{rez\_to\_tup}$ is defined to select only claims that were generated in 2014 and later (fiscal year based), extract patient ID, medical service type, patient age category, gender and medical expenditures and then summarize them into a spreadsheet. (2) $f_{tup\_to\_tup}$ is defined to aggregate average medical expenditures per capita for each age group, gender and medical service type. Figure 4 illustrates how a user can input the analytics logic into the framework and confirm its execution result. Figure 5 presents the analytics result of nation-scale medical expenditure breakdowns. This case study indicates that the proposed framework allows the user to perform a nation-scale analytics only by specifying the logic on the visual user interface without writing any programming language.

Function flow definition:

$$input \xrightarrow{f_{rez\_to\_tup}} R_1 \xrightarrow{f_{tup\_to\_tup}} output$$

Parameters of $f_{rez\_to\_tup}$**:**

selection:

$$e \mid (e \in C)(FiscalYear(e.RE_{<intervention\_year\_and\_month>}) = 2014)$$

conversion: *none*

extraction:

$$t.V_1 = e.RE_{<patient\_id>} \qquad t.V_2 = e.RE_{<age\_category>} \qquad t.V_3 = e.RE_{<gender>}$$

$$t.V_4 = \begin{cases} \text{'MED out-patient'} & \text{if } e.RE_{<receipt\_type>} = \text{'MED'} \\ & \quad \text{AND } e.RE_{<receipt\_kind>} \%2 = 0 \\ \text{'MED in-patient'} & \text{if } e.RE_{<receipt\_type>} = \text{'MED'} \\ & \quad \text{AND } e.RE_{<receipt\_kind>} \%2 = 1 \\ e.RE_{<receipt\_type>} & otherwise \end{cases}$$

$$t.V_5 = \sum_{r \in e.HO[]} \begin{cases} r_{<total\_points>} & \text{if } e.RE_{<receipt\_type>} \\ & \quad \in \{\text{'MED', 'DEN', 'PHA'}\} \\ & \quad \text{OR } (e.RE_{<receipt\_type>} = \text{'DPC'} \\ & \quad \text{AND } r_{<total\_point\_record\_flag>} = 1) \\ 0 & otherwise \end{cases}$$

Parameters of $f_{tup\_to\_tup}$**:**

selection: *none*

sql-based-conversion:

SELECT V_2, V_3, V_4, **SUM**(V_5)/**COUNT(DISTINCT** V_1**)**
GROUP BY V_2, V_3, V_4

Figure 3 – Example analytics logic

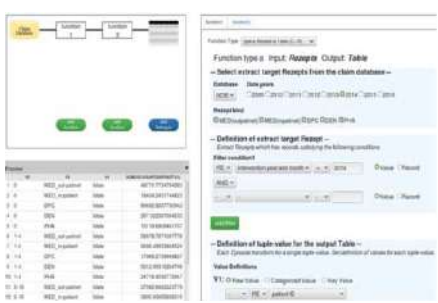


Figure 4 – User input and output

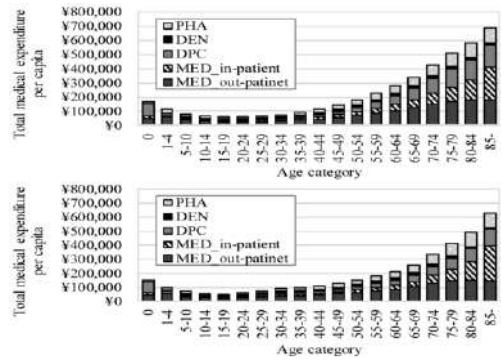


Figure 5 – Analytics result: male (upper), female (lower)

Conclusions

We have developed a novel analytics framework based on the formal description scheme that is useful to analyze nation-scale medical insurance claims database. The presented case study indicates that the framework helps the user to easily and intuitively describe an analytics logic and interprete the execution result.

Acknowledgements

This work has been in part supported by Health Labour Sciences Research Grant (Ministry of Health Labour and Welfare, Japan), Funding Program for World-Leading Innovative R&D on Science and Technology (Cabinet Office, Japan), Impulsing Paradigm Change through Disruptive Technologies Program (Cabinet Office, Japan) and ICT infrastructure establishment and implementation of artificial intelligence for clinical and medical research (Japan Agency for Medical Research and Development).

References

[1] Ministry of Health, Labour and Welfare, A role and analytics platform for NDB and care DB. https://www.mhlw.go.jp/file/05-Shingikai-12401000-Hokenkyoku-Soumuka/0000206673.pdf (Accessed on March 30, 2019) (Japanese).

[2] J. Sato, H. Yamada, K. Goda, M. Kitsuregawa, and N. Mitsutake, Enabling Patient Traceability Using Anonymized Personal Identifiers in Japanese Universal Health Insurance Claims Database, *AMIA 2019 Informatics summit* (2019) 345-352.

[3] Observational Health Data Sciences and Informatics, ATLAS. https://www.ohdsi.org/atlas-a-unified-interface-for-the-ohdsi-tools/ (Accessed on March 30, 2019).

[4] J. Sato, K. Goda, M. Kitsuregawa, and N. Mitsutake, A Preliminary Study on Description Scheme of Analytics Logic for Health Insurance Claims Data, *38th JCMI* (2018) 420-425. (Japanese).

Address for correspondence

Jumpei Sato. E-mail: jsato@tkl.iis.u-tokyo.ac.jp. Insititute of Industrial Science, The Univercity of Tokyo. 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505 Japan

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190544

Finding the Needle in the Hay Stack: An Open Architecture to Support Diagnosis of Undiagnosed Patients

Jannik Schaaf[a], Martin Boeker[b], Thomas Ganslandt[c], Christian Haverkamp[d], Tim Hermann[e], Dennis Kadioglu[a], Hans-Ulrich Prokosch[f], Thomas O.F. Wagner[g], Michael von Wagner[h], Johanna Schaefer[a], Martin Sedlmayr[i], Holger Storf[a]

[a] *Medical Informatics Group (MIG), University Hospital Frankfurt, Frankfurt, Germany*
[b] *Institute of Medical Biometry and Statistics, Medical Faculty and Medical Center – University of Freiburg, Freiburg, Germany*
[c] *Department of Biomedical Informatics, University Medicine Mannheim, Ruprecht-Karls-University Heidelberg, Mannheim, Germany*
[d] *IT-Department, University Hospital Freiburg, Freiburg, Germany*
[e] *Institute for Biometry and Medical Informatics, Otto-von-Guericke-University Magdeburg, Magdeburg, Germany*
[f] *Chair of Medical Informatics, Department of Medical Informatics, Biometrics and Epidemiology, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany*
[g] *Frankfurt Reference Center for Rare Diseases, University Hospital Frankfurt, Frankfurt, Germany*
[h] *Executive Department of Medical IT-Systems and Digitalization, University Hospital Frankfurt, Frankfurt, Germany*
[i] *Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine Technische Universität Dresden, Dresden, Germany*

Abstract

Clinical Decision Support Systems (CDSS) are promising to support physicians in finding the right diagnosis of patients with rare diseases (RD). The MIRACUM consortium, which includes ten university hospitals in Germany, will establish a diagnosis support system for RD. This system conducts a similarity analysis on distributed clinical data with the aim to identify similar patient cases at each MIRACUM site to offer the physician a hint to a possible diagnosis.

Keywords:

Rare Diseases, Clinical Decision Support, Data Science

Introduction

A big challenge with rare diseases (RD) is to find the correct diagnosis for a patient. A study by the EU showed that 25 % of the patients waited between 5 and 30 years for the correct diagnosis. In Europe, a disease is declared as “rare" if less than 5 out of 10.000 people are affected [1]. To tackle the problem of undiagnosed patients with RD, Clinical Decision Support Systems (CDSS) are promising. CDSSs provide clinically prepared and filtered information with the aim of achieving better health processes [2]. This paper focuses on a CDSS for RD based on distributed clinical data from ten university hospitals in Germany. The concept and the development of this system are part of the Medical Informatics for Research and Care for University Medicine consortium (MIRACUM), which is funded within the Medical Informatics Funding Scheme by the German Federal Ministry of Education and Research (BMBF) [3,4]. Within MIRACUM the university hospitals will establish Data Integration Centers (DIC) with the goal to improve collaborative research as well as clinical processes. A DIC will be established in the IT-infrastructure of each hospital which enables to exchange data among the partners in MIRACUM based on the principle of data federation. For query and analysis, the data of each hospital remains at the respective locations. To demonstrate the benefit of the evolving IT infrastructure, different use cases will be developed, including a use case about the diagnosis support for patients with RD [4]. In this use case, a similarity analysis on distributed clinical data is performed with the aim to identify similar patient cases at each hospital, which in turn can give the physician a hint to a possible diagnosis. In this paper we present a software-architecture based on distributed clinical data that can be devised in a CDSS for RD. The functionality of how similar patients are found in the data is not covered in this paper.

Methods

To perform a data analysis at each MIRACUM site, the data of each site needs to be harmonized. Harmonizing a very large amount of data from different previously established resources is a significant challenge [5]. To formally describe all data elements used for a similarity analysis, a meta data repository (MDR) is used, which is based on the international meta data standard ISO 11179 [6]. The description of data elements is moderated by clinical researchers and clinicians to define medical concepts based on meta data. All data in the DIC will be described at each site with their MDR. Used data elements will be mapped to a common dataset, which is available in the local databases [5]. The common dataset will be based on Common Data Models (CDM) such as the OMOP CDM, which includes a standard representation of common vocabularies for coding clinical concepts and allows to perform comparable analysis via different databases [7]. OMOP includes standardized vocabularies for representing data in the CDM

(e.g. SNOMED-CT) [8]. For data exchange between sites, HL7-FHIR-is used [9].

Results

The result of this work is a system architecture for a CDSS based on distributed clinical data (seen in Figure 1).

A physician formulates a query for similarity analysis to find similar patients. The web application DISERDIS (Diagnosis Support in Rare Diseases) will be available at each site. The application makes it possible to view the data of the undiagnosed patient and to perform a similarity analysis based on local data or other MIRACUM locations at one once. The patient data is stored in a OMOP database which is periodically updated with data from the respective hospital information system of the site. This involves the establishment of ETL-processes for mapping the data to the OMOP-CDM. When a similarity analysis is triggered at one site, a request is sent to the central MIRACUM search-broker which provides the request to the respective local FHIR server. The clinical hospital IT networks usually block external access to the data. Therefore, the search-broker is designed as a central request point allowing to forward the search request to the FHIR-Server which manages access to the data. The FHIR-Server is a REST API that can retrieve requests via corresponding FHIR resources. It retrieves the query from the search-broker and submits it to the similarity engine which performs the similarity analysis on the database. The data is available in the OMOP-CDM and must be transformed to the FHIR resources. The result is returned to the FHIR-Server which sends the result of the similarity analysis to the search-broker.

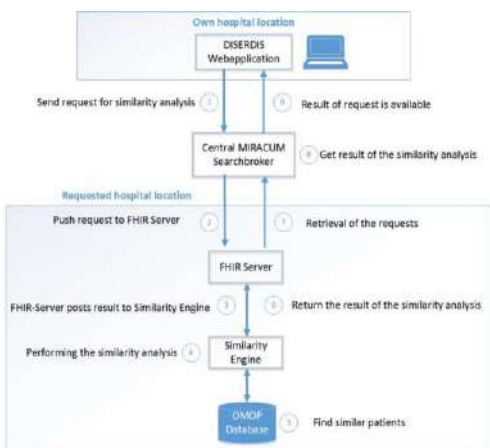


*Figure 1 – Detailed **A**rchitecture of **D**iagnosis **S**upport*

Conclusions

This paper demonstrates a concept of a system-architecture establishing a diagnosis support system for RD based on distributed clinical data.

Acknowledgements

MIRACUM is funded by the German Federal Ministry of Education and Research (BMBF) within the "Medical Informatics Funding Scheme" (FKZ 01ZZ1801A, 01ZZ1801B, 01ZZ1801C, 01ZZ1801E, 01ZZ1801H, 01ZZ1801L).

References

[1] R. Dragusin, C. Petu, C. Lioma, H. Jorgensen, I. Cox, L. Hansen, P. Ingwersen, and O. Winther, FindZebra: a search engine for rare diseases, *Int J Med Inform*. **82** (2013) 528–538.

[2] I. Sim, P. Gorman, R. Greenes, R. Haynes, B. Kaplan, H. Lehmann, and P. Tang, Clinical Decision Support Systems for the Practice of Evidence-based Medicine, *J Am Med Inform Assoc*. **8** (2001) 527–534.

[3] S. Gehring, and R. Eulenfeld, German Medical Informatics Initiative: Unlocking Data for Research and Health Care., *Methods Inf Med*. **57** (2018) 46–49.

[4] H. Prokosch, T. Acker, J. Bernarding, H. Binder, M. Boeker, M. Borries, P. Daumke, T. Ganslandt, J. Hesser, G. Höning, M. Neumaier, M. Marquardt, H. Lenz, H. Rothkötter, C. Schade-Brittinger, M. Sedlmayr, K. Sohrabi, and H. Storf, MIRACUM: Medical Informatics in Research and Care in University Medicine - A Large Data Sharing Network to Enhance Translational Research and Medical Care, *Methods Inf Med*. **57** (2018) 82–91.

[5] D. Kadioglu, B. Breil, M. Lablans, S. Mate, D. Schlue, H. Serve, H. Storf, F. Ückert, T. Wagner, P. Weingardt, and H. Prokosch, Samply MDR - A Metadata Repository and Its Application in Various Research Networks, *Stud Health Technol Inform*. (2018) 50–54.

[6] ISO/IEC 11179-1:2015, (2015). https://www.iso.org/standard/61932.html.

[7] C. Meier, L. Lang, P. Vormstein, R. Bieber, J. Bernarding, T. Hermann, C. Haverkamp, P. Horki, J. Laufer, G. Höning, H. Fritsch, K. Schüttler, T. Ganslandt, H. Prokosch, and M. Sedlmayr, Towards Implementation of OMOP in a German University Hospital Consortium., *Appl Clin Inform*. **9** (2018) 54–61.

[8] P. Stang, P. Ryan, J. Racoosin, J. Overhage, A. Hartzema, and C. Reich, Advancing the science for active surveillance: rationale and design for the observational medical outcomes partnership, *Ann Intern Med*. **153** (2010) 600–606.

[9] C. Mandel, D. Kreda, K. Mandl, I. Kohane, and R. Ramoni, SMART on FHIR: a standards-based, interoperable apps plattform for electronic health records, *J. Am. Med. Inform. Assoc.* **23** (2016) 899–908.

Address for correspondence

Jannik Schaaf
Medical Informatics Group
University Hospital Frankfurt
Theodor-Stern Kai 7, 60590
Frankfurt am Main, Germany
E-mail: schaaf@med.uni-frankfurt.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190545

Use of Nursing Interventions as an Indicator to Assess the Workload of Nurses in a Tertiary Care Surgical Ward Setting in Sri Lanka

S.N Silva[a], K.M. Mulleriyawa[b], S.R. Tissera[c], P.L. Jayawardene[d], K.A.P.T. Karunarathne[e]

[a] *Post Graduate Institute of Medicine, University of Colombo, Sri Lanka,*
[b]*General Surgery Section, National Hospital of Sri Lanka, Colombo, Sri Lanka*
[c] *Nursing Informatics Special Interest Group, IMIA NI SIG Country office, Sri Lanka*
[d]*School of Nursing, International Institute of Health Sciences, Sri Lanka*
[e]*School of Nursing, International Institute of Health Sciences, Sri Lanka*

Abstract

This research looked into the possibility of using the number of Nursing Interventions as a measure of Nurses workload in a Surgical ward setting. The quantification of the number of interventions was done using the Nursing Interventions Classification (NIC). This strategy could generate the average number of interventions required per patient and per shift as well as could assess the variation of the number of interventions under different factors.

Keywords:

Nursing Informatics, NIC Terminology, Workload

Introduction

Assessment of Nurses workload has been a challenging task in Sri Lanka as well as in hospitals worldwide. Nurses often complain of being overworked and declair work-related dissatisfaction. However, allocating nurses to wards as well as the assessmsnet of workload are done using indirect and crude measurements in Sri Lanka. The proper assessment of workload can have an added advantage of being able to quantify the need of physical resources necessary to carry out Nursing duties, in addition to better Human Resource Management.

The main share of work done by nurses in patient care is via Nursing Interventions. Therefore it is interesting to see how Nursing Interventions could be used as a direct scientific measure of quantifying the Nurses workload [1]. A standard approach in this regard would be to use the International Classification of Nursing Practices (ICNP), where Nursing Interventions are classified as NIC – Nursing Interventions Classification, along with the North American Nursing Diagnosis Association (NANDA) and Nursing Outcomes Classifications (NOC) [2].

However the use of a system such as ICNP is quite challenging in a setting like Sri Lanka, where the standard Nursing process or the use of care plans are not practiced in routine patient care. Therefore firstly the use of nursing care plans has to be established, where Nursing Interventions would then be a standard part of patient care. This should be done without disturbing the routine patientcare in a hospital setting [3]. For this purpose a General Surgical ward would be ideal out of the main specialities, because it is less complicated than a Medical or a Pediatric setup and will have a better variety than an Obstetrics and Gynaecology setting.

The main research question in this study was therefore to assess whether quantification of Nursing Interventions could be used as an indicator of Nurses workload.

The objectives of this study was therefore to assess the number of Nursing Interventions required per patient in a Surigal ward at the National Hospital of Sri Lanka and to describe the variability of it depending on some selected factors.

Methods

The project was carried out in the Surgical section of the National Hospital of Sri Lanka, the largest teaching hospital in the country. The hospital gets direct admissions as well as referals from all over the country. Therefore it has the widest range of General Surgical cases in a given ward in Sri Lanka. As the first phase of the study, Nursing care plans were implemented in 6 typical surgical wards over a period of 4 weeks in January 2018. The nurses were given a training by experienced tutors 2 weeks before the implementation. Special documentation files were provided with stuctured templates for care plans. Additional data such as medical complications and duration of stay were also collected. Reference documents were also provided to each ward. Third year Nursing students who are in their Surgical placement were used to assist the nurses during busy shifts.

Out of the 6 wards where the first phases were completed, two were selected where the implementation was most successful. Out of the 4 weeks, the best week was selected where nearly all of the presented patients had a care plan prepared for them. This resulted in a total of 131 care plans. The care plans were then coded according to the North American Nursing Diagnosis Association (NANDA) Taxonomy 1, Nursing Interventions Classification (NIC), and Nursing Outcomes Classifications (NOC) by 3 trained nurses who weren't involved in data collection. The paper based coded data were analysed using descriptive and inferential statistics.

Results

There were 131 completed and coded care plans for 131 patients. The mean age of a patient was 48.02 years. 45 had Diabetes, 38 had Hypertension and 14 had Ischemic Heart Disease. 81 were Emergency (Casualty) admissions and 49 were Elective admissions. These 131 patients had a total of 93 Nursing Diagnoses, 103 Nursing Observations and 1822 Nursing interventions.

The average number of Nursing Interventions performed per patient were 13.87 (SD 4.52). The table below shows the number of interventions categorized in to 5 groups and the number of patients who belonged to those categories.

Table 1– Number of patients under each Interventions group

Number of Interventions Group	Number of Patients	Percentage
Less than 5	3	2.3%
6-10	29	22.1%
11-15	55	42%
16-20	33	25.2%
More than 21	11	8.4%

The Average number of Nursing Interventions were most for those who had Diabetes at 15.25 on average and were 14.5 for those with Hypertension and 14.43 for those with Ischaemic Heart Disease. Those with Diabetes had a statistically significantly different mean number of Nursing Interventions than for those who did not (p=0.009). There seemed to be no statistically significant difference in the number of Interventions required by those with Hypertension compared to those without (p=0.31) and for those with Ischaemic Heart Disease, compared to those without (p=0,627).

Further, there seemed to be no statistically significant difference between the number of Nursing Interventions required by those admitted as Emergency admissions compared to the Elective admissions (p=0.121).

Increased length of stay does not necessarily seem to increase the average number of interventions required.

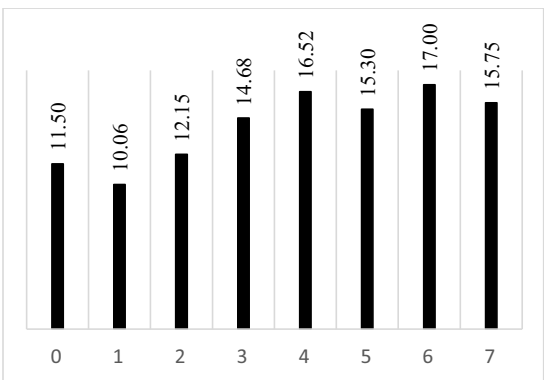


Figure 1– Average Number of Nursing Interventions required ccording to the Length of Stay in the Hospital

In total, there were 1822 Nursing Interventions practiced in these two wards during these two days. A rough assessment of the workload per nurse per shift could be made using this number. A ward would typically have 13 six hour shifts per day (6 in the morning, 5 in the afternoon and 2 in the night). Therefore the total number of shifts would be 182. This means per 6 hour shift a nurse would have to perform approximately at least 10 Nursing Interventions.

Discussion

The results seem to indicate that the Nursing Interventions Classification could be used well for approximating the Nursing work load under various situations.

Out of the Medical conditions concerned, Diabetes Mellitus seems to have a significant impact on the Number of Interventions required for a Surgical Patient rather than Hypertension or Ischaemic Heart Disease. However this situation is probably applicable only to a Surgical ward setting.

In a Medical ward, other co-morbidities could increase the number of Nursing Interventions required. Emergency admissions doesn't seem to require more number of interventions than routine elective admissions. Therefore these results bring to question some of the parameters used when deciding the number of shifts and when allocating staff in Surgical wards. Further evaluation of other factors using 'the Number of Nursing Interventions required' as the decisive indicator, could therefore reveal useful inteligence for Nurse Managers to allocate their staff in a more productive manner.

Similar studies could be done in all main sections of a Hospital such as Medical, Obstetrics and Gynecology, Pediatrics and Psychiatry. Unique patterns and factors that influence the number of Nursing Interventions required could then be identified. Such factors then should be considered alongside the different patient presentation patterns and seasonal variations before optimal staffing and shift allocations can be decided.

There are a few limitations of this study. The period concerned is not adequate to cover some patients who are in Surgical Wards. Therefore data collected over at least 4 weeks would produce far more accurate results. This study did not consider the number of Nursing Interventions according to the medical diagnosis, under a standard nomenclature or a classification. Such an addition would provide a more direct predictive ability of the work load required as per the seasonal pattners of patient presentations.

Conclusions

The study provides a very promising insight to scientifically validating and predicting the Nursing Work load using Nursing Interventions Classification. The use of "Number of Nursing Interventions Required" as the main indicator, with the consideration of its variation according to other confounding factors could provide a more scientific guideline to allocate Nursing Staff to meet the varying demands of each Surgical ward.

Acknowledgements

The Authors would like to Acknowledge the assistance provided by the Nursing staff and the Administration of the Surgical wards of the National Hospital of Sri Lanka and the School of Nursing of the International Institute of Health Sciences, Sri Lanka.

References

[1] J.F. Possari, R.R. Gaidzinski, A.F. Lima, F.M. Fugulin, and T.H. Herdman, Use of the nursing intervention classification for identifying the workload of a nursing team in a surgical center, *Revista latino-americana de enfermagem* **23** (2015), 781-788.

[2] M. Johnson, S. Moorhead, G. Bulechek, H. Butcher, M. Mass, and E. Swanson, *NOC and NIC linkages to NANDA-I and clinical conditions, Third Edition*, Elsevier Mosby, Maryland Heights, MO, 2012.

[3] V. Zamanzadeh, L. Valizadeh, F.J. Tabrizi, M. Behshid, and M. Lotfi, Challenges associated with the implementation of the nursing process: A systematic review, *Iranian journal of nursing and midwifery research* **20** (2015), 411-419.

Address for correspondence

Corresponding author : Dr. S. Nishan Silva, email : snishansilva@gmail.com.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190546

Implementation of a Terminology Server with SNOMED CT in Graph Databases

Elizabeth Silva Layes[a], Marcelo Bondarenco[a] , Daniel Machiavello[a], Fabián Frola[b], Martín Lemos[b]

[a] *Health Informatics Department, Sociedad Médico Quirúrgica de Salto, Salto, Uruguay,*
[b] *Development Department, Sociedad Médico Quirúrgica de Salto, Salto, Uruguay*

Abstract

This paper described the implementation of a terminology server performed at a health institution in Uruguay, whose architecture is based on SNOMED CT using graph databases (NoSQL). The aim of this development was to create an intuitive terminological service, making the most of SNOMED CT's ontology, and which can be used from a clinical, statistical, management, decision support and research point of view, among others, with good performance.

Keywords:
Systematized Nomenclature of Medicine; Vocabulary, Controlled; Databases, Factual

Introduction

Terminological services (TS) are essential components in the development of an Electronic Health Record (EHR), since their implementation strengthens the semantic quality of the data recorded during the care act. These services are fundamental tools for the management of controlled vocabularies in different instances of clinical events, whether for use in statistical data, management, decision support and research, or other areas. SNOMED CT is a global clinical terminology that covers a wide range of specialties, disciplines, and medical and care requirements. With SNOMED CT, clinical information is recorded with identifiers that refer to formally defined concepts as part of the terminology [1]. SNOMED CT allows the recording of clinical information with appropriate levels of detail through the use of relevant clinical concepts. This terminology structure allows information to be entered using synonyms that adapt to local preferences, while recording information in a consistent and comparable manner. SNOMED CT is also an ontology, which contains a poly-hierarchy model, meaning that a code can be grouped into different categories, forming directed acyclic graphs (DAG). Nonetheless, these graphs are difficult to implement in their entirety using relational or object-oriented databases, resulting in a loss of a wealth of information. SNOMED CT's continued expansion will make it more complex and model consistency will be more difficult to assure. Moreover, consumers of data will increasingly demand improvements to query functionality to accommodate additional granularity of clinical concepts without sacrificing speed. The above information and comparative retrieval characteristics of relational databases, triplestores, and general graph databases [9] determined the use of a graph database.

The emergence of Big Data has spurred development of new technologies and platforms, for example, Neo4j [2] which is a high-performance NoSQL database. We chose Neo4j because it had the best score of the graph databases [8]. This platform has already been used in different industries such as financial [3], social networks [4], in health [5], among others. Neo4j is a graphics database whose basic structure is composed of nodes, relationships and attributes. The nodes are designated as starting nodes and termination nodes, and two such nodes are connected by a relationship (edge). This database allows the representation of the SNOMED CT ontology, without loss of information, which makes it one of the most suitable platforms for this type of work. SNOMED CT has already been used in different works in conjunction with Neo4j [6] [7], but it has never been used as a terminology server to support clinical history in real time. Concretely, the objective of this work is the implementation of a terminology server in a health institution in Uruguay, using SNOMED CT on a graph-oriented database, in order to create an intuitive TS, making maximum use of the characteristics of the SNOMED CT ontology, and also to ensure usability from a clinical point of view.

Methods

In the first step, a graph database containing the SNOMED CT concept model was created. The implementation of the TS was done using the RF2 files of the International Release, Spanish Edition and the Uruguay extension. Currently, the TS is running with the Release International 2018-04-34. In terms of the graph database, the Neo4j Desktop Version 1.0.22 is used. Neo4j is a graph database platform based on Java that supports transactions with ACID (atomicity, coherence, isolation, and durability) properties. The content of SNOMED CT was extracted from the RF2 files and loaded into Neo4j using a series of scripts that were written in Python. The scripts extracted the following information from the RF2 files to create the nodes: "id", "term", and descType. For the creation of relationships, the following data is used: "sourceId", "destinationId", "relLabel", "typeId", "term", "descType." The nodes were created for all the basic SNOMED CT concepts, and the edges were created for all IS_A relationships, which is a relationship from a child concept to a parent concept. All nodes contain every aspect of a SNOMED CT concept such as concept ID, status, and the description of the concept. Queries on the terminology server are made using the declarative language of Neo4j (Cypher). In this language the query defined makes a lexicographical search using the descriptions (without accents and in minuscule) stored in the nodes, then a recursive search is made through the edges, using the IS_A relations (see Figure 1) but in the opposite direction (from parent to child). Finally, as a result of the query, the description and the ID of the obtained concepts is returned.

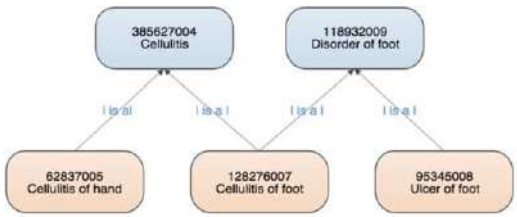


Figure 1 - (SNOMED IS_A Relationship) Obtained from https://confluence.ihtsdotools.org/

The Terminology Server is accessed by clients through a Glassfish application server that uses SOAP or REST protocols depending on the web services invoked. The server consults a Lexicon (under development) in the first instance (a linguistic tool with the morphological variations and grammatical uses of words), subsequently accessing the Terminology Server the Bolt protocol (see Figure 2).

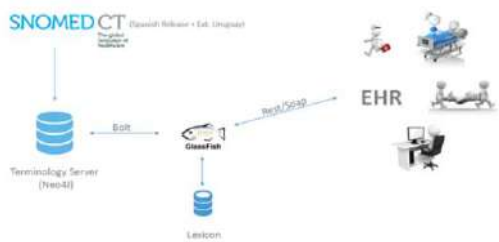


Figure 2 - Architecture

Figure 3 below shows the record form of the EHR, in which the TS is accessed through a web service to obtain terms associated with findings that will be linked to the patient as personal or family background. At present the TS is in production for use by internal medicine physicians.

Figure 3 - Background Record Form of the EHR

Terminology Server Features

In the present example (Figure 3), the query to the TS is made from the EHR and includes the following features:

- Analysis of the text entered (diabetes mellitus) by the lexicon.
- Search the set obtained from the lexicon in the database.
- Return of the found terms and the associated coding that was requested when making the query.

Conclusions

The development of a TS with a graph database architecture using SNOMED CT as the basis of the data model allows more comprehensive queries navigating through the database in a hierarchical manner than a relational database. We can navigate through the ontology choosing the levels of specificity in the terms retrievals, and the integration of extensions for use in a particular country can be included thanks to the characteristics of SNOMED CT. It is also noteworthy that the response times to the information are adequate, with an average of 500 milliseconds for the visualization of the collection of terms in the user interface of the EHR.

References

[1] "SNOMED CT Implementation".https://confluence.ihtsdotools.org/display/DOCTIG/2.+SNOMED+CT+Implementation. [accessed August 5, 2018].
[2] Neo4j. https://neo4j.com/. [accessed August 5, 2018].
[3] G.C.G. Van Erven, M. H., R.N. Carvalho, Á. Rocha, A. Correia, H. Adeli, L. Reis, S. Costanzo. "Detecting Evidence of Fraud in the Brazilian Government Using Graph Databases." Recent Advances in Information Systems and Technologies, WorldCIST 2017, 570.
[4] Nettleton, D. "Data Mining of Social Networks " *Computer Science Review*, **7** (2013), 1–34.
[5] Rajkumar Tekchandani, R. B., Maninder Singh. "Code clone genealogy detection on e-health system using Hadoop." *Computers & Electrical Engineering* **61** (2017), 15-30.
[6] W. Scott Campbell, J. P., James C. McClay, Praveen Rao, Dhundy Bastola, James R. Campbell. "An alternative database approach for management of SNOMED CT and improved patient data queries." *Journal of Biomedical Informatics* **57** (2015), 350-357.
[7] W. Scott Campbell, J. P., James R. Campbell (2015). "Graph database approach to management and use of SNOMED CT encoded clinical data." https://confluence.ihtsdotools.org/download/attachments/12780169/5_GraphDB_2015.pdf.
[8] "DB-engines ranking of graph dbms.", http://db-engines.com/en/ranking/graph+dbms. [accessed August 5, 2018].
[9] D.R. Schlegel, J. P. B., P.L. Elkin. "Comparing small graph retrieval performance for ontology concepts in medical texts." VLDB Workshop on Big Graphs Online Querying, Springer (2015), 32–44.

Address for correspondence

Corresponding author: Elizabeth Silva. Email: elizabeth.silva@smqs.com.uy
Corresponding author: Marcelo Bondarenco. Email: mbondarenco@smqs.com.uy

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190547

Named Entity Recognition in Prehospital Trauma Care

Greg M. Silverman[a],[b] Elizabeth A. Lindemann[c], Geetanjali Rajamani[d], Raymond L. Finzel[e], Reed McEwan[a], Benjamin C. Knoll[b], Serguei Pakhomov[e], Genevieve B. Melton[b,c], Christopher J. Tignanelli[b,c,f]

[a] Academic Health Center – Information Systems, University of Minnesota, Minneapolis, Minnesota, USA,
[b] Institute for Health Informatics, University of Minnesota, Minneapolis, Minnesota, USA
[c] Department of Surgery, University of Minnesota, Minneapolis, Minnesota, USA
[d] Stanford University, Stanford, CA, USA
[e] College of Pharmacy, University of Minnesota, Minneapolis, Minnesota, USA
[f] Department of Surgery, North Memorial Health Hospital, Robbinsdale, Minnesota, USA

Abstract

Natural language processing (NLP) methods would improve outcomes in the area of prehospital Emergency Medical Services (EMS) data collection and abstraction. This study evaluated off-the-shelf solutions for automating labelling of clinically relevant data from EMS reports. A qualitative approach for choosing the best possible ensemble of pretrained NLP systems was developed and validated along with a feature using word embeddings to test phrase synonymy. The ensemble showed increased performance over individual systems.

Keywords:

Emergency medical services, Labeling, Natural language processing

Introduction

More people die each year from trauma (10% of all deaths) than from malaria, tuberculosis and HIV/AIDs combined [1]. Inappropriate prehospital care is one of the largest contributors to preventable trauma mortality in the United States [2]. The use of natural language processing (NLP) named entity (NE) recognition (NER) for performance monitoring and quality improvement is a novel approach to bridge this gap. Early studies published in 2019 have begun utilizing NLP for prehospital stroke notes [3]. However, NLP for prehospital trauma notes represents a new domain of inquiry.

NLP techniques hold promise for circumventing current limitations of discrete element documentation of EMS reports by reducing the effort of manual data abstraction [4]. The efficacy of how existing systems can best be utilized for notes with specific sublanguage characteristics of prehospital trauma is not well characterized and is the subject of this study.

We leveraged a small corpus of EMS reports of motor vehicle collisions (MVC) to develop an objective measure for creating an optimized NLP ensemble as discussed by Finley, et al. [5]. We also used word embeddings as a feature, as discussed by Turian, et al. [6]; and applied methods as discussed by Pakhomov, et al. [7], and Meng and Morioka [8] to use this feature as a test for phrase synonymy.

Methods

This study was a pilot to classify clinically relevant phrases related to prehospital trauma care using NER. Results were derived through an evaluation of four clinical NLP annotation systems, on their own, and as an optimized ensemble.

Data Sources

This study utilized de-identified North Memorial Health Hospital prehospital EMS MVC reports.

Creation of a Gold Standard Corpus

The development of a gold standard corpus consisted of two main parts: (1) schema creation, and (2) manual text annotation. 37 entities were identified based on clinical guidelines and the NEMSIS 3 standards [9] and iteratively incorporated into an annotation schema. Three trained annotators individually annotated 25 reports to establish inter-rater agreement (0.89 kappa, 99% agreement). Following this step, the remaining reports used for this study were manually annotated.

Creation of System Generated Annotations

Given the paucity of manual annotations for use in supervised training of a statistical model, we utilized the Artifact Discovery and Preparation Toolkit (NLP-ADAPT) [10], which included the clinical NLP annotator systems: cTAKES, CLAMP, BioMediCUS and MetaMap [11:14], to annotate our corpus of EMS MVC reports.

Evaluation Methods

We partitioned the gold standard annotations into a set of 10 reports to determine the best-at-task system annotation types. The remaining 112 reports were used for evaluation of the selected best-at-task system annotation types.

Matches between manual and system annotated reports were determined using a relaxed rule as noted by Finley, et al. [7]. Precision, recall, and F_1 score for each entity and system annotation pairing were calculated based on matches.

Best-at-task Evaluation

To compare NLP systems with respect to NE capture, we ranked each system annotation types for each entity using the three measures of NER performance shown in Figure 1. The geometric mean of the rankings was calculated to classify the best-at-task system annotation type. The annotation type with the lowest geometric mean (GM) was deemed best-at-task.

$$\frac{TP}{FN} \qquad F_1 score \qquad \frac{TP}{\sqrt{n_{sys}}}$$

Figure 1– NER Performance Measures; Abbreviations: TP, true positive; FN, false negative; n_{sys}, total system annotations

For validation of best-at-task systems, we analyzed the entity Procedure Indication using the remaining 112 EMS reports by comparing them to their corresponding manually annotated reports. We then combined best-at-task systems as an ensemble

to evaluate performance and further test how a word2phrase model as an additional feature affected recall [6:8].

A threshold of cosine distance of 0.5 was chosen after qualitatively evaluating several terms (e.g., "unresponsive," "unconscious," "agonal," and "tachycardic") and their resultant set when processed through the word2vec distance function [7]. Using string alignment methods [8], we used the Levenshtein edit distance (LD) to estimate degree of synonymy by identifying system and manual annotation match pairs on the resultant set with the lowest LD value.

Results

42 manual annotations pertained to the Procedure Indication entity in the set of 10 notes. Three best-at-task system annotation types were identified for this entity (Table 1).

Table 1-Best-at-task Annotation Types for Procedure Indication; Abbreviation: GM, Geometric Mean

System/Type	F_1	Precision	Recall	GM
CLAMP/Sentence	0.03	0.01	0.50	1.59
cTAKES/Sentence	0.02	0.01	0.52	2.15
MetaMap/Phrase	0.01	0.00	0.60	2.30

931 manual annotations pertaining to Procedure Indication were noted in the 112 notes. Individually, the three systems performed similarly with respect to recall and precision compared to the gold standard. The union ensemble resulted in significant improved coverage (87%) (Table 2).

Table 2- Best-at-task Validation for Procedure Indication Compared to Gold Standard with Final Rank

System/Type	F_1	Precision	Recall	Rank
CLAMP/Sentence	0.04	0.02	0.53	1
cTAKES/Sentence	0.03	0.02	0.54	2
MetaMap/Phrase	0.02	0.01	0.54	4
Ensemble	0.11	0.06	0.87	

The top 2 best-at-task rankings were consistent during validation (Rank). As anticipated, the ensemble performed very well with respect to recall, but precision was still poor due to a high false positive rate. Also, the ensemble did not account for matching on synonymous phrases. For example, patients that are unconscious and have agonal respirations meet the procedural indication for intubation. Using our word2phrase resultant set, phrases identified in system and manual annotations were matched for synonymy (Table 3). We were able to identify 93% (104 of 112 notes) coverage (match) with mean LD value of 2.5 (range 0-19).

Table 3- Word2phrase and System Annotations; Abbreviation: w2p, word2phrase

Note	w2p phrases	w2p common token	Synonymous best-at-task annotation
1	agonal	shallow	tachypnic shallow
2	tachycardic	sbp	80's sbp.
3	tachycardic	lungs	lungs clear bilat
4	unconscious	scene	alert on-scene

Conclusions

The present work represents one of the earliest NLP studies conducted in the prehospital trauma domain. Here we describe an approach to create an optimized NLP ensemble that when supplemented with a word2phrase model allows for over 90% NE capture. While these results are encouraging, the development of an extensive prehospital trauma corpus is paramount to facilitate development of models with improved precision, and further validation and extension of our methods.

Acknowledgements

NIH NCATS UL1TR002494 and U01TR002062, NIGMS R01GM120079, and AHRQ R01HS022085 and R01HS024532.

References

[1] Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. https://www.cdc.gov/injury/wisqars/overview/key_data.html. Accessed on December 10, 2017.

[2] NHTSA Fatality Analysis Reporting System (FARS).

[3] R. Garg, C.T. Richards, A. Naidech, S. Prabhakaran, Predicting Cincinnati Prehospital Stroke Scale Components in Emergency Medical Services Patient Care Reports Using NLP and Machine Learning, *Stroke* **50** (2019), 296.

[4] American College of Surgeons Committee on Trauma. Resources for optimal care of the injured patient. The Committee, Chicago, IL, 2014.

[5] G.P. Finley, H. Liu, H. Xu, G.B. Melton, S.V.S. Pakhomov, Using ensembles of NLP engines without a common type system to improve abbreviation disambiguation, *AMIA Podium Abstract*, 2017.

[6] J. Turian, L. Ratinov, Y. Bengio, Word representations: A simple and general method for semi-supervised learning, ACL, *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 2010.

[7] S.V.S. Pakhomov, G.P. Finley, R. McEwan, Y. Wang, and G.B. Melton, Corpus domain effects on distributional semantic modeling of medical terms, *Bioinformatics* **32** (2016), 3635-3644.

[8] F. Meng, C. Morioka, Automating the generation of lexical patterns for processing free text in clinical documents, *J Am Med Inform Assoc* **5** (2015), 980-986.

[9] NEMSIS Data Dictionary. NHTSA v3.4.0 https://nemsis.org/technical-resources/version-3/version-3-data-dictionaries/ Accessed on May 1, 2019.

[10] R.L. Finzel, NLP-ADAPT. https://github.com/nlpie/nlp-adapt. Accessed on October 21, 2018.

[11] BioMedICUS. https://github.com/nlpie/biomedicus Accessed on November 1, 2018.

[12] MetaMap. https://metamap.nlm.nih.gov. Accessed on November 1, 2018.

[13] cTAKES. http://ctakes.apache.org. Accessed on November 1, 2018.

[14] CLAMP. https://clamp.uth.edu. Accessed on November 1, 2018.

Address for correspondence

Christopher Tignanelli, M.D.
Email: ctignane@umn.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190548

Predicting Disease-Free Lung Cancer Survival Using Patient Reported Outcome (PRO) Measurements with Comparisons of Five Machine Learning Techniques (MLT)

Jin-ah Sim[a], Young Ho Yun*[b]

[a] *Cancer Research Institute, Seoul National University College of Medicine, Seoul, Korea*
[b] *Department of Biomedical Informatics, Seoul National University College of Medicine, Seoul, Korea*

Abstract

The study was to develop the lung cancer patients' prediction model for predicting 5-year survival after completion of treatment by using Machine Learning Technology (MLT), adding patient reporting (PRO) measurements of lung cancer survivors to a variety of clinical parameters. Finally, the survival prediction models with the addition of lung cancer survivors' PRO measurements to the well-known clinical variables, based on diverse MLT, improved the predictive performance that explains 5-year disease-free lung cancer survival.

Keywords:

Machine Learning, Cancer Survivors, Lung Neoplasms

Introduction

As the number of cancer survivor increases [1, 2], one of the major challenges in surviving lung cancer is to predict the survival of lung cancer after the completion of treatment based on diverse patients' information.

In general, statistical approaches focused on inferring the characteristics of a population from sample data[3], MLTs are more suitable to develop the prediction model with dozens of parameters when more prognostic variables are included because standard statistics do not generally work in this situation [4].

In fact, many recent studies have suggested that patient-reported outcome (PRO) measurements can provide distinct prognostic information [5, 6], PROs may be utilized as predictive tools for cancer survival. Even a variety of prediction models for cancer mortality have been developed in the clinical setting based on the MLTs [7-9], there were fewer studies developing survival prediction model with PRO factors based on MLTs.

Therefore, we evaluated the predictive performances of the developed mathematical models, multiple survival prediction models with the combination of lung cancer survivors' well-known clinical variables and PRO measurements, based on diverse MLTs compared with the well-known traditional prediction model only constructed based on the clinical and demographic variables.

Methods

Data Acquisition

For this supervised learning, regular follow-ups were undertaken for the patients based on each hospital's registry after the completion of treatment after surgery. Among finally contacted 836 patients, excluded 27 subjects whose survival status was censored until December 31, 2011. Thus, a total of 809 patients were included in this study.

Candidate Variables for Predictive Model

The candidate variables for our model included demographics (age and sex), socio-economic status (marital status, educational level, monthly family income), and clinical data (cancer stage, local invasion of a tumor, regional lymph node metastasis, recurrence, number of comorbidities, treatment type, time since diagnosis). In addition, patient lifestyle such as Body Mass Index (BMI), and metabolic equivalents of task (MET)-hours per week for Physical activity (PA) were included. Baseline PRO measurements were taken in this modeling process.

Data Preprocessing

As most machine learning algorithms strictly induce knowledge from data, the quality of the knowledge extracted is largely determined by the quality of the underlying data[10]. Therefore, we first attempted to apply the missing values imputation based on some important candidate information available in the data set.

Then, we grouped the candidate prognostic factors (variables form socio-demographic, clinical factors and PRO factors) for the modeling process. Before the model construction, we elected over-sampling producer to reduce the error costs for the whole imbalanced data.[11] Therefore, we balanced the number of 'dead' and 'alive' cases by over-sampling 617 'dead' cases to make equivalent 713 'alive' and 'dead' cases.

Finally, data splitting was employed to avoid over-fitting in the model and to derive reliable estimates of the model performance.[12, 13] The holdout method randomly splits the whole data sample into two mutually exclusive training (80%) and testing (20%) sets. The training set was utilized to generate the prediction model and the remaining data was employed as a testing set to evaluate the model's predictive performance. We compared demographic and clinical characteristics between train and test set using Chi-square and t-test.

Machine Learning Algorithms

Five supervised MLTs based classification model were employed to build each multivariate model to predict 5-year survival rates for Korean lung cancer survivors in training set. Decision Tree (DT), Logistic Regression (LR), Random Forest (RF) and Ensemble learnings such as Bagging or AdaBoost were selected for the prediction modeling process. The performance of the derived predictive models based on MLTs were also internally validated by 5-fold cross-validation.

Model Validation and Comparison

Each of predictive performance was assessed by the area (AUC) under the receiver operating characteristic curve (ROC). AUCs with 95% Confidence Interval (CI) were calculated for performance comparison between proposed five MLT based models.

Results

Data proportions after data up-sampling, splitting

After missing imputation was conducted, not statistically significant difference between the training and test data set were found.

Comparisons of MLT based models' performance

Among the overall model performances, DT showed the lowest performance, RF showed the highest in accuracy (76.30% and 82.30%, respectively), while DT showed the lowest AUC and RF showed the highest AUC in models (0.800 and 0.918 respectively).

Conclusions

In conclusion, we found that the same proposed feature set can be applied into ensemble MLT algorithms (particularly random forest or Adaboost algorithms) to predict disease-free lung cancer survival.

Table 1. Performance comparison of data mining algorithms based on the MLTs (cross-validation)

Feature Set	Machine learning algorithm	Train Accuracy (N=912)	Test Accuracy (M=286)
1	DT	0.625	0.692
	LR	0.657	0.632
	Bagging	0.655	0.706
	RF	0.675	0.692
	AdaBoost	0.642	0.713
2	DT	0.758	0.745
	LR	0.814	0.825
	Bagging	0.783	0.773
	RF	0.918	0.941
	AdaBoost	0.932	0.948

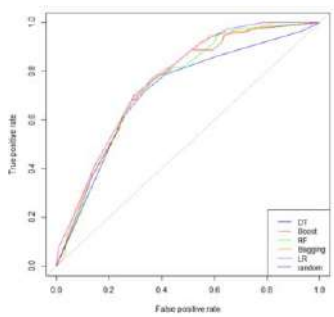


Figure 1. AUC curves comparison of models based on five MLTs using validation set (Model from the feature set 1)

Aknowledgements

This work supported by a grant of the National Cancer Center (NCC-0710410) and National Research Foundation (2016H1A2A1907839, 2018R1D1A1B07050704).

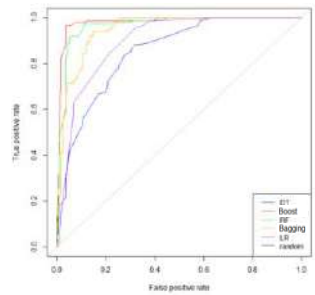


Figure 2. AUC curves comparison of models based on five MLTs using validation set (Model from the feature set 2)

References.

[1] Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2015.* CA: a cancer journal for clinicians, 2015. **65**(1): p. 5-29.

[2] Yun, Y.H., et al., *Needs regarding care and factors associated with unmet needs in disease-free survivors of surgically treated lung cancer.* Ann Oncol, 2013. **24**(6): p. 1552-9.

[3] Wang, P., Y. Li, and C.K. Reddy, *Machine learning for survival analysis: A survey.* arXiv preprint arXiv:1708.04649, 2017.

[4] Cruz, J.A. and D.S. Wishart, *Applications of Machine Learning in Cancer Prediction and Prognosis.* Cancer Informatics, 2006. **2**: p. 59-77.

[5] Gotay, C.C., et al., *The prognostic significance of patient-reported outcomes in cancer clinical trials.* J Clin Oncol, 2008. **26**(8): p. 1355-63.

[6] Montazeri, A., *Quality of life data as prognostic indicators of survival in cancer patients: an overview of the literature from 1982 to 2008.* Health Qual Life Outcomes, 2009. **7**: p. 102.

[7] Svensson, C.-M., R. Hübler, and M.T. Figge, *Automated classification of circulating tumor cells and the impact of interobsever variability on classifier training and performance.* Journal of immunology research, 2015. **2015**.

[8] Montazeri, M., et al., *Machine learning models in breast cancer survival prediction.* Technol Health Care, 2016. **24**(1): p. 31-42.

[9] Gao, P., et al., *Which is a more accurate predictor in colorectal survival analysis? Nine data mining algorithms vs. the TNM staging system.* PLoS One, 2012. **7**(7): p. e42015.

[10] Karim, M.N., et al., *Missing Value Imputation Improves Mortality Risk Prediction Following Cardiac Surgery: An Investigation of an Australian Patient Cohort.* Heart, Lung and Circulation, 2017. **26**(3): p. 301-308.

[11] Blagus, R. and L. Lusa, *SMOTE for high-dimensional class-imbalanced data.* BMC Bioinformatics, 2013. **14**: p. 106-106.

[12] Witten, I.H., et al., *Data Mining: Practical machine learning tools and techniques*. 2016: Morgan Kaufmann.

13. Kohavi, R. *A study of cross-validation and bootstrap for accuracy estimation and model selection*. in *Ijcai*. 1995. Montreal, Canada.

Address for correspondence

Young Ho Yun*, MD, PhD, Tel: +82-2-740-8417; Fax: +82-2-742-5947; E-mail address: lawyun@snu.ac.kr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190549

Computation of Brain Functional Connectivity Network Measures in Epilepsy: A Web-Based Platform for EEG Signal Data Processing and Analysis

Vimig Socrates[a,b], Arthur Gershon[a], Satya S. Sahoo[a,b]

[a] *Department of Population & Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA*
[b] *Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH, USA*

Abstract

Epilepsy is a serious neurological disorder that affects nearly 60 million individuals worldwide and it is characterized by repeated seizures. Graph theoretic approaches have been developed to analyze functional brain networks that underpin epileptogenic network. We have developed a Web-based application that enables neuroscientists to process high resolution Stereotactic Electroencephalogram (SEEG) signal data and compute various kinds of signal coupling measures using an intuitive user interface for study of epilepsy seizure networks. Results of a systematic evaluation of this new application show that it scales with increasing volume of signal data.

Keywords:
Electroencephalography, Workflow, Epilepsy

Introduction

Epilepsy is a serious neurological disorder marked by repeated seizures and it affects nearly 60 million individuals worldwide across racial, gender, and geographic boundaries [1]. More than 20% of these patients are refractory to anti-epileptic medication [2]. These patients are often considered for surgery for seizure freedom, which involves analysis of electroencephalogram (EEG) data for accurate characterization of the epileptogenic zone. Various measures for brain connectivity measures have been developed to provide new insights into the mechanism that underpins seizure onset and propagation. Many of these brain functional connectivity measures are derived from high-resolution EEG data, for example, signal data recorded from intracranial electrodes implanted using stereotactic techniques (also called SEEG signal data). The increasing availability of digital-era SEEG signal data from multi-institutional research studies requires the development of efficient data processing and analysis techniques that can effectively leverage the large volume of signal data for advancing brain connectivity research.

The European Data Format (EDF) is a widely used signal data format that was originally designed for recording of polysomnography signal data [3]. However, there are several challenges that impede the use of EDF for large-scale signal data analysis especially in the context of brain functional connectivity studies. In our previous work, we had developed a new model for signal data called the Cloudwave Signal Format (CSF) that addressed many of the limitations of EDF model [4], including the support for random read access to specific channels. In this abstract, we describe our work that builds on the CSF model to implement a Web-based workflow platform that allows users to process and analyze SEEG data to compute multiple measures of brain functional connectivity.

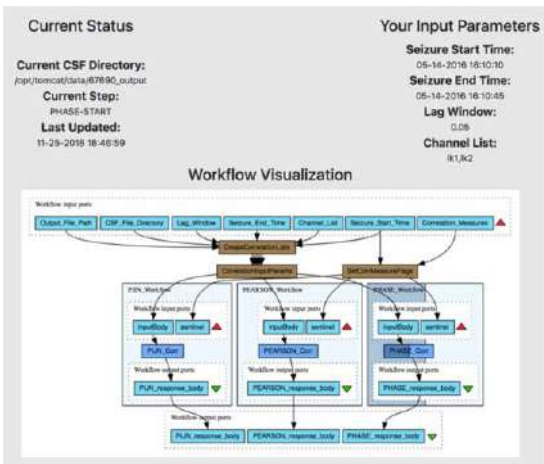


Figure 1 – User interface with visualization of the workflow module status. Light blue boxes are inputs, dark blue are REST calls, and brown are BeanShell scripts.

Related work. There has been a significant amount of work in the domain of brain connectivity [5]. Complex network analysis focus on computing graph theory-based measures, such as centrality and the detection of subgraph motifs, from network models of brain connectivity. For example, brain structural connectivity is used to study anatomical connectivity representing physical connections between brain regions whereas functional connectivity measures are used to determine functional coupling between brain regions [6]. These measures are used to analyze brain connectivity networks in healthy as well as persons suffering from neurological disorders such as epilepsy.

Method

There are three core components of our software tool, including: (1) a Representational State Transfer (REST) Application Programming Interface (API) for data processing; (2) a scientific workflow to integrate the API components as well as record metadata information; and (3) an intuitive Web-based user interface.

Component 1: RESTful Web Services

We developed a Java-based RESTful API for the conversion of EDF files to CSF files and computation of functional

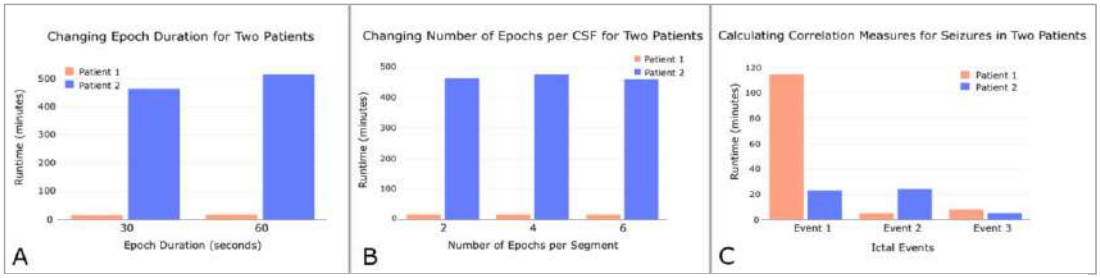


Figure 2 – Performance evaluation of data processing (Fig. 2a, 2b) and data analysis modules (Fig. 2c) using SEEG signal data

connectivity measures. Once CSF files are created, users are able to calculate channel-wise correlation values. Three correlation measures that have demonstrated efficacy in determining functional connectivity between brain regions have been implemented: Pearson's linear regression coefficient [7], a non-linear regression coefficient by Pijn [8], and a non-linear frequency-based measure called mean phase coherence [9]. The produced correlation matrices can be used to create adjacency matrices that describe a network graph.

Component 2: Scientific Workflow Tool

We used the open-source Taverna scientific workflow system to support large scale signal processing task and also allow users to visually track the entire signal processing pipeline [10]. Using the Taverna Workbench, we defined two workflow modules that take parameters as user inputs, which are invoked by the workflow module as REST calls. Figure 1 shows a screenshot of the user interface with a visualization of the workflow execution in the Taverna Workbench. We also use Taverna to record workflow provenance, including inputs, outputs, and intermediate results to support reproducibility of the data processing as well as analysis tasks.

Component 3: Intuitive User Interface

The user interface (UI) is designed using Python Django and Bootstrap. The UI has been designed to be intuitive and allow users to execute, validate, and monitor their workflow tasks. The UI consists of two forms to record user input for both data processing and data analysis. For example, the EDF-to-CSF data processing form allows users to specify input as well as output directories, epoch duration, and the number of epochs. On the backend, the server makes API calls to the various Taverna Server endpoints to start the workflow, and takes the user to a status page, with a real-time updating image of the workflow (Figure 1).

Results

In order to evaluate the scalability of the tool, we conducted a comparative analysis of the execution of the workflow on datasets with different sizes of signal data. We used de-identified data from two patients. The size of data from Patient 1 was 4.4 gigabytes (GB), and 122GB from Patient 2. The results of the comparative analysis are shown in Figure 2a and 2b for different input parameters to the data processing module, such as epoch duration and number of epochs per CSF file. Figure 2c shows the time taken to compute all three functional connectivity measures for different parameter values, including the length of the ictal event and number of channels analyzed.

Conclusions

In this paper, we describe the development of a Web-based user friendly software tool for large-scale signal analysis in epilepsy cohort studies to analyze functional brain networks. In future work, we plan to visualize network graphs and integrate computation of additional network connectivity measures, such as degree distribution and global as well as local efficiency measures.

Acknowledgements

This work is supported by NIH Big Data to Knowledge (BD2K) 1U01EB020955 grant, and NSF grant #1636850.

References

[1] World Health Organization, World Health Organization International Classification of Functioning, Disability and Health, Geneva, Switzerland, 2010.

[2] D.B. Smith, R.H. Mattson, J.A. Cramer, J.F. Collins, R.A. Novelly, B. Craft, & Veterans Administration Epilepsy Cooperative Study Group, Results of a nationwide Veterans Administration Cooperative Study comparing the efficacy and toxicity of carbamazepine, phenobarbital, phenytoin, and primidone, *Epilepsia*, **28** (1987), S50-S58.

[3] B. Kemp and J. Olivan, European data format 'plus'(EDF+), an EDF alike standard format for the exchange of physiological data, *Clinical neurophysiology* **114.9** (2003), 1755-1761.

[4] C. Jayapandian, A. Wei, P. Ramesh, B. Zonjy, S.D. Lhatoo, K. Loparo, G.Q. Zhang, and S.S. Sahoo, A scalable neuroinformatics data flow for electrophysiological signals using MapReduce, *Frontiers in neuroinformatics*, **9** (2015), 4.

[5] E. Bullmore and O. Sporns, Complex brain networks: graph theoretical analysis of structural and functional systems, *Nature reviews neuroscience*, **10** (2009), 186.

[6] B. Horwitz, The elusive concept of brain connectivity. *Neuroimage*, **19** (2003), 466-470.

[7] S.W. Scheff, *Fundamental Statistical Principles for the Neurobiologist: A Survival Guide*, Academic Press, London, 2016.

[8] J.P. Pijn, F. Lopes da Silva, Propagation of electrical activity: nonlinear associations and time delays between EEG signals, In *Zschocke S, Speckmann E-J, eds. Basic Mechanisms of the EEG,* Boston, MA, 1993.

[9] F. Mormann, K. Lehnertz, P. David, C.E. Elger, Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients, *Physica D* **144** (2000), 358-369.

[10] P. Missier, S. Soiland-Reyes, S. Owen, W. Tan, A. Nenadic, I. Dunlop, A. Williams, T. Oinn and C. Goble, Taverna, reloaded, *International conference on scientific and statistical database management*, (2010), 471-481.

Address for Correspondence

Satya S. Sahoo, PhD. Email: satya.sahoo@case.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190550

Korean Pharmacovigilance System Based on EHR-CDM

Nayeong Son[a], Bonggi Kim[a], Sooyoun Chung[a], Soonyoung Han[a]

[a] *Korea Institute of Drug Safety & Risk Management, Anyang-si, Gyeonggi-do 14051*

Abstract

An electronic health record (EHR) contains various clinical information for pharmacovigilance studies, but they remain difficult to use.. From 2016 to 2018, the ministry of food and drug safety and Korea institute of drug safety & risk management (KIDS) converted the EHRs of more than 9 million patients to a MOA common data model (CDM). KIDS developed the Medical record observation and assessment for drug safety network (MOA-Net), a web portal site to build a network between CDM data partners. Through MOA-Net, hospitals participated in pharmacovigilance studies and confirmed the usability and adequacy of the CDM.

Keywords:

Electronic health record, pharmacoepidemiology, pharmacovigilance

Introduction

Recent advances in health informatics has led to an increased interest in a new type of distributed database system known as a common data model (CDM). CDM is a database model with a standardized schema and vocabulary system. By using CDM, it became possible to perform simultaneous multi-center analysis using a reusable, parametrized query tool without exposing any kind of private information out of the medical institutes. In the Republic of Korea, the Ministry of Food and Drug Safety (MFDS) and the Korea Institute of Drug Safety & Risk Management (KIDS) developed an active pharmacovigilance system using CDM based on electronic health records (EHRs) since 2016. KIDS named this customized Korean model MOA (Medical record observation and assessment for drug safety)-CDM. During 2016-2017, MFDS constructed MOA-CDM at 9 hospitals in Korea. In 2018, KIDS selected five major hospitals including local representative drug safety centers as an attempt to expand the CDM. Also, KIDS developed MOA-Net, a web-based, multi-directional portal service to network government and data partners. Through MOA-Net, hospitals participated in pharmaco-vigilance studies and confirmed the usability and adequacy of the CDM.

Methods

Building MOA-CDM requires several steps such as extracting raw data from EHR, transforming the database schema and vocabularies into a common format, and loading. The standard schema of MOA-CDM includes essential parts of Sentinel CDM from the United States Food and Drug Administration (U.S. FDA)[1; 2] and the observational medical outcomes partnership (OMOP) CDM from observational health data science and informatics (OHDSI)[3; 4]. Since Korean hospitals mostly use Korean electronic data interchange (EDI) code and Korean standard classification of diseases (KCD) code in the National health insurance system, we mapped the local vocabularies into standard vocabularies. Local vocabularies are saved as source_value in OMOP CDM tables [4]. The standard vocabularies that we selected are as shown below [5; 6; 7; 8]:

Table 1. Vocabulary Mapping

Category	Local vocabulary (before mapping)	Standard vocabulary (after mapping)
Drug	EDI code	RxNorm
Condition	KCD-7	SNOMED-CT
Procedure	EDI code	EDI code
Measurement	Vary by hospital	LOINC

* SNOMED-CT: Systematized nomenclature of medicine-clinical terms, EDI: Korean electronic data interchange, LOINC: Logical observation identifiers names and codes

We designed the workflow of the Korean active pharmacovigilance system based on the CDM as follows: first, MFDS requests the analysis of domestic status to KIDS when a drug safety issue occurs. KIDS designs the analysisby referring to the literature review and experts advice. Next, KIDS distributes the analysis module through the MOA-Net and invites data partners to research. Participated data partners run the module using their own data and provide results to KIDS through the portal. Finally, KIDS integrates the results from the data partners and suggests answers for MFDS' safety concern.

We checked the adequacy and efficiency of the pharmaco-vigilance system based on CDM as follows. First, we performed abnormality detection and logical error test to enhance the quality of CDM data. We applied the Achilles heel rules from OMOP CDM [3; 4] as a standard of logical error test. We also checked the rate of concepts that are mapped into standardized vocabularies successfully. Second, we designed three nested, case-control studies about current pharmaco-vigilance issues in Korea - (1) Diclofenac beta di-methyl-aminoethanol & angioedema, (2) Clazpine & pleurisy, and (3) Allopurinol & thyroid stimulating hormone (TSH) increase. For every case, we randomly selected four controls by incidence density sampling method. We matched them for age, sex, and cohort-in-date. We aimed to calculate descriptive statistics of subjects, and unadjusted odds ratio (crude OR) as a signal statistics of adverse drug reaction. We developed analysis tools using structured query language (SQL) and R. Then, we distributed the tools and collected analysis of results from data partners through MOA-Net.

Results

In Korea, more than 9 million patient information records from 14 national representative hospitals were acquired to construct

a CDM database. The database converting EHR to CDM includes more than 127 kinds of clinical laboratory tests, and more than 5,000 drugs available in Korea. Every database entered the logical error test by the Achilles heel rule successfully. In the mapping rate test, we found some unusual concepts were not mapped into standard vocabularies. Most of the unmapped concepts turned out to be old drugs that are not currently used, or very new drugs in clinical trials. Continuous review and examination in mapping are still in progress to improve the quality of vocabularies by experts.

A total of 6 hospitals participated in three phamacovigilance studies. Among the three studies, we found two significant adverse drug reaction signals – clozpine & pleurisy, and allopurinol & TSH increase. For clozapine and pleurisy, we identified 81 cases of pleurisy and 308 matched controls. When all non-users were compared with users of clozapine, the risk of pleurisy (unadjusted OR) was increased by 8.46 fold (95% CI: 1.4-51.28). For allopurinol and TSH increase, we identified 10,663 cases of TSH increase and 42,631 matched controls. When all non-users were compared with users of allopurinol, the risk of TSH increase (unadjusted OR) was increased 3.52 fold (95% CI: 2.96-4.20). Although these statistics are not enough to conclude the relationship between drug use and adverse effect in the way that it is an unadjusted result, it still implies considerable association as a signal statistics. For diclofenac dimethylaminoethanol and angioedema, we couldn't assess the relationship because diclofenac use was considerably rare in our subjects. However, we identified significant differences in Non Steriod Anti Infammatory Drugs (NSAIDs), antibiotics use and history of urticaria which are all well-known risk factors in angioedema. The analysis result is as follow:

Table 2. Pharmacovigilance Results

Subject	Case (N)	Control (N)	OR	95% C.I.
Diclofenac-angioedema	1,719	6,876	1.33	0.14-12.85
Clozapine-pleurisy	81	308	8.46	1.4-51.28
Allopurinol-TSH increase	10,663	42,631	3.52	2.96-4.20

Conclusions

The Korean pharmacovigilance system based on MOA-CDM promises for active drug safety surveillance. As a result of performing several projects via MOA-CDM pharmacovigilance system, we found that the CDM system is faster and more efficient than analyzing health claims data to answer drug safety concerns. Also, it was possible to elaborate on research designs for adverse drug reaction using clinical information such as laboratory test results.

However, we need to overcome some difficulties in using CDM. First, we should consistently improve quality of data and vocabulary mapping. Also, hospitals should update their data periodically to assess the safety of new drugs. Therefore, automatic, real-time ETL (extraction-transformation-loading) and mapping engines should be developed by hospitals. Second, there is a risk of serious underestimation in pharmacovigilance studies using CDM because interhospital data sharing is impossible. For example, if a patient was prescribed medication elsewhere first and visited the hospital after they noticed an adverse drug reaction, an algorithm would classify him as a non-exposed case. On the other hand, if a patient was prescribed medication in a hospital and visited the hospital again when they noticed adverse drug reaction, an algorithm would classify him as an exposed control. Thus, researchers should be aware of the risk of bias and carefully interpret analysis result from CDM. Also, more medical institutional participation is requisite to enhance the effectiveness of this network.

Acknowledgements

This research was supported by a grant from the Korea Institute of Drug Safety and Risk Management in 2018.

References

[1] CURTIS, Lesley H., et al. Design considerations, architecture, and use of the Mini-Sentinel distributed data system. *Pharmacoepidemiology and Drug Safety* **21 (**2012), 23-31.

[2] Sentinel Initiative. https://www.sentinelinitiative.org/

[3] Observational health data science and informatics. https://www.ohdsi.org/

[4] OMOP common data model specifications version 5.0 (2014). Available at https://github.com/OHDSI/CommonDataModel/releases

[5] LIU, Simon, et al. RxNorm: prescription for electronic drug information exchange. *IT Professional* **7.5** (2005), 17-23.

[6] KOREA, Statistics. Korean Standard Classification of Diseases 7th Revision [Internet]. Daejeon: Statistics Korea.[cited 2016 Jan 20]. Available from: https://kssc. kostat. go. kr, 8443.

[7] DONNELLY, Kevin. SNOMED-CT: The advanced terminology and coding system for eHealth. *Studies in Health Technology and Informatics* **121, (**2006), 279.

[8] HUFF, Stanley M., et al. Development of the logical observation identifier names and codes (LOINC) vocabulary. *Journal of the American Medical Informatics Association* **5.**3 (1998), 276-292.

Address for correspondence

Bonggi Kim, Ph.D.

30 Burim-ro 169beon-gil, Dongan-gu, Anyang-si, Gyeonggi-do, Republic of Korea

E-mail: bgkim@drugsafe.or.kr

Phone: +82-2-2172-6770

Fax: +82-2-2172-6702

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190551

An Approach of Integrating Domain Knowledge into Data-Driven Diagnostic Model

Guanxu Su, Jianghui Wen, Zhaowei Zhu, Zhuo Liu, Wei Zhao, Xingzhi Sun, Gang Hu, Guotong Xie

Ping An Health Technology, Beijing, China

Abstract

A diagnostic model of general diseases could help general practitioners to decrease misdiagnoses and reduce workload. In this paper, we developed a neural network model that can classify potential diagnoses among 100 selected common diseases based on ambulatory health care data. We propose a novel approach to integrate domain knowledge into neural network training. The evaluation results show our model outperforming the baseline model in terms of knowledge consistency and model generalization.

Keywords:
Neural networks (computer), diagnosis

Introduction

Community hospital centers (CHCs) are primary clinical institutions in which general practitioners (GPs) provide services to people nearby. Diagnosing patients with various types of diseases requires GPs to grasp a wide range of clinical knowledge. However, CHCs in China are facing a severe shortage of skilled GPs, resulting in a considerable number of misdiagnoses [1,2]. To reduce the misdiagnoses and workload for GPs, a practical solution is to provide an auxiliary diagnosis service. There are two main types of diagnostic models in clinical practice: knowledge-driven models and data-driven models. Knowledge-driven models depend on domain knowledge and usually provide reliable recommendations. However, medical knowledge is hard to acquire and often too general to address individual cases. Alternately, data-driven models can extract finer hidden patterns or relationships from many patients' records and achieve a wide coverage of diseases. However, due to the overall optimization goal in the model building process, diagnoses provided by data-driven models could be unreasonable for some individual cases.

In this paper, we developed a diagnostic model that consists of both data and knowledge to address the specificity of individual patients and the use of general diagnoses. We build a data-driven neural network model as the baseline model based on the Centers for Disease Control and Prevention's (CDC) ambulatory health care data [3]. Then we integrate domain knowledge including a knowledge graph, hierarchy information, and symptom frequency into the neural network training by applying an iterative training method, doing multitask learning, and modifying neural network structures.

Methods

We extracted ambulatory health care data from the National Hospital Ambulatory Medical Care Survey (NHAMCS) and the National Ambulatory Medical Care Survey (NAMCS) provided by the CDC. We obtained over 2 million records from 1995 to 2015. The dataset contained patient information such as demographics (sex, age), medical history, symptoms, and diagnoses. Table 1 shows the feature vector of patient information extracted from the dataset. Diagnoses are indicated by international classification of disease 9th version (ICD-9) codes in the dataset. We filtered the diagnosis list and selected the top 100 common diseases as our target diseases.

Table 1- Input features.

Feature	Description
age	age of patient, integer
sex	sex of patient, dummy variable
Symptom (1-798)	the symptoms that the patient report, corresponding to 798 columns, with each being a dummy variable
past medical history (1-72)	24 kinds of disease history of the patient. Each disease history corresponds to 3 columns, with each represents 'yes', 'no', and 'NA', respectively

We considered the following types of domain knowledge.

1. Knowledge graph: The typical symptoms of diseases. We collect information on relations between diseases and symptoms from sources like OMAHA (http://www.omaha.org. cn/),Wanfang (http://www.wanfangdata.com.cn/index. html), and SNOMED CT (http://www.snomed.org/) to construct our knowledge graph.
2. Hierarchy information: The additional labels of diseases from different dimensions. We obtain the clinical departments of the target diseases by consulting qualified physicians.
3. Symptom frequency: Conditional probabilities of each symptom given diseases.

Figure 1 shows the final model on which we applied all domain knowledge including knowledge graph, department information, and symptom frequency to the baseline model. For the baseline model, inputs are feature vectors of patient information. The output gives confidence for each potential disease. The input layer and output layer are connected by two fully connected hidden layers. Hyperbolic tangent is used as the activation function for all the connections. Cross entropy loss is used for optimization.

To add knowledge to our data-based models, we first added the knowledge graph to the training procedure using an iterative training method. We defined two types of training steps. The first one is training steps on the training dataset. The second one is training steps on the knowledge graph. During the first type of training steps, all parameters would be optimized. During the

second type of training steps, we only trained the parameters related to symptoms covered in the knowledge graph. These are marked as blue nodes in Figure 1. This prevents other parameters from being misadjusted. In addition, we performed training steps on the knowledge graph for every 10 iterations of training steps on the training dataset.

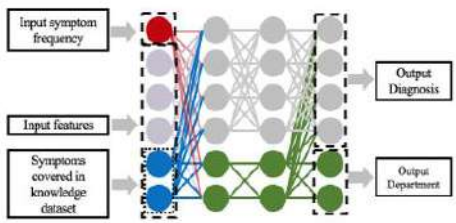


Figure 1 - Neural network Integrating all three types of domain knowledge.

Secondly, we made use of clinical department information. The green nodes represent the network for department prediction. The model for disease prediction shares nodes of the model for department prediction through the last hidden layer of the latter. During model training, the training step for the department prediction model is executed once after every training step for the disease prediction model to maintain the functionality of the hidden nodes for department prediction.

Finally, we added symptom frequency into the input layer. Aside from external informational like the knowledge graph and hierarchy information, there is information inside the data itself that can contribute to the classification model. The conditional probability of a symptom given each disease (referred to as symptom frequency) is such information inside the data that we can extract before training the neural network. We added that information into the input of the neural network, which is represented by the red node in Figure 1.

Results

The CDC dataset was randomly split into training and test datasets with a ratio of 8 to 2. We also retrieved data from Ping An Good Doctor, one of the largest disease diagnostic consulting online platforms providing real time connections between patients and doctors, as an additional test set to check model generalization [4]. We used metrics of micro and macro AUC [5], and accuracy on the test dataset, knowledge graph, and Ping An Good Doctor dataset as criteria to evaluate our five models. Accuracy at X (accu@X) is calculated as the ratio that the top X output diagnoses cover the true label.

Models 1-5 represent, respectively, the baseline model, model integrating the knowledge graph, department information, and symptom frequency, and the final model integrating all three above. As shown in Table 2, Model 4 achieved the best performance. It also can be seen that the five models didn't outperform each other significantly on the test dataset, which means that all five models fit the CDC dataset well. Table 3 shows the model performance of the Ping An Good Doctor dataset and knowledge graph. Model 2 and Model 5 were much more consistent with the knowledge graph compared to the other three models. It indicates that our method of iterative training on data and knowledge can successfully improve model consistency with the knowledge graph without sacrificing accuracy on the CDC test data. As for the accuracy of the Ping An Good Doctor dataset, all the modified models outperformed baseline Model 1. Among the modified models, Model 2 presented better generalization and higher accuracy on the Ping An Good Doctor dataset compared to the other models. Taking all results into consideration, Model 5, the model that integrated all the three types of domain knowledge, provides good knowledge consistency and high accuracy on both the CDC test data and Ping An Good Doctor dataset.

Table 2- Performance on test dataset.

Model	micro AUC	macro AUC	accu@1	accu@3	accu@5
1	97.09%	95.46%	47.93%	72.19%	81.24%
2	96.95%	95.10%	47.27%	71.66%	89.83%
3	97.28%	95.71%	48.38%	72.90%	81.79%
4	**97.34%**	**95.82%**	**48.38%**	**72.98%**	**81.71%**
5	97.22%	95.71%	47.38%	71.97%	81.14%

Table 3- Performance of the Ping An Good Doctor and knowledge graph datasets.

	Accuracy on Ping An Good Doctor data			Accuracy on knowledge graph
Model	**accu@1**	**accu@3**	**accu@5**	**accu@3**
1	21.16%	36.65%	45.31%	34.81%
2	23.30%	41.05%	50.57%	84.65%
3	23.30%	37.78%	47.44%	34.85%
4	22.87%	39.49%	49.29%	38.15%
5	21.31%	38.92%	50.85%	81.20%

Conclusions

In this study, we introduced an approach to integrate domain knowledge into a neural network for general disease diagnosis. The evaluation results showed that compared with the pure data-driven approach, our approach can lead to better performance in terms of knowledge consistency and model generalization on other data without sacrificing accuracy on the test data.

References

[1] China needs many more primary care doctors: The Economist. 2017. Available at: https://www.economist.com/china/2017/05/11/china-needs-many-more-primary-care-doctors, [accessed March 25, 2019].

[2] R.L. Phillips, L.A. Bartholomew, S.M. Dovey, G.E. Fryer, T.J. Miyoshi, and L.A. Green, Learning from malpractice claims about negligent, adverse events in primary care in the United States. *Qual Saf Health Care*. 13 (2004) 121-126. doi: 10.1136/qshc.2003.008029.

[3] Datasets and documentation. Available at: https://www.cdc.gov/nchs/ahcd/datasets_documentation_related.htm#data, [accessed March 25, 2019].

[4] Ping An Good Doctor showcases AI-powered, unstaffed clinics. https://www.mobihealthnews.com/content/ping-good-doctor-showcases-ai-powered-unstaffed-clinics, [accessed March 25, 2019].

[5] V. Van Asch, Macro- and micro-averaged evaluation measures. *Technical Report*, 2013.

Address for correspondence
Zhuo Liu
E-mail: lzh0201@126.com (prefered method of contact)

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190552

Early Nephrosis Detection Based on Deep Learning with Clinical Time-Series Data

Yohei Yamasaki[a], Osamu Sugiyama[b], Shusuke Hiragi[a,b], Shosuke Ohtera[a,b], Goshiro Yamamoto[b], Hiroshi Sasaki[b], Kazuya Okamoto[a,b], Masayuki Nambu[b], Tomohiro Kuroda[a,b]

[a] *Graduate School of Informatics Kyoto University, Kyoto-City, Kyoto, Japan,*
[b] *Kyoto University Hospital, Kyoto-City, Kyoto, Japan*

Abstract

Nephrosis is disease characterized by abnormal protein loss from impaired kidney. We constructed early prediction model using machine learning from clinical time series data, that can predict onset of nephrosis for more than one month. Long short-term memory capable of recognizing temporal sequential data patterns, was adopted as early prediction model for nephrosis. We verified our proposed prediction model has higher accuracy compared with those of baseline classifiers by 5-fold cross validation.

Keywords:
Decision support techniques, supervised machine learning, nephrosis

Introduction

Nephrosis is disease characterized by abnormal protein loss from impaired kidney. The disease sometimes results in acute kidney injury, venous thrombosis, or severe infection due to lowered oncotic pressure and immunoglobulin loss. Therefore, early detection is beneficial for patients.

In this research, we aimed to develop early prediction model using machine learning from clinical time series data that can predict rise of nephrosis. There were several issues to tackle when constructing the nephrosis prediction model, including selection of interpolation and classifier methodologies. As for interpolation, clinical time series data were sometimes partially collected, we must compensate missing values with consideration of temporal transitions. Selection of classifier also must consider temporal relationship of clinical time-series data.

To solve these problems, we adopted Recurrent Neural Network Model (RNN), which can recognize sequential data patterns, to deal with time series information. Subject examination data covered in this study was multi-dimensional, and number of examinations varied among patients. RNN is suitable for these kinds of data [1]. Unlike ordinary neural networks such as Multi-Layer Perceptron (MLP), RNN learns with consideration for relationship of continuous data by constructing weighted join that loops to hidden layer. In RNN architecture (simple RNN) composed only of simple weighted connections, vanishing gradient problem occurs at time of updating weights, and learning taking long-term time-series information into account is not guaranteed. To tackle this problem, Long Short-Term Memory (LSTM) had been proposed [2]. LSTM networks are well-suited to classifying, processing, and predicting based on time series data, since there can be lags of unknown duration between important events in a time series. In this study, LSTM was adopted as early prediction model, and results of prediction using simple RNN and LSTM model were compared.

As for interpolation methods, we selected basic mean-value interpolation and linear interpolation methods, and adopted the most accurate method.

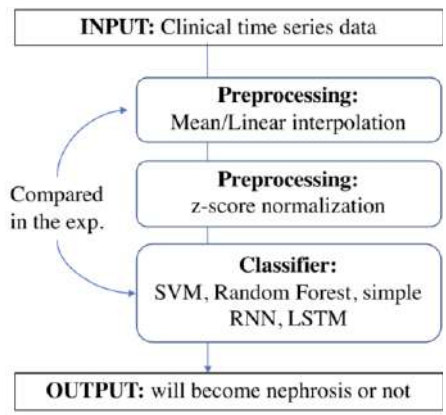


Figure 1 – Process of nephrosis prediction

Methods

Figure 1 illustrates steps of nephrosis prediction. The process of prediction can be divided into preprocessing and classification. In preprocessing, system first interpolates them with either mean-value interpolation or linear interpolation and then normalizes clinical time-series data with z-score normalization. After preprocessing, data is given to the classifier and acquire positive/negative value, meaning that patient may become nephrotic or not.

The prediction model was implemented with Python 3 and conventional classifiers, Support Vector Machine (SVM) [3] and Random Forest [4] were implemented with *scikit-learn* package; while simple RNN and LSTM were implemented with *keras* package (*tensorflow* backend). RNN and LSTM used in this experiment only had one hidden layer and number of units was 10. To prevent over-fitting, we added a dropout layer and set its probability as 0.3. Finally, softmax layer was applied to classify the positive/negative of future rise of nephrosis. The error function was categorical cross entropy, optimization method was Adam. The mini batch size was 16, and learning epoch was 50.

The clinical time-series data was collected from anonymous dataset of patients who visited Kyoto University Hospital over more than four times. Approximately 50,000 samples of 2,700 patients (377 patients were positive and rest negative) for five years were collected for the experiment. In addition, with

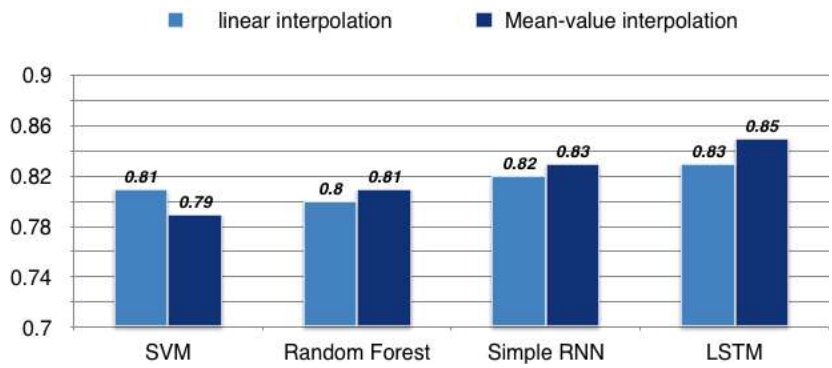


Figure 2– Accuracy comparison

Figure 3– Transition of accuracies in temporal domain

reference to opinion of the internal medicine kidney doctor, 12 examination items; *ALB, BUN, CRE, CRP, Ca, Cl, IP, K, Na, T-CHO, PCR and TP*, were extracted as values to be correlated with nephrosis.

Results

To verify our proposed early prediction model, we conducted two types of experiments, comparison with baseline methods and verification of the proposed model in temporal domain.

As for comparison with baseline methods, we verified the proposed nephrosis prediction model with combinations of two types of interpolation, which were mean-value interpolation and linear interpolation, and four types of classifiers, which were SVM, Random Forest, simple RNN, and LSTM. In total, we verified accuracies of eight kinds of identification methods by 5-fold cross validation (80% for training, 20% for validation). Figure 2 shows comparison of prediction accuracy of each method. LSTM had highest accuracy among prediction models. In comparing interpolation methods between RNN and LSTM, mean-value interpolation was more accurate than linear interpolation.

As for verification of proposed model in temporal domain, we collected another dataset 1, 2, 3, and 4 months respectively before detection of nephrosis and fed each ranged data to the LSTM classifier with mean-value interpolation. Figure 3 illustrates transition of accuracies of the proposed prediction model. From Fig. 3, we can see that the shorter the term, the more accurate the prediction was with the proposed model. Meanwhile, accuracy was not drastically decreased, even with the test-set from the entries of four months before.

Discussion

The comparison of baseline methods (Fig. 2) revealed that combinations of mean-value interpolation and LSTM has a higher performance with the experimental data. The results that mean-value interpolation has higher performance than linear interpolation indicate that transition of clinical data is non-linearly distributed and mean-value interpolation could be relatively better for the clinical time-series data. To apply more appropriate method to compensate for missing values, such as [5], would be future work.

Verification of the proposed prediction method with temporal domain (Fig.3) reveals that even with clinical time-series data that is collected four months before the detection of nephrosis, our proposed method has a good performance. This indicates that our proposed prediction model is suitable for early detection of nephrosis.

Conclusions

In this study, we aimed to develop a prediction model based on LSTM that enables early detection of nephrosis. Experimental results showed that the prediction model using LSTM with mean-value interpolation had the best performance among the comparison of two interpolation methods and four types of classifiers. The experimental results with temporal domain comparison reveals that the proposed prediction model could have a stable prediction rate even with the clinical data that was collected four months or less from the detection of nephrosis. The experimental results indicate that our proposed nephrosis prediction model is effective for the early detection of nephrosis both for accuracy and foresight.

References

[1] T. Mikolov and G. Zweig, Context dependent recurrent neural network language model, In: *IEEE Spoken Language Technology Workshop*, 2012, 234-239.

[2] S. Hochreiter and J. Schmidhuber. Long short-term memory, *Neural Computation*, **9** (1997), 1735-1780.

[3] B. Schölkopf, and A.J. Smola, Support vector machines and kernel algorithms, In: *Encyclopedia of Biostatistics 2 ed.*, John Wiley Sons, (2005), 5328-5335.

[4] A. Liaw and M. Wiener, Classification and regression by randomForest, *R news*, **2** (2002), 18-22.

[5] Z. Che, S. Purushotham, K. Cho, D. Sontag and Y. Liu, Recurrent neural networks for multivariate time series with missing values, *Sci Rep*, **8** (2018), 6085.

Address for correspondence

Osamu Sugiyama, sugiyama@kuhp.kyoto-u.ac.jp

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190553

Top-Level Design of a Normalized Chinese Clinical Terminology: An Integrated Application of National and International Data Standards and Terminologies

Haixia Sun[a], Yujing Ji[a], Panpan Deng [a], Junlian Li [a], Huiling Ren[a], Liu Shen[a], Ming Feng[b], Yi Wang[b]

[a] Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing, China
[b] Peking Union Medical College Hospital, Beijing, China

Abstract

This work describes the design and building of a Chinese clinical terminology (called CCTS). The terminology is similar to an ontology, and will promote the use of Chinese clinical data, such as indexing, retrieval and exchange. The terminology is a TOPL concept framework, which integrates hierarchical structures of Chinese and interenational reference terminology standards for health. Our framework includes 14 subtrees, 2286 classes and 65 relationships.

Keywords:
Vocabulary, Electronic Health Records, Reference Standards

Introduction

Wide implementation of electronic health record systems in China has made clinical data available electronically. However, little effort has been devoted to building a national clinical terminolgy, limiting the use of clinical data for indexing, retrieval, and exchange. Motivated by this, we developed a Chinese clinical terminology, named *Chinese Clinical Terminology System (CCTS)*. We divided this work into two phases: the first phase is the top-level design, constructing the top concept framework of CCTS; the second phase is the refinement and expansion, collection and organization of more terms and concepts according the top concept framework. In this paper we introduce the top-level design.

Methods

First, we analyzed main knowledge types in Electronic Medical Records (EMR) according to the Specification for Sharing Document of Electronic Medical Record and Basic Dataset of Electronic Medical Record released by National Health Commission of the People's Republic of China (NHCPRC). This analysis indicated the scope of the terminology. Domain experts then manually selected the 14 top classes.

Then, we extracted the top 5 level concepts from terminologies and code systems (Table 1). These concepts are recommended in Chinese and international health information standards. The 5 level concepts were extracted in the following steps:

1. Manually determine the best reference subtree for different classes by evaluating and ranking them for different classes
2. Merge other related subtrees by string matching and rules
3. Recommend preferred term for classes according to reference books published by the China National Committee for Terms in Sciences and Technologies (CNCTST)
4. Inspect hierarchical relationship manually
5. Define other medical semantic relationships among the first 5 level classes referring to semantic relationships in the Unified Medical Language System (UMLS)

Results

Our top-level design phase produced 2369 high level concept classes, organized into 14 subtrees (Table 2) with 65 kinds of semantic relationships. We recommend 136 core reference sources for lower subtrees. We are in the process of extracting and fusing lower subtrees from those sources.

Discussion

This paper describes the development of CCTS through a top-level concept framework that integrates hierarchical structures of Chinese and international reference standards for health terminology, as well as expert knowledge. Reference sources for each subtree to build lower subtrees were also recommended, which will be used to guide the construction of the terminology system.

Information model design for different concept classes and extracting Out-of-Vocabulary words from EMRs using medical natural language processing (MedNLP) technology will be used in future work. We anticipate that CCTS will cover common medical terms and concepts in various EMR systems in China, and can be used to greatly facilitate EMR searching, retrieval, clustering, and reasoning, by providing rich sets of synonyms and various clinical relationships. Correspondingly, it also can promote the interoperability of EMRs by intergating Chinese and international health reference standards.

Conclusions

Design of the top concept framework is the first step towards building a domain terminology system or ontology. Integrating multiple resources is an efficient way to achieve this goal, and domain experts are indispensable. We envision the top concept framework will be used to guide the next phase of CCTS.

Table 1– Some Core References for The 14 Subtrees

Top Classes	Core References
疾病诊断 (Disease Diagnosis)	Disease Classification and Code (GB/T14396, China); ICD 10, SNOMED CT, Disease Ontology
有机体(Organism)	MeSH
解剖部位(Anatomy)	FMA, MeSH
诊查对象(Observable and Examinable Object)	SNOMED CT
临床表现 (Clinical Manifestation)	Health Information Data Metadata Catalogue Part 6: Chief Complaints and Symptoms (WS 363.6, , China); SNOMED CT, HPO
诊疗项目、技术和方法 (Diagnosis and Treatment Item, Technique and Method)	Clinical test item classification and code (WS/T102, China), Health information data element catalogue Part 9: Laboratory examination (WS 363.9, China), ICD-9-CM-3, LOINC
化学药品和生物制品 (Chemical Drug and Biological Product)	ATC, Chinese Pharmacopoeia 2015
医用设备、器械和材料 (Medical Equipment, Instrument and Material)	Social Insurance Medical Service Classification and Code (LD/T01-2017, China), Health information data element value code Part 16: Drugs, equipment and materials (WS 364.16, China)
物质(Substance)	SNOMED CT
心理行为(Psychology and Behavior)	SNOMED CT
环境地理 (Environment and Geography)	World and Region Name Codes (GB/T2659-2000, China)
事件、事故和灾害 (Event, Accident and Disaster)	ICD-10, MeSH, SNOMED CT
人口学及社会经济学特征 (Demographic and Socioeconomic Characteristic)	Health Information Data Metadata Catalogue Part 3: Demographic and socioeconomic characteristics (WS 363.3-2011, China), Health information data element value code Part 3: Demographic and socioeconomic characteristics (WS 364,3-2011, China)
限定语(Qualifier)	SNOMED CT

Table 2– The Number of First 5 Level Classes

Subtrees	Level 2	Level 3	Level 4	Level 5
Disease Diagnosis	24	145	434	183
Organism	4	13	83	9
Anatomy	18	78	45	6
Observable and Examinable Object	27	-	-	-
Clinical Manifestation	5	38	54	1
Diagnosis and Treatment Item, Technique and Method	14	86	258	232
Chemical Drug and Biological Product	25	98	98	-
Medical Equipment, Instrument and Material	11	40	149	-
Substance	9	36	26	7
Psychology and Behavior	3			
Environment and Geography	7	4	10	2
Event, Accident and Disaster	3	25	-	-
Demographic and Socioeconomic Characteristic	2	27	6	-
Qualifier	24	-	-	-

Acknowledgements

This paper was funded by The Chinese Academy of Medical Sciences (Grant No. 2017-I2M-3-014)

References

[1] Fisher H M, Hoehndorf R, Bazelato B S, et al., DermO; an ontology for the description of dermatologic disease, *Journal of Biomedical Semantics 7* (2016), 38-46.

[2] Dramé K, Diallo G, Delva F, Dartigues JF, Mouillet E, Salamon R, Mougin F, Reuse of termino-ontological resources and text corpora for building a multilingual domain ontology: an application to Alzheimer's disease. *J Biomed Inform* 48 (2014), 171-82.

[3] Malhotra A, Younesi E, Gündel, Michaela, et al., ADO: A disease ontology representing the domain knowledge specific to Alzheimer's disease, *Alzheimer's and Dementia* 10.2 (2014), 238-246.

Address for correspondence
Corresponding author : Haixia Sun
Email: sun.haixia@imicams.ac.cn

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190554

The Acquisition of Structured Clinical Data from a Document-Based Electronic Medical Record System

Toshihiro Takeda[a], Dongyao Zhang[a], Shoya Wada[a], Akito Nakagawa[a], Kento Sugimoto[a], Shirou Manabe[a], Yasushi Matsumura[a]

[a] Department of Medical Informatics, Osaka University Graduate School of Medicine, Suita, Japan

Abstract

Our hospital stores all clinical records as Portable Document Formats (PDFs). These PDFs are delivered by each system with a document profile XML file. Using this interface, the items thought to be important for clinical studies are described in the document profile XML and delivered to the data warehouse (DWH). In case clinical data not stored in the DWH are needed, we extract the data from PDF documents. Even from scanned PDFs, the data can be extracted with high accuracy.

Keywords:

Electronic Health Records, Data Warehousing, Information Storage and Retrieval

Introduction

In parallel with the normal electronic medical record (EMR) system, our hospital operates a document-based EMR system called the Document Archiving and Communication System (DACS) [1]. With the DACS, the printed images of all medical records stored as Portable Document Formats (PDFs) are delivered to the DACS database with its document profile XML file. The DACS has a mechanism for promoting the secondary use of EMR data, especially those from departmental systems. In this paper, we outline the methods of transferring clinical data to the data warehouse (DWH) using the DACS.

Clinical data not stored as structured data are often needed in clinical retrospective studies. Since examination reports tend to be described in the same fixed format at each hospital, the data may be extracted from PDF documents. For the data extraction, Ortigosa et al. [2] tried to extract data from electrocardiograms. However, no other studies have evaluated the ease of extracting data from PDF examination reports, especially from scanned reports stored as PDFs. The secondary purpose of our study was therefore to verify the accuracy of the values extracted from examination reports stored as PDFs.

Methods

The Collection of Clinical Data from Departmental Systems

The DACS is configured with a document generator, deliverer, archiver and viewer. The document deliverer transfers the document profile XML to the document archiver. Using this interface, the items thought to be important for clinical studies (important items) are selected beforehand, described in the document profile XML and delivered to the DWH. To assess the current status of data stored through the document deliverer, we analyzed how many records are stored in our DWH in a one-month period (August 1-31, 2018).

Acquisition of Data from Examination Reports by Analyzing the Contents of PDF Files

In case clinical data not stored in the DWH are needed, the contents of the PDF files must be analyzed in order to acquire the data. Since examination reports often have characteristic layouts for each type of examination, the target result values can be identified by the fixed expression described in the examination report.

System Overview

We used the programming language Python to develop our program. From the Python library, we used pdfminer.six for operations involving the PDF contents. In the PDF file, the contents are defined by character strings and the coordinates of their four corners. Our program performs the following processes: identification of the target page, identification and acquisition of the target data, accuracy management of the acquired data, and writing out the acquired data.

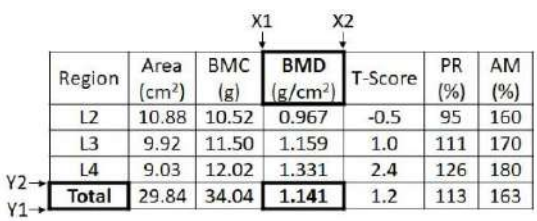
Identification of the Target Page

In order to identify the target page in which the target data is described from multi-page reports, we used the character string existing only on the target page as a marker.

If there was no single marker that existed only on the target page, pages were identified based on multiple markers or markers and their coordinate values.

Identification and Acquisition of the Target Data

If the target data is in the form of a character string in which the descriptions of the same column are invariant is used as a marker. For example, using "patient ID:" as a marker, the program searches for the target character string on the page, and the subsequent numerical value is acquired as the patient ID. If the target data are described in tabular form, a character string in the first column (title column) and in the first row (title row) are set as markers to determine the coordinate at which the target data are described.



Region	Area (cm^2)	BMC (g)	BMD (g/cm^2)	T-Score	PR (%)	AM (%)
L2	10.88	10.52	0.967	-0.5	95	160
L3	9.92	11.50	1.159	1.0	111	170
L4	9.03	12.02	1.331	2.4	126	180
Total	29.84	34.04	**1.141**	1.2	113	163

Figure 1 – Explanation of the method for acquiring tabular data.

Figure 1 shows a sample attempt at acquiring data using this method. The target value was the average bone mineral density (BMD) (target value = 1.141). We set “Total” in the first

column and "BMD" in the first row as markers. "Total" defines the coordinate of the Y-axis (Y1 and Y2), and "BMD" defines the coordinate of the X-axis (X1 and X2). The target data point is defined as the value from (X1, Y1) to (X2, Y2).

Data Acquisition from Scanned PDF Documents

For the scanned PDF documents, we used the optical character recognition (OCR) software program, the ABBYY Fine Reader 14 (developed by ABYY Company). The function ABBYY Hot Folder can convert images of printed text to machine-encoded text. In order to acquire the target data from scanned PDF documents, the device for marker selection, the setting of the acceptable range of coordinate values, and the accuracy management are needed due to the scanning condition and OCR accuracy.

Validation Target

In this study, we focused on a respiratory function test report and an X-ray BMD measurement report.

The respiratory function test report consists of a single page. The target data were the forced vital capacity percent (FVC(%)) and forced expiratory volume one second percent (FEV1.0%). The markers we set were "FVC" and "Ps/Pr" for the FVC(%) and "zissoku (described in Japanese *kanji* characters)" for the FEV1.0%.

The X-ray BMD measurement report is scanned document, so OCR must be used to acquire the target data. We attempted to acquire the BMD values of lumbar spine and femur and their peak reference (PR) values. The target data were the report name, patient ID, examination date, BMD and PR. The X-ray BMD measurement report consists of several pages. To identify the target page, we searched for the character string "f Lumber Spine", "f Left Hip" and "f Right Hip", all of which indicate the examination region, and "Total" with its coordinate. The target data of the report name, patient ID and examination date are described in the form of character strings, so we set the fixed characters "Scan Type", "Patient ID" and "Scan Da" as markers. The target data of the BMD and PR are described in tabular form, so we set "Total", "BMD" and "PR" as markers. To ensure the accurate management, we defined the "BMD" as a 5-digit number over 0.1 and "PR" as an integer over 10.

Results

The Current Status of Data Storage through the DACS

On August 31, 2018, the main system and 21 departmental systems were connected to the DACS, and 1,614 document types totaling 484,479 documents were registered to the DACS from August 1-31, 2018. A total of 9,291 kinds of important items were set in 697 documents from 13 departmental systems. From August 1-31, 2018, a total of 4,636,976 important items from 44,818 documents were stored on the data sharing server and in the DWH.

Acquisition of Data from Examination Reports by Analyzing the Contents of PDF Files

The Analysis of the Respiratory Function Test Report

We analyzed 147 respiratory function reports stored as PDFs. Our system was able to acquire the FVC(%) and FEV1.0% values from every report. We visually confirmed that all values had been accurately acquired.

The Analysis of the BMD Measurement Report

We analyzed the X-ray BMD measurement reports of 598 patients. A total of 2,735 pages were analyzed, with target data written on 1,057 pages. If all target data could not be obtained from the target pages, these pages were treated as errors by our system. The number of such pages that the system treated as error was 165; the reasons for the errors are misidentification of the target page (misrecognition of the marker used to identify the target page, 62 pages), misidentification of the target data (misrecognition of the marker used to identify the target data or of the entire character string of the target data, 73 pages) and the errors by the accuracy management (misreading of the target data by the program, 30 pages). These errors occurred mainly because the OCR software program misread the character string. We confirmed that the values obtained from the remaining 892 reports are all accurate.

Discussion

The EMR system must be designed to accumulate structured data in the DWH for the secondary use of clinical data. It is difficult to store data from every departmental system because of the increased costs. The infrastructure of the DACS to deliver PDF clinical documents from departmental systems to the DACS server with their respective document profile XML files could be used to store such clinical data in the DWH.

Despite our efforts to accumulate a substantial amount of data in the DWH, some clinical data are still not stored in the DWH because of the difficulty of the setting due to manufacturer's technical reasons. The findings of the present study concerning the acquisition of data from PDF reports show that our approach can be used to obtain data for clinical research.

However, markers for each kind of report must still be manually set using our system. Multicenter clinical studies will likely include data in different formats for each facility involved, and marker setting will be required for each format. If markers could be automatically set based on marked target values by machine learning, our program could be applied more widely.

Conclusions

Important items can be stored in our DWH from various departmental EMR systems using a document profile XML file based on the infrastructure of our document-based EMR system. Even data not directly stored in the DWH can be extracted by analyzing the clinical document PDF with high accuracy. The combination of these methods can promote the secondary use of clinical data.

Acknowledgements

This study protocol was approved by the Ethics Review Board of Osaka University Medical Hospital (Approval No. 17042, May 22th, 2017). The authors declare no conflicts of interest associated with this study.

References

[1] Y. Matsumura, N. Kurabayashi, T. Iwasaki, S. Sugaya, K. Ueda, T. Mineno, and H. Takeda, A scheme for assuring lifelong readability in computer based medical records, *Stud Health Technol Inform* **160** (2010), 91-95..

[2] N. Ortigosa and V.M. Gimenez, Raw data extraction from electrocardiograms with Portable Document Format, *Comput Methods Programs Biomed* **113** (2014), 284-289.

Address for Correspondence

Toshihiro Takeda, E-mail: ttakeda@hp-info.med.osaka-u.ac.jp, Facility: Department of medical informatics, Osaka University Graduate School of Medicine, Address: 2-2 Yamada-oka, Suita, 565-0871, Japan, Tel.: 81-6-6879-5900, FAX: 81-6-6879-5903

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190555

Evaluation of a Dental Diagnostic Terminology Subset

Heather L. Taylor [a], Zasim Siddiqui [b], Kendall Frazier [b], Thankam Thyvalikakath [b,c]

[a] *Richard M. Fairbanks School of Public Health, IUPUI, Indianapolis, Indiana, USA*
[b] *Indiana University School of Dentistry, Indianapolis, Indiana, USA;*
[c] *Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, Indiana, USA*

Abstract

The objective of this study was to determine how well a subset of SNODENT, specifically designed for general dentistry, meets the needs of dental practitioners. Participants were asked to locate their written diagnosis for tooth conditions among the SNODENT terminology uploaded into an electronic dental record. Investigators found that 65% of providers' original written diagnoses were in "agreement" with their selected SNODENT dental diagnostic subset concept(s).

Keywords:

Electronic Health Record; Electronic Dental Records; SNOMED-CT

Introduction

Recently, SNODENT (a subset of SNOMED CT since 2013) and DDS (an interface dental terminology) were harmonized [1-3]. The combination of these two terminologies led to the 2017 accreditation of *SNODENT, a dental diagnostic terminology* as the standard by the American National Standard Institute (ANSI) and the American Dental Association (ADA), and the creation of two subsets, SNO-DDS and SNO-DDS General Dentistry [2,3]. To date, *SNODENT, a dental diagnostic terminology*, or any of its subsets, have yet to be thoroughly evaluated for content coverage and completeness. The ADA Practice Institute developed a subset through expert opinion and consensus for use in the dental clinics of the School of Dentistry, University of Detroit Mercy, MI. The subset was developed to facilitate efficient documentation of common dental conditions seen in a general dentistry setting. Therefore, the objective of this study was to determine how well this subset of SNODENT met the needs of dental practitioners.

Methods

A subset of the SNODENT terminology (410 unique concepts), was uploaded into the training module of axiUm 6.x (Exan corporation, Vancouver, BC, Canada), the Indiana University School of Dentistry's electronic dental record (EDR).

We recruited a convenience sample of 20 participants, consisting of six faculty and fourteen third-year and fourth-year dental students. Participants were either full-time clinical faculty or third-year or fourth-year dental students to ensure they were familiar with documenting patient care in the EDR system.

Investigators selected a record describing a dental case-patient (herein referred to as "case-patient"), originally explained in a previous study [4]. The record included the case patient's health history information and oral findings related to periodontal disease and caries. Specifically, the case-patient had poor oral hygiene, generalized gingival inflammation, mesial and distal primary caries on anterior teeth, cracked teeth, secondary caries, extensive decay with pulp exposure, and a periapical radiolucency.

Upon enrollment into the study, participants were asked to review the record and to "think aloud" as they examined the case patient's medical history, dental findings, clinical photographs, and radiographs. The "think aloud" method includes the participant verbalizing their thoughts and actions as they carry out tasks [5-7] The entire session was audio-and video-recorded to capture participants' interactions and thoughts. Each participant thought aloud while reviewing the case-patient record and while writing dental diagnoses for eight teeth and the overall gingival health of the case-patient on a paper form.

Continuing to "think aloud," participants next worked within the treatment planning module of the EDR to locate and select the "best" SNODENT diagnostic terminology(s) for each of their written diagnoses. Afterward, participants rated their satisfaction on their selected SNODENT concept and its ability to represent their original written diagnosis. Specifically, each participant was asked to determine if they were "completely," "partially," or "not at all" satisfied with each SNODENT concept selected within the EDR.

In addition to participant observation, we recorded participants' interactions with the EDR using screen and voice capturing software Camtasia® (TechSmith Corporation, Okemos, MI, USA). At the end of the study session, each participant completed a questionnaire consisting of four Likert scale questions as well as two open-ended questions. The questionnaire was designed to assess the participants' opinions on the use and clinical value of the subset presented to them in the EDR.

Data Analysis

Two investigators (HT & ZS) independently compared each participants' written diagnoses and the corresponding SNODENT concept selected from the subset of SNODENT concepts uploaded into the EDR. Written diagnoses and selected SNODENT concepts were considered in "agreement" if they represented analogous clinical meanings. Conversely, if the written diagnosis and the selected SNODENT concept were different in meaning or intent, the match was labeled as "non-agreement" by investigators. Differences between the

two investigators' classification of "agreement" and "non-agreement" cases were resolved through discussion with a third investigator (TT). Inter-rater reliability between the two investigators was calculated using Cohen's Kappa coefficient. Further, the percentage of SNODENT concept selections deemed in "agreement" with the participant's written diagnoses was calculated. In addition, the percentage of agreement among subgroups (students versus faculty) was compared using a generalized estimating equation (GEE) model for logistic regression.

We calculated the overall percentage of participant satisfaction with their selected SNODENT concept. Students' satisfaction with their selections was compared to faculty' satisfaction using a GEE model for ordinal logistic regression. Regressions were performed with a 95% confidence interval at a p-value of 0.05. The software IBM SPSS Statistics Version 23 (SPSS, Inc., Chicago, IL, USA) was used to perform statistical analysis. Descriptive statistics were utilized to evaluate the Likert scale questions of the end questionnaire. Qualitative content analysis was used to determine any recurring themes in the open-ended responses of the end questionnaire. The recordings were evaluated to determine what barriers, if any, limited the participant's experience using the subset of SNODENT concepts within the EDR.

Results

Twenty participants selected a total of 251 SNODENT concepts (42 unique codes) to diagnose the specified dental conditions of the case-patient. Study investigators compared participants' written diagnosis and the selected SNODENT subset concept(s) for similarity in conceptual representation. Inter-rater reliability between the two investigators was 89%.

Investigators found that 162 (64.5%) of the written diagnostic concepts were in "agreement," and 89 concepts (35.5%) were in "non-agreement" with the participants' corresponding selected SNODENT concept(s).

The subgroup analysis revealed that students selected a total of 169 diagnostic concepts and had 75% "agreement" between their written diagnoses and their selected SNODENT subset concepts, whereas faculty selected a total of 82 diagnostic concepts and had 44% "agreement." The percentage of concept "agreement" among students was significantly higher than faculty (p=0.0270, odds ratio 3.1).

Participants' were asked to subjectively report their satisfaction with the SNODENT concept(s) and its representation of their original diagnosis. They reported "completely" satisfied with 155 (62%) of their selected SNODENT subset concepts, "partially" satisfied with 82 (32.5 %) and "not at all" satisfied with 14 (5.5%) of their selections. No significant differences were observed in satisfaction levels between students and faculty (p=0.54, odds ratio 1.5). Our end study questionnaire revealed that participants perceived value in this particular subset of the SNODENT terminology.

Analysis of the participants' open-ended responses to the end questionnaire revealed issues with the subset of SNODENT concepts and the EDR interface. Participants who were frustrated with the terminology reported that there were too many concepts to search through, too many options with similar meanings, and concepts missing. Regarding the EDR interface, participants noted their frustration with the categorization of certain concepts, the necessity to search for the "exact" concept with correct spelling (the EDR search toolbox offered no suggestions for misspelled words), and the time required to locate concepts.

Video analyses revealed that 75% (15) of participants experienced difficulty finding all of their written diagnoses within the subset of SNODENT concepts. To compensate, participants found substitute diagnostic concepts for the majority of their original diagnoses. For 11 searches, participants were not able to locate a substitute. Both students and faculty had difficulty in determining which SNODENT concept within the subset was the "best" selection, especially among concepts with similar meaning.

Conclusions

The majority of participants in this study agreed that a subset of SNODENT concepts within the EDR could add value to patient care and treatment planning. Investigators found that 65% of providers' original written diagnoses were in "agreement" with their selected SNODENT dental diagnostic subset concept(s). Our findings illuminate the need for continual improvement of dental diagnostic terminologies through revisions and updates. We recommend training on the use of dental diagnostic terminologies for documentation of dental diagnoses, and findings for all dental providers. In addition, an intuitive user interface has a major role in supporting accurate and complete documentation of diagnosis and findings using controlled terminologies.

References

[1] E. Obadan-Udoh, L. Simon, J. Etolue, O. Tokede, J. White, H. Spallek, M. Walji, and E. Kalenderian, Dental Providers' Perspectives on Diagnosis-Driven Dentistry: Strategies to Enhance Adoption of Dental Diagnostic Terminology, Int. J. Environ. Res. Public Health. 14 (2017). doi:10.3390/ijerph14070767.

[2] R.B. Ramoni, J. Etolue, O. Tokede, L. McClellan, K. Simmons, A. Yansane, J.M. White, M.F. Walji, and E. Kalenderian, Adoption of dental innovations: The case of a standardized dental diagnostic terminology, J. Am. Dent. Assoc. 148 (2017) 319–327.

[3] N. Shimpi, Z. Ye, R. Koralkar, I. Glurich, and A. Acharya, Need for diagnostic-centric care in dentistry: A case study from the Marshfield Clinic Health System, J. Am. Dent. Assoc. 149 (2018) 122–131.

[4] T.P. Thyvalikakath, M.P. Dziabiak, R. Johnson, M.H. Torres-Urquidy, A. Acharya, J. Yabes, and T.K. Schleyer, Advancing cognitive engineering methods to support user interface design for electronic health records, Int. J. Med. Inform. 83 (2014) 292–302.

[5] A.W. Kushniruk, and V.L. Patel, Cognitive and usability engineering methods for the evaluation of clinical information systems, J. Biomed. Inform. 37 (2004) 56–76.

[6] J. Joe, S. Chaudhuri, T. Le, H. Thompson, and G. Demiris, The use of think-aloud and instant data analysis in evaluation research: Exemplar and lessons learned, J. Biomed. Inform. 56 (2015) 284–291.

[7] T.P. Thyvalikakath, V. Monaco, H. Thambuganipalle, and T. Schleyer, Comparative study of heuristic evaluation and usability testing methods, Stud. Health Technol. Inform. 143 (2009) 322–327.

Address for correspondence

Thankam Thyvalikakath, DMD, PhD

Indiana University School of Dentistry

Department of Cariology, Operative Dentistry & Dental Public Health

Associate Professor & Director of Dental Informatics Core

1050 Wishard Boulevard, R2206

Indianapolis, IN 46202, Email: tpt@iu.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190556

Using SNOMED-CT to Help the Transition from Microbiological Data to ICD-10 Sepsis Codes

Iris Ternois[a], Typhaine Billard-Pomares[b], Etienne Carbonelle[b], Loriane Franchinard[c], Catherine Duclos[a,b]

[a] Univ Paris 13, Sorbonne Université, INSERM, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances pour la e-Santé, LIMICS, F-93019 Bobigny, France
[b] Hôpital Avicenne, AP-HP, Bobigny, France
[c] Foetal Medicine Department, Hôpital Armand Trousseau, AP-HP, Paris, France

Abstract

Assigning ICD-10 code of sepsis in regard of a pathogenic bacterium found in an haemoculture requires knowledge of microbiology because of the difference of granularity. The aim of this paper is to automate this coding thanks to the use of SNOMED-CT. A dichotomous classification of bacteria causing sepsis has been generated in respect of ICD-10. Our algorithm follows this and explores SNOMED-CT to assign the right ICD-10 code of the sepsis. Applied to a list of 164 bacteria, the system has an error rate of 1.22 %.

Keywords:

Medical coding, International Classification of Diseases, Systematized Nomenclature of Medicine

Introduction

Using medical terminologies to code health care data is an essential process, it allows the standardisation of this data, which is a necessary condition of their re-use in computer processings. The coding procedure is complex and it is difficult to achieve proper coding in terms of completeness or accuracy [1].

The cognitive process to code often implies to conciliate natural language data in clinical documentation and terminology concepts. It requires medical knowledge to find synonyms, to infer aggregation or specialization reasonings, and to get acquainted with the terminology usage rules [2,3].

In the field of infectiology, the available data is often a precise and textual description of the infection, with the identification of the pathogenic agent, easily trackable in lab results. In order to match this description into ICD-10 codes, there is a necessary operation to switch from a thin granularity (bacteria name) to a larger granularity (bacteria group).

To infer this transformation, a knowledge mobilization is necessary. This knowledge may be unavailable, or forgotten. In the domain of microbiology, a reasonning mediator may help to achieve this task [4] . For exemple, to assign the ICD-10 code A415 (Gram negative sepsis) from the result of a positive blood-culture containing Escherichia coli, knowledge about gram coloration is needed.

Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT) terminology covers every field of medicine and depicts bacteria into multi point-of-view hierarchies (cultural characteristic, taxonomy…). The hypothesis is that SNOMED-CT can create the necessary pivot to infer the correct affiliation to a bacteria to a more generic group as described in ICD-10.

The aim of this study is to developp a method to assign the right ICD-10 code for a sepsis, based on the name of the bacteria that caused it, found in the bacteriological results.

Methods

Constitution of a Gold Standard

A list of bacteria was created using an anonymous and agregated collection of positive blood-cultures extracted from the microbiological lab of Avicenne Hospital between January and July 2018. The positive blood-cultures not associated with any antibiotic resistance testing has been excluded. The list is composed of 164 distinct bacteria.

For each bacteria, two microbiologists gave an ICD-10 code for the potentially associated sepsis. The coding was later checked by a specialist in medical information coding. The discrepancies have been jointly corrected. This final list of ICD-10 codes constitutes the Gold Standard.

Construction of a classification of bacteria causing a sepsis according to ICD-10

The classification of spesis causing bacteria follows the logic behind the ICD-10 classification of sepsis. The two ICD-10 classes of interest are A40 (Streptococcal septicaemia) and A41 (Others septicaemia). They describe most cases of sepsis, except some exclusions.

The ICD-10 classification of these two categories have been studied, along with all the exclusions in order to build a dichotomous decision tree for the speis causing bacteria. Every leaf of the tree correspond to a ICD-10 code. This dichotomization allows to push as the last class the unprecise codes (A 408 : "Other streptococcal sepsis" or A418 : "Other specified sepsis"). Furthermore, the code A419 : "Sepsis, unspecified" is excluded because in our study the bacteria are identified from the blood sample, thus the "unspecified" code doesn't apply here.

ICD-10 sespsis coding algorithm

The principle is to find for each bacteria found in blood cultures sufficient information to map to the bacterial point of view of ICD-10 previously described.

For each bacteria tested, the algorithm (developed in Python) browses the preceding tree by verifying the matching with the leaf. There are three different situations.

Codes mentioning a bacteria defined by its species

The leaf describes a named bacteria. For instance, the code A410 is sepsis due to staphylococcus aureus. When the algorithm comes to the leaf "Staphylococcus aureus", it checks if the bacteria is a staphylococcus aureus, and assigns the code A410 if it is the case. This situation is the simplest, the assignation of the code is direct.

Codes mentioning a disease

The exclusions of the classes A40 and A41 can mention diseases associated with sepsis without the name of pathogenic bacteria (e.g. "Sepsis in tularemia"). SNOMED-CT is here used to find the causative agents of diseases. The PyMedTermino [5] module allows to explore these relationships. When the bacteria tested is part of the causative agents, the ICD-10 code is attributed.

Codes mentioning a bacteria defined by a grouping (genus, resiratory mode, etc.)

The leaf describes a group of bacteria. For instance, the code A415 is sepsis due to gram-negative bacteria. In order to check if the bacteria is a gram-negative bacillus, the algorithm exploration is managed with the PyMedTermino module.

First, the SNOMED-CT concept associated with the bacteria tested is retrieved. Then, the list of the parents of this concept is searched. The algorithm verifies the presence of the name of name of the group (e.g. "gram-negative") is found, the algorithm stops at this leaf. Else, the algorithm goes to the next leaf.

Evaluation

The results of the algorithm for the 164 bacteria have been confronted to the Gold Standard. Any difference of coding is a wrong assignation of the code by the algorithm.

Results

The 164 bacteria of the Gold Standard are split into 14 ICD-10 codes. 7 codes are represented by only one bacteria. The code A418 ("Other specified sepsis") contains 42 bacteria and A415 ("Other Gram-negative sepsis") contains 47.

A code has been found for each the bacteria in the list extracted from the lab. Only 2 bacteria were in the category "wrongly assigned code" : Actinobaculum schaalii and Sutterella wadsworthensis. The algorithm has an error rate of 1.22 %.

Sepsis caused by Actinobaculum schaalii was coded A414 by the algorithm, as a sepsis caused by an anaerobic bacteria. The Gold Standard says A418, which describes sepsis caused by gram-positive and not strictly anaerobic bacteria. No information about the aerobic status of this bacteria was found in SNOMED-CT. Sepsis caused by Sutterella wadsworthensis was coded A415 by the algorithm (gram-negative and aerobic bacteria), instead of A414 in the Gold Standard. No information about the respiratory status either anaerobic ou aerobic was found in SNOMED-CT.

Assignation mistakes are due to the absence of essential characeristics in SNOMED-CT for the classification. Furthermore, the concept of anaerobia is not clear and hard to discriminate from aerobia in some cases. For instance, the bacteria "Suterella wadsworthensis" is identified as a micro-aerophilic [6], and thus a sepsis would have been coded A415, like the algorithm returned but was classed as an anaerobic bacteria by the two microbiologists.

Conclusion

We developed a method allowing the coding of a sepsis from to the bacteria found in the haemoculture. This will dispense the coder to search the category of the bacteria and mobilize microbiology knowledge. A study of the impact on the precision of the coding has to be lead to verify the interest of the method, and its eventual scalability in other domains, presenting other caterogies of structured data (drugs treating a disease, procedures to treat a traumatism, etc.).

References

[1] L. Ahmadian, M. van Engen-Verheul, F. Bakhshi-Raiez, N. Peek, R. Cornet, and N. F. de Keizer, The role of standardized data and terminological systems in computerized clinical decision support systems : Literature review and survey, *Int J Med Inform 80* (2011), 81-93

[2] MF. Chiang, JC. Hwang, AC. Yu, DS. Casper, JJ. Cimino, and J. Starren, Reliability of SNOMED-CT Coding by Three Physicians using Two Terminology Browsers, *Amia Symp Proc* (2006), 131-135

[3] F. Bashki-Raiez, NF. de Keizer, R. Cornet, M. Dorrepaal, D. Dongelmans, and MWM. Jaspers, A usability evaluation of a SNOMED-CT based compositional interface terminology for intensive care, *Int J Med Inform 81* (2012), 351-362

[4] C. Duclos, J. Nobécourt, GL. Cartolano, A. Ellini, A. Venot, An ontology of bacteria to help physicians to compare antibacterial spectra, *Amia Symp Proc* (2007), 196-200

[5] JB. Lamy, A. Venot, and C. Dulcos, PyMedTermino : an open-source generic API for advanced terminology services, *Digital Healthcare Empowering Europeans* (2015), 924-928

[6] HM. Wexler, D. Reeves, PH. Summanen, E. Molitoris, M. McTeague, J. Duncan, KH. Wilson, and SM. Finegold, Sutterella wadsworthensis gen. nov., sp. nov., bile-resistant microaerophilic Campylobacter gracilis-like clinical isolates, *Int J Syst Bacteriol.*(1996), 252-8

Address for correspondence

Iris Ternois, LIMICS, 74 rue Marcel Cachin, 93017 Bobigny, France – iris.ternois@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190557

Representing Rules for Clinical Data Quality Assessment Based on OpenEHR Guideline Definition Language

Qi Tian[a,b], Zhexi Han[a,b], Jiye An[a,b], Xudong Lu[a,b,c], Huilong Duan[a,b]

[a] *College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, Zhejiang, China*
[b] *Key Laboratory for Biomedical Engineering, Ministry of Education, China*
[c] *School of Industrial Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands*

Abstract

Data quality assessments (DQA) reveal quality problems in electronic medical records (EHR) data. Generally, DQA methods describe quality rules in programming languages through hard-coding, which limits the implementation of DQA between heterogeneous systems and the interoperability of quality rules. To cover this gap, we conducted a case study applying Guideline Definition Language (GDL) in DQA to assess the quality of patient admission data in an EHR system of a hospital in China.

Keywords:

Data accuracy, electronic health records, programming languages

Introduction

Since electronic health record (EHR) data provide numerous opportunities for clinical research and decision-making, governments and experts advocate for the reuse of EHR data in such practices [1]. Unfortunately, the growing availability of EHR data is accompanied by a growing concern about the quality of EHR data [2]. Data quality assessments (DQA) are able to reveal data quality problems and indicate the level of data quality. Assessing the quality of EHR data is necessary before using such data in research and clinical practices.

Currently, hard-coding data quality rules and accessing data with query languages such as Structured Query Language (SQL) is the general way to evaluate data quality. However, to implement the rules corresponding to DQA requirements the evaluators must know the system's programming. Moreover, since EHR data structures are usually heterogeneous among different scenarios, the interoperability of DQA is difficult to achieve.

OpenEHR is an international health information standard focused on facilitating the universal interoperability of clinical data from EHRs and related systems [3]. The openEHR Guideline Definition Language (GDL) [4] is a language proposed by the openEHR community and used to represent clinical knowledge in clinical decision-making. As GDL is based on the openEHR archetype, a clinical concept model with standard clinical semantics, assessing data quality in GDL is a potentially promising means to facilitate semantic interoperability between different clinical scenarios. The feasibility of applying GDL in DQA, however, has not been investigated. To cover this gap, we conducted a case study applying openEHR GDL to represent the data quality constraints in DQA to assess the quality of patient admission data from an EHR database in China.

Methods

We initially analyzed the data quality requirements from a well-known DQA framework [5] and aligned it with the GDL to identify deficiencies to assessing data quality. Three deficiencies were identified and we extended the original GDL definition to address these deficiencies. First, we introduced new operators since the express operators in GDL were not adequate to describe data quality constraints. Second, new keywords were established in GDL to sufficiently describe data relationship such as *distinct*, *dependency*, and the coding system of an attribute. Third, the relationships of data records and data sets are inadequately expressed in GDL. For this deficiency, we defined new keywords to describe constraints of data sets, such as *count* and *sum*. The details of these extensions are shown in Table 1.

Since SQL is a widely-used query language which operates on data efficiently, we parsed the GDL data quality rules and converted them into SQL to obtain the target data. The workflow of parsing data quality rule is shown in Figure 1:

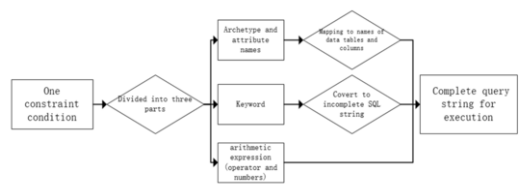


Figure 1. The workflow of parsing data quality rules.

For one data quality rule condition, we first divided it into three parts which are archetype or attribute name, keyword, and arithmetic expression including operator and numbers. Secondly, we mapped the archetype name and attribute name to the corresponding database table name and column name. At the same time, the keyword was converted into a corresponding incomplete SQL string. Finally, we assembled these parsed parts as a complete data query string to retrieve the satisfactory data records and calculate the counts of those records.

Results

The extending of GDL

The operators and keywords related in this case are shown below in Table 1.

Table 1– Instructions of operators and keywords

Type	Operator/ Keywords	Example	Description
Logical	||	RecorderName is not null ll RecorderCode is not null	Or
Logical	Not	Weight is not null	Not
Relational	== / is	WeightUnit is "cm"	Equal
Relational	<	Height < 200	Less than
Relational	<=	Height <= 200	Less than or equal
Relational	>	Heartrate > 0	Greater than
Relational	>=	Height >= 0	Greater than or equal
Keyword	Null	Weight is not null	Incomplete value or data value is null
Keyword	Unique	Entrynumber is unique	Not duplicate
Keyword	Exists in	Weight exists in Patient	Dependency of concepts.
Keyword	Code by	Gender is coded by T120	Assigning coding system of an attribute.
Keyword	Count	Count of patient >2000	Counts of the attributes
Keyword	Sum	Sum of cost <2000	Sum the values of an attribute

Assessing data quality of EHR in a hospital of China

We deployed the extended GDL to evaluate the patient admission data of an EHR database in a Chinese hospital. More than 88,000 person-time visits and 120,445 records of patient diagnosis are included in the extracted data. We evaluated 10 key data fields. The data quality rules and assessment results are showed in Table 2.

Table 2– the details of rules and executed results

Data field or data records	Rule description	Result(%)
Name	Name is not null	99.91
Patient ID	Patientid is not null	100
	Patientid is unique	100
Gender	Gender is not null	99.90
	Gender is coded by	100
Birthdate	Brithdate is not null	99.9
Date of visit	Adimit date/time is not null	100
Visit ID	Encounter identifiers is not null	100
Department of visit	Department is not null	93.94
	Department is coded by	99.9
Type of visit	Admission type is not null	100
	Admission type is coded by	100
Attending doctor	Attending doctor is not null	85.36
Date time of diagnosis	Date of diagnosis is not null	99.99
	Date of diagnosis > Adimit date/time	97.67
Content of diagnosis	Description is not null	97.29
Type of diagnosis	Class of diagnosis is not null	100
	Type of diagnosis is coded by	100
Patient demographics	Patient Demographics is unique	100
Admission	Admission is exists in Patient demographics	99.41
Diagnosis	Diagnosis is exists in Admission	99.56

Discussion

Assessing data quality with hard-coding in a programing language limits the interoperability of data quality rules. The GDL is based on openEHR, an international health information standard, which could promote the semantic interoperability of quality rules. Since the syntax is similar to natural language, GDL is also more readable to clinicians than professional programming languages. This case study shows that GDL is able to assess data quality in clinical practice. However, the quality requirements of this study are limited and this approach should be tested in other clinical scenarios.

Conclusions

In this study, we implemented an extended version of GDL successfully to assess the quality of patient admission data in a Chinese hospital EHR system. Future work should foucs on developing a systematic and professional domain-specific language for describing data quality constraints in clinical DQA.

Acknowledgements

This work was supported by the National Key R&D Program of China [grant number 2016YFC1300300].

References

[1] N.G. Weiskopf, and C. Weng, Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research, *J. Am. Med. Inform. Assoc.* **20** (2013) 144–151. doi:10.1136/amiajnl-2011-000681.

[2]M.G. Kahn, J.S. Brown, A.T. Chun, B.N. Davidson, D. Meeker, P.B. Ryan, L.M. Schilling, N.G. Weiskopf, A.E. Williams, and M.N. Zozus, Transparent reporting of data quality in distributed data networks, *eGEMs* **3** (2015) 1052. doi:10.13063/2327-9214.1052.

[3] What is openEHR?, (n.d.). https://www.openehr.org/what_is_openehr# (accessed October 8, 2018).

[4] Guideline Definition Language (GDL), (n.d.). https://www.openehr.org/releases/CDS/latest/docs/GDL/GDL.html (accessed October 8, 2018).

[5] M.G. Kahn, T.J. Callahan, J. Barnard, A.E. Bauck, J. Brown, D. Meeker, C. Weng, M.N. Zozus, and L. Schilling, A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data, *eGEMs* **4** (2016) 18. doi:10.13063/2327-9214.1244.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190558

NimbleMiner: A Novel Multi-Lingual Text Mining Application

Maxim Topaz [a,b]

[a] *School of Nursing & Data Science Institute, Columbia University, New York City, NY, USA*
[b] *Visiting Nurse Service of New York, New York City, NY, USA*

Abstract

This demonstration showcase will present a novel open access text mining application called NimbleMiner. NimbleMiner's architecture is language agnostic and it can be potentially applied in multiple languages. The system was applied in a series of recent studies in several languages, including English and Hebrew. The system showed good results in terms of text classification performance when compared to other natural language processing approaches.

Keywords:

Data mining; Medical informatics; Natural language processing,

Introduction

Natural language processing (NLP), is an important set of techniques aimed at extracting insignias from narrative health data. However, the vast majority of published NLP studies and existing NLP systems focus on processing information in English, while most of the languages around the world are described as low-resource languages in terms of NLP. More tools and approaches are urgently needed to overcome the resource barrier so advances in NLP can deliver more widespread benefits to health providers, researchers and patients.

Through this poster we will present a novel open access text mining application called NimbleMiner. NimbleMiner's architecture is language agnostic and can be potentially applied in multiple languages. So far, the system was applied in a series of studies in several languages. Examples include identifying fall in formation in clinical notes in English [1] and finding diabetes in clinical documents in Hebrew [2]. The system showed good results in terms of text classification performance when compared to other natural language processing rule-based or machine learning approaches. We will present the system and demonstrate the necessary steps in applying our pipeline. Examples in English and Hebrew will be provided throughout the presentation.

NimbleMiner's Architecture Overview

In general, NimbleMiner 's workflow is as follows (see Figure 1 for graphical depiction): **Stage 1- Language model creation** (Word embedding): The user selects a large corpus of clinical notes and defines word embedding model parameters including word window width (ranges between 1-10) and how many similar terms to present for every term entered by the user (ranges between 5-200). **Stage 2- Interactive rapid vocabulary explorer:** The user provides a query of terms of interest, and the system returns a list of similar terms identified as relevant (based on a combination of a cosine and Levenshtein distances) [1,2]. The user selects and saves the relevant terms. **Stage 3 - Labels assignment and review:** The system uses previously discovered similar terms to assign labels to narrative documents (while excluding notes with negations and other irrelevant terms). The user reviews and updates, when needed, lists of negated similar terms and other irrelevant similar terms. The user reviews the narrative documents with assigned labels for accuracy. **Stage 4 - Machine learning:** The user choses a machine learning algorithm that will be applied to create a predictive model. The model is then applied to predict which narrative document might contain the concept of interest. The user reviews the predicted documents and can repeat stages 2-4 to add new labels.

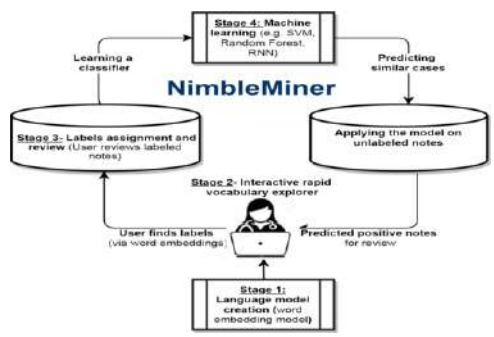


Figure 1– NimbleMiner's Architecture Overview

Follwing are further details on the system's architecture. Throughout the system description, we provide examples of words and phrases indicating patient's fall history, in accordance with a recent case-study [1]. The case-study aimed at identifying whether patients who received treatments from nurses in their homes had a history of falls. The narrative documents were nurses' clinical notes from each patient visit.

Stage 1: Language Model Creation (Word Embedding)

To prepare clinical notes for the word embedding model training, NimbleMiner pre-processes the notes to remove punctuation and lower-case all letters. Additionally, we convert frequently co-occurring words in the clinical notes into phrases with lengths of up to four words (4-grams). Next, NimbleMiner's workflow continues with generating a word embedding vector for all the clinical notes in the corpus. We use a skip-gram model (specifically the word2vec

implementation) to provide NimbleMiner users with an interactive way to rapidly identify words similar to the concepts they want to find in the text.

Stage 2: Interactive Rapid Vocabulary Explorer

The interactive rapid vocabulary explorer process consists of four steps.

Step 1: The user starts with inputting one or more keywords to the system, for example "fell" and "fall". The system suggests the most similar terms that appear in the same context, for example "tripped", "fell down", "had fallen", etc. In our approach, part of the potential similar terms are identified automatically based on the cosine similarity. Cosine similarity ranges between 0-1, with higher similarity being indicative of terms that are more similar. Terms are sorted based on the largest cosine similarity and a list of similar terms for each target word is generated. In our approach, the similar terms are called "simclins" (SIMilar CLINical terms). Simclins are defined as "words or phrases that have high positive predictive value in identifying the concept of interest". If a simclin is present, then the patient should almost always have the condition or a problem we are trying to find. For example, phrases like "pt collapsed" or "she fell down" are considered simclins indicating the presence of fall history. ***Steps 2-3****:* NimbleMiner suggests potentially similar terms and the user selects simclins. We ask the user to choose only definite synonyms (simclins) from the list of similar terms. For each simclin chosen by the user, the system presents the user with random sentences including this specific simclin. For each chosen simclin, NimbleMiner presents users with a list of additional 50 terms and asks to iteratively choose new simclins until the system cannot suggest any more new potentially similar terms. ***Step 4:*** Steps 2-3 are repeated until (1) the user cannot identify any additional simclins based on expertise or literature, and (2) the user finished reviewing all system suggested potential new similar terms. The process of simclin discovery stops at this point.

In summary, NimbleMiner's interactive rapid vocabulary explorer allows users to create large vocabularies of simclins in a very short time. The output of this stage includes a list of simclins curated and selected by the user.

Stage 3: Labels Assignment and Review

Next, NimbleMiner implements the positive-only labels learning framework to prepare the data for the machine learning stage. Previously discovered simclins are treated as positive labels. NimbleMiner uses simclins to identify and label all the clinical notes where the phenomenon of interest is described. Regular expressions are applied to identify vocabulary terms in the notes, labeling cases as "positive" when the term is present, and "unknown" when the term is absent. Negated simclins are labeled as "unknown" using the Negex vocabulary (e.g., notes with expressions like "no falls"). NimbleMiner allows users to add new negations beyond the exiting vocabulary and specify whether the negation should be searched for before or after the term (or both).

Finally, the user can review information and examples of all the clinical notes with simclins and negations. NimbleMiner also offers a summary statistics, including frequencies and percentages of each type of notes. The output of this stage is a list of clinical notes labeled as either positive or unknown. Based on the the positive-only labels learning framework, unknown narrative notes might contain negated simclins.

To create a full training set for machine learning, NimbleMiner extracts all positively-labeled clinical notes and an additional randomly selected corpus of clinical notes without the positive labels (labeled as "unknown"). Both positive-labeled and unknown-labeled corpora are then combined to create a training set that is processed by a machine learning classification algorithm of the user's choice. The output of this stage is a trained machine learning model that can be used to predict weather a narrative document, such as a clinical note, has a phenomenon of interest, for example information about patient's fall history.

Based on the resulting machine learning model, the system generates predictions for the remaining ("unknown") notes and the user can review the notes predicted as "positive" to identify additional simclins and then specify them using the interactive rapid vocabulary explorer (stage 2). New positively-labeled cases (if any) are added to the previous positively-labeled cases and the system goes through another machine learning step with an updated positively-labeled corpus. Users can decide to go through vocabulary explorer and machine learning phases until saturation is achieved (i.e., no additional training is needed, as perceived by the user). When learning is completed, the user can export the positively predicted/labeled cases for further research or clinical purposes.

Method Validation

NimbleMiner was validated and compared against other natural language processing approaches in a series of recent experiments. For example, NimbleMiner outperformed a well performing rule-based system in fall-related information classification tasks. NimbleMiner's overall F- score was 85.8% compared to 81% by the rule based-system [1]. In an additional ongoing project NimbleMiner was applied on a large sample of inpatient clinical notes in Hebrew to identify instances of diabetes mellitus (both insulin dependent type I diabetes and insulin resistant type II diabetes). NimbleMiner achieved high performance measures (average F-score = .94)[2].

Acknowledgements

This work would not be possible without Dr. Ludmila Murga, who created NimbleMiner's user interface. I am also thankful to Dr. Yoav Goldberg (Department of Computer Science, Bar Ilan University, Tel Aviv, Israel) who provided insights into improving NimbleMiner.

References

[1] M. Topaz, L. Murga, K.M. Gaddis, M. V. McDonald, O. Bar-Bachar, Y. Goldberg, and K. Bowles, Mining fall-related information in clinical notes: Comparison of rule-based and novel word embedding-based machine learning approaches, *J. Biomed. Inform.* (2019) 103103. doi:10.1016/J.JBI.2019.103103.

[2] M. Topaz, L. Murga, C. Grossman, D. Daliyot, S. Jacobson, S. Rozendorn, E. Zimlichman, and N. Furie, Identifying diabetes in clinical notes in Hebrew: A novel text classification approach based on word embedding [IN PRESS], in: *Medinfo Proccedings*, 2019.

Address for correspondence

Max Topaz, mt3315@cumc.columbia.edu.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190559

A Search Method to Support Temporal Transcriptome Analysis

Guenter Tusch[a], Shahrzad Eslamian[b]

[a] *Health Informatics and Bioinformatics Program, School of Computing and Information Systems, Grand Valley State University, Allendale, MI, USA,*
[b] *Blue Cross Blue Shield of Michigan, Detroit, MI, USA*

Abstract

Recently, there has been an increasing interest in mining time series databases with a focus on data representation. We propose a hybrid statistical and temporal logic model to yield search pattern that can be used for pattern matching and database search to identify novel and common patterns in temporal expression experiments. The method accounts for various challenges that can be found in publicly available gene expression databases.

Keywords:

Gene Expression Profiling, Time, Database

Introduction

Public gene expression databases make enormous amounts of transcriptomic data available to researchers through the Internet. The most popular databases are GEO (Gene Expression Omnibus) at the NCBI and ArrayExpress at the EBI. NCBI GEO contains more than 105,000 series of gene-expression data collected from over 2.7 million samples and over 19,000 platforms (as of November 2018) and ArrayExpress over 2.3 million assays and some 71,500 experiments. Researchers attempt to translate the wealth of information captured in these rapidly growing biomedical and genomic databases into actionable data for health care or prevention, however the proper treatment of temporal data remains a challenge.

Biologists typically rely on statistical measures to determine trends. Time-course experiments allow them to study the dynamics of transcriptomic changes in cells exposed to different stimuli by investigating pathways or "curated gene sets". Those can be found in the Gene Ontology (http://www.geneontology.org), a framework defining concepts to describe gene function and their relationships, and the Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg/), a database integrating information on genomic and systemic function.

Traditional approaches to find gene pattern across studies are clustering and meta-analysis. Clustering approaches typically are based on GLM and splines to model profiles [1], while a meta-analysis attempts to either analyzing pooled data sets or aggregating p-values (see, e.g., [2]. One popular clustering approach for temporal data is statistical model-based clustering as describes for instance in [1] assumes that data is generated from a mixture model of probability distributions or spline approximations of the data. The mixture parameters can be estimated by maximum likelihood estimation.

A different approach used to model complex time pattern in patient records is Knowledge Based Temporal Abstraction (KBTA) to transform the raw data into a qualitative representation of temporal change based on time intervals, not discrete temporal data. KBTA is the task of summarizing large amounts of time-oriented data using domain-specific knowledge. For KBTA genomic clustering see [2].

The goal of our study is to enable the user to compare his/her own data to public datasets, regardless of the platform used.

Methods

Finding patterns of interest in time series databases (Query by Content) can be described as follows: Given query time series and some similarity measure, find the most similar time series in a given database. The two primary difficulties in this kind of similarity search are time complexity and defining a similarity measure.

Time complexity refers to efficiently finding a sub pattern in a larger pattern sequence. The approach in this paper is based on temporal logic. Finding sequential patterns in a database is implemented through Horn-like rules that can be expressed in terms of Web Ontology Language (OWL) concepts and that can reason about OWL individuals. A query is represented in a logical clause structure and a pattern matcher is used to search for complex sequence patterns of interest in a given database. This could include logical interdependencies between the elements of a sequential pattern including potential constraints as combining individual profiles in pathway or integrative genomics, e.g., genome-wide microarray expression signatures.

With our approach there are some similarities between clustering and database search. A database search for similar temporal pattern can be viewed as finding a temporal cluster that includes the query sequence. The key difference is here the variety of data sources. While clustering is typically applied to data from one experiment and one single platform, the queries under consideration here might include data from different, for example, Affymetrix platforms. While genes from different species can be compared using ortholog databases like PANTHER (pantherdb.org), the data (expression profiles) itself most likely won't fit the same statistical model due to different measurement techniques. The problem exacerbates if we compare profiles stemming from different manufacturers like Ilumina, or even microarray and next generation sequence data.

For our method, in order to increase the signal-to-noise ratio, the first step is signal averaging, i.e., averaging a set of replicate measurements of gene expression values. After averaging and piecewise linear approximation, temporal modeling through KBTA allows for conversion of expression values into an interval-based qualitative representation in a similar way as the well-known SAX method [4]. Change in the proposed method is determined by statistical significance of the paired t-test in consecutive time points. If the difference is significant, the interval is labeled as increasing or decreasing depending on the direction of change. Therefore, for the proposed method the p-values and the direction of the change inform the temporal abstraction. Time

points in temporal gene expression data are not standardized but change from experiment to experiment. If a biological signal is present, it is assumed not to depend on the interval length. Our method allows to compare studies independently from time points or scaling in the time axis or amplitude.

Since gene expression studies measure up to 40K genes or gene products on a single array with very few replicates two issues are critical for our implementation, first, to increase the power of the statistical tests (paired t-test) by empirical Bayes methods, and second, p-value adjustment. A moderated t-statistic shrinks the polled variance by borrowing information across all genes of the particular chip. Since a multitude of genes has to be tested for differential expression on the same data set, the p-value has to be adjusted. We use the False Discovery Rate.

The voom transformation RNA-seq data to be handled the same way as microarray data.

Implementation

Most of the data are preprocessed and stored in a MySQL database for performance reasons. Our plan is to migrate to an Apache Hadoop environment using Apache Hive when the numbers of the available time series data increases significantly. As of November 2018, we found in NCBI GEO 4357 microarray data sets (or 21,320 arrays) and 323 RNA-seq datasets (or 15,539 arrays). We use the GDS format with annotation files, which are normalized, and then extract the data matrix for microarray data. For high throughput sequencing RNA-seq studies data are preprocessed and normalized using standard procedures. We use R Bioconductor (BioC - bioconductor.org) with limma throughout and additionally the voom transformation for RNA-seq data. The necessary databases are accessed using standard BioC tools like GEOmetadb. This implementation has been integrated into the SPOT web application via HTML, JavaScript and PHP that feeds into Protégé. For modelling complex time patterns, we use Protégé (http://protege.stanford.edu) with the help of Allen's temporal logic as has been described in an earlier publication. We used as an example a microarray study with differentially expressed genes grouped into three peak clusters.

Discussion and Conclusions

Our model differs from statistical model-based clustering as it only assumes the general linear model, but does not model the time courses explicitly. This approach allows to cluster gene expression profiles that stem from different platforms, even microarray and nextgen sequencing RNA-seq data sets while still taking the temporal information into account. Our approach also borrows from KBTA. One of the weaknesses of the KBTA approach is that arbitrary ("knowledge based") thresholds have to be determined. In our application thresholds would need to be determined for each platform, manufacturer and technology. The key advantage of our approach is that it implicitly estimates the thresholds from the data itself and uses statistical significance to generate biologically meaningful clusters. The temporal relationship is represented by comparing the means in each pair of groups.

Due to the complexity of the transcriptome and the limited knowledge so far, there are two main approaches to evaluate the feasibility of our method. The first is to rely on existing across platform studies especially with known homologous genes and use the results and determine some form of recall and precision. The second is to perform a simulation study and generate artificial gene expression sets with realistic features. There are a variety of cautionary points: Besides limma, there are other common algorithms to determine significance in time-series gene expression data, as for instance SAM, EDGE or BETR. Typically, a p-value is assigned to an entire gene set or time series, often incorporating a sophisticated weighting schema [4-7] Our approach could be adapted.

If a microarray study or RNA-seq study finds significant gene expression changes, those typically are verified through fluorescent, one-step reverse transcription-polymerase chain reaction (RT-PCR) or quantitative polymerase chain reaction (qPCR). Since so far there are several viable alternative approaches (esp. for RNA-seq data), each study potentially uses a different method to identify differentially expressed genes and similarly for clustering the gene sets as for example STEM, GQL [4-7] or TimeClust [2], the significant genes that our algorithm finds, may not be verified. One solution could be implementing different standard pipelines and giving the user the choice. Given, that we describe an exploratory and not modelling approach, we face the same challenges as other search tools in bioinformatics in terms of how to implement biological knowledge in a computer system.

We will compare ours to those published results and continue to evaluate our program on a representative set of sample studies.

References

[1] D. Lu, Y. Tripodis, L.C. Gerstenfeld, and S. Demissie, Clustering of temporal gene expression data with mixtures of mixed effects models with a penalized likelihood. *Bioinformatics,* **35(5)** (2019) 778–786.

[2] P.H. Sudmant, M.S. Alexis, and C.B. Burge, Meta-analysis of RNA-seq expression data across species, tissues and studies. *Genome biology*, **16 (**2015) 287.

[2] L. Sacchi, R. Bellazzi, C. Larizza, P. Magni, T. Curk, U. Petrovic, and B. Zupan, TA-clustering: Cluster analysis of gene expression profiles through Temporal Abstractions. *International Journal of Medical Informatics* **74** (2005), 505-17.

[4] Y. Shahar, A framework for knowledge-based temporal abstraction. *Artificial intelligence* **90** (1997), 79-133.

[5] C.W. Law, Y. Chen, W. Shi, and G.K. Smyth. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology* **15** (2014), R29.

[6] G. Tusch, O. Tole, and M.E. Hoinski, A Model for Cross-Platform Searches in Temporal Microarray Data. In *Conference on Artificial Intelligence in Medicine in Europe 2015 Jun 17*. Springer Internat. Publ. 2015. (pp. 153-158)

[7] F. Vázquez-Chona , B.K. Song, and E.E. Geisert Jr, Temporal changes in gene expression after injury in the rat retina. *Invest Ophthalmol Vis Sci* **45** (2004), 2737-46.

[8] Z. Bar-Joseph, A. Gitter, and I. Simon, Studying and modelling dynamic biological processes using time-series gene expression data. *Nature Reviews Genetics* **13** (2012), 552-64.

[9] L. Sacchi, C. Larizza, P. Magni, and R. Bellazzi, Precedence Temporal Networks to represent temporal relationships in gene expression data. *Journal of Biomedical Informatics* **40** (2007), 761-774.

Address for Correspondence

Dr. Guenter Tusch, One Campus Drive, D-2-238 MAK, Allendale, MI 49401, USA. Email: tuschg@gvsu.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190560

Design of Biomedical Informatics Framework for Personalized Medicine in Healthcare Organizations

Mohy Uddin

Executive Office, King Abdullah International Medical Research Center, King Saud bin Abdulaziz University for Health Sciences, Ministry of National Guard - Health Affairs, Riyadh, Kingdom of Saudi Arabia

Abstract

To implement personalized medicine effectively at organizational level, it is vital to identify, organize, integrate and leverage multi-dimensional patient data from heterogeneous and distributed resources within an organization. This paper presents the design of a novel informatics framework, to identify, organize and integrate patient's clinical, genomics and environmental data from existing clinical and biomedical resources, and to explore how this patient's data can be leveraged by informatics tools to achieve the goal of personalized medicine.

Keywords:
Personalized Medicine; Data Warehousing; Risk Assessment

Introduction

The new era of life sciences has brought many promising innovations to improve healthcare and personalized medicine is one of them. It is an advancing field of healthcare that is based on each patient's unique clinical, genomics and environmental profiles; thus providing individualized care and treatment using integrated, coordinated and evidence based approach. Using a case study of a hospital and research center, preliminary requirement analysis was performed. The needs of utilizing latest approaches of translational and personalized medicine were observed for the hospital to improve quality of patient care services; and the needs of improved biomedical data storage, analysis and interpretation methods were observed for the research center to develop a centralized biomedical research data platform. Based on these requirements, this paper presents the design of a novel biomedical informatics framework for Integrating Clinical, Genomics and Environmental Data (ICGED), which provides a roadmap for personalized medicine by integrating and utilizing patient's data from existing clinical and biomedical resources in an organization.

Framework

Objectives

Our framework is aimed to serve the following purposes:

- Identification, organization and integration of patient's data from diverse organizational resources
- Development of a centralized platform for data provision
- Data analytics, mining and interpretation
- Transformation of existing data into meaningful information
- Knowledge extraction from the available information

This framework is targeted for the following stakeholders:

- Healthcare professionals for finding patient's information to be used for individualized and tailored care and treatment.
- Biomedical researchers and investigators for utilizing existing information and using it for further research studies.

Information Resources Used by Framework

Our framework is built on holistic approach of utilizing and integrating following existing clinical and biomedical information resources within the organizational structure (see Figure 1). Some main resources identified using our case study includes, but are not only limited to, the following: electronic health records (EHR) / clinical information systems (CIS), ancillary / auxiliary information systems, disease registries, biobank, and bioinformatics research databases.

Categories of Patient's Data Used in Framework

Using available resources, our framework aims to extract following patient's data: clinical, demographics, lifestyle, phenotype and -omics (Figure 1).

Design of Tools Provided by Framework

Our framework suggests the following tools that can integrate and leverage multi-dimensional patient's data from various resources to provide approach of personalized medicine.

Data Integration and Analysis Tools (DIAT)

To implement the DIAT, Informatics for Integrating Biology and the Bedside (i2b2) [1] data warehouse (star schema) can be built by extracting and integrating the clinical and genomics data from different resources, such as clinical information system, biobank and disease registries (see Figure 1).

Risk Assessment and Prediction Tools (RAPT)

To implement the RAPT, data can be extracted from different resources, e.g. EHR and existing patient's disease and genetic registries, (see Figure 1). Using the machine learning algorithms [2], models for disease prediction can be constructed.

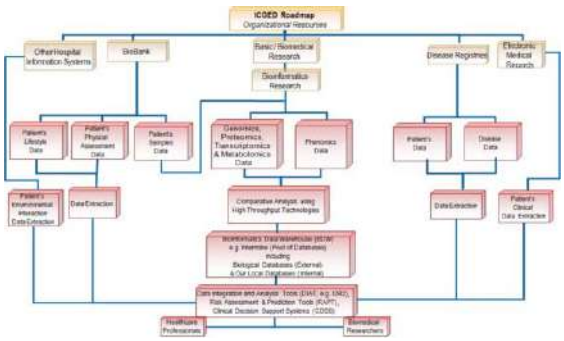


Figure 1 – Design of Informatics Framework for Integrating Clinical, Genomics and Environmental Data (ICGED)

Clinical Decision Support Systems (CDSS)

Using EHR, applications like SMART [3] and customized RAPT (see Figure 1), the decision support features can be implemented and used in the clinical practices.

Bioinformatics Data Warehouse (BDW)

The literature contains several examples of biological data warehouses, e.g. InterMine [4] that can be implemented by integrating various heterogeneous biomedical data formats.

Discussion

Though the future of personalized medicine is promising, but there are several challenges and barriers associated with it. In terms of implementation, there are always challenges of data standardization, quality, accuracy and ownership. Also, the technical issues must be dealt for data implementation aspects, such as extract, transform and load (ETL) process, data interoperability and data governance associated with data warehouses and analysis tools [5]. The revolution of big data is continuously changing the ways of analyzing and interpreting the data in precision medicine [6] that must be considered. In addition, the challenges of data interoperability, data harmonization and fine-grained data access are some other important aspects for future precision medicine [7]. In terms of economics, risk assessment, cost-benefit analysis, insurance policies and reimbursement issues should also be carefully weighed [8]. The consideration of ethical aspects [9], such as security, privacy and confidentiality related to the use of multi-dimensional patient's data is a sine qua non; moreover social and legal questions [9] in using genomic information in patient care practices are still a big challenge for personalized medicine and must be addressed by the future research.

Conclusions

This paper presented a biomedical informatics framework that utilizes and integrates patient's data from existing clinical and biomedical resources in an organizational setting. It provided a novel approach for identification of various patient information resources, extraction of different types of patient's data from these resources, and leveraging the patient's combined profiles through the DIAT, RAPT, CDSS and BDW tools in order to achieve the goal of personalized medicine.

References

[1] S.N. Murphy, G. Weber, M. Mendis, V. Gainer, H.C. Chueh, S. Churchill, and I. Kohane, Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2), *Journal of the American Medical Informatics Association* **17** (2010), 124-130.

[2] P.B. Jensen, L.J. Jensen, and S. Brunak, Mining electronic health records: towards better research applications and clinical care, *Nat Rev Genet* **13** (2012), 395-405.

[3] N. Wattanasin, A. Porter, S. Ubaha, M. Mendis, L. Phillips, J. Mandel, R. Ramoni, K. Mandl, I. Kohane, and S.N. Murphy, Apps to display patient data, making SMART available in the i2b2 platform, *AMIA Annu Symp Proc* (2012), 960-969.

[4] X.M. Fernández-Suárez and M.Y. Galperin, The 2013 Nucleic Acids Research Database Issue and the online Molecular Biology Database Collection, *Nucleic Acids Research* (2012).

[5] D.J. Foran, W. Chen, H. Chu, E. Sadimin, D. Loh, G. Riedlinger, L.A. Goodell, S. Ganesan, K. Hirshfield, L. Rodriguez, and R.S. DiPaola, Roadmap to a Comprehensive Clinical Data Warehouse for Precision Medicine Applications in Oncology, *Cancer informatics* **16** (2017).

[6] T. Hulsen, S.S. Jamuar, A.R. Moody, J.H. Karnes, O. Varga, S. Hedensted, R. Spreafico, D.A. Hafler, and E.F. McKinney, From Big Data to Precision Medicine, *Frontiers in Medicine* **6** (2019).

[7] D. Kalra, The importance of real-world data to precision medicine, *Personalized Medicine* **16** (2019), 79-82.

[8] S.P. Gavan, A.J. Thompson, and K. Payne, The economic case for precision medicine, *Expert Review of Precision Medicine and Drug Development* **3** (2018), 1-9.

[9] R. Chadwick, Ethical issues in personalized medicine, *Drug Discovery Today: Therapeutic Strategies* (2013).

Address for correspondence

Dr. Mohy Uddin

Executive Office, King Abdullah International Medical Research Center

Ministry of National Guard - Health Affairs

Riyadh 11426, Kingdom of Saudi Arabia

Email: drmohyuddin@yahoo.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190561

Characterizing VA Users with the OMOP Common Data Model

Benjamin Viernes, Kristine E Lynch, Brett South, Gregorio Coronado, Scott L DuVall

VA Informatics and Computing Infrastructure, Department of Veterans Affairs, Salt Lake City, UT, USA,
Department of Internal Medicine, Division of Epidemiology, University of Utah, Salt Lake City, UT, USA

Abstract

In 2015, the VA Informatics and Computing Infrastructure, a resource center of the Department of Veterans Affairs, began to transform parts of its Corporate Data Warehouse (CDW) into the Observational Medical Outcomes Partnership) Common Data Model for use by its research and operations communities. Using the hierarchical relationships within the clinical vocabularies in OMOP we found differences in visits, disease prevalence, and medications prescribed between male and female veterans seen between VA fiscal years 2000-17..

Keywords:

Electronic health ecords; systematized nomenclature of medicine; veterans

Introduction

The development and implementation of common data models (CDMs) address some of the logistical challenges of performing research on data generated from disparate healthcare systems by using standardized terminology to express clinical information unambiguously.[1] Even within a single healthcare system, a CDM can be advantageous as the standard representation of clinical concepts improves data quality, allows for rapid and reproducible execution of data management and analysis, and reduces barriers to entry for new researchers to complicated data systems such as electronic health records (EHR). The Veterans' Affairs (VA) is the largest integrated healthcare system in the United States and its EHR is especially complex, with data being sourced from what is analogous to >130 distinct medical systems. Theoretically, the same clinical concept (e.g., blood glucose lab test) may be represented more than 130 different ways, to represent each distinct VA medical facility.

In 2015, the VA Informatics and Computing Infrastructure (VINCI) began transforming parts of the VA's Corporate Data Warehouse (CDW) into the Observational Medical Outcomes Partnership (OMOP) CDM (hereinafter referred to as VA OMOP). VA OMOP version 5.2 contains a large portion of the CDW and allows for less complex, high-level views of many parts of the VA healthcare system. While OMOP is available to VA researchers, to date no one has provided a comprehensive clinical characterization of the VA population using OMOP. Thus, we sought to complete a characterization of veterans who utilize the VA for healthcare services using VA OMOP.

Methods

We compared male and female veterans with respect to visits, conditions, and medications. We utilized the Visit_Occurrence, Drug_Exposure, Condition_Occurrence, Death, and Person fact tables of the OMOP CDM for this analysis. Comparatively, these 5 OMOP tables comprised data transformed from >14 patient level fact tables and a large number of meta-data tables from the CDW data model.

Inclusion Criteria

Using the Visit_Occurrence table, we required patients to have an outpatient visit between fiscal years (FY) 2000 to 2017. Although VA enrollment data was not yet available in VA OMOP, we required patients to have been enrolled in the VA after 10-01-1999 (FY 2000) to ensure that we were only analyzing patients whose VA medical history began during our study period. Patients had to have a valid birthdate (i.e., be 18-114 years of age on the first day of the FY of their first visit), have a defined gender (male or female), and a date of death, if applicable, within or after the FY of their first visit.

Person and Visit Characteristics

We assessed basic patient characteristics using the Person table, a one row per patient table that contains information on demographic factors. We described the VA population using data representing characteristics of race, ethnicity, and deaths. Using the Visit_Occurrence table we calculated the number of outpatient visits, counts of new patients, and counts of returning patients, each overall and by year. The Visit_Occurrence table specified inpatient, outpatient, and emergency room visit types, and for this analysis, outpatient and emergency room visits were combined as outpatient visits. To understand the difference between genders, we compared mean visits per year by gender, percent of all new patients by year and gender, and percent of all patients who are returning patients by year and gender. Finally, we adjusted the mean visits per year per patient by creating rules for patient activity. Assuming that patient activity required constant system use, and based on distributions of visits, a FY was counted as active for a patient if a visit occurred during the year and there was no 1 year no gap preceding a year of no visits.

Condition Occurrence

We extracted the most common conditions by Systematized Nomenclature of Medicine—Clinical Terms (SNOMED CT)

classifications, OMOP's standard vocabulary for conditions. Within the SNOMED hierarchy, many levels were available to potentially evaluate the clinical concepts that described records in the OMOP Condition_Occurrence table. After identifying the most common overall SNOMED concepts in the Condition_Occurrence table, it was determined that the SNOMED 'Disease' concept (i.e., SNOMED code 64572001) and its related descendants was an optimal method for evaluating the burden of different diseases in the VA. This approach created disease classifications with minimal deviation from the standard mappings of International Classification of Disease diagnostic codes, the VA's condition vocabulary, to SNOMED. There were 89 immediate descendants, of the SNOMED disease hierarchy, of which only 74 had been observed in VA from FY 2000-2017. These 74 descendants helped to organize the 13,922 descendant concepts that described diagnoses in the VA OMOP.

We created a frequency table of the SNOMED disease child (e.g., mental disorder), the disease descendant (e.g. depressive disorder), and assessed differences between genders. The most prevalent diseases were then organized by overall patient prevalence during the study period, the most prevalent by each gender, and the comparative rankings for the opposite gender for highly prevalent diseases among each.

Drug exposure

RxNorm is the standard vocabulary to represent drug exposures. We used the RxNorm hierarchy that has been pre-processed and built into the OMOP Concept_Ancestor table to describe outpatient prescriptions at the level of ingredient. The ingredient can be derived from any RxNorm clinical drug form through the hierarchy of the RxNorm vocabulary. Of 7,612 unique RxNorm ingredient values present in the OMOP Concept table, 1,688 RxNorm ingredients had been prescribed in the VA with 1.1 million clinical drug descendants.For the most common prescribed ingredients, we assessed the patient frequency of prescription ingredients and gender differences.

Results

After applying our inclusion criteria, our final analytic sample contained 8,401,080 distinct Person_IDs with at least one outpatient visit from FY 2000-17. The majority (93%) were male, white (65%), and not Hispanic/Latino (77%). For the male population, 66.2% identified as white, and 10.9% identified as black. For the female population, 54.4% identified as white and 24.8% identified as black. Females were statistically significantly younger at age of first visit (39.75) compared to men (56.28). During the 18-year study period, 27% of males died compared to 7% of women.

The most common conditions as defined by SNOMED disease descendants were hypertension and hyperlipidemia with over 50% of all veterans having a recorded diagnosis (see Table 1 for comparison of conditions between genders).The order of most prevalent conditions in women (highest to lowest) was depressive disorder (32.9%), anxiety disorder (28.7%), hyperlipidemia and hypertension. Of the top ten prescribed medications, five were for pain management, two for blood pressure, one for cholesterol, one was an antibiotic, and one for gastroesophageal reflux disease (see Table 2). Female veterans had lower prevalence of many of the most commonly prescribed medications for chronic diseases. This makes sense intuitively due to the fact that on average VA women are younger than men and have lower prevalence of hypertension and diabetes. The top three drugs for women were for pain. Aside from aspirin, men had lower rates of prescription for all pain medications in the top ten.

Table 1. Ten Most Common SNOMED Diseases and Corresponding Rank and Percentage by Gender, VA 2000-17

Disease concept	Percent	Male	Female
1. Hypertension	53.3	1 (55.1)	4 (28.3)
2. Hyperlipidemia	47.9	2 (49.3)	3 (28.4)
3. Presbyopia	25.3	3 (25.7)	8 (19.6)
4. Sensorineural hearing loss, ilateral	22.5	4 (23.8)	62 (6.5)
5. Obesity	22.6	5 (22.2)	5 (27.3)
6. Type 2 diabetes	20.9	6 (21.8)	42 (8.8)
7. Gastroesophageal reflux disease	21.4	7 (21.7)	9 (18.5)
8.Tobacco dependence syndrome	20.9	8 (21.1)	11 (17.4)
9. Depressive disorder	21.9	9 (21.1)	1 (32.9)
10. Nuclear senile cataract	17.2	10 (17.9)	43 (8.6)

Table 2. Ten Most Common RxNorm Drug Ingredients and Corresponding Rank and Percentage by Gender, VA 2000-17

Ingredient	Percent	Male	Female
1.Acetaminophen	33.9	1 (33.5)	1 (39.0)
2.Simvastatin	30.3	2 (31.5)	33 (13.9)
3.Lisinopril	29.8	3 (31.0)	35 (13.4)
4.Omeprazole	26.2	4 (26.3)	4 (24.7)
5.Hydrocodone	22.6	6 (22.3)	3 (25.4)
6.Hydrochlorothiazide	22.3	5 (22.8)	25 (14.8)
7.Ibuprofen	18.4	7 (17.5)	2 (28.8)
8.Amoxicillin	17.5	9 (17.1)	7 (22.4)
9.Aspirin	16.8	8 (17.4)	62 (8.8)
10.Naproxen	16.5	11 (16.0)	6 (23.3)

Outpatient visits increased over time from FY 2000-17. This occurred for both genders, but for women, the number of visits had a higher rate of increase than men. In 2000, patients had an average of 4.84 visits per year with little difference between men and women. By FY 2017, that average had increased to 13.31 visits with women averaging almost three more visits per year than men. Although the majority of new patients each year were male, the relative proportion of new women veterans increased from 5% in 2000 to 9.4% in 2017.

Conclusions

A characterization analysis of VA users is important to best align research questions and methodological approaches to the VA EMR data, and aligns with the educational mission of VINCI. Additionally, by leveraging the hierarchical features of the OMOP CDM, we were able to understand gender differences in ways that would have been more difficult with VA CDW.

Acknowledgements

This work was supported using resources and facilities at the (VINCI), VA HSR RES 13-457.

References

[1] Voss EA, Makadia R, Matcho A, Ma Q, Knoll C, Schuemie M, DeFalco FJ, Londhe A, Zhu V, and Ryan PB, Feasibility and utility of applications of the common data model to multiple, disparate observational health databases, *J Am Med Inform Assoc.* **22** (2015), 553-564.

Address for correspondence

Scott L DuVall,VINCI, Director
Scott.Duvall@utah.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190562

Data Mining in Nursing: A Bibliometric Analysis (1990-2017)

Qian Xiao[a], Jiani Wang[a], Yanling Wang[a], Ying Wu[a]

[a] *School of Nursing, Capital Medical University, Beijing, China*

Abstract

To explore the current trends and future directions of data mining in nursing, we systematically searched English and Chinese databases (from 1990 to 2017) with data mining and nursing related keywords. We found 407 papers, which increased rapidly in the recent five years. Data mining was the most widely used in clinical nursing (50.6%). Chinese papers focused on exploring new nursing knowledge and rules, while English papers focused on promoting nursing practice by data mining.

Keywords:
Bibliometric analysis, Data mining, Nursing informatics

Introduction

Data mining is the process of unearthing information from large datasets using methods at the intersection of machine learning and statistics [1]. Data mining has become useful in assisting the discovery of nursing knowledge from big nursing data and transforming it into an understandable structure for further use [2]. This study aims to explore the current trends and future directions of data mining in nursing.

Methods

We systematically searched English databases (including Pubmed, Embase, and CINAHL), and Chinese databases (including China Biology Medicine, China National Knowledge Infrastructure and Wan Fang Database) with data mining and nursing related keywords. The database search covered the papers published between 1990 and 2017.

We analyzed the changing trend of literature, the main application areas of data mining in nursing, and the comparison between Chinese and international literature using bibliometrics analysis.

Results

We selected 407 papers, of which 112 (27.5%) were Chinese papers and 295 (72.5%) were English papers. The development of the literature presented an increasing trend annually, especially in the past five years (Figure 1). The number of papers from 2013 to 2017 accounted for 50.6% of the total papers. Data mining was the most widely used in clinical nursing (218, 53.6%), followed by community nursing and health management (61, 15.0%), nursing management (50, 12.2%), nursing education (37, 9.1%), and other aspects (41, 10.1%). The comparisons of the top 10 most cited articles showed that Chinese literature focused on exploring the new nursing knowledge and rules, while the English literature focused on promoting nursing practice by data mining. Table 1 lists the high-frequency keywords of data mining in nursing.

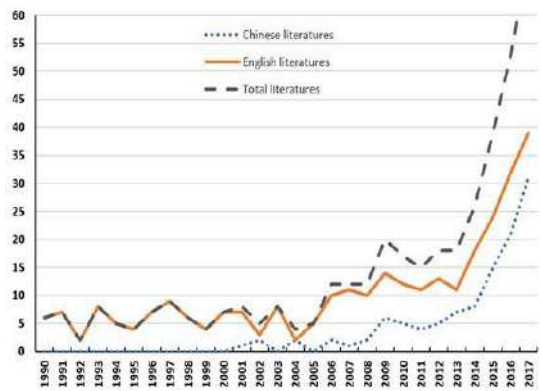


Figure 1– The trends of data mining papers published in English and Chinese databases

Table 1– The top ten high-frequency keywords of data mining in nursing

No.	Chinese literatures		International literatures	
	Keywords	Frequency n (%)	Keywords	Frequency n (%)
1	Data mining	112(100.0)	Data mining	295(100.0)
2	Nursing	98(87.5)	Nursing	233(79.0)
3	Medical informatics	52(46.4)	Nursing records	102(34.6)
4	Decision-making	43(38.4)	Decision-making	77(26.1)
5	Predictive model	36(32.1)	Nursing care	72(24.4)
6	Artificial intelligence	33(29.5)	Nursing informatics	70(23.7)
7	Nursing records	27(24.1)	Risk assessment	53(18.0)
8	Health policy	22(19.6)	Predictive model	51(17.3)
9	Knowledge discovery	17(15.2)	Elderly care	32(10.8)
10	Elderly care	13(11.6)	Nursing practice	11(3.7)

Discussion

The results are discussed in the context of data mining in clinical nursing. The authors offer their perspectives and insights on the results in this section. Potential impacts, plans, and recommendations for future work may also be presented here. It is important for authors to include a discussion of the limitations of their work and potential pitfalls in the interpretation of their results.

Conclusions

The data mining has penetrated different clinical nursing branches and has received attention progressively. The data mining concept is spread and accepted, and more efforts should be initiated to further strengthen the application of data mining in nursing.

References

[1] A. Rocha, R. Camacho, J. Ruwaard, and H. Riper, Using multi-relational data mining to discriminate blended therapy efficiency on patients based on log data, *Internet Interv* **12** (2018), 176-180.

[2] F. Almasalha, D. Xu, G.M. Keenan, A. Khokhar, Y. Yao, Y.C. Chen, A. Johnson, R. Ansari, and D.J. Wilkie, Data mining nursing care plans of end-of-life patients: a study to improve healthcare decision making, *Int J Nurs Knowl* **24** (2013), 15-24.

Address for correspondence

Ying Wu,
School of Nursing, Capital Medical University,
10 Xi-tou-tiao, You-an-men-wai Street, Feng-tai District,
Beijing 100069,
China,
Email: helenywu@vip.163.com,
Phone: +86-10-83911766

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190563

The Hotspots Analysis of Education and Management of Childhood Asthma Based on Cluster Analysis Method

Yana Xing[a], Yibo Wang[b], Wei Zhang[a], Hongmei Duan[a]

[a] *The School of Nursing, Beijing University of Chinese Medicine, Beijing, China*
[b] *Peking University Library, Peking University, Beijing, China*

Abstract

By the use of SPSS software and computer programs, this paper gives a co-word clustering analysis of the articles on education and management of childhood asthma worldwide published in the Web of Science before September 2018. Correlation and dissimilarity matrix and hierarchical clustering were constructed. Finally, 66 high-frequency keywords of 6147 papers were included, 9 hotspots in this field emerged by cluster analysis, which could provide some valuable information of hot research on this field.

Keywords:

Asthma; Co-word analysis; Cluster analysis

Introduction

Asthma is the most common chronic respiratory disease in children, affecting 1-18% of the population worldwide [1]. Recurrent attacks of asthma may lead to irreversible lung function damage and poor quality of life, causing substantial social and economic burden to the whole society [2].

Asthma can be controlled well through effective management and education, although it cannot be cured. However, uptake of treatment and management for childhood asthma remains disappointing, only 14.2% of patients were well controlled according to the Global Initiative for Asthma (GINA) criteria in major cities of China [3]. Suboptimal management and education are vital to asthma control for children.

The co-word analysis method is a form of content analysis technique in bibliometrics, with original information like keywords to include all published articles and produce a good sight into the development of scientific fields. It could reveal the co-occurrence relationships between a pair of keywords in the same context. The more similar the keywords are, the closer the distance [4]. This method was used to analyze the research hotspots and inner construction in the field of education and management of childhood asthma.

Methods

The articles about education and management of childhood asthma were retrieved from Web of Science database before September 2018, by using the subject term (Education OR management) AND (children OR pediatric OR adolescent OR infant OR childhood) AND asthma. As a result, a total of 6357 records were retrieved. After the removal of records of no keywords and abnormal data, 6147 papers were finally included.

Excel, computer program and SPSS 18.0 statistical software were mainly used to process and analyze the data below. In the first step, keywords frequency were counted by computer programs, some general keywords with no practical meaning such as asthma, education, etc. were removed to get more precise results; secondly, a series of mapping rules were formulated to merge the synonyms manually; thirdly, the keywords were replaced according to the above rules, and frequency of keywords were counted once more.

Next, high-frequency keywords were identified to yield the co-occurrence matrix. The value of the matrix represents the co-occurrence of a pair of keywords. Ochiia-coefficient was used to transform the co-occurrence matrix into the correlation matrix. Then, the dissimilarity matrix was constructed by subtracting the value of each element in the correlation matrix from 1. Finally, dissimilarity matrix was imported into SPSS18.0 software to perform hierarchical clustering analysis, a tree diagram was plotted.

Results

In total, 66 high-frequency keywords with frequencies ≥35 were identified after screening (Table 1). Based on characteristics of keywords, 66 high-frequency keywords were extracted for clustering and 9 clusters were formed at last (Figure 1).

Table 1– High Frequency Keywords of Literature（Part）

Keywords	Frequency	Keywords	Frequency
prevalence	794	self-management	331
adolescent	641	health	328
quality of life	638	risk	326
randomized controlled trial	527	guidelines	309
care	502	impact	301
united-states	442	diagnosis	297
symptoms	428	risk-factors	292
outcomes	366	severity	291
adherence	361	lung-function	290

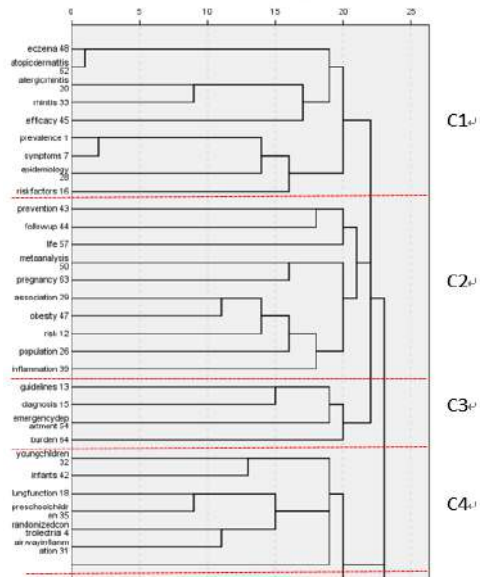


Figure 1–Clustering Results of High Frequency Keywords of Literature (Part)

Discussion

We summed up nine main research subjects in the field of education and management of childhood asthma worldwide, as shown below: epidemiology of allergic diseases in children, maternal obesity in pregnancy and the risk of asthma for offspring, burden of asthma, lung function in infants and preschool children, corticosteroid therapy for childhood asthma, development and validation of asthma related questionnaires, asthma morbidity for inner city children, self-management in adolescent and medication adherence, and food allergy management in children with asthma.

This field covers a wide range of sub-directions of research, involving the epidemiology, prevention, burden, lung function test, medication and food management of childhood asthma. Future research is required to investigate specific directions more deeply and guide the development of nursing practice and disease management.

Conclusions

In conclusion, this paper explored nine research hotspots in the field of education and management of childhood asthma by using co-word clustering analysis method. To some extent, these findings may provide new insight for understanding and predicting the dynamic directions of childhood asthma and could provide some reference and information support for clinical practice and theoretical basis for scientific research.

Acknowledgements

There is no declared of conflict in interest among all authors. Hongmei Duan was responsible for the design of this project and polish the manuscript, Yana Xing was responsible for the literature retrieval, project progression and primary paper writing, Yibo Wang was responsible for computer programming and Wei Zhang was responsible for data collection and subject extraction.

References

[1] Global Initiative for Asthma. 2018 GINA Report: Global Strategy for Asthma Management and Prevention. GINA [Internet]. 2018 [cited 2018 Nov 24]. Available from: https://ginasthma.org/2018-gina-report-global-strategy-for-asthma-management-and-prevention/.

[2] Ding B, DiBonaventura M, Karlsson N, Ling X. A cross-sectional assessment of the prevalence and burden of mild asthma in urban China using the 2010, 2012, and 2013 China National Health and Wellness Surveys. *J Asthma* 54(2017), 632-643.

[3] Ding B, Small M. Disease burden of mild asthma in China. *Respirology* 23(2018), 369-377.

[4] Courtial JP . A coword analysis of scientometricis. *Scientometrics* 31(1994), 251-260.

Address for correspondence

Hongmei Duan, PhD, MD

School of Nursing

Beijing University of Chinese Medicine

No. 11, Bei San Huan Dong Lu, Chaoyang District, Beijing, China

E-mail: duanhm@bucm.edu.cn

Tel: +86 13693124357

Postcode: 100029

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190564

Evaluation of Similar Term Definitions in Medical Device Adverse Event Terminology

Ayako Yagahara[a], Masahito Uesugi[b], Hideto Yokoi[c]

[a] Faculty of Health Sciences, Hokkaido University of Science, Sapporo, Hokkaido, Japan,
[b] Department of Medical Management and Informatics, Hokkaido Information University, Sapporo, Hokkaido, Japan
[c] Department of Medical Informatics, Kagawa University Hospital, kita-gun, Kagawa, Japan

Abstract

The purpose of this study is to extract similar term definitions used in the terminology of Japanese medical device adverse events. We employed Levenshtein and Jaro–Winkler distances as edit distances and Skip-gram, continuous-bag of words, and fast text to produce distributed representations in Word2Vec. A comparison of the accuracies of the models showed that Levenshtein distance had higher specificity whereas Skip-gram had higher sensitivity as compared to the other models.

Keywords:

Machine Learning; Vocabulary, Controlled; Equipment and Supplies, Hospital

Introduction

In Japan, medical facilities and medical device manufacturing companies are required to submit medical device adverse event reports (MDAER) to the Ministry of Health, Labor, and Welfare when medical device adverse events (e.g., catheter breakage) occur during a medical procedure. Since detailed descriptions of adverse events and problems caused to patients by medical devices are documented through free-text in MDAER, they are hard to categorize and conduct statistical analysis.

Therefore, to standardize the terms in MDEAR, medical device adverse event terminology (terminology of Japan Federation of Medical Devices Associations: JFMDA terminology) was published in March 2015 [1]. This terminology consists of 89 medical device terminologies developed by 13 industry groups that are members of JFMDA. Each terminology has three categories: medical device problem, patient problem, and component. In addition, each term in "medical device problem" and "patient problem" categories have definition sentences, synonyms, and FDA code from the Center for Devices and Radiological Health (CDRH) Terminology [2] as shown in Fig. 1. Internationally, between 2009 and 2011, ISO TC 210 worked in conjunction with Global Harmonization Task Force (GHTF) and FDA to produce ISO/TS 19218 - 1 Medical devices-Hierarchical coding structure for medical device adverse events, including both event codes (part 1) and evaluation codes (part 2). In 2011, International Medical Device Regulators Forum (IMDRF) was conceived to accelerate international medical device regulatory harmonization and convergence [3].

The 13 industry groups independently constructed each terminology using a bottom-up approach by gathering the terms used regularly in medical facilities to facilitate communication between medical staff and medical device manufacturers. We are now trying to map these terminologies to ensure consistency. One problem associated with mapping these terminologies includes cases where the notation of terms is considered to be the same concept but differed depending on the terminologies. Because there are about 3,500 terms related to medical device problems, manual verification requires great effort.

Thus, the purpose of this study was to detect synonyms automatically using definition sentences of the terms used in the JFMDA terminology.

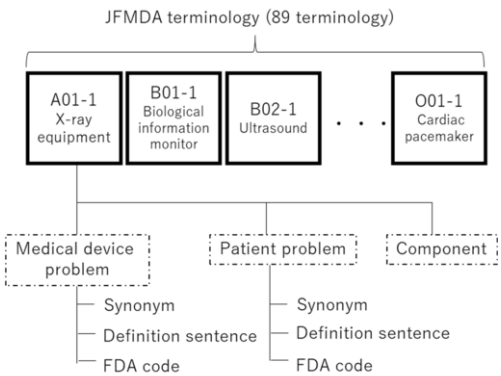


Figure 1–Overview of JFMDA terminology

Methods

Data

Terms used for describing medical device problems were the focus of this study. At first, we extracted the terms and their definition sentences. We made approximately 600,000 definition sentence pairs, of which 125 pairs were extracted arbitrarily.

Similarity Detection

We employed edit distance and Word2Vec as similarity detection methods. Edit distance is an algorithm for quantifying how two dissimilar strings are related to each other by counting the minimum number of operations required to transform one string into the other. Word2Vec is used to group vectors of similar words together into vector space to detect their similarities mathematically.

In this study, we employed Levenshtein [4] and Jaro–Winkler distances [5] as the edit distance and Skip-gram, continuous-bag of words [6], and fast text [7] to produce distributed representations in Word2Vec.

Japanese Wikipedia was used to create a distributed representation model of Word2Vec. In each model, the number of dimensions of the vector was set to 300.

Evaluation

As a gold standard, 50 similar definition pairs and 75 other definitions from the 125 pairs were identified by three experts in medical device safety. Receiver operating characteristic (ROC) analysis was carried out to evaluate the extraction accuracy of similar definition sentences, and area under curve (AUC) was calculated. The cutoff value was identified from the ROC curve using the Youden Index. Sensitivity and specificity were also calculated. ROC analysis was conducted using JMP 13.2.1.

Results

The AUC obtained from the ROC curve analysis and the sensitivity and specificity obtained from the cutoff value are shown in Table 1.

In comparing the models, we observed that both editing distance algorithms tend to have higher AUC and specificity compared to the three Word2Vec models. In particular, the AUC and specificity of the Levenshtein distance algorithm were highest. In Word2Vec, the sensitivity of Skip-gram had the highest score.

For the Levenshtein distance algorithm, the difference in the number of characters of definition sentence pairs was small, and the more common characters they had, the higher was the similarity between them. However, even though the concepts are the same, if the difference in the number of characters of a definition sentence pair is large, the similarity becomes low. Conversely, the accuracy of Skip-gram in Word2Vec did not depend on the number of characters. Therefore, to further improve accuracy, we believe that devising a method that is a combination of both Skip-gram and Levenshtein distance methods is necessary.

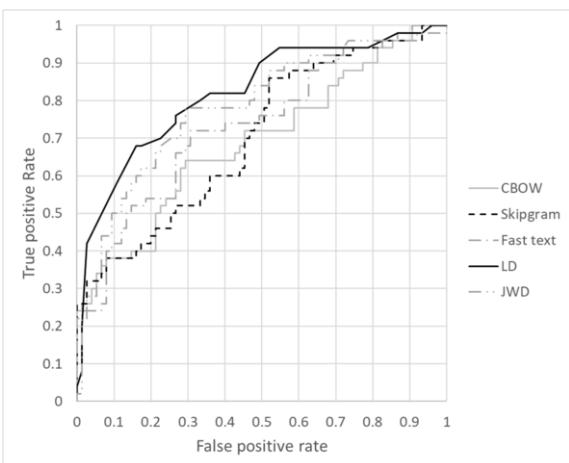


Figure 2–ROC curve

Conclusions

As a result of the experiments performed in this study to extract similar words using definition sentences, we drew the following conclusions: Skip-gram was the most sensitive and Levenshtein distance had the highest specificity and high AUC. To further improve accuracy, we suggest that devising a method that is a combination of both Skip-gram and Levenshtein distance methods is necessary.

Table 1–Comparison of the accuracy of each algorithm

Algorithm		Sensitivity	Specificity	AUC
Edit distance	Levenshtein distance	0.680	0.840	0.821
	Jaro–Winkler distance	0.780	0.707	0.782
Word2Vec	CBOW	0.640	0.707	0.690
	skip-gram	0.860	0.480	0.703
	fast text	0.720	0.693	0.734

Acknowledgements

This research is supported by the Research on Regulatory Science of Pharmaceuticals and Medical Devices from Japan Agency for Medical Research and development, AMED.

References

[1] Ministry of Health, Labour, and Welfare, Publication and Utilization of Medical Device Adverse Event Terminology (in Japanese), (n.d.). http://www.pmda.go.jp/files/000204082.pdf (accessed November 12, 2018).

[2] National Cancer Institute, Centers for Devices and Radiological Health (CDRH) Terminology Files, (n.d.). https://evs.nci.nih.gov/ftp1/FDA/CDRH/About.html (accessed November 12, 2018).

[3] International Medical Device Regulators Forum, About IMDRF, (n.d.). http://www.imdrf.org/about/about.asp (accessed November 12, 2018).

[4] V.I. Levenshtein, Binary Codes Capable of Correcting Deletions, Insertions and Reversals, *Soviet Physics Doklady*. **10** (1966) 707–710.

[5] W. Winkler, String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage, *Proceedings of the Section on Survey Research Methods*. (1990) 354–359.

[6] T. Mikolov, K. Chen, G. Corrado, and J. Dean, Efficient Estimation of Word Representations in Vector Space, in: Proceedings of ICLR Workshops Track, 2013.

[7] A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, and T. Mikolov, FastText.zip: Compressing Text Classification Models, *ArXiv:1612.03651 [cs]*. (2016).

Address for correspondence

Corresponding author: Ayako Yagahara
Mailing address: 7-Jo 15-4-1 Maeda, Teine, Sapporo, Hokkaido, Japan 006-8585
Email: yagahara-a@hus.ac.jp
Phone: +81-11-676-8504

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190565

A Graphical Representation Model for Electronic Health Records: A Preliminary Study

Lin Yang[a], Xiaoshuo Huang[a], Li Hou[a], Qing Qian[a], Jiao Li[a]

[a] Institute of Medical Information and Library,
Chinese Academy of Medical Sciences/Peking Union Medical College, Beijing, China

Abstract

To facilitate experts to find relevant clinical information models, we took a case study of openEHR. We proposed to use a graphical model to represent EHR archetype sets, aiming to optimize clincal information retrieval performance. In this study, we applied our graphic model to 523 OpenEHR archetypes and represented them as a graph with 5,008 nodes and 6,908 edges, which consists of 3,982 term nodes, 504 concept nodes, and 523 archetype nodes. On basis of the graphical model, it improved the performance for retrieving the clinical queries.

Keywords:
Electronic health record; Computer models; Information storage and retrieval

Introduction

A key requirement to achieve semantic interoperability is the standardization of clinical concept representation within electronic health record (EHR) data. The openEHR is a two-level modeling approach, which separates knowledge from information models. It consists of a Reference Model (RM), archetypes and templates. Archetypes define real world clinical concepts, such as "blood pressure", and concept-related data elements at knowledge level. Reuse of these clinical information models (archetype) could help interoperable information be created and shared in health records in a standard manner.

Archetype modeling methodology (AMM), proposed by Moner D[1], shows that searching for reusable archetypes from archetype repositories is an essential part throughout the development process. The openEHR community provides the Clinical Knowledge Manager (CKM)[2] to be a library of openEHR archetypes and templates. It supports the retrieval based on clinical concepts in different section of archetypes. It could help find reusable archetypes[3]. However, experts involved in modeling are mainly concerned about whether the concept name and core data items are covered[3-4]. For better results, end users usually need to do a large amount of preparatory work[3-4].

Since archetypes usually have their own hierarchical structures, and there are semantic relations between them, we considered that graphical representation of these potential knowledge may support clinical information models retrieval. Previous studies show that graphs could efficiently represent clinical knowledge[5]. Moreover, retrieval based on a graphical model, such as Bayesian networks, have been proved to be efficient and reliable[6]. In this paper, we aims to graphical represent of openEHR archetype sets to help identify archetypes that are relevant to a particular information need.

Methods

Archetype feature identification and extraction

As end users are mainly concerned about whether an archetype covers the concept name and core data items, we attempt to use clinical concepts and data elements to represent each archetype. An archetype is expressed in Archetype Definition Language (ADL), and mainly consists of three sections. Header contains a unique identifier for the archetype, and includes some descriptive information, such as concept name. Definition contains the main formal definition of the archetype, including all possible data elements that could be relevant for the clinical concept. Ontology contains the code that represents the meaning of nodes. Moreover, there is a relationship called specialization between archetypes. An archetype is a specialization of another archetype if it mentions that archetype as its parent. Thus, we extracted archetype ID, concept, data elements and parent archetype ID based on ADL files parsing as features (Table 1).

Table 1–Examples of Archetype Features

Archetype ID	**openEHR-DEMOGRAPHIC-ADDRESS.address-provider.v1**	**openEHR-EHR-CLUSTER.exam_urethra.v0**
Concept	Healthcare provider address	Examination of the urethra
Data elements	Address lines Building/complex sub-unit number Address site name Floor/level number ……	No abnormality detected Clinical description ……
Parent archetype ID	openEHR-DEMOGRAPHIC-ADDRESS.address.v1	Null

Graphical representation of archetypes

We used a directed acyclic graph (DAG), which called clinical resource graph, to represent the dependences among archetypes, clinical concepts, and data elements (Figure 1).

The clinical resource graph contains three layers. The first is the term layer. It contains the set of indexing data elements $T = \{T_i, i = 1...M\}$, M being the number of data elements from a given archetype collection. Each term node is linked to its corresponding concept node in the clinical concept layer. The second is the clinical concept layer. It contains the set of indexing concepts $C = \{C_j, j = 1...N\}$, N being the number of concepts. The third layer contains the set of archetypes $A = \{A_k,$

k = 1. . . K}, K being the total number of archetypes in the collection. If A_k is a specialization of another archetype A_p which defines Cj, there is a link joining the concept node Cj and the archetype node A_k.

Thus, this DAG has the set of variables V = TUCUA. The topology avoids connections between nodes in the same layer, and facilitates the inference process.

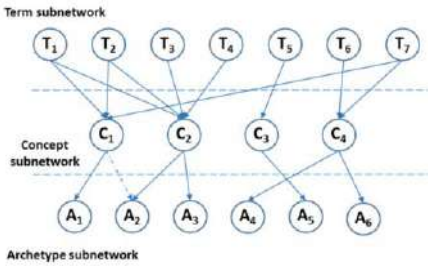


Figure 1–Topology of clinical resources graph

Data source

We downloaded 526 archetypes from CKM on August 30th, 2018. All files were in ADL format. We used the ADL parser (https://github.com/openEHR/java-libs/tree/master/adl-parser) to extract related features described above. In this collection, 3 archetypes did not use English as the description language, so the total number changed to be 523.

Results

Overview of clinical resources graph

We extracted 6504 object nodes from the collection, including 5981 data elements and 523 concept nodes. After removing the the duplicated ones, we got 3,560 and 504 respectively. Besides, there were 31 specialized archetypes, involving 20 parent archetypes while 8 of them were no longer in CKM. Therefore, the clinical resources graph had 5,008 nodes and 6,908 edges, which consisted of 3,982 term nodes, 504 concept nodes, and 523 archetype nodes. And 12 concept nodes had specialized relations with archetype nodes.

Evaluation of the performance

We formulated two test queries : Q-1 was “admit date/time, attending doctor, clinical description, discharge date/time”; and Q-2 was “report, test name, test results”. Then, we manually annotated all 523 archetypes, according to their relevance to each query, as the ground truth. To validate our clinical resources graph in supporting retrieval, we used retrieval model based on Bayesian Network proposed by Acid S[6]. Meanwhile, CKM and BM25F were selected as baseline methods. BM25F is an extension of the BM25 ranking function, which is applicable to structured documents consisting of multiple fields. We supposed that an archetype was decomposed into two fields: concept and data elements, and used the function proposed by Zaragoza H [7].

We evaluated the result in terms of Precision at cut-off points 1 (P@1) and 3 (P@3), as well as Recall (R). The result (Table 2) showed that our method was better than the two other methods. For instance, for Q-2, our method, BM25F and CKM achieved a P@3 of 1, 0.67 and 0.33, respectively. Besides, our mehtod could identified the specialized archetype “openEHR-EHR-COMPOSITION.report-result.v1”, which BM25F could not find. All these indicated that it was effective to use a graphical representation of relationships among data elements, concepts and archetypes to support search.

However, there were some important limitations. First, our approach could not solve the basic problem of semantic interoperability in EHRs, which must be solved from the perspective of the business domains the concepts originally belong to. Second, it was not clear that whether our method could be applied to other clinical information models. Third, our method could not find semantic associations when archetypes had few data elements, and terms used were totally different.

Table 2– Results of the experiment

	Q-1			Q-2		
Method	**P@1**	**P@3**	**R**	**P@1**	**P@3**	**R**
CKM	0	0	0	1	0.33	0.89
BM25F	1	0.67	1	1	0.67	0.16
Our method	1	0.67	1	1	1	1

Conclusions

To facilitate clinicians to find relevant clinical information models, we took a case study of openEHR. We tried to graphical representation of archetype sets to optimize existing retrieval results of CKM. The result showed that our method could support efficient search. This is a preliminary study. In future work, We will optimize the existing method according to its shortcomings, and validate its feasibility on other clinical information models.

Acknowledgements

This research is supported by Chinese Academy of Medical Sciences (Grant No. 2017PT63010, 2018PT33024), the National Key R&D Program of China (Grant No. 2016YFC0901901, 2017YFC0907503).

References

[1] Moner D, Maldonado J A, Robles M, Archetype modeling methodology. *J Biomed Inform* 79 (2018), 71-81.
[2] openEHR clinical Knowledge Manager. http://www.openehr.org/ckm/.
[3] Teodoro D, Sundvall E, João M J, et al, ORBDA: An openEHR benchmark dataset for performance assessment of electronic health record servers. *Plos One* 13 (2018), e0190028.
[4] Maranhã£O P A, Bacelar-Silva G M, Dng F, et al, Nutri-genomic Information in the openEHR Data Set. *Appl Clin Inform* 9 (2018), 221-231.
[5] Rotmensch M Halpern Y , Tlimat A , et al, Learning a Health Knowledge Graph from Electronic Medical Records. *Sci Rep* 7 (2017), 5994.
[6] Acid S, Campos L M D, Juan M. Fernández-Luna, et al, An information retrieval model based on simple Bayesian networks. *International Journal of Intelligent Systems* 18 (2003), 251-265.
[7] Zaragoza H, Craswell N, Taylor M, et al, Microsoft Cambridge at TREC 2004: Web and HARD track. *Proceedings of TREC-2004* 2004.

Address for correspondence

Jiao Li, Institute of Medical Information, Chinese Academy of Medical Sciences, 3rd Yabao Road, Chaoyang District, Beijing 100020, China. Email: li.jiao@imicams.ac.cn

This page intentionally left blank

II. Supporting Care Delivery

This page intentionally left blank

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190567

Using Inpatient Portals to Engage Family Caregivers in Acute Care Setting : A Literature Review

Bader Alshoumr, Ping Yu, Tingru Cui, Ting Song

Centre for IT Enabled Transformation, School of Computing and Information Technology, Faculty of Engineering and Information Sciences, University of Wollongong, Wollongong, NSW 2522, Australia

Abstract

This study aims to review the literature evaluating the design, acceptance, use and usability of inpatient portals for family caregivers. Fourteen articles were included. Information about patient medication, lab results and healthcare team was most important. Most caregivers were satisified and perceived the ease of use and usefulnesss of portals. Further research is needed to collect sufficient evidence with regard to the impact of inpatient portals on patient engagement and quality of care.

Keywords:
Patient portals, caregivers, Quality of Health Care

Introduction

An inpatient portal is a secure website application through which patients can access the health service information, perform online health- and administration-related transactions, communicate directly with providers, and access clinical information during hosptalization [1]. Family caregivers might need to act on behalf of patients during hospitalization due to the patient's disability or very young age. Therefore, they should also be offered access to an inpatient portal. In this review, we aim to assess the use and usability of the portal for family caregivers within an acute care setting. We also explore the information and design recommendations needed for effective engagement.

Methods

PubMed, Scopus, and CINAHL databases were searched for articles published between 2008 and 2018. We used three main terms to identify eligible articles: 1) engagement, 2) patient portal, and 3) acute care or hospital. We followed PRISMA guidelines to screen potential articles in the literature. We excluded articles which were either not written in English, were commentaries, or were disccussion papers.

Results

Of the 731 articles identified through database search and reference review, 14 articles were found to meet the final inclusion criteria (see Fig.1). The study design was a mixture of quantitative and qualitative methods. The main outcomes reported were family caregivers' information needs, portal design recommendations, acceptance, use and usability. Most studies were performed at multiple tertiary teaching hospitals while only four studies were performed at children's hospitals [2–5]. 78% of studies were conducted in the United States, two studies in Canada [2,6], and one study in South Korea [7].

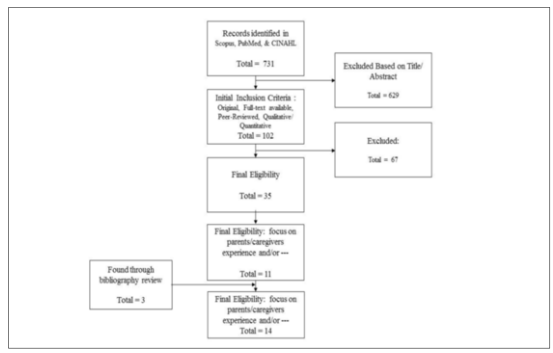


Figure 1– Article selection flow chart adapted from Kelly MM, Coller RJ, Hoonakker PLT. Inpatient Portals for Hospitalized Patients and Caregivers: A Systematic Review. J Hosp Med [8]

Caregivers' Information Needs

Three studies reported the desire of family caregivers to be offered real-time access to the patient's clinical information [8–10]. One study stated the importance of adding the patient's care plan on the portal during hospitalization [9]. Another study reported the preference of caregivers to view the health care team profiles online [10]. Two studies recommended offering detailed information about upcoming trials, uploading signed consent forms, and adding general information about the patient's transition to home care [10,11].

Portal Design Recommendations

Three studies provided design suggestions and recommendations. Caregivers suggested allowing them to customize the display of information based on patient's condition, presenting lab results as graphs instead of numbers [10] and adding functionality to notify them about upcoming clinical events and new lab results [8]. Others recommended to use the same themes, designs and functionalities in both, the inpatient and outpatient portals [12], to apply the bring your own device (BYOD) strategy inside the hospital, and also allow out-of-hospital access to the portal [12].

Caregivers' Portal Acceptance

Generally, family caregivers had positive perceptions toward the use of inpatient portal during their family members' hospitalization. They perceived a moderate satisfaction and

utility with the use of secure messaging in a rehabilitation care setting [6]. Parents of children who were admitted to either a medical or surgical unit at a pediatric hospital found the real-time visualization of the child's clinical information useful in tracking the child's progress and making critical decisions during hospitalization [5,13]. Some parents thought that the use of the portal during hospitalization might also enhance patient safety [5].

Caregivers' Portal Use

Three studies provided quantitative data about the use of an inpatient portal by caregivers. The first study reported that 90% of 300 caregivers at a tertiary children's hospital accessed the portal and used all of its functionalities [5]. Another study reported that 84% of 120 caregivers at a tertiary hospital used a web-based inpatient portal for less than 4 days, and 16% continued to use it for 5 to 10 days [9]. The last study found that the average caregivers' use of the portal was 2.5 times/month during 14 months of the study at a children's rehabilitation hospital [6].

Portal Usability

Two studies performed a usability evaluation test for caregivers in two different contexts; oncology [2] and neonatal care [9]. The tools examined were a web-based toolkit for oncology and intensive care [2] and a parent decision support tool for neonatal care [9]. Findings from both studies indicated that these tools were useful, efficient, and effective for engaging caregivers in the patient care process.

Conclusions

The overall findings of this review indicate that an inpatient portal has a great potential to enhance patient engagement in health care during hospitalization. However, study in this area is still in its early phases. Previous studies were mainly focused on exploring the design requirements or assessing the feasibility of the portal. Future work should evaluate the impact of using the portal with particular focus on caregiver's engagement, patient safety, and the quality of care.

References

[1] D.K. Ahern, S.S. Woods, M.C. Lightowler, S.W. Finley, and T.K. Houston, Promise of and potential for patient-facing technologies to enable meaningful use, *Am. J. Prev. Med.* **40** (2011). doi:10.1016/j.amepre.2011.01.005.

[2] S.A. Weyand, M. Frize, E. Bariciak, and S. Dunn, Development and usability testing of a parent decision support tool for the neonatal intensive care unit., *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2011** (2011) 6430–3. doi:10.1109/IEMBS.2011.6091587.

[3] L. V Grossman, S.W. Choi, S. Collins, P.C. Dykes, K.J. O'Leary, M. Rizer, P. Strong, P.-Y. Yen, and D.K. Vawdrey, Implementation of acute care patient portals: recommendations on utility and use from six early adopters, *J. Am. Med. Informatics Assoc.* **25** (2017) 1–10. doi:10.1093/jamia/ocx074.

[4] M.M. Kelly, S.M. Dean, P. Carayon, T.B. Wetterneck, and P.L.T. Hoonakker, Healthcare Team Perceptions of a Portal for Parents of Hospitalized Children Before and After Implementation, *Appl. Clin. Inform.* **8** (2017) 265–278. doi:10.4338/ACI-2016-11-RA-0194.

[5] M.M. Kelly, P.L.T.T. Hoonakker, and S.M. Dean, Using an inpatient portal to engage families in pediatric hospital care, *J. Am. Med. Informatics Assoc.* **24** (2017) 153–161. doi:10.1093/jamia/ocw070.

[6] G. King, J. Maxwell, A. Karmali, S. Hagens, M. Pinto, L. Williams, and K. Adamson, Connecting families to their health record and care team: The use, utility, and impact of a client/family health portal at a children's rehabilitation hospital, *J. Med. Internet Res.* **19** (2017) e97. doi:10.2196/jmir.6811.

[7] B. Ryu, S. Kim, K.-H. Lee, H. Hwang, and S. Yoo, Inpatient satisfaction and usage patterns of personalized smart bedside station system for patient-centered service at a tertiary university hospital, *Int. J. Med. Inform.* **95** (2016) 35–42. doi:10.1016/j.ijmedinf.2016.09.003.

[8] S.K. Bell, S.D. Roche, A.C. Johansson, K.P. O'Reilly, B.S. Lee, K.E. Sands, D.S. Talmor, and S.M. Brown, Clinician perspectives on an electronic portal to improve communication with patients and families in the intensive care unit, *Ann. Am. Thorac. Soc.* **13** (2016) 2197–2206. doi:10.1513/AnnalsATS.201605-351OC.

[9] A.K. Dalal, P.C. Dykes, S. Collins, L.S. Lehmann, K. Ohashi, R. Rozenblum, D. Stade, K. McNally, C.R.C. Morrison, S. Ravindran, E. Mlaver, J. Hanna, F. Chang, R. Kandala, G. Getty, and D.W. Bates, A web-based, patient-centered toolkit to engage patients and caregivers in the acute care setting: a preliminary evaluation., *J. Am. Med. Inform. Assoc.* **23** (2016) 80–7. doi:10.1093/jamia/ocv093.

[10] E. Kaziunas, D.A. Hanauer, M.S. Ackerman, and S.W. Choi, Identifying unmet informational needs in the inpatient setting to increase patient and caregiver engagement in the context of pediatric hematopoietic stem cell transplantation, *J. Am. Med. Informatics Assoc.* **23** (2016) 94–104. doi:10.1093/jamia/ocv116.

[11] Y.S. Lee, C. Garfield, N. Massey, S. Chaysinh, and S. Hassan, NICU-2-HOME: supporting the transition to home from the neonatal intensive care unit using a mobile application, in: Proc. CHI EA '11, 2011: pp. 2257–2262. doi:10.1145/1979742.1979947.

[12] S.A. Collins, R. Rozenblum, W.Y. Leung, C.R. Morrison, D.L. Stade, K. Mcnally, P.Q. Bourie, A. Massaro, S. Bokser, C. Dwyer, R.S. Greysen, P. Agarwal, K. Thornton, and A.K. Dalal, Acute care patient portals: a qualitative study of stakeholder perspectives on current practices, *J. Am. Med. Informatics Assoc.* **24** (2017) 9–17. doi:10.1093/jamia/ocw081.

[13] M. Maher, E. Kaziunas, M. Ackerman, H. Derry, R. Forringer, K. Miller, D. O'Reilly, L.C. An, M. Tewari, D.A. Hanauer, and S.W. Choi, User-Centered Design Groups to Engage Patients and Caregivers with a Personalized Health Information Technology Tool, *Biol. Blood Marrow Transplant.* **22** (2016) 349–358. doi:10.1016/j.bbmt.2015.08.032.

Address for correspondence

Ping Yu, Centre for IT-enabled Transformation, School of Computing and Information Technology, Faculty of Engineering and Information Sciences, *University of Wollongong, Wollongong, New South Wales, 2522, Australia,* Email: ping@uow.edu.au.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190568

Quick Cognitive Impairment Test for Cancer Patients Using Emotional Stroop Effect

Eiji Aramaki[a], Chihiro Honda[a], Shoko Wakamiya[a], Akira Sato[b], Isao Myashiro[b]

[a] *Nara Institute of Science and Technology (NAIST), Nara, Japan,*
[b] *Osaka International Cancer Institute, Osaka, Japan*

Abstract

Recent studies have attributed impaired cognitive function in cancer patients, or Cancer Related Cognitive Impairment (CRCI), to various causes. CRCI screening is vital for guiding important decisions about treatment options. This study investigates the emotional Stroop-test-based CRCI screening, examining response time when naming the colors of negative emotional words. Cancer patients (n=17) participated in two tests: (1) the Stroop task; (2) State-Trait Anxiety Inventory. Results suggest that Stroop-based CRCI screening is feasible.

Keywords:

Cancer, Cognitive Assessment Screening Instrument, Stroop Test

Introduction

With advancing technology in the medical field and increasing treatment options available, it is important for patients to engage in decision making with intact cognitive processes. However, recent reports have described Cancer Related Cognitive Impairment (CRCI), a phenomenon in which the cognitive function of cancer patients might be impaired. The risk factors for CRCI remain unclear, but research suggests causes such as chemotherapy, surgery, radiotherapy, hormone therapy, cancer-related inflammation, physical and mental stress, and attentional fatigue [1]. Although various pathways have been proposed, little is known about CRCI mechanisms. Studies also indicate that the CRCI is a common phenomenon: 75% of cancer patients potentially have CRCI [1].

To elucidate CRCI mechanisms, this study proposes a quick measurement based on a Stroop task [2]. A Stroop task is a widely known psychological method named after John Ridley Stroop. The Stroop phenomenon illustrates the difficulty of naming the color of a word when a mismatch exists between the color and the word, such as the word GREEN printed in red color. Many variations on the Stroop task use stimuli other than colored words. One variation applied to clinical areas uses the word set related with the anxiety of participants. This Stroop task is called an Emotional Stroop task. Such a test works by assessment of the response time of the participant when naming colors of negative emotional words. For example, depressed participants are slower at naming the color of depressing words than doing so for non-depressing words. Non-clinical subjects have also been shown to name the color of an emotional word (e.g., "war," "cancer," "kill") more slowly than when naming the color of a neutral word. This effect is called an emotional Stroop effect [3-6]. Particularly, the cancer-related emotional Stroop effect is called the cancer Stroop effect [7-9].

The proposed quick cognitive impairment screening is based on the cancer Stroop effect, specifically, the degree to which a small word set can capture the cancer Stroop effect.

Materials and Methods

We designed a three-minute screening method using 20 words (10 for cancer-related words and 10 for natural words).

Participants: Cancer patients were recruited at the venue of a classical live music concert held at the Osaka International Cancer Institute on May 12, 2018. Seventeen completed all the trials and questionnaires before and after the concert. All participants were Japanese and fluent Japanese language speakers.

Evaluation Metrics: This study investigated patients to ascertain their (1) anxiety and (2) attention.

- **Anxiety: State-Trait Anxiety Inventory:** The State-Trait Anxiety Inventory (STAI) developed by Spielberger et al [10] is a broadly used measure to assess the states of anxiety (A-State) and trait anxiety (A-Trait). Toccafondi et al [11] found that cancer patients' A-State scores in this scale improved after attending an hour-long live music concert, although A-Trait did not change significantly. For this study, A-State, translated into Japanese [12], was used to assess participants' states of anxiety. The 20-item questionnaire used 4-point scales (1-4).
- **Attention: Cancer Stroop Task:** The Cancer Stroop Task consisted of 10 cancer-related words and 10 neutral words chosen for earlier research [7] investigating attentional biases of women who had a family history of breast cancer. This study examines patients without specific cancer type. Therefore, breast-cancer-specific words used in the earlier study were replaced with general cancer-related words. Our preliminary experiments found no significant difference between word types in terms of word length, frequency, or character strokes. A Stroop task application was developed (Apple iPad; Apple Corp.). The words were presented at the center of a black screen in one of four colors: red, blue, green, or yellow. While the word was presented, four gray boxes at the bottom of the screen displayed the name of each color written in black in each box (*Figure 1*).
- Participants were instructed to ignore the meaning of the displayed words and tap one box that corresponded to the color of the word as quickly as possible. In addition, to begin each trial at a fixed distance to the color options, participants were asked to tap a gray circle with the word 'Next' appearing at the center of the screen after answering each trial. The word order of the presentation was randomized. The reaction time for each word and incorrect answers were recorded and saved in the device as a screenshot at the end of each session.

Table 1. Cancer Words and Normal Words

	CANCER WORDS	NEUTRAL WORDS
1	生検 (Biopsy)	足首 (Ankles)
2	がん (Cancer)	小冊子 (Booklet)
3	化学療法 (Chemo)	噴水 (Fountain)
4	白血病 (Leukemia)	家具 (Furniture)
5	腫れ (Lump)	公平 (Impartial)
6	悪性 (Malignant)	雑誌 (Magazine)
7	緩和ケア (Palliative Care)	ピッチャー (Pitcher)
8	医療用麻薬 (Medical drug)	粉 (Powder)
9	放射線治療 (Radiation)	着実 (Steady)
10	腫瘍 (Tumor)	衣装部屋 (Wardrobe)

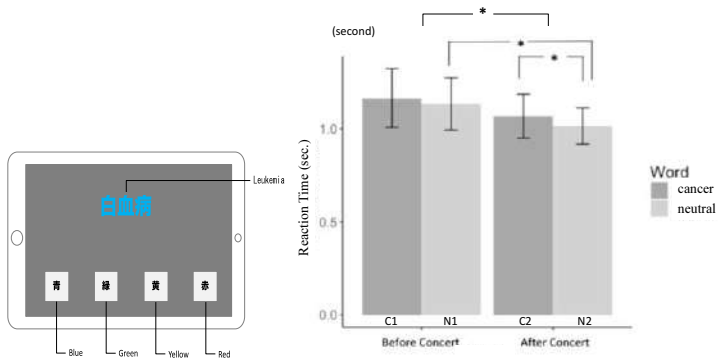


Figure 1. Stroop application made for use on an Apple iPad (left) and Interaction between word type and time condition (right).

Procedure: The Stroop task and STAI A-State questionnaire were completed before and after the monthly live music concert held at the Osaka International Cancer Institute on May 16, 2018. Before the concert, participants completed STAI A-State questionnaire after providing written informed consent to participate. Then, participants received instruction and practiced the Stroop task, doing 20 trials consisting of 10 practice words. After the practice session, participants completed the main trials (80 trials; 20 wrods * 4 times). They attended an hour-long classical music concert, after which they completed the questionnaire and main Stroop trials again.

Paired student's *t*-tests were used to compare mean STAI A-State scores before and after the concert. Latency exceeding 3 SDs from the mean were excluded from analysis. Study protocol was approved by the Medical Ethics Committee of Osaka International Cancer Institute (#18009) April 27, 2018.

Results and Discussion

The participants were all cancer patients (they provided their hospital identification numbers). Pparticipants included 2 men and 15 women, with ages of 45–65 (mean=69.1, SD =9.23). Anxiety before and after (A1 vs. A2): Anxiety after the concert (mean=25.5, SD=5.8) was significantly lower than that before the concert (mean=33.4, SD=9.1) in *t*-test. Music concerts significantly reduced the patients' anxiety. We investigated the music concert effects and the Stroop effect. Various comparisons exist for the Stroop effects (*Figure 1*).

Total comparison between results obtained before and after (C1+N1 vs. C2+N2): Average latency (or Reaction Time; RT) for cancer words was significantly longer (mean=1.14, SD= 0.28) than that for neutral words (mean=1.04, SD=0.20). The Neutral word comparison between results obtained before and after (N1 vs. N2): The neutral words RT after the concert (mean=1.01, SD=0.18) was significantly shorter than that before the concert (mean=1.13, SD=0.26). Comparison between after cancer and after neutral (C2 vs. N2): After the concert, the cancer word RT (avg.=1.06, SD=0.22) was significantly longer than the neutral word RT (mean=1.01, SD=0.18) (*Figure 1*). No other relations were statistically significant.

Acknowledgments

Supported by JST CREST (JPMJCR16E2), JSPS (JP16H06395 and JP16H06399) Japan.

References

[1] Janelsins, M.C., et al., Prevalence, mechanisms, and management of cancer-related cognitive impairment. *Int Rev Psychiatry*, 2014. **26**(1): p. 102-13.
[2] Stroop, J.R., Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 1935. **28**: p. 643-662.
[3] Toh, W.L., D.J. Castle, and S.L. Rossell, Attentional biases in body dysmorphic disorder (BDD): Eye-tracking using the emotional Stroop task. *Compr Psychiatry*, 2017. **74**: p. 151-161.
[4] Wittekind, C.E., et al., Performance in a blocked versus randomized emotional Stroop task in an aged, early traumatized group with and without posttraumatic stress symptoms. *J Behav Ther Exp Psychiatry*, 2017. **54**: p. 35-43.
[5] Witthoft, M., et al., Neuronal and behavioral correlates of health anxiety: results of an illness-related emotional Stroop task. Neuropsychobiology, 2013. **67**(2): p. 93-102.
[6] Yaffe, B. and D.J. Walder, Schizotypy and specificity of negative emotions on an emotional Stroop paradigm in the general population. *Psychiatry Res*, 2016. **239**: p. 291-300.
[7] DiBonaventura, M.D., et al., A computerized Stroop task to assess cancer-related cognitive biases. *Behav Med*, 2010. **36**(2): p. 37-43.
[8] Erblich, J., et al., Biased cognitive processing of cancer-related information among women with family histories of breast cancer: evidence from a cancer stroop task. *Health Psychol*, 2003. **22**(3): p. 235-44.
[9] Custers, J.A., et al., Selective attention and fear of cancer recurrence in breast cancer survivors. *Ann Behav Med*, 2015. **49**(1): p. 66-73.
[10] Spielberger, C.D. and S.J. Sydeman, State-Trait Anxiety Inventory and State-Trait Anger Expression Inventory, in the use of psychological testing for treatment planning and outcome assessment, M.E. Maruish, Editor. 1994, Lawrence Erlbaum Associates.
[11] Toccafondi, A., et al., Live music intervention for cancer inpatients: The Music Givers format. Palliat Support Care, 2017: p. 1-8.
[12] Iwata, N., et al., Positive and negative affect in the factor structure of the State-Trait Anxiety Inventory for Japanese workers. *Psychol Rep*, 1998. **82**(2): p. 651-6.

Address for correspondence:

Eiji Aramaki, PhD; eiji.aramaki@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190569

How to Measure Circadian Rhythms of Activity and Their Disruptions in Humans Using Passive and Unobtrusive Capture of Phone Call Activity

Timothée Aubourg[a,b], Jacques Demongeot[a,c], Félix Renard[a], Hervé Provost[b], Nicolas Vuillerme[a,c]

[a] *Université Grenoble Alpes, AGEIS, Grenoble, France*
[b] *Orange Labs, Meylan, France*
[c] *Institut Universitaire de France, Paris, France*

Abstract

Monitoring circadian rhythms of social activity is crucial for preserving the health and wellness of ill elderly people. In this paper, we assess the ability of phones to be used as a temporal and social daily activity sensor from a passive and unobstructive measure of phone call activity. To this end, we introduce a methodology specifically designed to automatically measure both persistence and disruptions in circadian rhythms of phone call activity with 26 adults older than 65 years.

Keywords:
Circadian Rhythm; Cell Phone; Elderly

Introduction

Biologically, circadian rhythms are cyclical processes characterized by a near-to-24-hour period according to the individual [1]. The analysis of their expression is crucial for preserving the daily health and wellness of the older person at home by capturing its routines of activity and their probable disruptions in order to anticipate potential health issues [2]. Unfortunately, it must be acknowledged that, although innovative approaches as actigraphy have shown promising results for capturing the physical activity of the individual, the sensors needed often show limitation for collecting other types of data about the individual's activity as social interactions [2,3]. This point is problematic insofar as social disruptions in the older person's habits may directly reflect severe issues occurring with the aging process, like social isolation or depression for instance.

Thus, whether and how circadian rhythms of social activity and their disruptions could be measured in older adults on a passive and unobtrusive way remain to be investigated. The present paper is specifically designed to address this issue.

Methods

We introduced the phone device as a temporal and social sensor of the individual's daily activity, and we used a passive and unobstructive measure of phone call activity. We analysed the outgoing phone call detail records (CDRs) of 26 volunteers older than 65 years on a period of twelve consecutive months, and we focus on two specific phone call parameters, namely (1) the date and (2) the hour of call. We further describe a methodology specifically designed to automatically assess (1) the existence of circadian rhythms in outgoing phone call activity by leading persistence analysis that was developed in recent system complex literature [4], and (2) the existence of significative disruptions in these rhythms by calculating their monotony signatures [5].

Calculating Circadian Rhythms

We followed the descriptive approach recently proposed in [4] for circadian rhythm estimation in phone call activity. This approach consists in (1) coarse-graining the time dimension into a unique day divided into 24-one-hour time slots, and (2) calculating the average phone call frequency for each time slot.

Assessing the Consistence of Circadian Rhythms

We followed an analytical approach of persistence [4]. For each individual, at each 4-month period of time, T1, T2 and T3 :

We note D_{self} a measure such as:

$D_{self}(i,T,T+1) = \sqrt{JSD(P_i^T, P_i^{T+1})}$, where P_i^T is the discrete probability distribution of ego's i calls ratio at time period T.

We noted $< D_{self} >$, the average of D_{self} such as:

$< D_{self} > = \left(\frac{1}{N_T}\right) . \sum_{T=1}^{N_T-1} D_{self}(i,T,T+1)$, where N_T is the number of time periods, here 3.

Then, we noted D_{ref} the reference scale such as:

$D_{ref}(i,j,T) = \sqrt{JSD\left(P_i^T, P_j^T\right)}$, where P_i^T is the discrete empirical probability distribution of ego's i calls ratio at time period T and P_j^T is the discrete probability distribution of ego's j calls ratio at the same time period T with $i \neq j$.

We noted $< D_{ref} >$, the average of D_{ref} such as:

$< D_{ref} > = \left(\frac{1}{n-1}\right) . \sum_{i=1}^{n-1} D_{ref}(i,j,T)$, where n is the number of individuals, here 21.

Finally, persistence of circadian rhythms of a given ego is validated if and only if $< D_{self} > / < D_{ref} > < 1$.

Assessing the Existence of Disruption in Circadian Rhythms

To assess the existence of disruption(s) in circadian rhythms, we used the stochastic monotony signature of a function that indicates if the function is increasing-or-constant or decreasing through a sequence of, respectively, sign + and sign -. In practice, we assessed the existence of a significant disruption between two successive, and independent, circadian rhythms X and Y after testing the similarity of these signatures due to a common causality between X and Y (hypothesis H1) against a

random choice of the values of the successive sgnXi's (hypothesis H0) [5].

Results

Existence of persistent circadian rhythms

The consistence of circadian rhythms was confirmed by a persistence analysis (see Methods section). Results stand out all the egos have their self-distance lower than their reference distance, which means their pattern tends to retain their shape through time. Numerically, averaging results for the whole egos, $<D_{self}>$ = 0.24 (± 0.06), whereas $<D_{ref}>$ = 0.38 (± 0.07).

Disruption in circadian rhythms

The stochastic monotony signature comparison tests highlighted the existence of variations between two successive time periods of the year in the way the individual allocates his time of communication throughout the day. Figure 1 illustrates these results by mapping the monotony signatures variations of circadian rhythms in outgoing phone call activity for each individual into a heat-map.

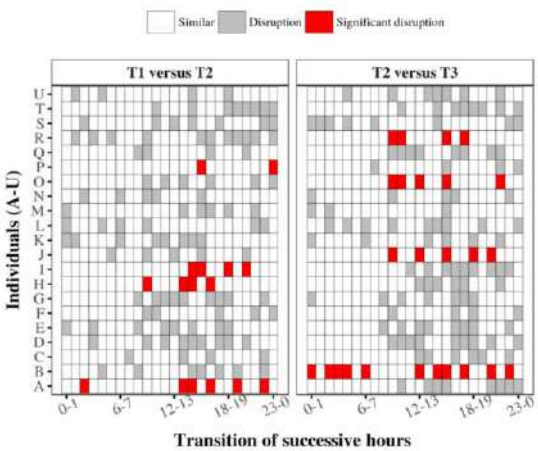


Figure 1– Existence of disruption(s) between two successive circadian rhythm calculated from two successive periods of time.

Discussion

Our results suggest the persistence of daily social interactions in phone communication in the older population, and disruption(s). These features could be advantageously used for completing traditional health monitoring approaches using circadian rhythms analysis by giving daily data of social interactivity. However, a number of caveats and limitations have to be taken into account.

First, any rapid generalization of our results might be avoided because of rather small sample size (n = 21). Then, we insist on the fact that analysed circadian patterns concerns only the outgoing phone call activity. It could be relevant to enhance robustness of such circadian estimation by combining outgoing call pattern with corresponding social interaction data as incoming phone calls, text messages, location data and, more generally, in leading multi-dimensional analysis based on temporal active and passive data from phones [6].

Conclusions

In this papier, we assessed the use of phone device as a temporal and social sensor of the daily social activity of the elderly individual for a health monitoring purpose. To this end, we analyzed the outgoing phone call detail records of 26 volunteers older than 65 years on a period of twelve consecutive months. We give a methodology to automatically assess (1) the existence of circadian rhythms in outgoing phone call activity by leading persistence analysis, and (2) the existence of significant disruptions in these rhythms by calculating their monotony signatures. Within our population, we showed that older adults have consistent circadian rhythms in phone call activity, which is reflected by their ability for persisting in time. Then, we demonstrated that, despite this persistence, disruptions could occur through time in circadian rhythms of outgoing phone call activity by comparing their monotony signatures.

References

[1] Farhud D, Aryan Z. Circadian Rhythm, Lifestyle and Health: A Narrative Review. Iran J Public Health. 2018 Aug;47(8):1068-1076.

[2] Grandin LD, Alloy LB, Abramson LY. The social Zeitgeber theory, circadian rhythms, and mood disorders: Review and evaluation. Clin Psychol Rev. 2006;26(6):679–94.

[3] Skeldon AC, Phillips AJK, Dijk D-J. The effects of self-selected light-dark cycles and social constraints on human sleep and circadian timing: a modeling approach. Sci Rep. 2017;7:45158.

[4] Aledavood T, López E, Roberts SG, Reed-Tsochas F, Moro E, Dunbar RI, Saramäki J. Daily rhythms in mobile telephone communication. PLoS One. 2015;10(9):e0138098.

[5] Demongeot J et al. Stochastic monotony signature and biomedical applications. Comptes rendus biologies.2015;338(12):777-783.

[6] Quadri C, Zignani M, Capra L, Gaito S, Rossi GP. Multidimensional human dynamics in mobile phone communications. PLoS One. 2014;9(7):e103183.

Address for correspondence

Timothée Aubourg: timothee.aubourg@orange.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190570

Clinical Decision Support System for Evaluation of Patients with Musculoskeletal Disorders

Lecian C. Lopes, Sayonara de Fátima F. Barbosa

Graduate Program of Health Informatics, Universidade Federal de Santa Catarina, Florianopolis, Brazil

Abstract

Clinical evaluation of the patient and the follow-up of the rehabilitation process are pillars of physical therapy care. The SIAVA-FIS decision support system is presented, which accesses graphical evaluation and evolution data, allowing physical therapists to follow the results of the therapeutic procedures in a mobile and web platform. Results indicate that SIAVA-FIS meets physical therapists´ needs and that usability presents effectiveness, efficiency and satisfaction in the use of tasks evaluated by users.

Keywords:
Clinical Decision Support Systems; Rehabilitation; Diagnosis

Introduction

The field of health informatics has emerged as an interdisciplinary science that evaluates how health information and knowledge can be effectively used for clinical decision making[1]. Multidisciplinary care is an integrated team approach to healthcare and involves different healthcare professionals, like physical therapists, that are movement experts who provide services that help restore function, improve mobility, relieve pain, and prevent or limit permanent physical disabilities in patients with injury or disease[2]. Physical evaluation, tests and measurements are means for collecting information and the physical therapist uses them to identify signs or symptoms of changes in health conditions, diagnosis, prognosis and treatment decision making[3]. A clinical decision support system favors the reduction of costs by searching for alternatives for displacement, technological resources, information in the diagnosis and evolution of conduits[4].

Methods

The method used in this study was based on the development and evaluation of a web-based system and mobile application (app) for physical therapy assessment of outpatients with musculoskeletical disorders. It was composed by two phases. The first phase was the development. The system was developed using free software, in the PHP language, MySql database, with the IONIC framework for the hybrid application, and the Angular Material for the web framework. Unified Modeling Language (UML) was selected. For the hosting services in the cloud, Hostgator Brazil was used. To capture users requirement, we conducted two workshops with 06 physical therapists from the service. In the first workshop, we asked users about specific requirements for the application to fit into their practice, including clinical parameters and subjective data to be considered for assessment. Participants presented their suggestions and discussed in the group. One week after the first workshop, the group met again and presented a synthesis of previous suggestions along with suggestions obtained from review from textbooks, that discussed and revised.

The second phase was usability evaluation. The system was evaluated by physical therapy students and professors at a private university in Brazil. They answered a questionnaire that was developed according to NBR-9244-11, with 17 questions, distributed in three criteria: system use (four questions), system content (eight questions) and system interface (five questions). These criteria were assessed based on a five point Likert scale (1-5), with the following response categories: 1 = strongly disagree, 2 = disagree; 3 = neutral; 4 = agree, and 5 = strongly agree. The top three points (3,4,5) were considered to have indicated criteria approval. This study was approved by the Research Ethics Committee.

Results

The system was named SIAVA-FIS (Assessment and Reassessment System in Physical Therapy) and is intended to assessment and follow-up of patients through the measurement of data collected. SIAVA-FIS decision support system comprises 2 modules: the first is a web module, where the data is registered and the items collected are configured, as well as the possibilities of printing and alerts of the system; the second module is a mobile application, that will collect the selected information and send it to the web system to generate the various types of graphics. As a result, the system presents clinical parameters such as vital signs (blood pressure, temperature, heart rate and respiratory rate), body mass index as well as other assessment measures according to the segment assessed, such as goniometry, muscle strength, perimetry, muscle tone, and pain, among others. The respective alerts were also included when were not in agreement with the current and updated reference values.

The SIAVA-FIS web module has a login screen for each registered user and is hosted on a server in the cloud at the URL www.siavafis.com.br. As every physical therapist make their own planning in consultation with their patients, after the identification login, the initial screen presents the patients that were already registered in the clinical system by the physical therapist, and their respective appointments, or no information when no patient has been registered until that moment.

In addition, through the web system it is possible to visualize the data collected by the mobile application and verify in the web module its graphical presentation, presenting the data as specified by the physical therapist, such as values and instructions on clinical parameters.

The SIAVA-FIS mobile application is an interface to the web module. It does not allow double entry by physical therapists, as they do not exchange patients with each other, unless they change their working schedule with each other.

Usability Evaluation

The system was evaluated in different criteria. In the system use criterion, consisting of four items, averages were between 4.27 (± 0.59) and 4.53 (± 0.74), with an overall mean of 4.4666 and the standard deviation 0.677, demonstrating that the evaluators agree partially with the use of the system according to the application of the concepts defined in the methodology. In relation to the system content criterion, composed of 8 items, the general average of this criterion was 4.36 and the concept was partial agreement in the evaluation. In the system interface criterion, the five items were evaluated with averages between 4.13 (± 0.640) and 4.73 (± 0.210) with an overall mean of 4.51 and the standard deviation 0.55, demonstrating that the users agree fully with the system interface. The average of the majority of questions was positive, with a score higher than 4.

Discussion

Physical therapists need to manage a vast amount of information to make clinical decisions, and the use of information technology for systematically processing data and supporting information and knowledge[1] can improve their practice. Clinical decision systems (CDS) help front-line staff to select interventions for patients with musculoskeletal disorders, and may be classified into computer-based tools/ questionnaires, treatment algorithms/models, and clinical prediction rules/classification systems. CDS tools, especially those employing rapidly advancing computer technologies, are under development and of potential interest to health care providers, case management organizations and funders of care[5]. The use of a decision support system may also help to standardize outcome measures collection, build standardized measurement into clinical workflows, and show the process and outcome of patient care, and the use of standardized functional measurements that can be used across disciplines [6]. Physical therapists are fundamental to define specific requirements on the practical use of software which cover the administration, documentation and evaluation of the entire therapy process, adapted individually by the therapists [7,8]. The system was developed considering the data that are part of the physical therapist assessment, and it was planned so that information can also be inserted remotely, at the point of care, allowing the graphic visualization of all the consultations. In this way, although the use of SIAVA-FIS can make it possible to identify patterns and trends of abnormalities, allowing a quick and accurate identification so that the physical therapist assessment can occur in a more precise way [7], it must be subjected to further validation before can be recommended for large-scale implementation [8].

Conclusions

The SIAVA-FIS presents good usability, as well as efficiency and satisfaction in the use of the tasks and can contribute to the evaluation of quality indicators and improvement of the physiotherapeutic care.

References

[1] S Wilkinson , J Chevan , D Vreeman. Establishing the centrality of health informatics in physical therapist education: if not now, when? *Journal of Physical Therapy Education*. 24(2010),10-15.

[2] American Physical Therapy Association-APTA (2011) Today's physical therapist:a comprehensive review of a 21st-Century health care profession.

[3] American Physical Therapy Association. *Guide to Physical Therapist Practice 3.0*. Alexandria, VA, 2014.

[4] S Nimkar. Promoting individual health using information technology: trends in the US health system. *Health Education Journa*l. 75(2016), 744-52.

[5] DP Gross, S Armijo-Olivo, WS Shaw, KWilliams-Whitt, NT Shaw, J Hartvigsen , Z Qin, C Ha, LJ Woodhouse, IA Steenstra. Clinical decision support tools for selecting interventions for patients with disabling musculoskeletal disorders: a scoping review. *J Occup Rehabil*. 26(2016), 286-318.

[6] KN Daley, D Krushel, J Chevan. The physical therapist informatician as an enabler of capacity in a data-driven environment:an administrative case report. *Physiother Theory Pract*. 27(2018), 1-11.

[7] K Messer-Misak, R Egger. Specific requirements of physiotherapists on the practical use of software in the therapeutical process. S*tud Health Technol Inform*. 223(2016), 158-66.

[8] MD Westby, A Klemm, LC Li, CA Jones. Emerging role of quality indicators in physical therapist practice and health service delivery. *Phys Ther*. 96(2015), 90-100.

Address for correspondence

Sayonara Barbosa

sayonara.barbosa@ufsc.br

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190571

Brazilian National Service of Telediagnosis in Electrocardiography

Maria Beatriz Alkmim[a], Cláudia B. G. Silva[b], Renato M. Figueira[a], Daniel V.V. Santos[a], Leonardo B. Ribeiro[a], Maria Cristina da Paixão[a], Milena S. Marcolino[a], Jailton C. Paiva[c], Antonio Luiz Ribeiro[a]

[a]Telehealth Center from the University Hospital of Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[b]Departamento de Gestão da Educação na Saúde, SEGTES, Ministry of Health, Brasília, Brazil
[c]Laboratório de Inovação Tecnológica em Saúde (LAIS), Universidade Federal do Rio Grande do Norte, Natal, Brazil

Abstract

Access to specialized care remains unsolved in Brazil. The National Telediagnosis Project was created to expand successful telediagnosis experiences nationwide. The Telehealth Network of Minas Gerais (TNMG) was selected as a reference for tele-electrocardiogram (ECG). We aim to describe the experience of TNMG of developing and implementing the Brazilian National Service of Telediagnosis in Electrocardiography. Implementation planning includes discussion of workflows, standard procedures, responsibility definition for stakeholders, and adaptation of TNMG telediagnosis system. Tele-ECG has been implemented in 79 municipalities in 5 states. In a survey with 152 health professionals, 57% noted that ECG was not available in the local public health system before, 63% indicated tele-ECG service utilization ≥3 days per week, 96% considered the service very useful and 89% were very satisfied with it. In conclusion, the service fills a gap in specialized care in the public system and can improve access to a basic exam in remote and underserved regions.

Keywords:

Electrocardiography, Telemedicine, Access to Health Care

Introduction

Brazil is a country with continental dimensions, divided into 5 regions, including 27 states and 5,507 municipalities. There are important economic, social and cultural differences among the regions. Although the universal public health system, created in 1988, operates nationwide, with a broad coverage of Primary Care, access to specialized care is still an unsolved problem. These inequalities made the country a promising field for the implementation of telehealth. Therefore, in 2007 the Brazilian Ministry of Health (BrMoH) created the Brazilian National Telehealth Program, to support family health teams through teleconsultation, telediagnosis and teleducation performed by Telehealth Centers (TC) established in different states of the country. Currently, there are 25 TC in operation at different development levels.

The Telehealth Network of Minas Gerais (TNMG) is a collaborative network of seven public universities in the state of Minas Gerais, Southeast Brazil, coordinated by the University Hospital of *Universidade Federal de Minas Gerais* [1]. It was implemented in 2005 and currently provides telehealth care for 813 municipalities in Minas Gerais, mainly in primary health care (PHC) centers, but also in emergency departments, ambulances and hospitals. Being one of the first TC to participate in the Brazilian National Telehealth Program, it has a long and wide experience in telediagnosis in cardiology. Over 3.8 million tele-electrocardiogram (ECG) reports and over 124,000 teleconsultations have already been performed, as well as tele-retinography, Holter and tele-education activities, with quality assured by regular audits.

In 2017, the BrMoH established a National Telediagnosis Project as part of the Brazilian National Telehealth Program, in order to expand successful telediagnosis regional experiences to a nationwide scale.

The preliminary telediagnosis experiences chosen comprehended ECG, retinography, dermatology and spirometry. The TNMG was chosen to be one of the first National Telediagnosis Project participants, helping BrMoH in its conception and offering tele-ECG reports.

The present work describes the experience of developing and implementing the Brazilian National Service of Telediagnosis in Electrocardiography (BrNSTE) during its first year of operation, under the framework of the BrMoH National Telediagnosis Project.

Methods

The BrMoH National Telediagnosis Project was conceived as an innovative process to perform exams using telemedicine tools in an organized way, with quality and low cost, to significantly improve the availability of certain exams, especially to remote regions with low human development index (HDI) and restricted access to specialized care.

The Project consists of a partnership of BrMoH, State Health Department, Municipal Health Department, specialized TC and regional TC (RTC). The criteria used by the BrMoH to choose a specialized TC to offer certain exam are based on the TC previous proven experience and capacity to significantly expand its offer at low cost. BrMoH only selects states to implement the project which have a RTC in operation with ongoing funding by BrMoH.

A National Telediagnosis Platform, developed by BrMoH demand to *Universidade Federal do Rio Grande do Norte* (UFRN), was integrated to TNMG telediagnosis system in order to manage the project. The TNMG telecardiology service, besides performing the ECG reports, also provides synchronous teleconsultation to local medical doctors to discuss the severe clinical cases. Educational material for tele-education in cardiology is available on TNMG website to support local health professionals.

Implementation planning includes flows and standard procedures, role and responsibility definition for stakeholders, adaptation and customization of TNMG telediagnosis system. The implementation process starts with a presencial two-day training of the RTC team at TNMG. The two first local

implementations are made by RTC team together with the TNMG team to guarantee the success of the implementation.

The impact of implementation of BrNSTE on local healthcare was assessed in a survey with health professionals, health unit coordinators and health managers from 35 towns in Acre, Bahia and Mato Grosso in June and July 2018. A Likert-scale questionnaire with 8 questions was developed. Professionals were contacted at last 3 times by telephone or email. There were attempts to contact 409 professionals, but email addresses and telephone numbers were wrong for some of them. We tried to get at least one professional of each category (doctors, nursing staff, coordinators) in each health unit.

Results

In September 2017, the BrNSTE was launched in the remote town of Xapuri (state of Acre), in the Amazonian region. Since then, the service is being expanded to other Brazilian states, as Bahia (November 2017), Mato Grosso (February 2018), Ceará and Roraima (October 2018). Other states have been chosen by the BrMoH to receive the Project for the next months.

From September 2017 to October 2018, the BrNSTE delivered through the TNMG 33,178 tele-ECGs for 79 municipalities.

TNMG telecardiology service works 7 days / 24 hours. It has a response time below 10 minutes for emergency ECGs and between 2 and 4 hours for elective ones. On average, each town performs 90 ECGs/month. Utilization ratio (defined as number of towns using the system divided by number of towns with the system implemented) is monitored monthly, and is always above 90%.

With regards to the questionnaire, from the 152 respondents 57% informed that, before the implementation, the ECG was not available in the local public health system; and even when it was available, no standardized report was provided. After implementation of BrNSTE, 63% indicated tele-ECG service utilization ≥3 days per week; 94% considered as adequate the time to receive the report. In addition, 96% of respondents reported the tele-ECG service very useful, 98% indicated much benefit for patients and 89% were very satisfied with it. There were not significant differences in responses according to professional category.

Discussion

The present study describes the deployment of a large and complex telehealth service, developed under the guidelines of a national policy with the collaboration of different stakeholders, including a TC with expertise in tele-ECG, regional TCs, responsible for local implementation, an University (UFRN) that provides support for the BrMoH activities and the BrMoH itself. The results reported above demonstrated that the project was successful, reaching 79 towns in 5 different Brazilian states, including some very remote locations, in the Amazon region and in the Northeast of the country. A large number of ECG from these places have been performed and analyzed, with a utilization ratio of more than 90%. For most of these places, an ECG service was not available before the implementation of the BrNSTE. Health professionals, coordinators and managers have shown to be satisfied with the service.

Cardiovascular diseases are the most important causes of death in Brazil and in most developing countries. ECG is a basic method of recognition of cardiovascular diseases, useful also for risk stratification and evaluation of prognosis. For some cardiac diseases, as myocardial infarction and arrhythmia, ECG is the basic diagnostic tools and treatment is not possible without rapid and immediate availability of the exam. Since BrNSTE provides timely and accurate diagnosis 24/7, it has an enormous potential of improving the access of patients from these remote and resource-constrained localities to adequate health care, reducing acute and long-term morbidity and mortality.

There are other experiences of large tele-ECG services, in Brazil [2] and other developing countries [3], but many projects are devoted to acute coronary syndrome or have a more restricted area of coverage. The present report describes a nationwide initiative that seems to be unique, with high utilization ratio and excellent satisfaction of the users.

The study also opens the possibility of universalization of the access to medical diagnostic methods through the use of simple, low-cost telehealth systems. Indeed, the possibility of scaling-up the service provided, as well as other diagnostic methods, to the whole Brazilian population points out for a solution for some of the access barriers that currently hasten the effectiveness of the Brazilian Universalized Health System. Additional challenges include the full integration of this service to the electronic health records provided by the BrMoH and the sustainability of the BrNSTE as a permanent policy of the country.

Conclusions

In conclusion, a Brazilian National Service of Telediagnosis in Electrocardiography was implemented in 2017, based in the experience of the state of Minas Gerais, and it is now established in 5 other Brazilian states, with several other states already waiting for implementation. The service fills a gap in specialized care in the Brazilian public health system and can improve access to a basic exam in remote and underserved regions, with a potential to decrease morbidity and mortality related to cardiovascular diseases, the most common cause of death in Brazil.

Acknowledgements

The Brazilian National Service of Telediagnosis in Electrocardiography is funded by the Brazilian Ministry of Health. ALPR was supported, in part, by CNPq (research fellowship number 310679/2016-8) and FAPEMIG (PPM-00428-17), and is a member of the National Institute of Science and Technology for Health Technology Assessment (IATS) (CNPq/465518/2014-1).

References

[1] Alkmim MB, et al. Improving patient access to specialized health care: the Telehealth Network of Minas Gerais, Brazil. *Bull World Health Organ* **90** (2012), 373-378.

[2] Giuliano Ide C, Barcellos Junior CL, von Wangenheim A, Coutinho MS. Issuing electrocardiographic reports remotely: experience of the telemedicine network of Santa Catarina. Arq Bras Cardiol. 2012 Nov;99(5):1023-30.

[3] Mars M. Telemedicine and advances in urban and rural healthcare delivery in Africa. Prog Cardiovasc Dis. 2013 Nov-Dec;56(3):326-35.

Address for correspondence

Antonio Luiz Ribeiro, tom@hc.ufmg.br. Centro de Telessaúde do HCUFMG - telessaude@hc.ufmg.br. Av. Professor Alfredo Balena, 110 - 1º Andar - Ala Sul - Sala 107 30130-100 - Belo Horizonte - MG - Brazil - +55(31) 3409-9201

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190572

Accuracy of Self-Reported Weight Collected Through a Web-Based Platform in a Weight Loss Trial: Validation Study of the POEmaS Clinical Trial

Alline M Beleigoli[a,b,c,d], Andre Q Andrade[e], Maria de Fátima H Diniz[a], Roberta S Alvares [a], Marina H Ferreira[a], Leticia A Silva [a], Marcia C Rodrigues [a], Luma Jacomassi[a], Amanda G Cerqueira [a], Antonio L Ribeiro [a, d]

[a] Department of Internal Medicine, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil,
[b] Flinders Digital Health Research Centre, Flinders University, Adelaide, South Australia, Australia,
cDepartment of Medicine, University of Adelaide, Adelaide, South Australia, Australia
d Centre of Telehealth, Hospital das Clinicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
e Quality Use of Medicines and Pharmacy Research Centre, School of Pharmacy and Medical Sciences,University of South Australia, Adelaide, Australia

Abstract

Self-reported anthropometric data in web-based weight loss interventions may be inaccurate. We studied the agreement between online self-reported and measured weight in the course of the POEmaS randomized controlled trial. Measured weight was not different from reported one (- 0.4 kg; 95%CI -0.93 to 0.12). 95.6% of the cases were within the limits of agreement (Bland-Altman method). Self-reported weight collected online was accurate, which suggests that interventions and outcomes assessment can rely on these data.

Keywords:
Body weight, ehealth, data accuracy

Introduction

Delivering weight loss interventions via the web have been increasingly more common. Accessibility 24hours/7days, anonymity and cost-effectiveness are some of the advantages when web-based is compared to face-to-face interventions.

Weight self-monitoring is one of the key components of weight loss interventions. Regular self-monitoring has been associated with better outcomes for weight loss and maintenance. Accurate and reliable anthropometric data are essential to intervention planning and progress monitoring. Discrepancies between measured and self-reported anthropometric data can lead to misclassification of weight status and in the context of weight loss trials can bias the outcomes. Therefore, it is fundamental to have validation studies of self-reported web-based data.

The aim of this study was to validate self-reported weight collected via a web-platform for weight loss adult participants of the POEmaS randomized controlled trial.

Methods

POEmaS trial

Students (current and past) and staff of the Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil, were recruited from September to October, 2017. Inclusion criteria were age between 18 and 60 years, self-reported BMI >=25 Kg/m^2, intention to lose weight by changes in lifestyle habits, internet access [1]. Subjects who reported comorbidities that demanded specific dietary or physical activity recommendations participants of other weight loss programs were excluded.

1,298 participants were randomized by a stratified randomized block to one of three groups: (1) Minimal intervention for 24 weeks with subsequent access to the platform; (2) Platform group -received the weight loss program delivered by the web-based platform for 24 weeks; (3) Platform weight loss program for 24 weeks plus online dietitian coaching for 12 weeks.

Sample and procedures for the validation study

We randomly selected 12.5% of the POEmaS' population balanced across study groups to guarantee that the validation study sample is representative of the study population.

Participants were contacted by email for the validation study and instructed to report weight on the specific questionnaire on the web platform. Weight (kg) was measured using calibrated Welmy® scales (to the nearest 100 g. Participants were wearing light clothes and without shoes during the measurements. A minimum 4-hour fasting and voiding 5-10 minutes before the measurements were recommended to all participants.

Ethics

All participants signed a specific consent form for the validation study. The Federal University of Minas Gerais (UFMG) Ethics Research Committee approved the POEmaS trial and the validation study (CAAE: 73545717.5.0000.5149).

Statistical analysis

Frequencies and mean (standard deviation) were used to describe qualitative and continuous variables, respectively. A paired t-test was used to determine the differences between reported and measured weight. To evaluate agreement between paired anthropometry we applied the Bland–Altman method.

Results

A total of 159 individuals predominantly females (129; 80.1%) participated in the validation study. Mean age was 36.5 years (11.1). Other characteristics are displayed in Table 1.

Table 1 Characteristics of the validation study´s participants

Characteristic	Validation study (n=159)
Gender	
Female	128 (80.5)
Male	31 (19.5)
Age, years	36.5 (11.0)
BMI, kg/m^2	29.4 (4.1)
Education	
Undergraduate	39 (24.5)
Graduated/post graduated	120 (75.2)
Group	
Minimal intervention	59 (37.0)
Platform	42 (26.4)
Platform plus online dietitian	58 (36.6)

Differences between reported and measured weight were not significant (Table 2).

Table 2 Measured and self-reported weight and difference

	Reported	Measured	Difference (measured – reported)	P value
Weight, kg	80.0 (77.8 to 82.2)	79.8 (77.6 to 82.0)	- 0.40 (-0.93 to 0.12)	0.13

The difference between measured and reported weight plotted against their mean stayed within the limits of agreement for 152 of the 159 cases (95.6%), which shows high agreement between measured and reported data (Figure 1).

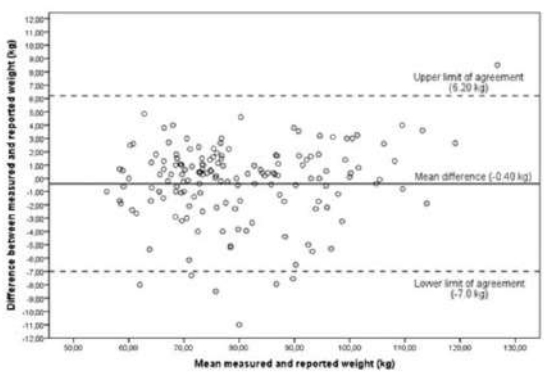


Figure 1 - Bland Altman plot for self-reported and measured weight (in kg)

Discussion

In this validation study with participants of a randomized controlled trial for weight loss, we found good agreement between measured and self-reported weight through the web-based platform.

Validation studies of self-reported anthropometric outcomes in web-based weight loss interventions are scarce. Contrarily to our findings, Harvey-Berino et al. [2] found a significant mean difference of -0.86kg, whereas Jerome et al. [3] found a progressive underreporting from -0.5kg at 6 months to -1.1kg at 24 months. A particular feature of our population that might explain this difference is the extremely high education level - 75.2% were graduated or post-graduated.

One limitation of our validation study is that, due to organizational issues, it was not possible to guarantee weight measurement and online reporting at the same time. Weight variation during the day under certain conditions, such as fasting, might have influenced our results.

Conclusions

Weight self-report through web-based questionnaires in weight loss interventions can be accurate. This suggests that individual interventions planning and outcomes' assessment can rely on these data.

Acknowledgements

This study was supported with grants from the National Institute of Science and Technology for Health Technology Assessment (IATS) – CNPq/Brazil. AM Beleigoli was supported by Prêmio CAPES de Teses. AL Ribeiro was supported by CNPq (research fellowship number 310679/2016–8) and FAPEMIG (PPM-00428-17).

References

[1] A.M. Beleigoli, A. Queiroz de Andrade, M.F. Haueisen Diniz, R.S. Alvares, and A.L. Ribeiro, Online platform for healthy weight loss in adults with overweight and obesity - the "POEmaS" project: a randomized controlled trial, *BMC Public Health* 18 (2018), 945.

[2] J. Harvey-Berino, R.A. Krukowski, P. Buzzell, D. Ogden, J. Skelly, and D.S. West, The accuracy of weight reported in a web-based obesity treatment program, *Telemed J E Health* 17 (2011), 696-699.

[3] G.J. Jerome, A. Dalcin, J.W. Coughlin, S. Fitzpatrick, N.Y. Wang, N. Durkin, H.C. Yeh, J. Charleston, T. Pozefsky, G.L. Daumit, J.M. Clark, T.A. Louis, and L.J. Appel, Longitudinal accuracy of web-based self-reported weights: results from the Hopkins POWER Trial, *J Med Internet Res* 16 (2014), e173.

Address for correspondence

Alline Beleigoli

1 Sturt Buildings Flinders University, Sturt Rd, Bedford Park SA 5042, Adelaide, Australia

abeleigoli@gmail.com

+61 8 8201 3303.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190573

Cloud-Driven Application for Measurement of Wound Size

Suryaprakash Kompalliy[a,b], Vivek Bakarajuy[a], SB Gogia[c]

[a] *International School of Engineering (INSOFE), Hyderabad, India*
[b] *ECE Department, Carnegie Mellon University, Rwanda*
[c] *Society for Administration of Telemedicine and Healthcare Informatics, New Delhi, India*

Abstract

A web-based tool is described to store images of ulcers and measure ulcer size automatically. The web tool enables doctors and caregivers to upload images from any device. The ulcer size is automatically extracted using image processing algorithms.

Keywords:

Wound size, internet

Introduction

Measurements are key to management, especially for long term problems like ulcers. Size variation helps, but accuracy of measurement has issues related to depth, body, and curvature. We hereby describe an open source web application for measurement of ulcer size for consistent and comparative documentation with device and lighting independence. Application features and usage methodologies are suitable for standalone, team-sharing and tele-monitoring scenarios.

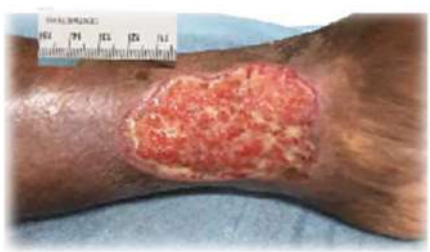
Prior Methods

Existing methods of measuring wounds include scale/ruler measurements (linear measurements), acetate paper measurements [1,2], depth gauges, and ultrasound imaging. (Figure 1). Tracing the wound perimeter with a marker on clear acetate can be more accurate but is complicated by condensation related fogging and obscuration of wound margins. Contact with the wound area can lead to patient discomfort and even infection. Automated techniques have been developed [1] to overcome these drawbacks.
Stereophotogrammetry reconstructs a 3D image through twin cameras [3]. A variant is to use a single camera with Image J™, [4] a target sheet for calibration [5], laser based triangulation [6], and structured light [7,8], to map the 3D surface of a wound and then segment out the wound region which can be more accurate but expensive.

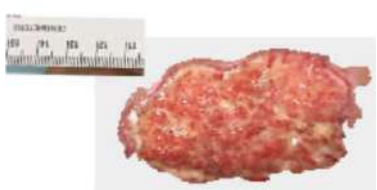
Methodology

Our technique matches [4,5] and involves the following broad steps:

1. User takes image of the wound region, with a known scale placed in background,
2. The scale and wound are extracted from the image.
3. Measurements on the scale are used to determine how many pixels correspond to one mm^2,
4. The image is loaded after login on to the free access website – http://med.insofe.edu.in:8282/(Username: test, Password: testp@ss1234),
5. At the site, the image becomes visible along with an online drawing tool which allows the user to broadly mark out the area of interest by drawing two quadrilaterals – one outside the wound area, and the other inside the wound,
6. Image processing is used to convert the image to an identifiable number of pixels corresponding to this marked region which are used to determine the wound area in mm^2.



(a) Image of wound with scale



(b) Wound and scale extracted

(c) Extent and size measured using algorithms

Figure 1 - Workflow

Discussion

You cannot manage what you cannot measure. Wounds and ulcers need daily assessment which includes the size and depth of

wounds. We offer an easy method which can be used by clinicians. It is currently free for use by anybody.

References

[1] R. Khoo and S. Jansen. The evolving field of wound measurement techniques: a literature review. *Wounds* **28** (2016), 175-181.

[2] M.E. Wendelken, W.T. Berg, P. Lichtenstein, L. Markowitz, C. Comfort and O.M. Alvarez. Wounds measured from digital photographs using photo-digital planimetry software: validation and rater reliability. *Wounds* **23** (2011), 267-275.

[3] C.J.K. Bulstrode, A.W. Goode, and P.J. Scott, Stereophotogrammetry for measuring rates of cutaneous healing: a comparison with conventional techniques. *Clin Sci* **71** (1986), 437-443.

[4] R. Shetty, H. Sreekar, S. Lamba, and A.K. Gupta. A novel and accurate technique of photographic wound measurement. *Indian J Plast Surg* **45** (2012), 425-429.

[5] H.A. Thawer, P.E. Houghton, M.G. Woodbury, D. Keast, and K. Campbell. Computer-assisted and manual wound size measurement. *Ostomy Wound Manage* **48** (2002), 46-53.

[6] U. Pavlovčič, J. Diaci, J. Možina, and M. Jezeršek. Wound perimeter, area, and volume measurement based on laser 3D and color acquisition. *Biomed Eng Online* **14** (2015), 39.

[7] T.A. Krouskop, R. Baker and M.S. Wilson. A noncontact wound measurement system. *J Rehabil Res Dev* **39** (2002), 337-346.

[8] P. Plassmann and T.D. Jones. MAVIS: a non-invasive instrument to measure area and volume of wounds. *Med Eng Phy* **20** (1998), 332-338.

Address for correspondence
Dr. S. B. Gogia, SATHI, 28/31 Old Rajinder Nagar, New Delhi 110060, INDIA.
email gogia7@gmail.com, radhamadhavi.neelamraju@insofe.edu.in, vivek.bakaraju@insofe.edu.in, surya.kompalli@insofe.edu.in, kompalli@cmu.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190574

Towards the Definition of an Intelligent Triage and Continuous Monitoring System for Hospital Emergency Departments and Clinics

Antonis Billis[a], Maria Zouka[b], Petros Nicopolitidis[c], Paraskevas Lagakis[a], Evangelos Logaras[a], Nancy Karanasiou[d], Alexis Fourlis[d], John Gialelis[e], Dimitrios Kallergis[f], Georgios I Papadimitriou[c], Christos Douligeris[f], Theodosis S. Papavramidis[g], Maria Krizea[e], Panagiotis D. Bamidis[a]

[a] *Lab of Medial Physics, Medical School, Aristotle University of Thessaloniki, Greece,*
[b] *Department of Anaesthesiology and Intensive Care, AHEPA University Hospital, Greece*
[c] *Department of Informatics, Aristotle University of Thessaloniki, Greece*
[d] *VIDAVO S.A., Thessaloniki, Greece*
[e] *Industrial Systems Institute/ATHENA RC, Platani, Patras, Greece*
[f] *Department of Informatics, University of Piraeus, Piraeus, Greece*
[g] *1st Propedeutic Department of Surgery, AHEPA University Hospital, Aristotle University of Thessaloniki, Greece*

Abstract

Recent statistics have demonstrated that Emergency Departments (EDs) in Greece lack in organization and service. In most cases, patient prioritization is not automatically implemented. The main objective of this paper is to present IntelTriage, a smart triage system, that dynamically assigns priorities to patients in an ED and monitors their vital signs and location during their stay in the clinic through wearable biosensors. Initital scenarios and functional requirements are presented as preliminary results.

Keywords:

Triage, Clinical Decision Making, Wearable Electronic Devices

Introduction

Screening and monitoring of patients in Greek hospitals' EDs and clinics is poorly organized and seriously understaffed. An exception is the AHEPA General Hospital, which uses a 5-level triage system, using the ESI (Emergency Severity Index) algorithm [1]. This system provides patient priority data through monitors which then appear to the corresponding doctor, however it does not provide an evaluation based on the patient's vital signs. This paper proposes the design of a system that i) implements dynamic prioritization of patients within EDs and ii) improves the 24-h monitoring efficacy of patients in a clinic. The novelty of this approach lies in the fact that real-time screening and continuous monitoring of ESI level 3 incidents, is automated via intelligent bio-monitoring sensors and customized machine learning algorithms. In addition, through an appropriate network infrastructure, it is possible to trace the patients routes. In this way, administrative coordination between individual hospital departments is strengthened as it will be possible to extract statistical data on patient flows within the hospital. Similar attempts have been recently documented in literature [2-3].

Methods

The proposed system, the overall concept of which is shown in Figure 1, combines collecting measurements from body sensor networks, transferring them to a central point and processing these measurements to draw conclusions about the severity of the medical incident and the appropriate notifications for the medical staff. More specifically, IntelTriage consists of the following subsystems: i) Portable Subsystem of Biosensors and Tracking Devices, ii) Medical Decision Support, iii) Interoperability, iv) Networking and v) Security.

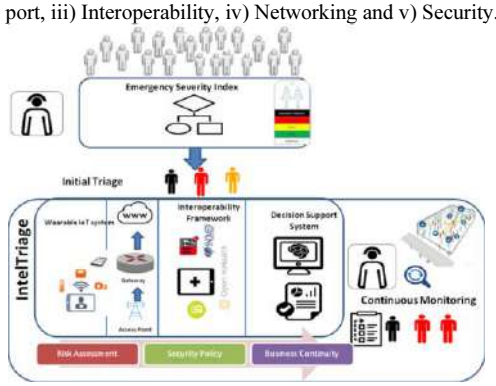


Figure 1–Conceptual diagram of IntelTriage solution

Portable Subsystem of Biosensors and Tracking Devices

The portable system consists of a Recording Platform with Body Sensors. This subsystem will gather information such as heart rate, breath rate, oxygen saturation, and transmit them to the central system database. In addition, the portable system will send beacons, which will be received from fixed 802.11 Access Points (APs).

Medical Decision Support Subsystem

Bio-parameters and Wi-Fi signals will be analyzed to assist the physician managing critical patients. The integration of machine learning algorithms and statistical models will provide dynamic monitoring of the progress of patients' health, using time series analysis and real-time changes detection, aiming at optimal triage of cases and reprioritization on the fly.

Interoperability Subsystem

While developing the monitoring, processing, and medical support subsystems, appropriate user interfaces (APIs) and RESTful Web Services will be developed to support the exchange of medical data. Based on international e-health standards, such as Health Level 7 (HL7), ICD-10 and SNOMED,

the conceptual interoperability of the end-to-end system will be ensured.

Networking Subsystem

The use of optic network in the wired section will be studied, with Bandwidth Dynamic Assignment and effective handling of the different motion priorities. The same priority system will be used for the wireless section as well. Parameters such as information priority, performance and energy consumption will be considered.

Security

Regarding data security, the key issues to address are: (a) Protection of confidentiality in the transmission, storage and processing of medical data; (b) Protecting the integrity of data (c) Availability of the data to authorized system users and (d) Detailed access control based on the need-to-know principle.

Results

Scenario

As an example of the operation of the system, the following scenario of use is provided. The patient GV enters the specially designed area of ED, where he is assigned to two nurses and one doctor. The patient has a symptom of acute appendicitis and, based on the ESI, he is categorized as Level 3 with a surgical need. The supervising physician decides to use the IntelTriage system to monitor the patient. The information received from the wearable system is available in the patient's medical records. As the patient goes through laboratory examinations, a significant increase in heart rate (> 120 / min) and respiratory rate (20% increase beyond normal) is observed. The Decision Support System then classifies the patient at Level 2, automatically alerting the medical staff about the issue and the location of the patient, in order to take appropriate actions.

Initial list of functional requirements

Based on the scenario provided by clinicians and a literature review, the initial list of functional requirements that IntelTriage shall meet are listed in Table 1.

Table 1– IntelTriage functional requirements initial list

No	Requirement	Ref.
1	Recording data from a biosensor device, incl. heart rate and oxygen saturation	[4]
2	Algorithmic breathing rate calculation	[5]
3	Local storage of data in the biosensor device	[6]
4	Wearable functionality in case it is out of WiFi range	[6]
5	Encoding data to be send	[7]
6	Data transmission taking into account energy consumption and data loss	[8]
7	Patient indoor location tracking	[6]
8	Automatic recognition of abnormalities in vital signs	[8]
9	Set a threshold for acceptable values of vital signs	[9]
10	Send notifications to the physician about the patient under monitoring	[9]
11	Visualization of indoor location patterns	[8][9]

Discussion & Conclusions

This paper presents an initial attempt towards the definition of functional requirements of the main subcomponents of the IntelTriage system, a smart triage and hospital monitoring system. However, during the lifetime of the project several new versions of functional requirements and implementation refinements will take place based on the active involvement of both clinicians and patients in small scale pilots either in the lab or the actual field of application, the ED and the clinics of the hospital. The system will be installed at the AHEPA ED and 1st Surgical Clinic premises, for a 6-10-month evaluation period. The following outcome measures are expected to be collected:

- Reaction times of nursing and medical staff when an event occurs
- Burn-out levels of nurses and doctors after a shift
- Changes in morbidity and mortality
- Level of patient satisfaction with the services provided

Acknowledgements

The IntelTriage project has received funding from the Ministry of Economy & Development, Special Secretary for ERDF & CF, Managing Authority of EPAnEK Competitiveness Entrepreneurship Innovation operational program, under grant agreement No. T1EΔK-02489.

References

[1] D. R. Eitel, D. A. Travers, A. M. Rosenau, N. Gilboy, and R. C. Wuerz, "The Emergency Severity Index triage algorithm version 2 is reliable and valid," *Acad. Emerg. Med.* **10** (2003), 1070–1080.

[2] M. Dehghani Soufi, T. Samad-Soltani, S. Shams Vahdati, and P. Rezaei-Hachesu, "Decision support system for triage management: A hybrid approach using rule-based reasoning and fuzzy logic," *Int. J. Med. Inform.* **114** (2017), 35–44.

[3] C. Beck and J. Georgiou, "A wearable, multimodal, vitals acquisition unit for intelligent field triage," in *Proceedings - IEEE International Symposium on Circuits and Systems*, (2016), 1530–1533.

[4] J. Welch, B. Kanter, B. Skora, S. McCombie, I. Henry, D. McCombie, R. Kennedy, and B. Soller, "Multi-parameter vital sign database to assist in alarm optimization for general care units," *J. Clin. Monit. Comput.* **30** (2016), 895–900.

[5] J. Hernandez, D. McDuff, and R. Picard, "BioWatch: Estimation of Heart and Breathing Rates from Wrist Motions," *Proc. 9th Int. Conf. Pervasive Comput. Technol. Healthc.*, 2015.

[6] M. P. Farry, "Sensor networks for social networks," *Ieee Trans. Instrum. Meas.* (2006), 1–9.

[7] W. Guédria, E. Lamine, and H. Pingaud, "Enterprise Interoperability," **144** (2013).

[8] O. S. Albahri, A. S. Albahri, K. I. Mohammed, A. A. Zaidan, B. B. Zaidan, M. Hashim, and O. H. Salman, "Systematic Review of Real-time Remote Health Monitoring System in Triage and Priority-Based Sensor Technology: Taxonomy, Open Challenges, Motivation and Recommendations," *J. Med. Syst.* (2018).

[9] A. P. Monitoring, T. Gao, T. Massey, W. Bishop, D. Bernstein, L. Selavo, A. Alm, and D. White, "Integration of Triage and Biomedical Devices for ' T '," *Pulse* (2006), 34–39.

Address for correspondence

A.S. Billis, Lab of Medical Physics, Medical School, Aristotle University of Thessaloniki, PO Box 376,. 54124, Thessaloniki, Greece, tel: +30-2310-999310, fax: +30-2310-999702, email: ampillis@med.auth.gr.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190575

StudyAlert: From eCharts to Modern Messengers

Florian Brenck[a], Achim Michel-Backofen[b], Christian Katzer[a], Nathan Smykalla[a], Constantin Bott[b], Christian Koch[a], Michael Sander[a]

[a] *Department of Anesthesiology, Operative Intensive Care and Pain Therapy, Justus-Liebig-University Giessen, Germany*
[b] *Department of Clinical Data Management, Justus-Liebig-University Giessen, Germany*

Abstract

We developed a tool to detect patients possibly elegible for clinical studies by analysing the HL7-message-stream of the patient management system and notifying study investigators by email or a common messenger service via secured communication channels.

Keywords:

Electronic Health Records, Clinical Trial, Health Level Seven

Introduction

Electronic patient charts are a widespread method of clinical documentation across hopsitals in Europe and Northern America. As documentation is easy configurable in these systems it is a great resource for screening hospitalised patients for the ability of being included into clinical studies [1-2]. We developed a tool for scanning all messages sent by the clinical documentation software ICUData to the server. These messages are sent as HL7-2. Our tool is adjustable to every other patient data management system (PDMS) using this international communication standard [3-4].

Methods

Using Java we developed a graphical user interface (GUI) to allow setting up screening protocols. Investigators could be registered together with their contact information. In this GUI elements of interest within the patient chart could be defined. It is possible to scan HL7-messages for any item that is available in the patient chart and is chosen/filled out or even provisional. Furthermore items can be defined to exceed or fall below defined cut-offs before raising an alert. Messages from the PDMS are read by the StudyAlert-application server and the bitestream is analysed to meet any of the defined criteria. If an item is found the preferred method of notification is selected and the registered investigator is informed about this hit.

This will be done by either email or the threema(R) messenger service. Complying with the strict security measures of a university hospital we employed a bridge communication between the hospital-IT-infrastructure and the Internet. All necessary information about a hit are sent from the StudyAlert-Server to the StudyAlert-Notifier Server via email from the hopsital-network to the university-IT-network where the latter server is installed. Using SMTP email communication the investigator is informed when chosing email. If the messenger service is used we employed the threema-API to send out messenger notifications stating patient-id and name of study for which this patient might be eligible according to the scanning criteria. A third server process receives delivery and read notifications using http-POST-requests so that the whole process of scanning, hit, sending notification and read status can be traced.

Results

We developed a fully scalable screening tool for possible study patients employing only standard technologies and allowing to record a complete tracing protocol.

Discussion

Scanning clinical information entered in PDMS or HIS in real-time to raise alert upon certain conditions is a developing use case in these systems [5]. Screening for study patients in large university hospitals is often hard to manage because of the number of paitents and the various criteria to be met. Analysing the complete HL7-Data traffic is a sufficient and time safing way, using a common messenger service for notification of the study personall allows a complete trace documentation about notification process.

References

[1] R. W. Carlson, S. W. Tu, N. M. Lane, T. L. Lai, C. A. Kemper, M. A. Musen et al., Computer-based screening of patients with HIV/AIDS for clinical-trial eligibility, *Online J Curr Clin Trials* (1995), Doc No 179.

[2] D. A. Cook, F. Enders, P. J. Caraballo, R. .A. Nishimura, F. J. Lloyd, An automated clinical alert system for newly-diagnosed atrial fibrillation, *PloS one* 10 (2015), e0122153.

[3] J. C. Camacho Rodriguez, S. Staubert, M. Lobe, Automated Import of Clinical Data from HL7 Messages into OpenClinica and tranSMART Using Mirth Connect, *Stud Health Technol Inform*, 228 (2016), 317-21.

[4] J. W. Kuo and A. M. Kuo, Integration of Health Information Systems Using HL7: A Case Study, *Stud Health Technol Inform* 234 (2017), 188-94.

[5] R. Bellomo, M. Chan, C. Guy, H. Proimos, F. Franceschi, et al., Laboratory alerts to guide early intensive care team review in surgical patients: A feasibility, safety, and efficacy pilot randomized controlled trial, *Resuscitation*, 133 (2018), 167-172.

Address for correspondence

Dr. Florian Brenck
Department of Anesthesiology, Operative Intensive Care and Pain Therapy
Justus-Liebig-University Giessen
Rudolf-Buchheim-Strasse 7, 35392 Giessen, Germany
Tel +49 641 985 444 94
Fax +49 641 985 444 99
florian.brenck@chiru.med.uni-giessen.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190576

Assessment of Kuwait Health System Towards Telemedicine Readiness & Adoption: Organizational and Technical Issues

Ali Buabbas [a], Hamza Alshawaf [b]

[a] *Community Medicine & Behavioural Sciences, Faculty of Medicine, Kuwait University*
[b] *Health Informatics & Information Management Department, Faculty of Allied Health Sciences, Kuwait University, Kuwait*

Abstract

This study aims to assess the organisational readiness to predict the success of telemedicine system implementation in Kuwait. Semi-structured interviews were conducted with: (1) IT managers to assess the readiness of the technical infrastructure of Kuwait health system; and (2) policy makers to assess the organisation's ability to receive telemedicine. A set of policies and strategies were developed to demonstrate how the Kuwaiti health system could be ready for telemedicine services.

Keywords:
Adoption, health planning, telemedicine

Introduction

New and significant changes are expected in healthcare organisations upon the adoption of telemedicine technology. Implementation of information system (IS) in health care sector is not easy; it is a multi-dimensional process, wherein the failure rate of US projects can reach at least 40% [1, 2]. This failure deters managers and policy makers from easily accepting change. A literature review showed how the adoption of telemedicine depends on acceptance by different stakeholders [3]. Therefore, the success of an IS project relies on meeting the requirements of the users. The importance of readiness appears often in the literature due to the impact that the current abilities of an organisation can exert to inhibit or support telemedicine adoption. In the Telemedicine Maturity Adoption Model by HIMSS, the first level requires providers to have a telemedicine strategy in place, with assurances of regulatory compliance and security [4].

The aim of the research study was to assess the organisational readiness to predict the successful implementation of telemedicine adoption in Kuwait health care system. The research objectives were: (1) To explore and identify the factors that support or hinder the adoption of telemedicine systems from the perspective of the organisation and the technical issues; (2) To develop a policy for overseas referral patients to complement the existing policy; (3) To develop a strategy for optimum implementation of a telemedicine system in Kuwait.

Methods

A qualitative study was used through conducting semi-structured interviews with: (1) Six information technology (IT) managers who were responsible for Information and Communication Technology (ICT) implementation in the six health regions at the Ministry of Health; (2) and seven policy makers who were responsible for overseeing telemedicine adoption.

Our main thematic groups for the interview questions were based on the literature and considered the research setting. Each thematic group was sub-divided into sub-topics: (1) Current referral abroad system, which included size of budget and patients' issues; (2) ICT and telemedicine context, which included ICT use and potential of telemedicine; (3) Telemedicine adoption, which included factors of inhibiting and supportive, and the impact of culture; and (4) Decision makers and organizational readiness, which included existing and new polices, and the role of the Ministry of Health.

Regarding the interviews of IT managers, two main thematic groups were identified with associated sub-topics: (1) Readiness of ICT infrastructure, which included the currant capability of ICT, telemedicine technical requirements, and policy of ICT use; and (2) telemedicine adoption, which included institutional plans and managerial and cultural factors

Interviews were 30 minutes in duration. The interviews were transcribed manually, and thematic analysis was performed, wherein sub-topics were developed and confirmed by the two researchers.

Results

The demographics of the participants (IT managers and policy makers) consisted of all being male and Kuwaitis. All the policy makers were physicians with a minimum of ten years of administrative experience. IT managers were computer engineers and had a minimum experience of five years.

The policy makers confirmed that the Kuwait Ministry of Health had concerns over the size of the budget that had been allocated annually to refer Kuwaiti patients abroad for treatment. They also stated that there were psychological pressures on patients who travelled abroad for treatment, such as emotional effects from leaving their home. Limited awareness and abscence of a leader specialist in using ICT in medicine was attributed to the lack of consideration for telemedicine. All of the participants agreed that telemedicine would be useful and cost effective if implemented properly. Staff's resistance to change, work overload, and legality issues were identified as inhibitors specific to the Kuwait health system. The participants agreed that the cultural and social concerns of the patients should be carefully considered, including supporting privacy and confidentiality through a secure telemedicine system. Policy makers agreed that to avoid misuse new policies should be developed to organise the teleconsultation process.

IT managers stated that available equipment (i.e. computers, networks, servers, printers) in the six health regions in Kuwait were compatible and reliable. However, electronic health

record system implementation has not been completed, and although there are future government plans to enhance internet connection, it currently is not always available. Participants were not aware about the concept of telemedicine. They reported the obvious bureaucracy of the main administration, where its centralization was attributed to a slow decision-making process. Most of the participants indicated that no policy existed for purchasing and use of ICT.

Discussion

Strategies for the implementation of a telemedicine system in the Kuwaiti healthcare system include complete implementation of health information systems and electronic patient records in the hospitals, and awareness campaigns for telemedicine use among the medical departments. [5]

A policy has been developed to demonstrate how telemedicine can be integrated with the Kuwaiti healthcare system. A limitation of this study is the few numbers of participants.

Conclusion

The qualitative approach succeeded in assessing the current technical infrastructure of ICT use in the Kuwaiti healthcare system, as well as the organisational readiness.

References

[1] K. Kimiafar, M. Sarbaz, A. Sheikhtaheri, and A. Azizi,The impact of management factors on the success and failure of health information systems. ***Indian Journal of Science and Technology*** **8** (2015), ISSN 0974 -5645.

[2] B. Kaplan and K.D. Harris-Salamone, Health IT success and failure: Recommendations from literature and an AMIA workshop. *J Am Med Inform Assoc,* **16**(3) (2009), 291–299.

[3] Abouzahra, M. Causes of failure of health IT projects. 3rd International Conference on Advanced Management Science, *IPEDR, IACSIT Press* **19** (2011).

[4] N. Sokolovish and B. Fera, Advancing telemedicine through an adoption model, in: HIMSS15 Conference, Chicago, Illinois, 2015. http://assets.fiercemarkets.net/public/healthit/himss15eytelemed.pdf

[5] Buabbas A. *Investigation of the adoption of telemedicine technology in the Kuwaiti health system: Strategy and policy of implementation for overseas referral patients*, PhD thesis, Brunel University, 2013.

Correspondence:

Ali Buabbas, Community & Behavioral Medicine,
Faculty of Medicine, Kuwait University,
ali.buabbas@HSC.EDU.KW

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190577

Impact of the Performance Gap Between Interactive Alerts and Quality Metrics

Kari Bunkers[a,b], Daniel Cronk[b], Marcia Core[c], David Blair[a], Mark Parkulo[b,d], Pedro J. Caraballo[b,d]

[a]*Department of Family Medicine*
[b]*Center for the Science of Health Care Delivery,*
[c]*Department of Information Technology,*
[d]*Department of Medicine. Mayo Clinic, Rochester, Minnesota, USA.*

Abstract

Interactive alerts are used to enhance compliance with primary prevention and have been shown to improve quality metrics. However, the degree of impact of these alerts is controversial and there is concern with excessive alerting. Our objective is to develop reliable processes to assess the direct impact of interactive alerts on clinical performance. Here we present preliminary finding related to the evaluation of the performance gaps between alerts and clinical practice.

Keywords:

decision support systems, quality indicators, prevention and control.

Introduction

A plethora of clinical research and national guidelines supports the implementation of multiple primary prevention interventions. Modern electronic health records use clinical decision support (CDS) as integrated on-screen interactive alerts to facilitate ordering these interventions and improve compliance with related quality metrics [1]. However, alerts are not usually implemented in isolation and their direct impact has been difficult to measure. We need reliable tools to critically assess the direct and indirect impact of interactive alerts. Here we present our preliminary work assessing clinical practice interaction with alerts, and the clinical outcome gap as a tool to easily detect opportunities for improvement.

Methods

Interactive alerts allow ordering suggested tests directly from the alert and are available during the entire primary care encounter. Providers can also order the tests outside the alert using regular orders. We retrospectively (1/2018 to 9/2018) collected data: 1. Use of the alerts to order recommended tests, 2. Orders placed outside the alerts. 3. Performance of related quality metrics. We represent these data as proportions of all eligible adult primary care encounters (Figure 1) and defined as performance gaps: Gap 1, the difference between ordering the test inside the alert (dark grey) and outside the alert (light gray) and Gap 2, the difference between ordering the test (dark + light gray) and the quality metric (red line).

Results

Evaluation of 4 interactive alerts (depression screening, breast cancer screening, colon cancer screening and lipid panel follow up) showed variable gaps that are easily visualized in a graphical representation (Figure 1). These gaps vary among interventions and can guide better understanding of the true impact of the alerts and related workflows. For adult depression, the quality metric can be explained by all the orders (Gap 2 is zero). However, only approximately 50% of the orders were done using the alert (Gap 1). Conversely, the other 3 interactive alerts showed a poor performance. The orders done with the alert among all the orders (Gap 1) for breast cancer, colon cancer and lipids were 28%, 50% and 27%, respectively. But more important, Gap 2 is also large, 56%, 70%, and 75%, respectively. These large gaps are likely related to direct outreach to patients via portal and mail using health maintenance registries to identify patients with gaps and bypassing orders done during the clinical encounter and data captured in the CDS transactional system. Although, these proportions are approximations, these preliminary results require additional evaluation of the workflow and consideration to change the alerts

Conclusions

Systematic assessment of the performance gaps of individual interactive alerts could be an important step to enhance clinical impact and manage potential alert fatigue. Additional evaluation is underway to better characterize the different types of gaps and define follow up interventions to correct the gaps including decommissioning of ineffective alerts.

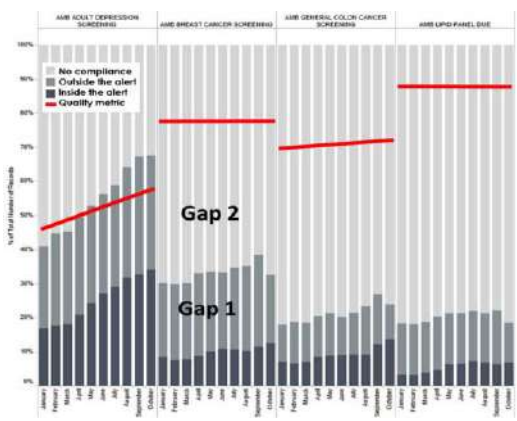


Figure 1– Performance of four interactive alerts comparing the actions taken using the alert, outside the alert and the overall performance of the clinical practice (quality metric).

References

[1] McCoy AB, Thomas EJ, Krousel-Wood M, Sittig DF. Clinical decision support alert appropriateness: a review and proposal for improvement. *Ochsner J*. **14** (2014), 195-202.

Address for correspondence

Pedro J. Caraballo.

Mayo Clinic.

200 First Street SW. Rochester, MN 55905. USA.

Phone: 507-284 8586.

Fax: 507-2845370.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190578

First Feasibility Analysis of Ballistocardiography on a Passenger Flight

Marie Cathrine Wolf[a], Nico Jähne-Raden[a], Henrike Gütschleg[a], Ulf Kulau[b], Mario Kallenbach[c], Klaus-Hendrik Wolf[a]

[a] *Peter L. Reichertz Institute for Medical Informatics, Hannover Medical School, Hanover, Germany*
[b] *Institute of Computer Engineering, Technische Universität Braunschweig, Germany*
[c] *Institute of Flight Guidance, Technische Universität Braunschweig, Germany*

Abstract

The project SCARAB² (Scalable, Robust and Adaptive on Board Ballistocardiography) aims to use Ballistocardiography (BCG) to monitor flight passengers. In order to show that recorded BCG data from flights give evaluable information even in the noisy environment of an airplane, we monitored a heart-healthy passenger using BCG. Furthermore, we show that there can be a conclusion to heart activities from the recorded ballistocardiogram by comparing the data to a concurrently recorded electrocardiogram (ECG).

Keywords:

Ballistocardiography, Micro-Electrical-Mechanical Systems, Accelerometry

Introduction

A significant number of the world's population suffer from cardiovascular diseases, out of which about 17.7 million people die as a result [1]. If a heart-related issue occurs, fast and professional help is needed to prevent worse consequences. However, on an airplane flight, it is indeed hard to handle such an emergency.

Within the SCARAB² project, we try to develop a small, lightweight system to monitor heart-related signals from the passengers to detect potential problems early and provide information for a professional diagnosis, as well as first possible treatment options by the crew.

The chosen method for this is the Ballistocardiography (BCG). With the BCG the cardiac contraction forces transferred via the body's surface on other structures like seats can be used to gain information about the heart's functionality. This method derives from early space measurements, without external disturbances [2; 3]. Unlike space, there are external disturbances such as noisy environment on earth, while talking, walking, and in cars or buses. BCG sensors measure all the acceleration and not only the ones made by the heart. It is, therefore, necessary to know the characteristics of the surrounded noise by measuring them as well as the heart's vibrations on the surface. Nevertheless, this technology has found its place in clinical diagnostics with the help of researching groups, while having nearly noise-free surroundings [4; 5].

To proof the general feasibility of usable BCG measurements from passengers while on an aircraft, we planned a first non-laboratory trial.

The motivation for such an environment is the ever growing air traffic and passenger volume, therefore, the number of medical emergencies during a flight is growing accordingly [6].

Since the aircraft can be perceived as a system that has only the frictional resistance of the air during a flight as an external factor and otherwise represents a self-contained system, we have opted for this measurement environment.

For example, when the passenger is already suffering from cardiac diseases, the stress and the changed pressure or the anxiety can lead to a deterioration of the health state.

Monitoring the passenger's health status during the flight, or in the best case before departure (e.g. during boarding), supports flight attendants to initiate counter procedures. Unlike traveling by car, bus or train, the airplane is not able to stop at need and such incidents cause flight abortion and huge financial loss for the airline and a lot of stress for the passenger [7].

With these motivations in mind, we worked on a method to validate the hypothesis, if it is possible to measure vital parameter of a subject during a flight.

Methods

For the real environment trial, one male, healthy subject with an attached sensor system [shimmer3 ExG] including a 4-lead ECG unit and a triaxial accelerometer was measured during a regular three hours passenger flight with an average passenger airplane (Airbus A330).

The Shimmer is equipped with a Freescale MMA7361 accelerometer sensor with a range between 1.5 and 6g, for the measurement, the 1.5g setup with a sample rate of 204.8Hz was used [8].

The Shimmer was placed on the sternum due to the fact that it has been shown as one of the best points to measure the heart's contraction on the chest surface, using the built-in accelerometer of the Shimmer sensor [9].

Figure 1 - Picture of the Shimmer3 ExG.

Furthermore, the ECG electrodes for the ECG were attached at the predetermined locations on the upper body of the subject, as shown in *Figure 2*. The electrodes are attached to the body using the Einthoven's triangle electrocardiography derivations. Furthermore, one-fourth derivation was measured using the right leg position.

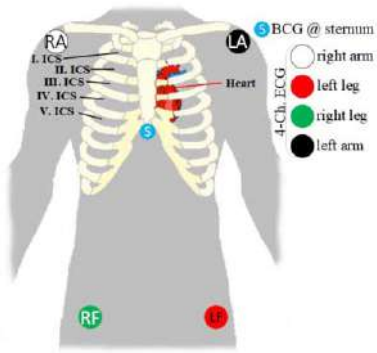


Figure 2 - Sketch of the attached Shimmer 3 ECG Unit (BCG) and the positions of the electrodes on the upper body.

To gain a comparable data stack, the digital output of the accelerometer was calibrated, using the 3D-ellipsoid equation [10]. The further preprocessing was realized using the statistical computing language R (v3.5.1).

The first step of analysing was the detection of the QRS-complex within the ECG reference data by peak detection (package pracma, v1.9.9), which exposed the ECG-R-peaks.

In order to be able to analyze the accelerometer data in relation to the encountered R-peaks, the accelerometer signal had to be filtered, so inadvertent frequencies, e.g. the movement caused by respiration and vibration noise caused by the aircraft itself, could be filtered out. After spectrum analysis via Fourier transformation, we used second order Chebychev band-pass filter with the selected lower and upper frequency of 25 Hz and 60 Hz.

The graph was evaluated by searching for a periodic pattern in the ballistocardiogram in a range of time after the QRS complexes. The delay between the ECG (QRS complex) and the BCG measured contraction derives from the time of muscle stimulus to actual movement.

Results

The heart of an adult human beats with a frequency of approximately one beat per second (Pulse at rest of an adult is between 50 to 90 beats per minute). Following, the localisation of the BCG sensor on the subjects thorax, the frequency of the measurements must be half of the pulse rate, in order to be able to measure the contraction of the left and the right ventricle; so called seismocardiography effect.

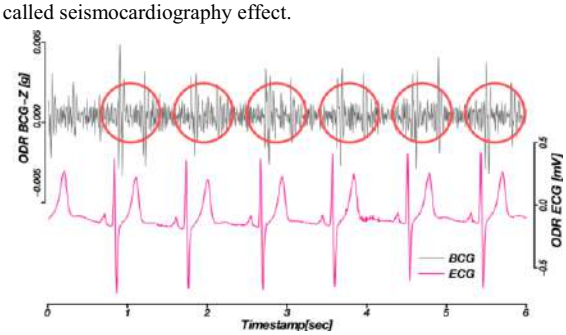


Figure 3 - Electrocardiogram (ECG) and Ballistocardiogram (BCG) with periodic pattern marked with circles.

In Figure 3 the BCG data (only z axis) related to the ECG data can be seen for a signal snippet of 7 seconds. It becomes apparent that periodic pattern in the ballistocardiogram is well identifiable in a range of time after the QRS complexes, which is known to be in average 120ms after R peak [3].

Conclusions

With this work, we present first analyzed inflight measured BCG data from a digital accelerometer attached at the sternum position including an ECG reference signal. It becomes apparent that peaks in the ballistocardiogram are well identifiable and an assessment of acceleration data recorded on a passenger flight is possible without a high amount of data processing.

Furthermore, the graphical representation of the recorded data reveals that a statement over heart activities from the acceleration data is possible. We presented that in the ballistocardiogram, peaks are identifiable in a range of time after the QRS complex in the electrocardiogram.

With this knowledge, we currently prepare further steps, first to increase the accuracy of the data output and to later develop a BCG health monitoring system for passenger flights [7].

References

[1] WHO, Global Hearts Initiative, https://www.who.int/cardiovascular_diseases/global-hearts/en/ [accessed May 10, 2019]

[2] J.M. Zanetti and K. Tavakolian, Seismocardiography: past, present and future, Conf Proc IEEE Eng Med Biol Soc 2013 (2013), 7004-7007.

[3] O.T. Inan, P.F. Migeotte, K.S. Park, M. Etemadi, K. Tavakolian, R. Casanella, J. Zanetti, J. Tank, I. Funtova, G.K. Prisk, and M. Di Rienzo, Ballistocardiography and seismocardiography: a review of recent advances, IEEE J Biomed Health Inform 19 (2015), 1414-1427.

[4] M.D. Rienzo, E. Vaini, and P. Lombardi, Wearable monitoring: A project for the unobtrusive investigation of sleep physiology aboard the International Space Station, in: 2015 Computing in Cardiology Conference (CinC), 2015, pp. 125-128.

[5] A.M. Carek and O.T. Inan, Robust Sensing of Distal Pulse Waveforms on a Modified Weighing Scale for Ubiquitous Pulse Transit Time Measurement, IEEE Trans Biomed Circuits Syst 11 (2017), 765-772.

[6] D. Silverman and M. Gendreau, Medical issues associated with commercial flights, Lancet 373 (2009), 2067-2077.

[7] N. Jahne-Raden, H. Gutschleg, M. Kallenbach, T. Feuerle, and U. Kulau, Poster: SCARAB² - Scalable, Robust and Adaptive on Board Ballistocardiography, in: 2018 14th International Conference on Distributed Computing in Sensor Systems (DCOSS), 2018, pp. 111-112.

[8] www.shimmersensing.com [accessed March 22, 2019]

[9] N. Jähne-Raden, T. Märtin, M. Marschollek, K. Heusser, and J. Tank, BCG-mapping of the thorax using different sensors: First experiences and signal quality, in: 2016 IEEE SENSORS, 2016, pp. 1-3.

[10] M. Gietzelt, K.H. Wolf, M. Marschollek, and R. Haux, Performance comparison of accelerometer calibration algorithms based on 3D-ellipsoid fitting methods, Comput Methods Programs Biomed 111 (2013), 62-71.

Address for Correspondence

Nico Jähne-Raden, Peter L. Reichertz Institute for Medical Informatics, Hannover Medical School, Carl-Neuberg-Str. 1, D-30625 Hanover, Germany, Nico.Jaehne-Raden@plri.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190579

Responses of Staff Nurses to an EMR-Based Clinical Decision Support Service for Predicting Inpatient Fall Risk

Insook Cho[a], Insun Jin[b]

[a] *Department of Nursing, Inca University, Incheon, Republic of Korea,*
[b] *Department of Nursing, National Health Insurance Service Islam Hospital, Gyeonggi-do, Republic of Korea*

Abstract

Wide spread of electronic medical records provide an opportunity to use time-variant longuitudinal data near real time. Hospital nurses would benefit greatly from the ability to use such data to predict adverse event risks of individual patient. We have developed an clinical decision support service to predict inpatient falling using machine learning and clinical big data approach. This study reports the initial responses of nurses to the service in an acute care setting.

Keywords:

Clinical Decision Support Systems, Accidental Falls, User–Computer Interface

Introduction

One-third of inpatient falls are preventable, but actually preventing them remains difficult and complex due to several common problems,[1] among which the authors noticed on inadequate assessments with limits of current fall-risk assessmnet tools. We have developed an evidence-based prediction model for the individual-level risk of falling based on more than 70 longitudinal data elements that are routinely captured by nurses in electronic medical records (EMR).[2] The model uses a Bayesian network inference and machine learning techniques, and it produced high values of the c-statistic of 0.96 and 0.99 in a cross-site validation study conducted at two large hospitals.[2] The performance was compared with those for the Hedrich II fall risk assessment tool (c-statistic=0.69) and the STRATIFY tool (c-statistic=0.65), respectively. The EMR data included in the model were medication, patient classification (KPCS), the fall risk assessment tool, and the nursing-process (assessment, diagnoses, and intervention), demographics, and administrative data. In implementing the prediction model in a clinical decision support service (called IN@SiGHT, Intelligent Nursing @ Safety in Guidance of Health Technology) at a hospital of 800-bed size, the model was customized along with the request of department of nursing of the hospital, asking they want to no longer use of STRATIFY tool. Without the data of STRATIFY tool, we readjusted the model and have gotten the performance of 0.87 for c-statistics. In 2017, the service was integrated with the EMR system of the hospital, and have provided nurses with information about which individual patients are at risk of falling, along with recommendations for planned nursing interventions. The IN@SiGHT service was initially introduced to six medical-surgical nursing units. We have monitored the performance of the service including the changes of response and behavior of its users prospectively. Here we report the short-term responses of nurses in practice over the 6 months since the IN@SiGHT service was started.

Methods

We used a repeated survey approach for data collection and analysis to understand the responses and experiences of nurses. There were 209 nurses working at the 6 included units. We randomly selected 3 nurses from each unit—a senior nurse (experience >5 years), a junior nurse (1 year < experience ≤5 years), and a newer nurse (experience ≤1 year)—and analyzed the total of 18 nurses twice. Repeated sampling of nurses was conducted.

The first survey was performed in May 2017 after launching the service. A training session that introduced the service with guidance on how to use it was provided by the unit for 1 week. The second survey was completed in a self-reported manner 9 months later.

The survey questionnaire consisted of five sections: (i) user experience with the service, (ii) perceived effect on fall prevention tasks, (iii) overall satisfaction, (iv) perceived advantages and disadvantages of the service, and (v) knowledge about fall prevention, including how this is affected by demographics. User perception was measured by the Korean version of the System Usability Scale (SUS),[3] with overall satisfaction scored on a global scale ranging from 0 to 100. Knowledge was measured by the AHRQ Fall Knowledge Test tool.[4] The perceived effect and perceived advantages and disadvantages were queried using open-ended questions. Quantitative data were analyzed as frequency and ratio values, while thematic data analysis was applied to qualitative data. This study was approved by the hospital IRB review board (IRB # 2016-08-005-005).

Results

The average ages of the participants were 28.7 and 27.9 years at the two surveys, and the average lengths of their clinical experience were 5.0 and 5.4 years, respectively. Their level of knowledge about fall prevention, SUS score, and overall satisfaction score all increased during the study. (Figure 1)

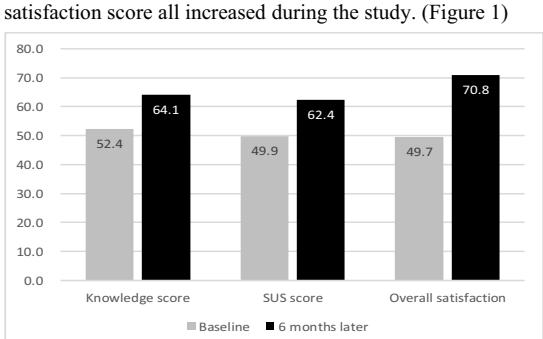


Figure 1. Comparisons of knowledge score, Sysyem Usability Scale (SUS) score, and overall satisfaction

Regarding the perceived effects on fall prevention tasks, most of nurses said that they became very careful about documentation due to the thoughts that the data they entered will be used to infer risks of falling. In addition, the care plan suggested along with fall risks makes their work be efficient. But, on the other hand, they felt burdensome to enter assessment data. These results were consistent at the answers on the questions about the advantages and disadvantages of the CDS service. (Table 1) As the advantages of the service, overall helpfulness and trustfulness were the frequent themes and the number of mentioning also increased. Regarding the disadvantages, several nurses said that the CDS information of risk was different from their own judgment, which decreased at six months later. The number of nurses stating that the burden of documentation was 31% (= 9/ 29) and decreased to 28% (= 7/25). Several nurses expressed concerns on the increase of just entering nursing interventions which were suggested by the system, but not conducted by them.

Table 1. Responses of the nurses about the advantages and disadvantages of the clinical decision support service (IN@SiGHT)

		Frequency(%)	
Responses		**Start point**	**6 months later**
Adva-ntages	The CDS service was helpful overall	14 (52)	18 (42)
	The CDS information helped them pay attention to risk patients and their factors	9 (33)	17 (39)
	The information of CDS was trustful	4 (15)	8 (19)
	Subtotal	27 (100)	43 (100)
Disadv-antages	Sometimes the CDS information was different from their own judgment	13 (45)	9 (36)
	They became more careful for documentation with increased burden on it	9 (31)	7 (28)
	Concerns for increased documentation of not conducted interventions	5 (17)	6 (24)
	Hard to understand the system algorithm	2 (7)	3 (12)
	Subtotal	29 (100)	25 (100)

Discussion

The nurses had attitudes that were neutral or even slightly negative when the IN@SiGHT risk prediction service was first introduced. Some of the experienced nurses reported that the predictions were different from their judgment. Nurses clinical judgment depends on individual experience and expertise, and bedside nurses have different levels of nursing expertise, which tends to lead to variation in clinical judgment and decision-making. However, the clinical judgment combined with objective observations and measurements is the most accurate way to predict safety risks. Some of junior and newer nurses were more cooperative and had more-positive attitudes associated with their curiosity about the service. Most of the nurses showed better knowledge, satisfaction, and attitudes after the 6-month intervention. On the other hand, we found that some nurses were not familiar with predictive analytics and might not be able understand or interpret the analysis results.

Conclusions

The widespread adoption of EMR is making clinical big data a frequently discussed topic that presents new opportunities in practice. To achieve it more efficiently, we need to assess users understanding, knowledge, and attitude on such approaches and consider their sociotechnical aspect's concerns including workflow and workaround patterns.

Acknowledgements

This study was supported by a grant of the Korea Healthcare Technology R&D Project, Ministry for Health, Welfare & Family Affairs, Republic of Korea (No. HI17C0807) and National Research Foundation of Korea (No. NRF-2019R1A2C2007583)

References

[1] The Joint Commission. Sentinel event alert 55: Preventing falls and fall-related injuries in health care facilities. Sentinel Event Alert; 2015. Available at: http://www.jointcommission.org/assets/1/18/SEA_55.pdf (accessed March 2017).

[2] Cho I, Boo E-H, Chung E, Bates DW, Dykes P. Novel Approach to Inpatient Fall Risk Prediction and Its Cross-Site Validation Using Time-Variant Data. Journal of medical Internet research. 2019;21(2): e11505.

[3] Brooke J. SUS: a quick and dirty usability scale. Usability Evil Ind. 1996;189(194):4–7.

[4] Ganz DA, Huang C, Saliva D, Shier V. Preventing falls in hospitals. Prepared by RAND Corporation, Boston University School of Public Health, and ECRI Institute; 2013. Available at: http://www.ahrq.gov/professionals/systems/hospital/fallpxtoolkit/index.html (accessed March 2016).

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190580

Adverse Drug Event Reporting Rates After the Implementation of an EHR Integrated Reporting System

Eunice Correa, Oscar Ignacio Jauregui, Julia Frangella, Verónica Peuchot, Carlos Otero, Daniel Luna

Department of Health Informatics, Hospital Italiano of Buenos Aires, Buenos Aires, Argentina

Abstract

Under-reporting of adverse drug events (ADEs) is a common issue across healthcare systems, and lack of integration with clinical workflow and systems are among the leading causes of this problem. We sought to describe the development of an ADEs reporting system within an EHR that represents user needs and captures relevant data. We compared periods before and after the implementation, and describe the corresponding reporting rates.

Keywords:

Drug-Related Side Effects and Adverse Reactions; Pharmacovigilance; Electronic Health Records

Introduction

Adverse drug events (ADEs) are a leading cause of emergency department visits and hospitalization. They tend to occur in healthcare facilities even if there are means to prevent them, because ADEs information transfer is inadequate causing unintentional re exposures to the substance [1]. To ensure pharmacovigilance activities, robust ADE reporting mechanisms should be in place to deliver timely and comprehensive data to assess safety of used drugs [2,3].

Underreporting of ADEs to internal or external pharmacovigilance entities is a major limitation to drug safety surveillance systems, estimated in 80% to 95% [4,5]. Among the reasons are time constraints and duplication of the documentation, since most reporting systems are not integrated with EHRs. In that matter, ADEs reporting systems that are embedded in the clinical workflow, that make use of existing data to avoid duplication of documentation and capture the complex nature of adverse drug events may address this problem and increase ADEs reporting rates [6].

The aim of this study is to compare the ADEs reporting rates before and after the implementation of an EHR integrated electronic reporting system, and to describe the design and development of such functionality.

Methods

Setting

HIBA is a non-profit healthcare academic center that has a network of two hospitals with 750 beds (200 for intensive care), 41 operating rooms, 800 home care beds, and 25 outpatient clinics located in Buenos Aires city and its suburban area. Since 1998, the HIBA has run an in-house-developed health information system, supporting both clinical and administrative operations [7]. It has been recently certified by the HIMSS as level 7 in the Electronic Medical Record Adoption Model, being the first hospital in Argentina and the second in Latin America reaching this stage.

The HIBA has an informatic tool for the reporting of ADEs to the pharmacovigilance section that can be accessed within the incident reporting form, located in the hospital's intranet. The report can be done in a free text form, and has no integration with patient or professional master indexes, nor with the Hospital's EHR.

Reporting System Design and Development

An interdisciplinary team was formed, which carried out interviews with physicians, nurses and pharmacists that were users of the ADEs reporting systems, and with members of the pharmacovigilance section to better understand the causes of the under reporting of events. Then an interface mockup was design taking into account the elements that should be present. A series of user testing sessions were performed, and usability flaws were detected and solved prior to the design of the high-fidelity mockup, which was again tested by end users in an iterative process. Finally, the software was developed, and prior to the implementation a series of education material was released to users by email and in the institution´s intranet.

Substance Controlled Subset Development

The HIBA has a terminology server that is offered as a service for applications inside and outside of the institution achieving semantic interoperability [8]. In that context, a substance valueset was created within the terminology server by automatic mapping all of the substances present in SNOMED- CT database.

Data Collection and Analysis

The study design was a before and after study. The inclusion criteria were adult hospitalized patient with an ADEs report during hospital stay.

Observation period was divided into two:

- Pre-implementation group was defined as patients hospitalized between August and October 2017. During this period the main tool for ADEs reporting was not integrated into EHR.
- Post-implementation group was defined as patients hospitalized between August and October 2018. During this period the developed tool to report from EHR for ADEs reporting was available.

Statistical analysis was performed using Chi-squared tests in the STATA software.

Results

ADEs Report Interface

The reporting interface consisted in a pop-up menu that could be triggered from different modules inside the EHR. A "substance" and a "clinical manifestation" fields were created, in which the user could select a term from a specially created subset of terms.

Then the user had to select the severity of the reaction, the date if known, and a checkbox to report the reaction to the hospital's pharmacovigilance section. The generated adverse reactions would be loaded as a problem into the patients' problem list. Also, an email would be sent to the pharmacovigilance section if the corresponding checkbox was selected by the user.

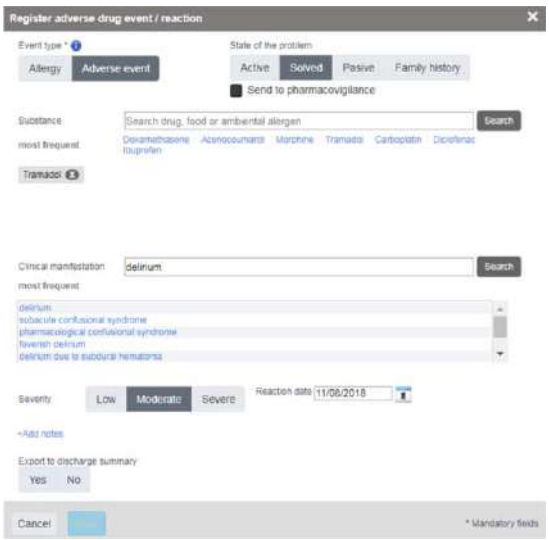


Figure 1 – ADEs Reporting System Interface

ADEs Reports

The number of reported ADEs in both periods are presented in Table 1. The number of reports per hospital day was 0,9/10000 in the first period and 1.9/10000 in the second period. Table 2 depicts the characteristics of the reported cases in both periods.

Table 1– ADEs Reports Before and After the Intervention

	Before	After	p-value
Admitted patients	12,810	12,848	
Total inpatient days	75,853	71,712	
ADEs reports *			
Total	7(0.5‰)	14(1.1‰)	0.128
By incident report	7	5	0.561
By EHR tool		9	
Inpatient day &	4 (2-6)	3 (2-6)	0.116

*Absolute frequency (Relative rate per 1000 patient)
&Median (Interquartile Range)

Table 2– ADEs Report Cases Characteristics by Period

	Before (n=7)	After (n=14)	P value
Inpatient day &	5 (2-6)	6.5 (4-20)	0.226
Patients age	73 (66-79)	58 (30-71)	0.03
Inpatient day &	4 (2-6)	3 (2-6)	0.116

&Median (Interquartile Range)

Conclusions

After our analysis we arrive to the conclusion that having the possibility of reporting adverse reactions integrated into the EHR, seems to facilitate the reporting and to improve communication with the pharmacovigilance sector. However, it is necessary to explore larger periods of implementation of the tool and to carry out the educational stage to users.

References

[1] P.J. Zed, R.B. Abu-Laban, R.M. Balen, P.S. Loewen, C.M. Hohl, J.R. Brubacher, K. Wilbur, M.O. Wiens, L.J. Samoy, K. Lacaria, and R.A. Purssell, Incidence, severity and preventability of medication-related visits to the emergency department: a prospective study, *CMAJ.* **178** (2008) 1563–1569.

[2] S. Suissa, and E. Garbe, Primer: administrative health databases in observational studies of drug effects--advantages and disadvantages, *Nat. Clin. Pract. Rheumatol.* **3** (2007) 725–732.

[3] B. Carleton, Post-Marketing Pharmacosurveillance in Canada: A Background Paper Prepared for the Working Conference on Strengthening the Evaluation of Real World Drug Safety and Effectiveness, 2005.

[4] L. Hazell, and S.A.W. Shakir, Under-reporting of adverse drug reactions : a systematic review, *Drug Saf.* **29** (2006) 385–396.

[5] E. Lopez-Gonzalez, M.T. Herdeiro, and A. Figueiras, Determinants of under-reporting of adverse drug reactions: a systematic review, *Drug Saf.* **32** (2009) 19–31.

[6] C.M. Hohl, S.S. Small, D. Peddie, K. Badke, C. Bailey, and E. Balka, Why Clinicians Don't Report Adverse Drug Events: Qualitative Study, *JMIR Public Health and Surveillance*. **4** (2018) e21.

[7] F. Plazzotta, D. Luna, and F. González Bernaldo de Quirós, [Health information systems: integrating clinical data in different scenarios and users], *Rev. Peru. Med. Exp. Salud Publica*. **32** (2015) 343–351.

[8] D. Luna, G. Lopez, C. Otero, A. Mauro, C.T. Casanelli, and F. González Bernaldo de Quirós, Implementation of interinstitutional and transnational remote terminology services, *AMIA Annu. Symp. Proc.* (2010) 482–486 .

Address for correspondence

Oscar Ignacio Jauregui.
Hospital Italiano de Buenos Aires.
Juan D. Perón 4190 (C1181ACH)
Buenos Aires, Argentina
Tel.(5411) 4959-0200
oscar.jauregui@hospitalitaliano.org.ar

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190581

Patient Health Information Technology Designed for Shared Decision Making: If We Implement It, Will It Become Normal Clinical Practice?

Selena Davis

University of Victoria, School of Health Information Science, Victoria, Canada

Abstract

Personal health records designed for shared decision making have the potential to engage patients in self-management decision making. This pre-implementation evaluation was underpinned by Normalization Process Theory and utilized mixed methods to describe the cognitive and behavioral work of implementation, and its potential to integrate into practice. Participants invest in sense-making, commitment and appraisal processes with strong agreement for positive impact on engagement. Future implementation success will depend on a systemic investment of collective action.

Keywords:

Personal Health Records, Decision Making

Introduction

Today there is increased interest in health information technologies (HIT) that engage patients in decision making as part of health self-management. Health decisions are important, and an optimal and collaborative approach is shared decision making (SDM) between the patient and care provider. SDM that results in a treatment decision and care plan at a specific point in time combines the best available evidence and patient values and preferences [1]. The process of SDM is modelled [2][3] to include four core elements:

1. Acknowledge – awareness that a decision is needed, and choice exists;
2. Consider – receive and interpret options, including benefits and risks;
3. Decide – explore preferences, values, and goals and incorporate them into the making of the decision; and
4. Act – record the decision and track outcomes.

SDM is fundamental to person-centered care, increases patient and provider satisfaction, improves quality of life and clinical outcomes, and fosters a better patient–provider relationship [4]. However, implementation into practice has been difficult [5].

Personal health records (PHR), a patient-facing HIT, is a promising technology to overcome barriers for integrating SDM [6]. PHRs offer patients access to their health information, the ability to contribute information to their health record, health management and decision support tools, and the means to communicate with their providers. Patients' experiences with accessing their PHR are positive and offer feelings of empowerment and engagement [7] and the use of a PHR improves communications with care providers, as well as a sense of self-management [8]. Yet PHRs have not seen widespread adoption, often a result of system architecture or functionalities [9].

*e*nhanced-*PHR* System (*e-PHR*)

A conceptual framework for PHR technology designed to enable SDM (*e-PHR*) indicated 22 PHR functions to enable the four core elements of SDM [10]. For example, the SDM core element Decide is enabled by the PHR core functions of *Access health information* and *Communicate with others,* and comprises 11 sub-functions, such as access to provider clinical notes and care plan and message care team. The scoping literature study [10] also indicated the environment for *e-PHR* is architected as an interconnected PHR – i.e. it gathers and auto-populates patient data from multiple health information systems and applications.

Designed for function and the broader digital health ecosystem, PHRs present an opportunity for improvement in SDM and patient engagement. To implement *e-PHR*, a process evaluation is useful since the literature informs us that the number of implemented interventions that become 'normalized' is limited (i.e. fits in with the work of individuals and the context of practice and no longer requires additional effort). Normalization Process Theory (NPT) seeks to understand the cognitive and behavioral process work people do, individually and collectively, to integrate an intervention in its social context [11].

Guiding Theoretical Framework

NPT provides a framework to analyze four mechanisms and their constructs known to influence implementation success. Coherence is the sense-making work required by individuals in teams. Cognitive Participation is the work done as they anticipate roles and tasks to accomplish the new ways of doing things. Collective Action is the work of operationalizing the intervention. Reflexive Monitoring is comprehending the effects. NPT has been used to aid implementation planning, and evaluating and understanding implementation processes [11].

This research aimed to understand: (i) the integration work of *e-PHR* that participants do, individually and collectively; and (ii) the potential for *e-PHR* to integrate into practice to engage patients in self-management decisions.

Methods

This mixed methods investigation utilized a triangulation convergence study design. This design concurrently collects and equally-weights quantitative measurement instruments and surveys, and qualitative semi-structured individual interview data. The approach offered a deeper level of understanding and gave a voice to the multiple participant types (patients, care providers, and organizational HIT and clinical leaders).

The first quantitative measurement instrument of NPT, the NoMAD [12], describes the level of agreement of participants with

statements of the four NPT mechanisms and related constructs. Utilizing the same five-point Likert scale as the NoMAD instrument, a fixed survey of practice-related outcomes scored the level of agreement with potential outcomes, such as engagement in self-management decision making. A semi-structured interview, guided by NPT, was used to describe the work of integrating *e-PHR*. The concurrently-gathered qualitative and quantitative data were collected to thematic saturation. Survey data was analyzed using R statistical software for descriptive statistics. Interview data were transcribed and imported into Atlas.ti and coded for evidence of the constructs of NPT via a deductive approach. Finally, emergent descriptive themes were identified, along with quotations that best illustrated the themes. The analyzed quantitative and qualitative data were amalgamated to present findings.

Results

Twenty-seven participants in British Columbia, Canada participate: patients (*n=8*), care providers (*n=11*) and organizational leaders (*n=8*). The qualitative results produced a rich explanation of the implementation work. The quantitative results offered mean scores and indicated a direction of agreement with measures of normalization and potential outcomes.

Congruency was observed between the quantitative and qualitative data. Qualitative results indicated that *e-PHR* made sense as explained by two themes for Coherence: a *game changing technology* and *sensibility of change*. To participants, *e-PHR* is a supportive approach to healthcare for patients and would formalize collaborative relationships, provide access to a comprehensive set of data, and offer timely and convenient communications. Meaning was also linked to concerns of workload and workflow, and a limited shared understanding of purpose. Participants strongly agreed (4.6/5) with Cognitive Participation processes. This was explained qualitatively by two themes: *sharing ownership of the work* and *enabling involvement*. To participants, *e-PHR* is the right direction for healthcare. Participants expressed an openness to new ways of working individually and together, as well as some fear of change and lack of systemic ownership of the change. Weak agreement (3.6/5) was observed with Collective Action processes. This implementation mechanism was best explained qualitatively by one theme: *uncovering the challenge of building collective action*, and three sub-themes: *assessing fit*, *adapting to change together*, and *investing in the change*. The effort to enact *e-PHR* would require an upskilling of care providers, a shared accountability among patient and care providers, and sufficient leadership and financial investment. Participants expressed a potential disruption of current relations and a concern for a systemic willingness to adapt to change together; as well as, challenges with team work and some lack of trust and confidence in others ability to carry out tasks. Participants appraised *e-PHR* as explained by two themes for Reflexive Monitoring: *reflecting on value*, and *monitoring and adapting*. To participants, *e-PHR* would remove barriers to care and increase care efficiency, but outcomes must be measured, and benefits demonstrated. Finally, participants strongly agreed (4.5/5) that *e-PHR* would positively affect engagement in self-management decision making and agreed (3.9/5) it would become routine work, offering two descriptive themes for the practice-related outcomes: *care is efficient,* and *care is person-centered*.

Conclusions

The state of SDM in clinical practice is not a question of whether we should do it, rather it is a question of successfully integrating it within today's evolving EHR–PHR ecosystem and person-centered care approach, and with tomorrow's mobile, interconnected, and ubiquitous technology environment to engage patients in self-management decisions. Using the NPT framework, findings from this pre-implementation process evaluation indicated participants invest in sense-making, commitment and appraisal work of e-PHR; but, integration into practice will only be attained when systemic effort is invested to enact it. Further research is needed to explore this gap to inform priorities and approaches for a future implementation.

Acknowledgements

Research supported by Diabetes Action Canada, a strategic patient-oriented research network in diabetes and its related complications, part of CIHR SPOR in Chronic Disease.

References

[1] V. M. Montori, "Shared Decision Making." [Online]. Available: http://shareddecisions.mayoclinic.org/.

[2] G. Elwyn et al., "Shared decision making: a model for clinical practice.," *J. Gen. Intern. Med.*, vol. 27, no. 10, pp. 1361–7, Oct. 2012.

[3] G. Makoul and M. L. Clayman, "An integrative model of shared decision making in medical encounters," *Patient Educ. Couns.*, vol. 60, no. 3, pp. 301–312, Mar. 2006.

[4] L. A. Shay and J. E. Lafata, "Where Is the Evidence? A Systematic Review of Shared Decision Making and Patient Outcomes," *Med. Decis. Mak.*, vol. 35, no. 1, pp. 114–131, Jan. 2015.

[5] F. Légaré and H. Whitteman, "Shared decision-making: examining key elements and barriers to adoption into routine clinical practice," *Health Aff.*, vol. 32, no. 2, pp. 276 – 284, 2013.

[6] A. G. Fiks et al., "Parent-Reported Outcomes of a Shared Decision-Making Portal in Asthma: A Practice-Based RCT," *Pediatrics*, vol.135(4), pp. e965–e973, Apr. 2015.

[7] J. Moll et al., "Patients' Experiences of Accessing Their Electronic Health Records: National Patient Survey in Sweden," *J. Med. Internet Res.*, vol. 20, no. 11, p. e278, Nov. 2018.

[8] J. J. Rief et al., "Using Health Information Technology to Foster Engagement: Patients' Experiences with an Active Patient Health Record," *Health Commun.*, vol. 32, no. 3, pp. 310–319, 2017.

[9] T. Lee, *Information Technology for Patient Empowerment in Healthcare*. Walter de Gruyter GmbH & Co KG., 2015.

[10] S. Davis et al., "Shared decision-making using personal health record technology: a scoping review at the crossroads," *J. Am. Med. Informatics Assoc.*, vol. 24, no. 4, pp. 857–866, Jul. 2017.

[11] C. R. May et al., "Using Normalization Process Theory in feasibility studies and process evaluations of complex healthcare interventions: a systematic review," *Implement. Sci.*, vol. 13, no. 1, p. 80, Dec. 2018.

[12] T. L. Finch et al., "NoMAD: Implementation measure based on Normalization Process Theory. [Measurement instrument]," 2015. http://www.normalizationprocess.org.

Correspondence:

Dr. Selena Davis, daviss@uvic.ca

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190582

How to Improve Local-Level Data Use Culture at Each Level of the Health System? An Implementation Science Study

Kassahun Gashu[a], Alemayehu Teklu[a], Arielle Mancuso[b], Ashenafi Tazebew[a], Berhanu Endehabtu[a], Zeleke Mekonnen[a], Binyam Tilahun[a]

[a] College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia
[b] World Health Organization, Geneva, Switzerland

Abstract

Health programs are reliant on complex decision-making to efficiently utilize limited resources, but local-level data use is still challenging. This study aimed to assess barriers to fostering local-level data use culture at each level of a health system. Results show that awareness gaps, lack of motivation, inconsistent supervision, poor community engagement, and lack of accountability are major bottlenecks. Establishing an accountability system and capacity building on health data use could improve its implementation.

Keywords:
Health information systems, Ethiopia

Introduction

Health programs are reliant on a complex decision-making process to efficiently utilize limited resources at the local level. Health systems require quality data from health information systems to plan and ensure that the workforce is fully funded and equipped with the necessary commodities, infrastructure, resources, and policies to deliver services [1]. Health data enables health planners and managers to make decisions regarding the effective functioning of health facilities. At higher levels, health information is needed for strategic policy-making and resource allocation [2]. The relationship between data quality, demand for data, and data use creates a cycle that leads to improved health programs and policies [3]. Evidence shows that organizational and technical factors can affect data use culture [4].

A study in Uganda showed that health care data use was 24% [5]. A study in Ethiopia indicated that data use for decision-making was not adequate at lower levels and feedback mechanisms were weak [6; 7]. A data quality and information use assessment in Ethiopia showed a limited culture of data use for decision-making. Only 37% of the facilities based discussions and decisions on findings from routinely collected health information [8]. The use of health data for decisions and actions to improve the quality of health services and achieve performance goals is also vital to improve shared accountability within organisations [9].

Existing evidence indicates that data quality and the culture of data use in Ethiopia is poor, however little is known about the major barriers that affect data use at local levels of the health system. This study aimed to identify barriers to establishing local-level data use practices that could help to inform the creation of interventions to address bottlenecks.

Methods

This study was conducted in North Gondar Zone, Amhara region, Northern Ethiopia. The zone has 22 Districts and 557 Kebeles (lowest unit of administration). A qualitative study was conducted from January to September, 2017. Data was gathered through key-informant interviews using a semi-structured question guide. Data producers, data users, community members, and decision makers were engaged for key informant interviews. Community leaders, Health Develoment Armies (HDAs), and health extension workers (HEWs) from the community were involved. Facility managers, health information technicians (HIT) and case team leader were selected from health facilities. Managers, plan and program experts, and one selected expert were purposively selected from District Health Offices, Zonal Health Departments, and Regional Health Bureaus. Questionnaires were adapted from the Performance of Routine Information System Management (PRISM) conceptual farmework [4]. Document review of monthly, quarterly, and annual reports and supervision feedback were also assesed using checklists. A thematic content analysis was undertaken.

Results

Characteristics of study participants

This study engaged 21 key informants: 7 from the community, 6 from Health Centers, 6 from Woredas (districts), 1 from a Zonal Health Department, and 1 from a Regional Health Bureau. The majority of the participants were males 13 (62%). Eleven (52%) of them were aged below 34 years.

Data use culture

Community level

All community-level decision makers believed that information use has a positive role in supporting data quality, though the practice was not routine. A 27-year-old HEW mentioned that:

> *"The culture of information use is very poor at all levels. The more we use data, the better we know its function. Then we take care of quality. The more we ignore data use, we can't understand its benefit then we don't care about its quality."*

Community leaders reported that community engagement in local decision-making was limited only to planning and performance reviewing.

Health facility level

This study found that Performance Monitoring Teams (PMT) were in place at all health facilities, but only 25% of them were functional. Among all of the respondents at different levels, 11 (52.4%) strongly agreed and 7 (33.3%) somewhat agreed that decisions were made based on evidence. In health facilities where PMTs were functional, participants reported unfocused meetings. One Health Management Information System (HMIS) focal person mentioned that:

> *"….Yes, we want to have the meetings regularly, but we don't have clear guideline and expected results. People talk many unrelated and non-useful things in the meetings …"*

District and higher levels

At the decision-maker's level, all respondents reported that data use can ensure shared accountability. A 29 female respondent noted that:

> *"…Yes, data use improves accountability, it's only the presence of data that somebody to be accounted for both failure/success."*

Overall, respondents agreed that increased data use can lead to better data quality and better health service delivery at the lower levels and better health coverage at the decision-maker level. One District manager said:

> *"If we are just pipeline for the data to pass through us without use, we will not improve data quality and our decisions will not be based on evidence."*

Barriers to effective data use culture

Table 1 shows identified barriers to local-level data use culture.

Table 1 - Major Barriers to local-level healthcare data use

No.	Barriers	Health system level
1	Awarness gaps	All levels
2	High staff turn-over	Health facilities
3	Lack of motivation	Health facilities
3	Poor community engagement	Community
5	Inconsistent supervisions	All levels
6	Poor accountability system	All levels

Discussion

The value of health data is determined by its utilization in decision-making. Quality data provide accurate and timely information to manage services and aid in prioritizing and ensuring the best use of resources. This study shows that using data for decision-making can improve data quality. Evidence indicates that increased use of data will help to improve its quality, which will in turn lead to more data use [1; 4].

The national strategy established PMTs to conduct monthly meetings for performance review, problem analysis, and solution planning [3; 6; 10]. While PMTs were established at all health facilities in this study, only 25% of them ran PMT functions. Similar studies also report weak data use [3; 6; 10].

Fabricated reports and reporting data to the next level with minimal or no usage were the most common challenges [8, 10, 11]. All respondents agreed that the awareness gap experienced by health professionals was a significant obstacle to using data for decision-making. One female nurse respondent said:

> *"We have awareness gaps and there is no motivations… Health workers are overloaded and negligent to data use."*

High turnover of staff is the largest challenge that regional health bureaus face in terms of data management and information use for decision-making. One mentoring and evaluation team leader mentioned that:

> *"...For example, last year we trained all focal persons [HMIS focal] and when we did quick assessment at the end of the year, around 62% have left from their work. When untrained health workers were hired, they face difficulties to manage the data properly."*

Shared accountability leads to better use of health data for decision-maiking and actions to improve the coverage and quality of health services [9]. All respondents reported that there is no accountability system or written documents for data quality and data use.

Conclusions

In this study, the culture of local data use for decision-making at the local level was low. Awareness gaps, lack of motivation, inconsistent supervision, poor community engagement, and the lack of accountability were major bottlenecks. Establishing accountability systems and capacity building on health data use could improve its implementation.

Acknowledgements

The authors would like to thank the WHO and the University of Gondar for financial and material support for this study.

References

[1] T. Nutley and H.W. Reynolds, Improving the use of health data for health system strengthening, *Global health action* **6** (2013), 20001-20001. doi: 10.3402/gha.v6i0.20001

[2] C. AbouZahr, T. Boerma, Health information systems: the foundations of public health, *Bull World Health Organ* **83** (2005), 578-583.

[3] K. Foreit, S. Moreland, A. LaFond, Data Demand and Information Use in the Health Sector: Conceptual Framework. Chapel Hill, NC: MEASURE Evaluation, Carolina Population Center (2006), 1-17.

[4] A. Aqil, T. Lippeveld, and D. Hozumi, PRISM framework: a paradigm shift for designing, strengthening and evaluating routine health information systems, *Health policy and planning* **24** (2009), 217-228. doi: 10.1093/heapol/czp010

[5] A. Aqil, D. Hotchkiss, T. Lippeveld, E. Mukooyo, S. Asiimwe, Do the PRISM framework tools produce consistent and valid results? A Uganda study. Working Paper. National Information Resource Center, Ministry of Health, MEASURE Evaluation, Uganda, (2008).

[6] FMOH, Health sector transformation plan, (2015).

[7] S. Mengiste, Analysing the Challenges of IS implementation in public health institutions of a developing country, (2010).

[8] G. Mesfin, H. Mohamed, H. Tesfaye, N. Worku, D. Mamo, N. T. Fatoumata, Data quality and information use: A systematic review to improve evidence in Ethiopia, *African Health Monitor* **14** (2012), 53-60.

[9] M. Adamki, D. Asamoah, K. Riverson, Assessment of Data Quality on Expanded Programme on Immunization in Ghana: The Case of New Juaben Municipality, *J Health Med Informat* **6** (2015), 196. doi: 10.4172/2157-7420.1000196

[10] FMOH, Information revolution road map, (2016).

Address for correspondence

Kassahun Dessie Gashu

Email: kassahundessie@gmail.com; Tele: +251912071344

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190583

Variation in National Clinical Audit Data Capture: Is Using Routine Data the Answer?

Dawn W. Dowding[a], Natasha Alvarado[b], Lynn McVey[b], Mamas Mamas[c], Rebecca Randell[b]

[a] *Division of Nursing, Midwifery and Social Work, University of Manchester, Manchester, UK,*
[b] *School of Healthcare, University of Leeds, Leeds, UK*
[c] *Institute of Primary Care & Health Sciences, Keele University, UK*

Abstract

National Clinical Audit (NCA) data are collected from all National Health Service providers in the UK, to measure the quality of care and stimulate quality improvement initatives. As part of a larger study we explored how NHS providers currently collect NCA data and the resources involved. Study results highlight a dependence on manual data entry and use of professional resources, which could be improved by exploring how routine clinical data could be captured more effectively.

Keywords:
Clinical Audit; Quality Improvement; Information Systems

Introduction

In the UK, there are over 100 National Clinical Audits (NCAs) that are either centrally developed and managed through the National Clinical Audit and Patient Outcomes Programme (NCAPOP) or by independent professional organisations. The aim of the NCAs is to provide data for measuring the quality of care provided by National Health Service (NHS) providers, as well as stimulating quality improvement (QI) initiatives [1]. There is evidence that NCAs have led to improvements in the quality of patient care [2]. However there have also been reports of variation in how NHS Trusts engage with the NCA data, with reports of a lack of resources and variations in data quality impacting on their value as feedback on performance [3].

We are currently undertaking a study to develop and evaluate QualDash, an interactive web-based quality dashboard that supports clinical teams, quality subcommittees, and NHS Trust Boards to understand and make use of NCA data. This poster reports a subset of the findings from Phase 1 of the study, focusing on how NCA data are currently collected across NHS Trusts, and the resources involved. It then reflects on the implications of these findings, in terms of the utility of using electronic health record data for capturing NCA and other audit data.

Methods

The study focuses on two NCAs; the Myocardial Infarction National Audit Project (MINAP) and the Paediatric Intensive Care Audit Network (PICANet), with evaluation of cardiology departments and Paediatric Intensive Care Units (PICUs). MINAP provides data on the management of ST-elevation myocardial infarction (STEMI) and non-ST segement elevation myocardial infarction (NSTEMI), compared to national and international standards to participating hospitals and ambulance services. PICANet records admission details and treatment provided to all critically ill children in PICUs.

A total of 5 NHS Trusts (providers) participated in the first phase of the study; all 5 take part in MINAP and 3 of the 5 have PICUs and also take part in PICANet. The number of staff at each site varied from 20,000 (site 4) to 8,000 (sites 2 and 3), and patients seen per year from 700,000 (site 3) to 1.5 million (sites 1 and 4). Semi-structured interviews were conducted with 54 individuals across the NHS Trusts and the wider NHS organisational regional structure, including NCA leads, members of NHS Trust quality and safety committees, Trust Boards, and clinical commissioning groups (CCGs). CCGs are the statutory NHS bodies responsible for planning and commissioning health care services in localities in England.

Interviews focused on participants' role and their experience of, and involvement with, NCAs. We also explored what data were collected for each NCA, how data are captured and how they are used in the NHS Trust and by whom. All interviews were audio-recorded, transcribed and entered into a qualitative software program (*NVivo 10*) for indexing. Interview data were analysed using framework analysis. After familiarisation with the data, a thematic framework was developed to index the data before interpretation of key themes.

Results

Interview data were collected between November 2017 and June 2018. Participants, 30 of whom were female and 24 male, worked in both clinical and non-clinical roles. Twenty-eight of the participants were clinicians (14 doctors/surgeons and 14 nurses), 22 had non-clinical managerial or support roles (including senior managers who were members of Trust Boards and quality and safety committees; some of these participants had clinical backgrounds and some did not), and 4 participants worked within CCGs. An overview of the key results are summarised in Table 1.

Resources used to support NCA data collection, entry, submission, validation

There was considerable variation in the resources used by each site to support the various NCAs, both within NHS Trusts and across Trusts. For **MINAP**, 3 of the sites had designated

clinicians who were responsible for data collection, entry, submission, and validation. This varied from having a full-time nurse with overall responsibility for collection and evaluation of data (Site 1) to having nurses collecting data alongside their clinical responsibilities (Sites 2 and 3). Site 4 had mixed usage of non-clinical and clinical resources, with a non-clinical primary PCI project assistant collecting data for STEMI patients and 2 acute chest pain specialists collecting and entering data for NSTEMI patients. Site 5 used non-clinical personnel to collect and enter MINAP data, employing a non-clinical cardiology information analyst, assisted by another team member.

The picture for **PICANet** across the 3 NHS Trusts that participated in this NCA was different. In all three of the sites clinicians initially fill out the PICANet forms, with non-clinical staff (database managers, audit-coordinators and audit clerk) then collating and checking the information, identifying missing and inaccurate data.

Table 1 – How NCA data are captured

	MINAP	**PICANet**
Resources		
Clinical	Full-time nurse (site 1) Nurse + clinical resp (sites 2, 3) Specialist nurse (site 4)	Clinicians fill out the form (sites 1, 4, 5)
Non-Clinical	Project assistant (site 4) Information analyst (site 5)	Database manager, audit coordinator/clerk – (sites 1, 4, 5)
Systems		
Database	Sites 1, 2, 4, 5	Sites 1, 4
Excel	Site 1	Sites 1, 4
Directly into portal	Site 3	Site 5

Systems for data collection, and entry

Similarly, there was considerable variation in the systems' NHS Trusts used to collect and enter the data into the NCA web portals. For MINAP, all of the sites apart from Site 3, had some form of in-house database that was also used to record data (Site 1 also used Microsoft Excel spreadsheets) before uploading to the web portal. Site 3 was the only site to enter data directly into the web portal. All sites relied on manual data entry by an individual into the database/web portal.

For PICANet two sites used in-house databases/spreadsheets for data collection that were then uploaded to the web portal. At site 5, data were copied and pasted from the Trust's patient data management system directly into the PICANet online portal retrospectively after a patient's discharge.

Use of NCA data

The way in which NCA data were captured by NHS Trusts affected how they were used. Issues raised included the accuracy of the data (with some Trusts highlighting concerns about data reliability) and timeliness. The way NCA data were both collected by Trusts (sometimes retrospectively) and then reported by suppliers (often yearly) meant that data were considered by some Trusts to be out-of-date and not useful to inform practice.

Conclusions

There is variation in how NHS Trusts capture NCA data, with organisations dedicating expensive resources (such as highly trained professionals, like nurses) to ensure data are collected, uploaded, and checked for reliability. All of the NHS Trusts in our study collected data manually, with some automating the uploading process to web portals. With the growing implementation of electronic health records in the NHS, emphasis should be given to how to more effectively use routine data for data capture. Considering the burden of NCA data collection and processing in the NHS (over 100 NCAs currently), this could free valuable professional resource for care elsewhere in the healthcare system.

Acknowledgements

This research is funded by the National Institute for Health Research (NIHR) Health Services and Delivery Research (HS&DR) Programme (project number 16/04/06). The views and opinions expressed are those of the authors and do not necessarily reflect those of the HS&DR Programme, NIHR, NHS or the Department of Health.

References

[1] K. J. Phekoo, J. Clements, and D. B. Bell, "National Clinical Audit Quality Assessment - Overview of the self-assessment survey: 'audit of audits,'" Healthcare Quality Improvement Partnership, London, UK, 2014.

[2] J. S. Birkhead, L. Walker, M. Pearson, C. Weston, A. D. Cunningham, and A. F. Rickards, "Improving care for patients with acute coronary syndromes: initial results from the National Audit of Myocardial Infarction Project (MINAP)," *Heart*, **90** (2004), 1004–1009.

[3] A. Taylor, J. Neuburger, K. Walker, D. Cromwell, and O. Groene, "How is feedback from national clinical audits used? Views from English National Health Service trust audit leads," *J. Health Serv. Res. Policy*, **21** (2016), 91–100.

Address for correspondence

Dawn W Dowding, Professor of Clincial Decision Making, Division of Nursing, Midwifery and Social Work, Univeristy of Manchester, Oxford Road, Manchester M13 9PL, UK.

Email: dawn.dowding@manchester.ac.uk

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190584

A Tale of Two Databases: The DoD and VA Infrastructure for Clinical Intelligence (DaVINCI)

Scott L. DuVall[a], Michael E. Matheny[b], Ildar R. Ibragimov[c], Trey D. Oats[c], Jay N. Tucker[a], Brett R. South[a], Augie Turano, Hamid Saoudian[a], Casey Kangas[d], Keith Hofmann[d], Wendy Funk[d], Chris Nichols[c], Albert Bonnema[c], Louis Ferrucci, Jonathan R. Nebeker[a]

[a]VA Informatics and Computing Infrastructure, VA Salt Lake City Health Care System and University of Utah, UT, USA,
[b]VA Informatics and Computing Infrastructure, VA Tennessee Valley Healthcare System and Vanderbilt University, TN, USA,
[c]Defense Health Agency, West Falls Church, VA, USA,
[d]Kennell and Associates, Falls Church, VA, USA

Abstract

The Department of Defense (DoD) and Department of Veterans Affairs (VA) Infrastructure for Clinical Intelligence (DaVINCI) creates an electronic network between the two United States federal agencies that provides a consolidated view of electronic medical record data for both service members and Veterans. This inter-agency collaboration has created new opportunities for supporting transitions in clinical care, reporting to Congress, and longitudinal research.

Keywords:

Electronic Health Records, Health Information Exchange, Government Agencies

Introduction

The DoD and VA Infrastructure for Clinical Intelligence (DaVINCI) was sponsored as a Joint Incentive Funded (JIF) inter-agency project. It built on existing data and governance processes to reduce cost and most immediately provide return on investment. On the VA side, the Veterans Health Administration Health Services Research and Development VA Informatics and Computing Infrastructure (VINCI) formed the core infrastructure. On the DoD side, the Defense Health Agency Information Delivery Division Health Services Data Warehouse (HSDW) and the Military Health System Data Repository (MDR) formed the core infrastructure. DaVINCI provided the first DoD-VA full population data integration and brought together medical records, claims, and benefit information from both agencies. Combining records from multiple data sources presented several challenges, but the potential benefits for properly credenetialed teams to access and seamlessly utilize relevant information from multiple source systems in both agencies provided motivation for DaVINCI.

Methods

The goal of DaVINCI is to increase DoD-VA data integration for interagency collaboration and resource sharing in support of the Joint Strategic Plan and each agency's healthcare and benefits missions [1]. DaVINCI creates a cornerstone of informatics and data sharing that support health analysis, operations, and other secondary uses of health care data. The special relationship between DoD and VA (that is, all those eligble for care in VA were first armed forces service women or men in the DoD) means that the data sharing and overlapping populations are distinct from most other overlapping healthcare systems [2, 3]. Three distinct populations are included in DaVINCI. The DoD Only population contains persons who have electronic medical records from DoD, but not from VA. This includes active duty, guard, reserve personnel, and those who have completed service but who have not yet received care in a VA facility. The VA Only population consists of those who have electronic medical records from VA, but served before DoD used computerized record systems. The DoD and VA population contains electronic medical records from both agencies.

Instead of building interfaces with each source system, DaVINCI built on the many years of work aggregating and cleaning source systems into data warehouses (VINCI, HSDW, and MDR). Crosswalk lists are generated from the respective data warehouses and then combined and compared. This process removes employees, service member dependents, and others that are outside the DaVINCI populations. Following best practices, the data from both systems were transformed into a common data model [4] called Observational Medical Outcomes Partnership (OMOP) [5]. This allows VA and DoD data to be joined seamlessly in a longitudinal record regardless of which source agency or source system the data originates.

Once the crosswalk is complete, data files from each agency are prepared and transmitted securely. The governance process transfers stewardship of the data as well, meaning that VA medical record data becomes part of the service member's DOD medical record and vice-versa. In this way, the service member or Veteran electronic medical record follows them in subsequent care episodes.

Results

To date, DaVINCI contains more than 33 billion rows of electronic medial record data including VA data for 22,763,277 Veterans and DoD data for 6,947,102 service members. Of these, 4,039,155 (17.8% of Veterans and 58.1% of service members) have medical records in both VA and DoD.

The magnitude of current DoD and VA databases is considerable. These data are diverse and include not only healthcare untilization including inpatient, outpatient, pharmacy, immunizations, vitals, but also data on eligiblity, service theater, and demographic information. The push for precision medicine and the future influx of genomics data could be included in future data transfers to make meaningful conclusions to improve outcomes.

DaVINCI has been used to plan and evaluate mental health care utilization after military service has ended, report to Congress, reply to inquiries from top VA and DoD leadership, and measure health outcomes and utilization during transition from military service.

DaVINCI deploys and evaluates solutions for governance, data integration, data security, analysis, and supercomputing to provide a solid foundation for integrated clinical and business intelligence platforms. By allowing a more complete read-only view into a service member or Veteran record, DaVINCI can facilitate improved coordination of polytrauma patients and accelerated disability evaluations. By allowing more rapid analysis of many service member and Veteran longitudinal records, this program can facilitate more accurate and efficient screening for syndromes such as traumatic brain injury and planning for joint activities such as inter-agency resource balancing of specialty-care.

Over the next decade, all VA and DoD facilities will transition to a joint-hosted instance of a commercial electronic medical record system. This move will drastically increase the interoperability of care and data sharing between VA and DoD. It is also likely that in the future, a shared data warehouse containing all data from VA and DoD would be created. In this scenario, not only is the current DaVINCI project an effective bridge providing the only combined-population analytics solution while the new infrastructure is fully implemented, but has the potential to speed the transition process. The work of standardizing and combining VA and DoD data together can mean that initial loads of a combined data warehouse can take advantage of the linking and mapping that is already done. This may reduce the need to extract from all source systems and even allow both institutions data sets to be loaded together.

The dynamic computing environment created through this collaboration not only supports the magnitude of data needed to be stored but also utilizes parallel computing to support quick responses for queries. The end result is a solution which can handle the current data requirements, can easily expand with increasing data, and supports quicker query requirements.

Conclusion

The DaVINCI project has already proven useful in determining which resources are needed on the VA side for separating service members and coordinating test results already performed on the DoD side to reduce duplication. It has been used to determine immunization status and overlap in utilization in areas where DoD and VA share medical facilities. Finally, it has supported several research projects studying conditions ranging from hearing loss to traumatic brain injury and heart disease to prostate cancer.

The opportunities that DaVINCI creates for caring for and studying the overlapping patient populations of VA and DoD will continue to benefit the health of active duty personnel and Veterans.

Acknowledgements

Funding support provided by Department of Defense and Department of Veterans Affairs Joint Incentive Fund JIF13157-DAVINCI. This work was supported using resources and facilities at the VA Salt Lake City Health Care System with funding from VA Informatics and Computing Infrastructure (VINCI), VA HSR RES 13-457.

References

[1] S.L. DuVall, M.E. Matheny, I.R. Ibragimov, T.D. Oats, et al. The Department of Defense (DoD) and Department of Veterans Affairs (VA) Infrastructure for Clinical Intelligence (DaVINCI). *Proc American Medical Informatics Association* (2018).

[2] E.Vayena, A. Blasimme. Health Research with Big Data: Time for Systemic Oversight. *The Journal of Law, Medicine & Ethics* **46**(1) (2018), 119–129. https://doi.org/10.1177/1073110518766026

[3] J. Polonetsky, O. Tene, J. Jerome. Benefit-Risk Analysis for Big Data Projects. Future Privacy Forum. Washington, DC (2014), 1–13.

[4] S.T. Rosenbloom, R.J. Carroll, J.L. Warner, M.E. Matheny, J.C. Denny. Representing Knowledge Consistently Across Health Systems. *Yearb Med Inform* **1** (2017), 139-147, Aug26.

[5] https://www.ohdsi.org/data-standardization/

[6] E. Brignone, J.D. Fargo, R.K. Blais, A.V. Gundlapalli. Chronic Health Conditions Among US Veterans Discharged from Military Service for Misconduct. *Prev Chronic Dis*. **15** (2018), Oct 11, E122. doi: 10.5888/pcd15.180238.

[7] A.F. Mohanty, L.M. McAndrew, D. Helmer, M.H. Samore, A.V. Gundlapalli. Chronic Multisymptom Illness Among Iraq/Afghanistan-Deployed US Veterans and Their Healthcare Utilization Within the Veterans Health Administration. *J Gen Intern Med*. **33**(9) (2018), 1419-1422. doi: 10.1007/s11606-018-4479-6.

Address for correspondence

Scott L. DuVall, PhD, Scott.Duvall@va.gov

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190585

Analysis for the Annual Text Amount of Electronic Medical Records

Kenichiro Fujita[a, b], Osamu Sugiyama[a], Shusuke Hiragi[a], Kazuya Okamoto[a], Tadamasa Takemura[b], Tomohiro Kuroda[a]

[a] *Division of Medical Informatics and Administration Planning, Kyoto University Hospital, Kyoto, Japan,*
[b] *Graduate School of Applied Informatics, University of Hyogo, Kobe, Japan*

Abstract

The amount of text of electronic medical records and its changes over time are not clear. In designing an electronic medical records system, prediction of the amount of text is important. We analyzed the number of characters described in the electronic medical records. As a result, it became clear that the annual text quantity of electronic medical records follows the lognormal distribution, and also the amount has been increasing year by year.

Keywords:
Electronic Medical Records, Prediction Model, Medical Records.

Introduction

In addition to digitization of clinical records centered on electronic medical records (EMR), expansion from information system for each medical institution to electronic health records (EHR) is proceeding. As a result, very large amounts of data are accumulated in EMR and EHR. In considering system design and UX, the amount of data stored in the system is a very important factor. If it is possible to predict the amount of text of EMR, it can contribute to system design and UX.

There is a prior study that the page number of the paper chart approximates the lognormal distribution [1]. There is also a report that the amount of information for each outpatient visit increases due to electronic medical record [2]. However, it has not been clarified about the amount of text of the entire EMR including outpatient and hospitalization and the change over time of EMR.

In this research, for the purpose of constructing a prediction model of the amount of text of the electronic medical record, we analyze the amount of text of the electronic medical record.

Methods

We need to consider the method of measuring the amount of text stated in EMR. If EMR over multiple years are simply aggregated on a patient basis, changes over time can not be analyzed. In addition, it is inappropriate to compare them each other because the start point of the medical treatment is different between patients.If simply counting on a unit of time (year, month, etc.), although temporal changes can be considered, it becomes impossible to identify patients included therein. It is necessary to examine the correlation between patient attributes and the amount of EMR described in order to construct a prediction model, which is the purpose of this study. Therefore, it is not appropriate to simply aggregate only in units of time. From the above discussion, we take both the patient and the time unit as parameters, and we counted characters of EMR for each patient per year. In this study, Each separated EMR is named as “patient annual EMR”, and also numbers of characters written in a patient annual EMR is named as "patient annual EMR text amount".

In this study, we will use EMR data of Kyoto University Hospital written from 2006 to 2016 as a data set. It contains no image data but text data by medical staff. Character counting was performed using SQL on IBM DB2 10.5. Statistical analysis was conducted using R 3.4.2.

Results

Characteristics of All Periods

The data set has 883,312 patient annual EMRs. Table 1 summarizes the patient annual EMR text amounts over the target period. Figure 2 shows the histogram of the patient annual EMR text amounts. The difference between the mean and median and the histgram show the patient annual EMR text amounts not following normal distribution.

Table 1– The Patient Annual EMR Text Amounts

	Total	Outpatient	Inpatient
Mean	25,647	10,401	88,940
Median	4,285	3,833	28,342
Maximum	8,209,843	839,453	8,203,040

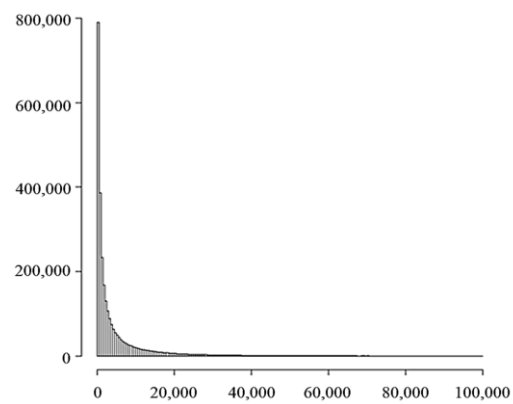


Figure 1 – Histgram of the Patient Annual EMR Text Amounts.

Table 2 summarizes the results of the logarithms of the patient annual EMR text amounts over the target period. Figure 3 shows the histogram of the logarithms of the patient annual EMR text amounts. Results of Kolmogorov–Smirnov test and Shapiro-Wilk test and the histogram show that the patient annual EMR text amounts follows the lognormal distribution.

Table2- Logarithm of the Patient Annual EMR Text Amounts

	Total	Outpatient	Inpatient
Mean	8.412	8.167	10.227
Median	8.363	8.251	10.252
p-value of Kolmogorov–Smirnov test	0.772	0.2209	0.8638
p-value of Shapiro-Wilk test	0.4889	0.6805	0.477

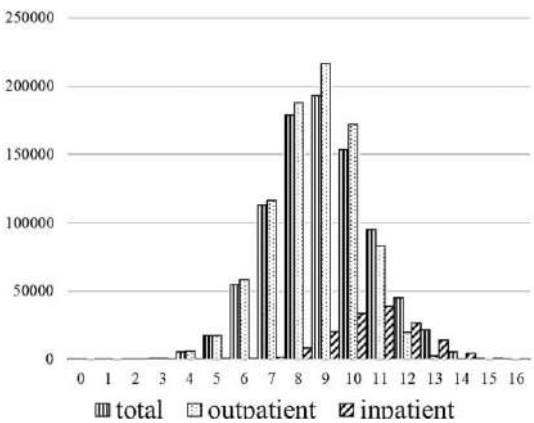


Figure 2 –Histgram of Logarithm of the Patient Annual EMR Text Amounts.

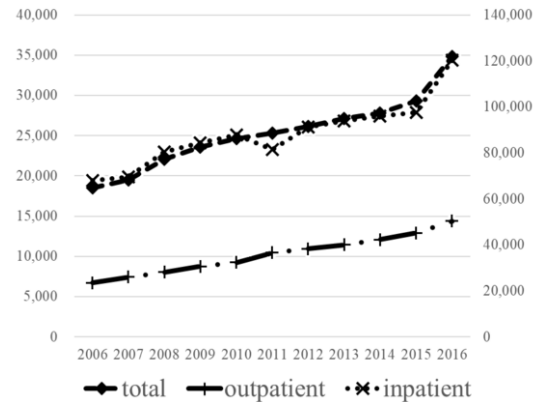


Figure 3 –Mean Trend of the Patient Annual EMR Text Amounts.

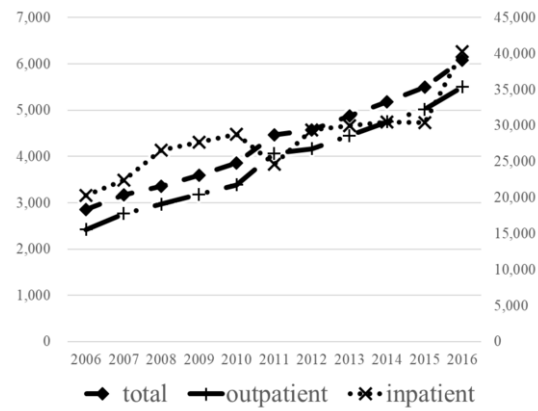


Figure 4 – Median Trend of the Patient Annual EMR Text Amounts.

Chronological Change

Figures 3 and 4 show the trends in the mean and median of the patient annual EMR text amounts for each year in the target period. For total, outpatients, and inpatients, the patient annual EMR text amounts continued to increase in both the mean and the median.

Conclusions

In this research, we analyzed the characteristics of the patient annual EMR text amounts in the target period and its change over time. As a result of the analysis, it became clear that the patient annual EMR text amounts follows the lognormal distribution, and the amount has been increasing year by year. The data set of this study is EMR of one hospital, and it is necessary to investigate EMR of multiple hospitals for validating the above conclusion.

In future research, we will investigate factors that increase the amount of EMR text and search for explanatory variables of the prediction model.

References

[1] N. Ouyang, M. Ikeda, Z.X. Chui, and K. Yamauchi, An analysis of medical record data in an optical disk filing system, *Medical Record Administration* **8** (1996), 41-43.

[2] E. Aramaki, S. Masukawa, E. Yamada, K. Waki, and K. Ohe, Introduction of electronic health record system increases quantity of chart, *Japan Journal of Medical Informatics* **30 Suppl**. (2010), 1046-1049.

Address for Correspondence

Kenichiro Fujita, 54 Kawaharacho, Shogoin, Sakyo-ku, Kyoto, 606-8507, Japan. Email: kfujita@kuhp.kyoto-u.ac.jp

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190586

Measuring the Intention of Using Augmented Reality Technology in the Health Domain

Parisis Gallos[a], Joseph Liaskos[a], Charalabos Georgiadis[b], Enkeleint Aggelos Mechili[c], John Mantas[a]

[a]Health Informatics Laboratory, School of Health Sciences, National and Kapodistrian University of Athens, Athens, Attica, Greece,
[b] School of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece,
[c] Department of Healthcare, Faculty of Public Health, University of Vlora, Albania

Abstract

Augmented Reality technology can provide useful tools and devices to support healthcare services. The aim of this study is to investigate the intention of IT and health care scientists' to use Augmented Reality technology in Healthcare. A survey was conducted using a questionnaire based on a theoretical research model. According to the results, the participants seem to have positive perception about using the Augmented Reality technology in health domain, and they intend to use it.

Keywords:
Technology Assessment, Delivery of Health Care, Augmented Reality

Introduction

Augmented reality (AR) can be defined as an interactive visualization system (e.g. a head-mounted display, a computer, a game console, a smartphone, a tablet) allowing the merging of digital content with the real environment sourrounding the user, and blending both "real-world" elements with "virtual" ones [1]. The evolution in the area of mobile and wearable devices, has led to a rapid growth of AR technologies. AR can provide useful tools combining innovative devices and software, which can contribute positively to the domain of healthcare [1-3].

In this context, much scientific work is taking place on the assessment of the AR technology [4,5]. The acceptance of this technology is crucial and promising [6-8], especially in the health domain [9].

The aim of this ongoing study is to investigate the intention of information technology (IT) and healthcare scientists' to use AR technology in healthcare.

Methods

For the aim of our study, a questionnaire was developed based on a theoretical research model. The proposed model was a combination of Davis' Technology Acceptance Model (TAM) [10] and other related research work [11,12]. Our model included six dimensions with at least two questions (items) per dimension. All items used a 7-item Likert scale from "strongly disagree" to "strongly agree". The questionnaire was anonymous and distributed through Google forms to Greek IT and healthcare scientists during 2018. Specific relationships between the dimensions were tested, based on the following hypotheses (Figure 1):

- H1: Does "Perceived Usefulness" positively affect "Attitude Toward Use".
- H2: Does "Perceived Ease of Use" positively affect "Attitude Toward Use".
- H3: Does "Attitude Toward Use" positively affect "Behavioral Intention to Use".
- H4: Does "Compatibility" positively affect "Behavioral Intention to Use".
- H5: Does "Relative Advantage" positively affect "Behavioral Intention to Use".

The sample consisted of 81 IT and healthcare professionals. Descriptive statistics and structural equation modeling data analysis was performed using SmartPLS 2.0 M3 to conduct partial least squares path modeling [13].

Results

The participants of this study were 34.2% males and 65.8% females. The average age of the participants was 27.6 years old. According to the descriptive statistics, the mean and median value for all items values were above 4, which reveals the positive attitude of the participants (Table 1).

Table 1– Survey Results

Items	Mean	St. Dev	Median	Individual Item Loadings
ATT1	6.07	1.00	6	0.88
ATT2	5.68	1.20	6	0.82
BI1	5.95	0.93	6	0.76
BI2	5.67	1.10	6	0.86
BI3	5.40	1.29	6	0.75
COMP1	5.98	0.98	6	0.84
COMP2	5.79	0.98	6	0.70
PEU1	5.53	1.24	6	0.93
PEU2	4.48	1.53	4	0.70
PU1	6.06	1.00	6	0.86
PU2	5.85	1.07	6	0.89
RA1	5.57	1.09	6	0.73
RA2	5.94	1.03	6	0.79
RA3	5.83	1.01	6	0.81

Based on the partial least squares analysis, the measurement and the structural models were examined. To test the validity and the reliability of the models, individual item loadings (for indicator reliability), internal consistency, convergent and discriminant validity, were tested and produced reliable results [14].

The coefficient of determination (R2) was 0.612 for the "Behavioral Intention to Use" and 0.456 for "Attitude toward Using". All relationships between the dimensions (as described

in the hypotheses, Figure 1) were found to be statistically significant. The structural model was validated by applying a bootstrapping technique (with 5000 resamples) based on a two-tailed t-test (with α=0.05 or 0.01). All the hypotheses were confirmed (Figure 1), with at least 95% significance (p-values<α).

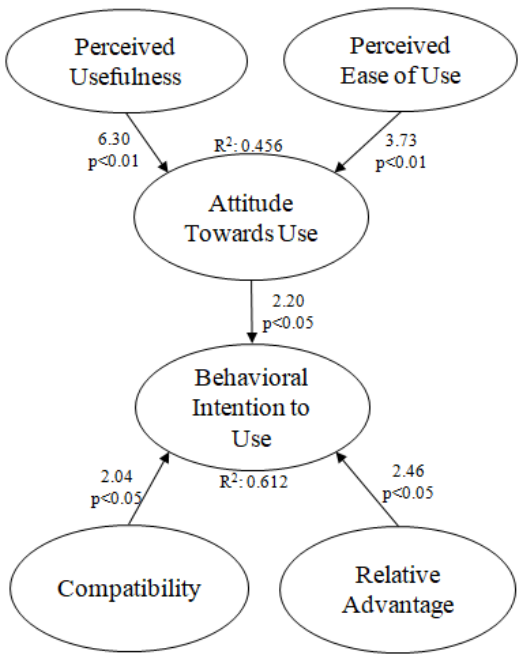


Figure 1– Theoretical research model

Discussion

According to the study results, the participants seem to have positive perception about the implementation of the AR Technology in health domain, and they intend to use it. Perceived Usefulness and Perceived Ease of Use positively affect the participants' Attitude towards Using AR technology in Healthcare. Additionally, Relative Advantage and Compatibility, as well as Attitude towards Using positively affect the Behavioral Intention to Use AR technology. These findings present the positive attitude of the paerticipants towards AR technology, their positive view as they consider it as useful and usable technology, and their intention to use it in their professional healthcare enviroment.

Conclusions

The survey results indicate that the participants have a positive attitude towards the Augmented Reality Technology in health domain and they express their intention to use it. Limitations of this survey may be the young age of the participants, as younger professionals might be more familiar with new technologies compared with older professionals. Future work may include further investigation of behaviors and attitudes of IT and health care professionals towards virtual and augmented reality technologies being used in healthcare.

Acknowledgements

Authors would like to thank the participants, who formed the population of the survey.

References

[1] G. Riva, R.M. Baños, C. Botella, F. Mantovani, and A. Gaggioli, Transforming experience: The potential of augmented reality and virtual reality for enhancing personal and clinical change, *Front Psychiatry* **7**(SEP) (2016).

[2] J.W. Yoon, R.E. Chen, E.J. Kim, O.O. Akinduro, P. Kerezoudis, P.K. Han et al., Augmented reality for the surgeon: Systematic review, *The International Journal of Medical Robotics and Computer Assisted Surgery* **14** (2018), e1914.

[3] P. Gallos, C. Georgiadis, J. Liaskos, and J. Mantas, Augmented Reality Glasses and Head-Mounted Display Devices in Healthcare, *Stud Health Technol Inform* **251** (2018), 82-85.

[4] N. Bourdel, T. Collins, D. Pizarro, A. Bartoli, D. Da Ines, B. Perreira, and M. Canis et al., Augmented reality in gynecologic surgery: evaluation of potential benefits for myomectomy in an experimental uterine model, *Surgical endoscopy* **31**(1) (2017), 456-461.

[5] C. Moro, Z. Štromberga, A. Raikos, and A. Stirling, The effectiveness of virtual and augmented reality in health sciences and medical anatomy, *Anatomical Sciences Education* **10**(6) (2017), 549-559.

[6] M. Tom Dieck et al., A theoretical model of mobile augmented reality acceptance in urban heritage tourism, *Current Issues in Tourism* **21**(2) (2018), 154-174.

[7] A. Baktash et al., Adoption of Augmented Reality Technologies in Tourism: Visitors' Acceptance of Smart Glasses: An Abstract, *Academy of Marketing Science Annual Conference*, Springer, Cham, 2017.

[8] F. Bonetti, G. Warnaby, and L. Quinn, Augmented reality and virtual reality in physical and online retailing: a review, synthesis and research agenda, *Augmented Reality and Virtual Reality,* Springer, Cham, 2018, 119-132.

[9] C. Mather et al., Helping Hands: Using Augmented Reality to Provide Remote Guidance to Health Professionals, *Stud Health Technol Inform* **241** (2017), 57-62.

[10] F.D. Davis, R.P. Bagozzi and P.R. Warshaw, User Acceptance of Computer Technology: A Comparison of Two Theoretical Models, *Management Science* **35** (1989), 982-1003.

[11] P. Gallos, J. Minou, F. Routsis, and J. Mantas, Investigating the Perceived Innovation of the Big Data Technology in Healthcare, *Stud Health Technol Inform* **238** (2017), 151-153.

[12] P. Gallos and J. Mantas, The Intention of Big Data Technology Usage in Healthcare, *MIE 2018, Gothenburg, Sweden* (2018).

[13] C.M. Ringle, S. Wende and A. Will, SmartPLS 2.0 (M3), Available from: http://www.smartpls.de, Hamburg, 2005.

[14] K.K.-K. Wong, Partial Least Squares Structural Equation Modeling (PLS-SEM) Techniques Using SmartPLS, *Marketing Bulletin* **24** (2013), Technical Note 1.

Address for correspondence

Parisis Gallos, Papadiamantopoulou 123, Athens, 11527, Greece.
e-mail: parisgallos@yahoo.com, tel:+302107461448.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190587

The Ligurian HIV Network: How Medical Informatics Standards Can Help Clinical Research

Barbara Giannini[a], Sara Mora[a], Roberta Gazzarata[b], Antonio Di Biagio[c], Giovanni Cenderello[d], Chiara Dentone[e], Maurizio Setti[f], Daniela Fenoglio[f,g], Giovanni Cassola[d], Claudio Viscoli[h], Mauro Giacomini[a,b,g]

[a] *Department of Informatics, Bioengineering, Robotics and System Engineering, (DIBRIS), University of Genoa, Italy*
[b] *Healthropy, Savona, Italy*
[c] *Infectious Diseases Unit, IRCCS AOU S. Martino-IST, National Institute for Cancer Research, Genoa, Italy*
[d] *SC Malattie Infettive, Ospedali Galliera, Genova, Italy*
[e] *Infectious Diseases Unit, ASL-1 Imperiese, Sanremo, Imperia, Italy*
[f] *Department of Internal Medicine, University of Genoa; IRCCS AOU San Martino IST Genoa, Italy*
[g] *Centre of Excellence for Biomedical Research, (CEBR), University of Genoa, Italy*
[h] *Department of Infectious Disease, IRCCS AOU San Martino IST, (DISSAL), University of Genoa, Italy*

Abstract

Integrating evidence from systematic research in daily clinical practice is one of the pillars of evidence-based medicine. Electronic data capture tools simplify data collection from different centers and supports the management of multicenter clinical trials. The Ligurian HIV Network (LHN) is one such tool, originating from a regional effort to integrate clinical trial capabilities for HIV and other chronic infectious diseases. In order to manually collect a complete report of all clinical tests on patients enrolled in a trial, a strenuous human effort and the allocation of great resources would be necessary. Moreover, the risk of error in a manual system is very high. The proposed system automatically extracts clinical data from the EHR of three hospitals of the LHN in a standardized way, and enhance their re-use in clinical trials. Through dedicated questionnaires, physicians reported a strongly positive feedback about the efficacy of the platform in supporting clinical research.

Keywords:

Electronic Health Records, Health Level Seven, Clinical Trial

Introduction

Clinical Trials are indispensable tools for Evidence-Based Medicine [1]. To collect the large amounts of data needed for clinical research, the implementation of a system, which coordinates the conduction of scattered clinical trials, is required. The interchange of data coming from different and heterogenic medical structures has to be managed. The ARCA (Antiretroviral Resistance Cohort Analysis), ICONA (Italian COhort Naïve Antiretroviral) and CISAI (Italian Coordination Study for Allergies and Infections from HIV) databases constitute the most important Italian clinical cohorts for HIV studies.

In 2013, a regional research network called the Ligurian HIV Network (LHN) was implemented by some of the authors [2]. The LHN was originally conceived as a mere web platform to enable the collection of data from HIV patients using a web interface, in order to perform multi-centric clinical trials at a regional level [3]. For the first year, the platform was consistently used for its initial scope, supporting about half a dozen local studies. Routine use of the platform, together with the intention to address the well-known problem of errors due to manual data input into web interfaces, led to a consecutive gradual evolution of the system. The leading idea was that clinical data, already available in digital format in the LIS (Laboratory Information System) and other components of the Hospital Information System (HIS), could be extracted and automatically exported to the LHN database. In this way, patients' laboratory data could be updated daily without human intervention. Clinical data was chosen since laboratory data is the largest data set used in clinical research concerning infectious diseases. Once this first process was completed, having reached a fully updated database, the authors decided to automatically connect the repository to the national HIV clinical studied mentioned above.

Methods

The LHN technical implementation choices are guided by the Electronic Source Data Interchange (eSDI) group recommendations. The scenario adopted and the architectural approach of the developed solution are thoroughly described in [4], while the technical approach is similar to those described in [5][6]. In order to protect patients' privacy, only indispensable personal information is managed such as year of birth, sex and nationality. The patient hospital code, that cannot be directly connected to the identity of the patient, makes it possible to track each patient through various hospital departments. If a patient moves to another hospital, explicit medical coordination is required.

A tool for the direct extraction of laboratory information was implemented, so that the data can be conveyed from the Laboratory Information System (LIS) towards the LHN. Depending on the hospital's level of IT support, two different scenarios arise: 1) the hospital does not have a LIS in which digital data are stored or it has LIS but does not allow external agents to access data; 2) the hospital has a LIS and allows external agents to access data. Hospitals belonging to the first scenario could participate in the LHN by inserting data manually from the web interface, while those belonging to the second scenario could exploit the paradigm of data reuse. Automatic clinical data flow from a hospital LIS towards the LHN is allowed by a Windows Console Application that encapsulates clinical data in standard Health Level 7 Clinical Document Architecture (HL7 v3 CDA r2) documents and sends them through a web service that is responsible for document validation and data recording. During this process, the web service also involves a terminology service, in conformity with the standard CTS2 (Common Terminology Service Release2), which allows the translation of local hospital terms to LOINC (Logical Observation Identifiers Names and Codes)[7]. This ensures a

correct understanding of the data exchanged between different laboratories that use their own terminology. A trigger launches the console application program once upon a night, searching for registered patients who have done laboratory exams during the day. When the program encounters for the first time a patient just registered on the platform, all historical data are retrieved.

Results

The LHN currently involves the Infectious Diseases Departments of eight hospitals in Liguria and Piedmont. Three hospitals (IRCCS AOU San Martino IST, Galliera and Sanremo) use the second scenario while the others use the first.

The three hospitals in the second scenario have two different informative systems, so two different methods for sharing clinical data were developed. The Galliera Hospital offers several services to access patients' clinical data, including a data management service to access laboratory test results. After authentication, data are available on demand in XML or JSON proprietary format. On the other hand, IRCCS AOU San Martino IST and Sanremo hospitals do not share data through services, thus the only way to access data is to extract them directly from the LIS, installing the console application inside the hospital firewall. Currently 4338 patients from eight different hospitals are involved in the LHN. Among them, data from 1851 patients from San Martino hospital, 1241 from Galliera hospital and 267 from Sanremo hospital are automatically updated using the described tool. The EHRs of the Galliera and San Martino hospitals started working at operating speed in 2008. There is a consistent difference among hospitals belonging to the first and the second scenarios in terms of quantities registered. In the data for Sanremo hospital, this trend is not yet appreciable, since the automatic connection with the LIS was only set up recently. Automatized centers insert 10 times lab test per year more than non-automatized ones (10679 vs 1910) and also a greater number of parameters per test is considered (41 vs 31). Using this tool, anonymized data were automatically transferred from the Electronic Health Records (EHR) towards the Clinical Data Management System (CDMS). This enhanced clinical trials conduction on chronic viral infections, the project has recently been expanded to include HBV and HCV diseases. Presently nine regional studies are supported by the platform as they are listed in the left menu of the web site (https://reteligurehiv.it/Default.aspx), pressing each button a short description is available.

This fully updated database is a reliable source of information that could be reused to fill national databases, like ARCA, CISAI and ICONA. At present, 1243 LHN patients are registered in the ARCA database, 590 in CISAI and 458 in ICONA. A program has been developed to continuously update the ARCA database with CD4+, HIVRNA and genetic sequence data for all patients included in the LHN. The CISAI study also involves HCV infected subjects and collects a greater amount of data. A collaboration with the ICONA data manager has not yet been arranged. Temporarily, a tool for data extraction in Excel format, which is available on the platform, helps in collecting data to be sent to the ICONA database.

Statistical reports and charts are prepared from the data collected using tools available on the web interface, in order to show continuously updated information. An extraction tool has been set up on the web interface to enable physicians to export selected information from the database to spreadsheet files. The extraction engine is flexible and optimized in timing. Patients, time range and the parameters of interest are customizable by the user.

Discussion and Conclusion

The LHN was initially set up to provide physicians with a web-tool to administrate data from HIV-positive patients in primary-care and to re-use the collected clinical information to perform CTs in Northern Italy. In this paper, an extension of the former architecture is presented, which provides the development of a system that automatically transfers clinical information from the hospitals' EHR towards the LHN database. A further extension enables direct communication between the developed system and the main HIV national databases.

The proposed system works properly for the two illustrated scenarios. In particular, for the second scenario, data can be automatically extracted in a standardized way, and conveyed towards the central repository for their re-use in clinical trials.

In order to manually collect a complete report of the laboratory activity for all the patients involved, a strenuous human effort and the allocation of great resources would be necessary. Moreover, the risk of error in a manual system is very high. Instead, a tool able to automatically transfer all data, such as the one described in this paper, is essential to rapidly collect all laboratory information in a reliable manner. The use of medical standards ensures possible future integrations with additional health care structures, enhancing the expansion of the project. Currently this system manages and elaborates data coming from regional hospitals, but in the near future, it would be possible to expand it, involving extra-regional facilities, thanks to interoperability efforts.

Statement on ethics vote

The research was approved by the Liguria Ethics Committee (directed by prof. Manlio Ferradini) on August 28th 2013,

References

[1] T. Fahey, S. Griffiths, and T. J. Peters, "Evidence based purchasing: understanding results of clinical trials and systematic reviews," *Bmj*, vol. 311, no. 7012, pp. 1056–1059, 1995.

[2] P. Fraccaro and M. Giacomini, "A web-based tool for patients cohorts and Clinical Trials management," *Stud Heal. Technol Inf.*, vol. 180, pp. 554–558, 2012.

[3] P. Fraccaro *et al.*, "The Ligurian human immunodeficiency virus clinical network: A Web tool to manage patients with human immunodeficiency virus in primary care and multicenter clinical trials," *Med. 2.0*, vol. 2, no. 2, 2013.

[4] R. Gazzarata, B. Giannini, and M. Giacomini, "A SOA-Based Platform to Support Clinical Data Sharing," *J. Healthc. Eng.*, vol. 2017, p. 27, 2017.

[5] R. Gazzarata, F. Vergari, T. Salmon Cinotti, and M. Giacomini, "A standardized SOA for clinical data interchange in a cardiac telemonitoring environment," *IEEE J Biomed Heal. Inf.*, vol. 18, no. 6, pp. 1764–1774, 2014.

[6] M. Bonetto *et al.*, "I-Maculaweb: A Tool to Support Data Reuse in Ophthalmology," *IEEE J. Transl. Eng. Heal. Med.*, vol. 4, pp. 1–10, 2016.

[7] R. Gazzarata, M. E. Monteverde, E. Vio, C. Saccavini, L. Gubian, and M. Giacomini, "A CTS2 compliant solution for semantics management in laboratory reports at regional level," in *Informatics for Health*, 2017.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190588

Linking Care Coordination Measures to Care Outcomes to Improve Care Quality

Rima Artonian Gibbings[a], Nilmini Wickramasinghe[b]

[a] University of Illinois in Chicago, Chicago, USA,
[b] Swinburne University and Epworth HealthCare, Melborne, Victoria, Australia

Abstract

Currently, poor patient experience coupled with too many instances of missed care, repeated treatments/tests and poor quality all detract significantly from high value care delivery and are plaguing the healthcare systems globally and especially in the US. A care co-ordination solution enabled by technology is proffered as an appropriate remedy in this research in progress.

Keywords:
Patient care, Treatment outcomes

Introduction

Isolated healthcare information systems based in hospitals, provider practices, pharmacies, and laboratories are resulting in fragmented care delivery systems [1] and inefficiencies in care outcomes. Joining these disintegrated information systems efficiently to one another becomes imperative when caring for patients suffering with one or more chronic conditions. In the U.S., the number of patients with chronic diseases, functional limitations or disabilities exceeds 120 million [1] and these Medicare beneficiaries with seven or more conditions on average see eleven physicians including three primary care physicians and eight specialists [2]. Each of these providers could potentially own and operate stand-alone information systems which could potentially result in lack of coordinated care and become a major contributor to complications which occurs during various stages of the patient journey [3] and thus, to address this growing problem we proffer a technology enabled care co-ordination solution.

Methods

To demonstrate the role of the proffered technology solution and establish fidelity and usability, we adopt a multi-phase mixed method study. Phase 1 consists of a systematic literature review to identify all key factors which then leads to the developed conceptual model. Phase 2 includes a two-arm trial at healthcare facilities in the US and Australia where identified critical care measures are tracked by clinicians as compared to standard care approaches. By choosing to look at the same technology solution in different healthcare systems, i.e., US and Australia, we anticipate uncovering initial insights regarding macro level issues. The technology solution used to enable this process is the OneView Point-of-Care system. Phase 3 includes analyzing the data gathered using a mixed methodology. In Phase 4, the technology solution care protocols are revised as required.

Results

Enhancing care coordination has been associated with improved care quality and the patient experience [4]. To do this effectively, it is essential to accurately/precisely detect care coordination measures that are associated with certain healthcare outcomes. Categorizing patient care activities under coordination measures and associating these measures with care cycles (outcomes) will facilitate the formation of strategic care plans that emphasize accountability and efficiency. Our research in progress addresses this key need by using technology as an enabler. We present the theoretical framework of our systematic literature review in Figure 1.

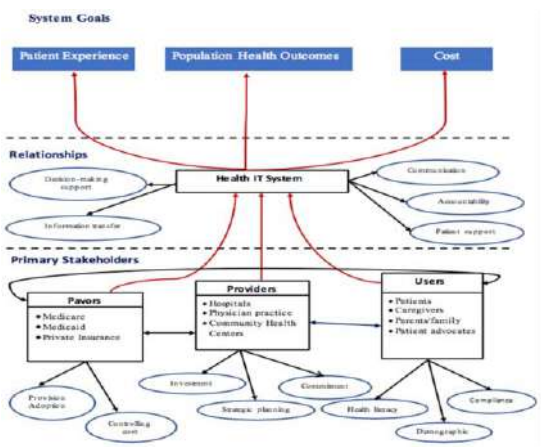


Figure 1. The theoretical framework of the literature review.

Conclusions

A sustainable solution that ensures superior patient experience, high quality care outcomes and at the same time minimizes healthcare costs by limiting duplicated care and/or unplanned readmissions requires high level of care coordination. In today's healthcare context, this necessitates a technology solution. Assessing coordinated measures within current health information systems must include all determinants that could impact healthcare tasks. These measures will transfer care facility processes from a reactive practice into an integrated and collaborative one. We believe our proffered solution will provide a superior patient experience coupled with adherence to a value-based care paradigm.

References

[1] Bodenheimer, T. (2008, March 6). Coordinating Care - A Perilous Journey through the Health Care System. The New England Journal of Medicine.

[2] Pham, H. H., Schrag, D., O'Malley, A. S., Wu, B., & Bach, P. B. (2007, March 15). Care Patterns in Medicare and Their Implications for Pay for Performance. The New England Journal of Medicine.

[3] Rudin, R. S., & Bates, D. (2013, Juen 19). Let the Left hand know what the right is doing: a vision for care coordination and electronic health records. *JAMIA*.

[4] Schultz, E. M., Pineda, N., Lonhart, J., Davies, S. M., & McDonald, K. M. (2013, March 28). A systematic review of the care coordination measurement landscape. BMC Health Service Research.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190589

Implementation of a Tool for the Assessment of the Frail Elderly in a Robotic Agent

Blanca Gonzalo de Diego[ab], Marta Domínguez del Campo[a], José Mª Santamaría García[ab], Mª Lourdes Jiménez Rodríguez[a], Marta Fernández Batalla[ab], Enrique Monsalvo San Macario[ab], Adrián Santamaría Pérez[c], Jesús Pinto Freyre[c], Lydia Madariaga Casquero[a].

[a] Grupo de Investigación MISKC, Universidad de Alcalá, Madrid, Spain,
[b]Centro de Salud Meco (SERMAS), Madrid, Spain,
[c] Lapicero Blanco, Guadalajara, Spain

Abstract

The concern for aging, chronic illness, and dependence is relevant in today's society. The Nursing discipline is responsible for its approach to care. Furthermore, this profession must develop technologies to assist. The main objective is the development of a robotic system for assessment of fall risk and adherence to pharmacological treatment. The magnitude of this prototype design and its validation by experts warrant the attention to support care for the vulnerable and fragile with technology.

Keywords:
Frailty; Nursing Care; Robotics

Introduction

The present study starts with the following research aim: *Design of a robotic agent that assesses fragility by the following two measures: risk of falls and pharmacological adhesion.*

The current paradigm places us before the reality of aging, its chronic illnesses, vulnerability, and fragility, which are far-reaching in our society [1]. It is therefore necessary to identify potential risks and prevent them. [2] In this research we will focus on the approach of two care problems that can usually be seen in the frail elderly: the risk of falls and the difficulty of pharmacologic adherence. Both care problems are related to each other.

Assessment of the Risk of Falls in the Frail Person

In Spanish territory, the modified Downton Scale has been used to evaluate the risk of falls in people based on five criteria [1].

Adherence to Pharmacological Treatment in the Fragile Person

There are studies that show that older and fragile people are more susceptible to the adverse medication effects [3]. Not to mention, they usually take a greater amount of medication (polypharmacy) that increases risks, including the risk of falls [3].

In Primary Health Care, in the Community of Madrid, the test used for the assessment of pharmacological adhesion is the "Morisky-Green Test" [4]. Some articles show the advantages of using this questionnaire [4-6].

Robotics in Health Care for People

The repercussion of the new advances is evident in society: the incorporation of Knowledge-Based Systems [7] in devices managed by people. Health professionals need to undertake new strategies in the development of technology to care for patients [8-11]. Recent research in the European framework stress the importance of paying attention to the daily life of people, especially in the context of fragility in relation to robotics [6,12-19].

Methods

The present study employs a deductive methodology using a knowledge extraction technique and the learning of that knowledge. The research study was carried out from March to May 2018 in the Computer Science Department of the Polytechnic School of the University of Alcalá. The study consisted of the following phases:

- Phase I: Extraction of knowledge about the research problem
- Phase II: Development of the prototype
- Phase III: Implementation of the algorithm
- Phase IV: Verification and validation by experts

Results

The study shows the following results for each planned phase:

- Results for Phase I: **The knowledge extracted led to the creation of** the Downton Scale Modified Downsizing and Morinsky-Green Test prevention algorithm.
- Results for Phase II: Hardware elements were used for the following functions: execution of algorithm, data storage, bluetooth communication, transmission of information to the user. The robotic system interacted with mobile devices allowing a human-robot interaction experience.
- Results for Phase III (Included General and Specific Description):
  - Product Perspective - The system has been designed for the assessment of the frail elderly people, identifying potential risks for the subsequent prevention of them through health care. The product was designed to work on mobile devices via Bluetooth connection.
  - Characteristics of Users - Users of the tool were nursing professionals and caregivers.
  - External Interfaces - The user interface implemented was an Android App, and the

hardware and software interface was the robot and the software Arduino.

- Results for Phase IV: The results obtained were presented in different congress and international publications. The experts' validation carried on with the testing of the tool. The objective was to verify its validity for the international expert community and to link its relationship to the multidisciplinary and interdisciplinary communities with computer and health knowledge through a group of experts (Doctors in Nursing and Computer Science). It is expected that the tool will be validated at the community level through piloting in health and non-healthcare centers.

Discussion

At present, there are multiple studies that contemplate the importance of using robotic agents and computer tools in the care of the frail elderly. As a result of this research, the use of this technology for vulnerable groups, such as children and the elderly, are now being explored. Except, in addition, one should focus on creating a new care model based on the indispensable mechanisms for the maintenance of life that represents human needs and responses of the population.

Conclusions

The robotics of care proposes to include assessment systems for the care of vulnerable people (15,16). Its use in the field of Nursing favors a new vision applicable in healthcare and home settings and promotes the approach of a new sustainable model of care robotics that adapts to different communities.

Acknowledgements

This study had been financed by the Illustrious Nursing Association of Madrid. The Management about Information and Standard Knowledge of Care Group at the University of Alcala supported this work.

References

[1] Ministerio de Sanidad, Servicios Sociales e Igualdad. Informes, estudios e investigación. Estrategia de Promoción de la Salud y Prevención en el SNS. *Documento de consenso sobre prevención de fragilidad y caídas en la persona mayor*. 2014 Jun 11.

[2] World Health Organization. *WHO Global Report on Falls Prevention in Older Age.* (2007). http://www.who.int/ageing/publications/Falls_prevention7March.pdf

[3] W. Thompson, C. Lundby, T. Graabaek, D. S. Nielsen, J. Ryg, J. Sondergaard, A. Pottegard. Tools for Deprescribin in Frail Older Persons an Those with Limited Life Expentancy: A Systematic Review. *The American Geriatrics Society (JAGS)* (2018), 1–9.

[4] M. A. Rodríguez Chamorro, E. García-Jiménez, P. Amariles, A. Rodríguez Chamorro, M.J. Faus. Review of the Test Used Measuring Therapeutic Compliance in Clinical Practice. *Atención Primaria* **40** (2008), 379–431.

[5] D.E. Orem. Conceptos de enfermería en la práctica, 4th Edition, Masson-Salvat, Barcelona, (1993).

[6] Gonzalo de Diego B, Herrero Jaén S, Monsalvo San Macario E, Madariaga Casquero L, Jiménez Rodríguez ML, Santamaría García JM, Ramírez Sánchez SC, Vialart Vidal N, Condor Camara DF. Robotic Implementation of the Necessary Mechanism for the Human Characteristics Simulation: Approach from Self-Care Conceptual Modeling. *Stud Health Technol Inform*. **250** (2018), 115-120.

[7] M. L. Jiménez, Sistema basado en conocimiento para la ayuda en el diagnóstico del cansancio del desempeño del rol del cuidador [tesis], Universidad de Alcalá, España, 2006.

[8] S. Russell, P. Norvig. Inteligencia artificial, un enfoque moderno, Pearson Education, Madrid, 2004.

[9] A. Barr, E.A. Feigenbaum, Handbook of Artificial Intelligence, Computer Science Department, Stanford, 1979, 121-199.

[10] Ministerio de Sanidad, Servicios Sociales e Igualdad. Estrategia para el Abordaje de la Cronicidad en el Sistema Nacional de Salud Estrategia aprobada por el Consejo Interterritorial del Sistema Nacional de Salud el 27 de junio de 2012. Madrid, 2012.

[11] J.M. Santamaría García. *Investigación deductiva, representación lógica e implementación computacional sobre las limitaciones de acción del Autocuidado según el modelo de Dorothea Orem*. PhD Thesis. School of Computer Engineering. Department of Computer Science. University of Alcala, 2008.

[12] A. García-Soler, D. Facal, U. Díaz-Orueta, L. Pigini, L. Blasi, R. Qiu. Inclusion of service robots in the daily lives of frail older users: A step-by-step definition procedure on users' requeriments. *Archives of Gerontology and Geriatrics* **74** (2018), 191-196.

[13] P. Simou, A. Alexiou, K. Tiligadis. Artificial humanoid for the Elderly People. *Adv Exp Med Biol.* **821** (2015), 19-27.

[14] V. Faucounau, Y. H. Wu, M. Boulay, M. Maestrutti, A. S. Rigaud, QuoVADIS project. Caregivers' requeriments for in-home robotic agent for supporting community-living elderly subjects with cognitive impairment. *Technology and Health Care* **17** (2009), 33-40.

[15] Y.H. Wu, V. Cristancho-Lacroix, C. Fassert, V. Faucounau, J. de Rotrou, A.S. Rigaud. The Attitudes and Perceptions of Older Adults With Mild Cognitive Impairment Toward an Assistive Robot. *Journal of Applied Gerontology* (2014), 1-15.

[16] B. Gonzalo de Diego, A. González Aguña. Diseño e implementación robótica de un sistema de atención en cuidados en niños. A propósito de la disfagia. *ENE, revista de enfermería: La Ciencia del Cuidado*, **10** (2016).

[17] F. Nieto Franco. Design and Implementation of a Robotic Software for the Assessment of Pain in Children. *Nursing Informatics* (2018), 250.

[18] B. Gonzalo de Diego. Los problemas de salud crónicos en las personas: un análisis del impacto desde la agencia y la demanda de autocuidado. *ENE, Revista de Enfermería* **10** (2016).

[19] M. Fernández Batalla, E. Monsalvo San Macario, A. Gonzále Aguña, J.M. Santamaría García. Diseño de un método de análisis para el cálculo de la vulnerabilidad como predictor de la fragilidad. *Revista ENE de Enfermería* **12** (2018).

Address for correspondence

Contact: Blanca Gonzalo de Diego. Research Group MISKC, University of Alcalá. Alcalá de Henares, Madrid (Spain). Email: blanca94gd@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190590

Using Contemporary e-Learning Tools to Teach Staff How to Use New Online Documentation Systems

Jeffry S. Gordon, Ryan E. McNew, Elizabeth E. Weiner, and Patricia Trangenstein

Vanderbilt University School of Nursing, Frist Nursing Informatics Center, Nashville, TN, USA

Abstract

Medical centers, hospitals, and health care practices are implementing electronic documentation systems and regularly adding new features to these systems. Training users to effectively use these systems in a scalable way has been a challenge. Learners learn at different rates and have different needs. While traditional face to face instruction has been the gold standard, eLearning technologies can provide acceptable alternatives. This poster will demonstrate three eLearning methodologies for training.

Keywords:
Education, Nursing, Documentation, Staff Development

Introduction

Clinician education and training on new EHR systems is one of the biggest hurdles encountered when implementing a new documentation system. Omaya, Cipriano, et al., claim that the frustration experienced by clinicians with their inability to use the EHR efficiently leads to significant burnout, particularly since up to 50% of a clinician's time can be spent in clinical documentation [1]. Roney reports that, "Training sessions for complex EHR systems can be overwhelming and time consuming." According to Andres Jimenez, MD, CEO of ImplementHIT, "Many hospitals may not receive optimal outcomes from EHRs because hospital staff and physicians are not trained effectively" [2].

Kim, Rodriguez, et al. state that competencies in using an EHR are a critical skill for health care professionals [3]. Even our own informatics users have identified training existing hires on new EHR features and new hires on the system are among their biggest problems, but unfortunately, according to Karimzadeh and Hoffman, "Academic bioinformatics curricula rarely train users in documentation." They further claim that this failure results in "documentation debt" where more time is spent answering questions one on one than it would have taken to develop the right documentation in the first place [4].

Face to Face Instruction vs. Distance Learning

The traditional standard for training has been face-to-face education where users are often seated at computers and taught how to navigate the system. Learning theory is clear. Thalheimer claims that without appropriately spaced practice and repetition, learning is not as effective as it can be. He recommends widely spaced repetition as more effective than narrowly spaced repetition or no repetition at all [5]. Single session face-to-face training cannot accomplish that.

Providing the bulk of the content asynchronously via a distance platform provides several key benefits. First, employees can learn at their own rate. Secondly, employees do not have to be taken, en masse out of production. Third, there is no travel or meal expense for either the users or the instructor. Fourth, because of no travel or removal from production, widely distributed repetition is an attainable learning practice. Fifth, there is commonality of experience. Sixth, learners model an expert.

Methods

We taught our students in our Informatics program each of these techniques over a two semester course. They insisted on this type of content to help them meet their training and documentation needs.

Rapid turn-around just in time documentation

Rapid turn-around just in time documentation is that type of text-based materials with screen captures that can be produced quickly in response to a specific instructional need or question/problem from a user. Text-based materials are generally the easiest to produce with few resource requirements and they can be very effective in getting new content out to large numbers quickly. The advantages here are speed of development and speed of distribution, thus the name, "rapid turn-around documentation". Gerchev recommends a flexible approach to documentation that includes a variety of the approaches described here. He emphasizes capturing images and making sure the documentation is up to date with the latest product, which is easily accomplished with rapid turn-around documentation [6].

One of the more popular contemporary platforms for these text-based files is Wordpress. Wordpress provides an organizational structure that can be used to later search for stored text documentation. Typically, these searches are based on keywords embedded within document itself. By providing links to documents in a Wordpress environment, changing the document there only once means everyone with the link to the document always sees the latest version. We currently have over 230 documents in our Wordpress site.

Narrated screen recordings

Screen recordings are an excellent way of presenting dynamic content to demonstrate a specific procedure on the computer. Yearwood states in an interview by Robert Kelly, "It is almost like having your personal tutor whose message never changes regardless of how many times you rewind or review the content" [7].

Three major advantages to screen recordings are that the user can rewind the video and review sections that were originally confusing. Secondly, the instructor can use mouse highlighting to put the focus on an exact the area of the screen he wants to emphasize. Thirdly, depending on how these videos are stored, they can be accessed at any time from any place with an internet connection. Camtasia is our product of choice.

Creating an online simulation/tutorial

Krueger and Barnes demonstrated that regular continual practice helps retain skills [8]. One approach is to develop a simulated application using an eLearning development tool such as Lectora, from Trivantis, to duplicate many of the features inside an EHR. These applications can provide online guided instruction when the system detects the user needs help. Zhang, et al, refers to this as the Virtual Mentor (VM). The VM contains multimedia integration, just in time knowledge acquisition, interactivity, self-directivity, flexibility, and intelligence [9].

The instructional development team uses the product to create simulated cases from a set of contrived cases that cover the concepts and skills on the EHR the user needs to learn. A script is developed based on learning outcomes. This script allows the developer to capture screens from the actual simulated EHR and sets the sequence with which those screens will be displayed.

Comparison of features and requirements

Rapid turnaround documentation with screen capture	Narrated screen recordings of the Application	Tutorial-simulation with Real Screen Captures
Quick to develop and inexpensive	Quick to develop; daunting for stage fright instructors	Takes time to develop and is relatively more expensive
Good for quickly needed response	Good for visual learners	Allows users to develop "muscle memory"
Can be emailed or stored on any server	Requires a media server to deliver	Requires a web server with database storage capabilities

Movement through the simulation is governed by the developer of the application using the button and logic capabilities available in the eLearning tool and match very closely to the actual EHR. Areas in the EHR the developer does not want the user to wander in will be disabled. Total control of where the user goes is up to the developer using the software tool. This

way the lesson is focused only on those areas that are part of the specific set learning outcomes. The user cannot go down rabbit holes that will not get them to their final goal. Responses and response rates can be stored for future discussion with a live instructor trainer.

Results

Student course evaluation feedback from the Technology Components for Informatics class was that all seven students responding reported they strongly agreed with the statement that this content was relevant to their course of study. The six students responding in the Web Development for Healthcare Applications class reported they agreed or strongly agreed that the online simulation/tutorial content was relevant to their course of study. Typical comments were, "A valuable learning approach by having us create our own learning module using the tools required." Another said, "It significantly increased my confidence and skills in using this type of platform for creating interactive eLearning modules. Learning, "How to use Lectora to create interactive EPIC training".

Conclusion

The techniques described here can solve many of the problems that face to face instruction cannot solve. Our students advocate the need for individualized pace instruction, rapid response to questions, common experience, repetitive learning, modeling an expert, reducing travel costs, and reducing the need to take people out of production for training.

References

[1] A. K. Ommaya, P. F. Cipriano, D. B. Hoyt, K. A. Horvath, P. Tang, H. L. Paz, et al, Care-centered clinical documentation in the digital environment: solutions to alleviate burnout: discussion paper, *NAM Perspectives* (2018), 1-13.

[2] K. Roney (2012), 3 Solutions to Improve EHR Training, Outcomes, Retrieved November 8, 2018 from http://www.beckershospitalreview.com/healthcare-information-technology/3-solutions-to-improve-ehr-training-outcomes.html.

[3] J. Kim, H. Rodriquez, P. Hector, K. Estlin and C. Morris, Impact of longitudinal electronic health record training for residents preparing for practice in patient-centered medical homes, *Perm J* **21** (2017), 16-122.

[4] M. Karimzadeh, and M. Hoffman, Top considerations for creating bioinformatics software documentation, *Brief Bioinform* **19** (2004), 693-699.

[5] W. Thalheimer, (2006), Spacing learning events over time: what the research says, https://www.worklearning.com/wp-cotent/uploads/2017/10/Spacing_Learning_Over_Time__March2009v1_.pdf, (accessed November 8, 2018).

[6] Gerchev, I. Is your product's documentation good enough? (2014), http://www.sitepoint.com/products-documentation-good-enough/, (accessed November 8, 2018).

[7] Kelly, R. Using screen capture software to improve user learning, (2011), http://www.facultyfocus.com/articles/teaching-with-technology-articles/using-screen-capture-software-to-improve-user-learning/, (accessed November 8, 2018).

[8] J. Krueger, and K. Barnes, Continual reinforcement and practice enhance learner success, (2011), http://olc.onlinelearningconsortium.org/effective_practices/continual-reinforcement-and-practice-enhance-learner-success, (accessed November 8, 2018).

[9] D. Zhang, J. Zhao, L. Zhou, and J. Nunamaker, Can e-learning replace classroom learning?, *Commun ACM* **47** (2004), 75–79.

Address for correspondence

Jeffry Gordon, Vanderbilt University School of Nursing, 461 21st Ave. S., Nashville, TN 37240-1104, USA; E-mail and Skype for Business: jeff.s.gordon@vanderbilt.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190591

Introduction of a Program to Improve the Information Sharing System of Food Allergy Patients

Ji Su Ha, Su Hyun Kim, Sang Hee Lim, Tae Hoon Ko, Sae Won Choi, Hae Young Lee, Kyung Hwan Kim

Department of Medical Informatics, Seoul National University of Hospital, Seoul, Republic of Korea

Abstract

Although the symptoms of food allergy are diverse and sometimes dangerous to life, we have not been able to effectively share information on patients' food allergies. In this study, we developed a program to prescribe the allergic formula for each patient and to establish a series of processes for safe allergenic meal delivery based on the standards and guidelines of food allergy management. We then assessed the utility of the "introduction of food allergy program" by comparing the number of allergic prescriptions before and after the program application for inpatients. Through the development and introduction of the program, all hospital staffs, including medical staff and dieticians, can share information on food allergy patients. Systematic management of food allergy patients from doctor's prescriptions has provided the basis for safe meal preparation.

Keywords:

Food allergy, Hospital Information Systems, Prescriptions

Introduction

Food allergy is an adverse reaction caused by immunity after ingestion of food [1]. Food allergy is a common disease experienced by 6 ~ 8% of children and 2 ~ 3% of adults [2]. According to an epidemiological survey conducted in 2000, 43,045 children aged 6-15 years at the Pediatric Allergy Respiratory Society, the prevalence of diagnosed as food allergy more than once was 4.7% [3]. In a study of 25,000 children whose ages were 6-12, "self-reported" food intake was 10.9% [4]. The food allergy usually grows naturally as it grows [5]. This is an increasing trend [6].

Although food allergies vary in their severity, severe reactions such as anaphylaxis, angioedema, hypotension, and asthma attacks may pose a life threat if they fail to receive early first aid treatment [7]. At the Seoul National University Hospital, there was a case in which complaints occurred due to proximity errors and side effects due to missing food allergy information by patient when issuing meal orders, but there is no systematic program for sharing food allergy information.

Methods

Since June 2017, the Allergy Internal Medicine, Feeding and Nutrition Department, Nursing Headquarters, QA Team, and Information Development Team have gathered and organized a process to share food allergy information. Based on this, we developed and introduced food allergy information sharing program in the Hospital Information System (HIS).

To evaluate the effectiveness of the food allergy program, we surveyed the status of food allergy information sharing among patients admitted to Seoul National University Hospital from March 1, 2018 to August 31, 2018.

In addition, the number of prescriptions for allergic meals in July 2017 before the introduction of the program and the number of prescriptions for allergic meals in July 2018 after the introduction of the program were compared.

Results

The information sharing and coping process for food allergy patients were prepared as follows.

1) Establish criteria to be managed and guidelines for food allergy management.

2) Categorize allergy information and systematically classify and input information when hospitalized.

3) Physicians will prescribe the allergic formula for each patient, and provide the allergic formula in the nutrition course, explain the treatment formula and adjust the meal plan.

4) Rapid first aid and patient monitoring based on [Alert] information from doctors, nurses, and food and nutrition department in case of food allergy symptoms.

According to the above work process, the food allergy program in HIS was developed as follows.

1) In order to systematically share allergy information, the food allergy item of the hospital nursing information survey system was changed from the free text to the categorized check item. One thing to consider is ‘Mackerel / pupa / lacquer / hemp / taro’ are not likely to be classified by item, but collecting items that are not provided in patient formula improves the convenience of registration (Fig. 1).

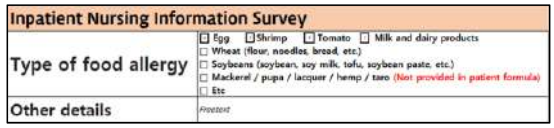


Figure 1

2) When you check the food allergy item on the inpatient nursing information search page, it linked to 'Food Allergy' item of [Alert] (Fig. 2).

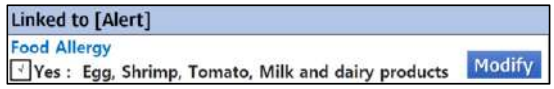


Figure 2

3) In case of adding or modifying the items of food allergy since the creation of the inpatient nursing information survey system, add or modify through the 'Food Allergy' item of [Alert] (Fig. 3).

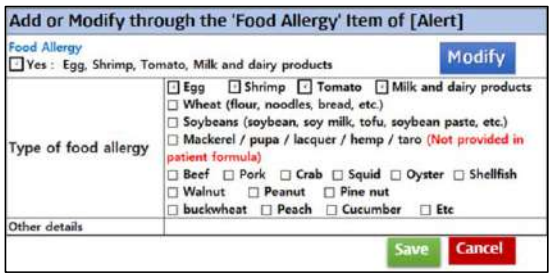


Figure 3

4) The primary care physician issued an allergic meal order (the name of 'other treatment formula').

5) The nurse issues an allergic meal order as prescribed by the doctor.

6) Dietitians provide allergic meals by patient through food management program of food allergy patients.

7) At the time of hospitalization after the patient is discharged from hospital [Alert], the food allergy information is linked to the inpatient nursing information survey.

8) Dietitians are able to manage food allergy categories, manage food allergy categories.

From March 1, 2018 to August 31, 2018, 2,033 patients (5%) were enrolled in a total of 37,429 hospitalized patients with food allergies. As a result of analyzing the number of registered food allergies, the most frequent item among all 2,666 food allergy items was mackerel / pupa / lacquer / hemp / taro item, and total 321 (12%).

Table 1- Number of food allergies registered by Seoul National University Hospital (March to August 2018)

Type of food allergy	Number of registrations (times)
Mackerel / pupa / lacquer / hemp / taro	321(12.0%)
Peach	296(11.1%)
Shrimp	202(7.5%)
Milk and dairy products	122(4.6%)
Egg	106(4.0%)
Pork	97(3.6%)
Chicken	60(2.3%)
Crab	58(2.2%)
Wheat (flour, noodles, bread, etc.)	58(2.2%)
Other (crustaceans)	44(1.7%)

Of the total 76,623 meals prescribed by physicians in July 2017 before the introduction of the program, 34 (0.04%) were allergic. It was analyzed that 2,303 (3%) out of 76,623 of the total meal prescriptions were provided to the patients, and 98.5% of the allergic diets provided through the intervention of nurses and food and nutrition. After the introduction of the program, the physician's diet was revised to be linked to the actual meals provided. In July 2018, the total number of allergic prescriptions and recipes out of 80,766 prescribed by physicians was 2,784, or 3.4%.

Discussion

The purpose of this study is to evaluate the utility of food allergy program based on the process to improve the information sharing system of food allergy patients.

Through the reflection of the computerized system and the data analysis, we obatained the following positive evaluation.

1) The proportion of patients enrolled in a specific period of time with food allergy is 5%, which is similar to that reported in adults with a prevalence of approximately 3-4% [5].

2) By categorizing the items subject to food allergies, it is possible to provide more systematic food allergy meals.

3) Prior to this program, there was no program to share the fact that all hospital staffs such as medical staff and nutritionist were food allergy patients. It is possible to share food allergy patient information through [Alert] 'Food Allergy' as an effective opportunity to treat patients.

4) The number of allergies provided after the introduction of this program was about 3% and 3.4%, respectively, compared to the one month before and after the introduction. However, the rate of providing food allergy was 0.04% and 3.4%. This means that doctors, nurses, and nutritionists are able to manage food allergy patients in an integrated manner through the introduction of computerized programs.

Conclusions

In conclusion, the development and introduction of the program allowed all hospital staffs, including medical staff and dieticians, to share information on food allergy patients, and managed the food allergy patients in an integrated manner to lay the groundwork for a safe meal recipe.

Acknowledgements

The author thanks for the development of the program based on this study, the allergy medicine, the food and nutrition department, the nursing headquarters, and the QA team are united and the improvement plan is derived. In addition, I would like to thank the team members / developers who helped me to develop the Food Allergy Program.

References

[1] National Institute of Allergy and Infectious Diseases (July 2012). "Food Allergy An Overview". 4-5.
[2] Ministry of Food and Drug Safety. https://www.mfds.go.kr
[3] Oh, J.W., Pyun, B.Y., Choung, J.T., Ahn, K.M., Kim, C.H., Song, S.W., et al. Epidemiological change of atopic dermatitis and food allergy in school-aged children in Korea between 1995 and 2000. J Korean Med Sci 2004;19: 716-23.
[4] Shek, L. P. C., & Lee, B. W. (2006). Food allergy in Asia. Current opinion in allergy and clinical immunology, 6(3), 197-201.
[5] "Food Allergies and Intolerance". Health Canada. December 6, 2007. Archived https://www.canada.ca/en/health-canada/services/food-nutrition/food -safety/food-allergies-intolerances.html
[6] Branum, A. M., & Lukacs, S. L. (2008). Food allergy among US children: trends in prevalence and hospitalizations. P.1.
[7] Sicherer, S. H., & Sampson, H. A. (2014). Food allergy: epidemiology, pathogenesis, diagnosis, and treatment. Journal of Allergy and Clinical Immunology, 133(2), 291-307.

Address for correspondence:

Jisu Ha. smile_jisu@snuh.org.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190592

Alicanto Online Latin American Maternal Informatics Community of Practice

Juan Henao[a], Yuri Quintana[a,b], Charles Safran[a,b]

[a] *Division of Clinical Informatics, Beth Israel Deaconess Medical Center, Boston, MA, USA*
[b] *Harvard Medical School, Boston, MA, USA*

Abstract

Most cases of maternal deaths could be avoided with timely access to quality healthcare, but a key challenge in addressing quality of care in maternal health is the lack of accurate data and analytics. Implementing online communities of practice is a way to resolve this, but in low--and middle-income countries this is particularly challenging. We discuss the design of the Alicanto Online Latin-American Community of Practice that focuses on both outcomes and process indicators.

Keywords:

Information Technology, Pregnancy, Quality Improvement

Introduction

The health of pregnant women is critical for global development. The Sustainable Development Goals (SDG) agenda and the Global Strategy for Women's, Children's, and Adolescent's Health 2016–2030 aim to reduce maternal and newborn deaths [1]. Most cases of maternal deaths could be avoided with timely access to quality healthcare, but a key challenge in addressing quality of care in maternal health is the lack of accurate data. Accurate data improves our knowledge of whether the effective interventions are being received by pregnant and post-partum women and enables the targeting of resources to those most in need. However, diverse outcomes of interest, variable standards of care, quantity and quality of data can be drastically different, as well as cultural and other differences; can substantially modify measurements depending on the outcomes of interest studied.

To improve maternal outcomes in Latin America (LA), we are implementing an online community of practice that can deliver evidence-based guidelines and collect priority indicators, while providing spaced learning medical education and guidance for the measurement and monitoring of maternal and health. This paper describes the challenges in maternal data collection in low- and middle-income countries, and the design of the Alicanto Online Latin-American Maternal Informatics Community of Practice.

Challenges in Global Maternal Outcome Networks

In many low-income countries, maternal deaths go uncounted and frequently the cause of death is unknown or not recorded correctly and the maternal care process is equally poorly registered or not registered at all [2]. Many patient registration systems and electronic health records in low resource settings have problems with non-standardized record-keeping techniques which result in missing records, inconsistencies, poor data quality, and inaccuracies and hence undermine evidence-based decision making in healthcare service delivery [3]. This makes it difficult for national health programs to allocate resources where they are needed the most. To achieve this goal is necessary the integration and harmonization of high amounts of heterogeneous medical data that is stored in different health information systems. Such a task is challenging in both developed and developing countries [4]. Comprehensive database applications for a domain can reduce such variation within this domain. The Netherlands established national domain information models to support electronic information exchange, using cases from perinatology as a national pilot [5]. They found that in some instances, additional agreements are necessary about the preferred vocabulary because the professional organizations need to harmonize their materials and the limitations are reached for what should be part of the standard, and what professional organizations should develop and maintain within their realm. Another challenge is data aggregation and overlap. For maternal care, clinical data is often generated from various sources (prenatal screenings, primary care providers, midwives) and the health information may exist in both paper-based and computer-based systems at institutions located in different geographical locations. The overlap across systems introduces the potential for data variation through duplication of data entry and differing concept definition or context of use. Studies show that redundant and inconsistent records lead to errors, extra effort, misdirected data, over-reliance on the spoken word, inaccuracies, information loss, limited standardization, miscommunications and limited outcomes evaluations being a major cause of medical incidents [6]. Research has also shown how coordination and communication among clinicians and across settings resulted in greater efficiency and better clinical outcomes [7]. In general, the burden on individual providers of collecting data has been well documented, as has the lack of use of data collected at such great cost [8], which breaks the feedback mechanism whereby monitoring and review can result in improved delivery of interventions.

Alicanto Network Objectives and Approach

The network objectives are to strengthen and enhance the existing national data systems and evaluation programs through information and communications technology (ICT). The project builds on existing regional communities of practice in maternal health, in which we will leverage technical expertise across LA to foster the exchange of lessons learned, building a sustainable and reproducible capacity of Human Resources for Health (HRH). These HRH can provide high-quality maternal care interventions for developing countries, particularly for hard-to-reach populations and vulnerable groups, maximizing the use of resources and guiding policymaking, in an attainable, informed and sustainable manner. At this point, the stakeholders of the network include hospitals and regional governments, member clinics at Argentina, Chile and Colombia

and the Division of Clinical Informatics (DCI) at Harvard Medical School (HMS), which has previously signed a framework agreement with PAHO to collaborate in supporting the advancement of eHealth in Latin America and the Caribbean. The network members are involved in co-creating the system and educational content for management of high-risk pregnancies relevant to their needs and agreeing on data encoding standards. The initial courses focus on pre- eclampsia and hemorrhage, which are the causes of most maternal deaths. This co-creation uses S.M.A.R.T goals and Bloom's Taxonomy [9] and is being done via weekly web conference meetings, sharing data dictionaries and database designs to achieve consensus on common data formats for outcomes comparison and establishing methods to capture process metrics.

Strategies aimed at reducing maternal deaths need to address inequities in access to good quality maternal health interventions, which are based on learning and applying best medical practices protocols, step-by-step labor, delivery and immediate postpartum period management guidelines. We will deliver those using web-based and mobile app tools (eLearning, mhealth) inside Alicanto Integrated Mobile Health Network, but also the platform provides an online consultation system (telehealth) for obstetric emergencies between rural HRH and experts at reference hospitals, allowing for appropriate, coordinated and timely patient referral to a higher level of care if needed. Participating centers will also have access to an asynchronous discussion forum to discuss with colleagues the guidelines and lessons learned from the cases; using validated, structured, and process-focused frameworks to reflect and learn from experiences, given the fact that it has been showed that medical knowledge, job satisfaction, and self-efficacy do not increase by only using an SMS-based continuing medical education (CME) intervention that fails to stimulate lateral learning i.e. learning from your peers [30]. Since processes variables are included in our design, these can be used in a feedback mechanism to identify areas for improvement and allocating resources in areas likely to have the greatest impact.

The first pilot is being conducted between Fundación Valle del Lili, a major teaching hospital in Cali, Southwest Colombia and a small rural clinic in the West Coast Pacific region. Despite recent efforts in Colombia to improve the number of health professionals in rural locations and implement a red code for postpartum hemorrhage in accordance with international standard obstetric protocols, Maternal mortality ratio (MMR) remains high and lags behind Latin American countries at a similar level of economic development (e.g., Mexico and Argentina). The isolated and discontinuous Colombian Pacific coastal lowlands have a majority of population with a uniquely African genetic heritage (90%) but there are also populations of mixed race (6%) and Native Americans (4%). This region has an MMR similar to 2010 African countries, like Ethiopia, Ghana, and Ruanda, or similar to Haiti, the country with the highest MMR in Latin America and the Caribbean [10]. The Alicanto platform will provide a scalable mobile solution to support rural clinics across Colombia with expansion to other Latin American centers. The second pilot is scheduled for Southern Chile and third pilot in northern Argentina, both in locations where experts are far away from major teaching hospitals.

Conclusion

By having co-creation of the network, we are achieving a standardized set of common goals and data metrics that can be useful at the local level while being also useful at a regional level for data comparing. By having harmonized sets of data and tools, according to site characteristics ranging from minimal to optimal infrastructure and clinical conditions, it will be possible to compare outcomes across similar centers. This will help to create a policy development platform that introduces a systematic approach to policymaking informed by evidence and collaboration between institutions and nations. The indicators and methodology can be reused by projects that map to the same core dataset and will be available in an online public repository and a community of practice will support the use and adaptation of the tools in the repository.

Acknowledgements

Funding sources for the work and other relevant acknowledgements are noted here. Authors may also present disclosures or disclaimers to their work in this section.

References

[1] S. Kuruvilla, F. Bustreo, T. Kuo, C. Mishra, K. Taylor, H. Fogstad, G.R. Gupta, K. Gilmore, M. Temmerman, and J. Thomas, The Global strategy for women's, children's and adolescents' health (2016–2030): a roadmap based on evidence and country experience, *Bulletin of the World Health Organization* **94** (2016), 398.

[2] S. Aziz and M. Rao, Existing record keeping system in government teaching hospitals of Karachi, *Journal-Pakistan Medical Association* **52** (2002), 163-173.

[3] S. Ismail, M. Alshmari, K. Latif, and H.F. Ahmad, A granular ontology model for maternal and child health information system, *Journal of Healthcare Engineering* **2017** (2017).

[4] Z.A. Bhutta and A. Hafeez, What can Pakistan do to address maternal and child health over the next decade? *Health Research Policy and Systems* **13** (2015), S49.

[5] W.T. Goossen, M.J. Jonker, K.U. Heitmann, I.C. Jongeneel-de Haas, T. de Jong, J.W. van der Slikke, and B.L. Kabbes, Electronic patient records: domain message information model perinatology, *International Journal of Medical Informatics* **70** (2003), 265-276.

[6] C. AbouZahr and T. Boerma, Health information systems: the foundations of public health, *Bulletin of the World Health Organization* **83** (2005), 578-583.

[7] A.L. Bhasale, G.C. Miller, S.E. Reid, and H.C. Britt, Analysing potential harm in Australian general practice: an incident-monitoring study, *Medical journal of Australia* **169** (1998), 73-76.

[8] World Health Organization, A rapid assessment of the burden of indicators and reporting requirements for health monitoring, *Geneva: World Health Organization* (2014).

[9] T. Jarboe and S. Sadler, *Bloom's Taxonomy and Depth of Knowledge: Learning Charts*, ABC Schoolhouse, 2012.

[10] A. De La Torre, Z. Nikoloski, and E. Mossialos, Equity of access to maternal health interventions in Brazil and Colombia: a retrospective study, *International journal for equity in health* **17** (2018), 43.

Address for Correspondence

Juan Henao; juanfdracc1@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190593

Open and Linkable Knowledge About Management of Health Information Systems

Konrad Höffner[a], Franziska Jahn[a], Anna Lörke[a], Thomas Pause[a], Birgit Schneider[a], Elske Ammenwerth[b], Alfred Winter[a]

[a] Leipzig University, Institute for Medical Informatics, Statistics and Epidemiology, Leipzig, Germany
[b] University for Health Sciences, Medical Informatics and Technology, Institute of Medical Informatics, Hall in Tirol, Austria

Abstract

Given a care delivery organization, its health information system can be defined as the part of the organization that processes and stores data, information, and knowledge. There is an enormous number of frameworks, textbooks and articles that describe the scope of health information system management from the perspective of medical informatics. Transforming this knowledge to Linked Open Data results in a structured data representation that is accessible for both humans and machines, the Semantic Network of Information Management in Hospitals (SNIK). We present interfaces that are useful for researchers, practitioners and students, depending on their objectives and their Semantic Web skills.

Keywords:

Semantic Web, Information Management, Health Information Systems

Introduction

Given a care delivery organization, its health information system (HIS) can be defined as the part of the organization that processes and stores data, information, and knowledge. It usually consists of a large number of different application systems, computers, and network components (Winter et al. 2011). Managing an HIS comprises planning, monitoring and directing activities. Due to the complexity and the unique conditions in health care, HIS management is an exceptionally challenging task. There is an enormous number of frameworks, textbooks and articles describing the scope of HIS management from the perspective of medical informatics. However, the disciplines of business informatics and information systems (IS) provide an even broader view on information systems and their management. A holistic view on HIS management comprises knowledge from different scientific disciplines and requires well-defined links between these different worlds. These links help researchers and students connect their existing knowledge with further knowledge from other sources during research and learning.

In order to integrate different knowledge sources and to provide the knowledge in a structured, machine-readable data format, we extracted knowledge about HIS management from three textbooks and other sources (see Table 1) and converted it to RDF [4]. The combination of this knowledge results in SNIK, the Semantic Network of Information Management in Hospitals ("Krankenhaus" in German), which is freely and openly accessible by open tools. In order to encourage and enable other researchers, students and health informatics professionals to use available knowledge of HIS management, we interlink classes from different sources, and present and compare the interfaces of SNIK for different target audiences.

Methods and Materials

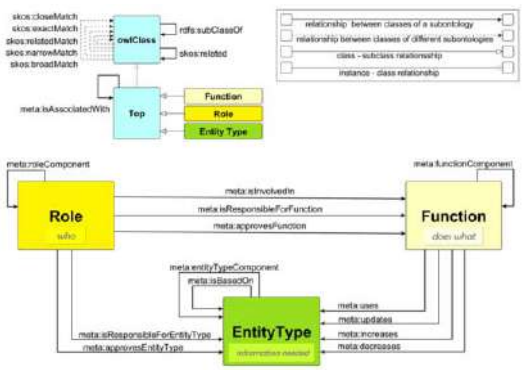


Figure 1—The SNIK Meta Model

SNIK is modelled as a modular (see Table 1) OWL 2 DL ontology. The Meta Model is the central module and provides a common vocabulary for the domain of HIS management (see Figure 1). It contains classes (rectangles in Figure 1), which represent concepts of HIS management and relations (the labelled edges in Figure 1), which represent possible interactions between them. For example, a role and a function can be connected by "is involved in", "is responsible for" or "approves". The Meta Model is extended by five subontologies (see Table 1). At the head of the class hierarchy is the "Top" class, which has exactly three disjunctive subclasses. Following the Meta Model, each class has to be a subclass of exactly one of them. The superclass of a new concept can be found by answering the question: "Who ("Role") does what ("Function") and which information ("EntityType") is needed to perform this function?"

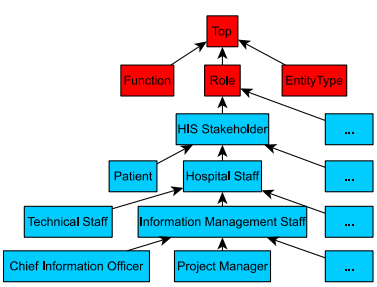


Figure 2—Excerpt from subclass hierarchy of the Meta Model and the subontology

Results

SNIK v0.8 contains 4729 classes, 329 properties, 713 interlinks and 112747 triples and is available under the CC BY-NC-SA 4.0, using open standards over interfaces with different compromises between expressivity and accessibility.

RDF Dump

In line with the *Linked Data* principles [2], we assign a unique URI to each class and follow the specifications of the Resource Description Framework (RDF), which states how facts about concepts can be modeled as *triples*. Each triple contains a subject, property and an object. There are multiple RDF serializations, such as Turtle [3], which allow abbreviating URIs using prefixes, see Table 1. SNIK is available as Turtle files at http://www.snik.eu/download/snik-0.8.zip.

RDF Browser

Users can look up a resource using the HTTP protocol to receive a description of the class, using the standards RDF and SPARQL. The LodLive RDF browser at http://www.snik.eu/ontology offers a human-readable description. Triples between URIs can be followed so that new information can be discovered.

SPARQL

The SPARQL Protocol and RDF Query Language (SPARQL) [3] allows us to manipulate and query SNIK as RDF. We offer free read access to the SPARQL endpoint (query service) at http://www.snik.eu/sparql both for manual queries and as an API. The SNIK project uses the endpoint as an API in several applications, both custom made, such as SNIK Graph (see Figure 3), and adapted software, such as the RDF browser and the OntoWiki ontology editor, see [4]. The endpoint presents the most expressive interface but requires knowledge of both the SPARQL syntax and the SNIK Meta Model.

Table 1—Modules of SNIK

Ontology	Prefix	Source
http://www.snik.eu/ontology/meta	meta	All Sources
http://www.snik.eu/ontology/bb	Bb	Textbook
http://www.snik.eu/ontology/ciox	ciox	CIO Interview
http://www.snik.eu/ontology/ob	Ob	Textbook
http://www.snik.eu/ontology/he	He	Textbook
http://www.snik.eu/ontology/it4it	It4it	Standard

SNIK Graph

SNIK Graph, available at http://www.snik.eu/graph, visualizes the structure of SNIK by modelling each class as a node and each RDF triple and OWL restriction as an edge.

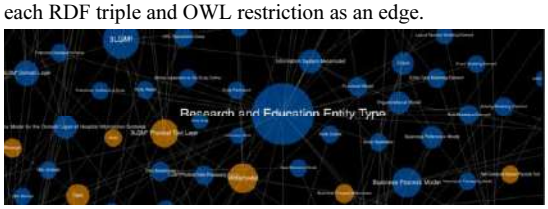


Figure 3—The SNIK Graph Visualization

Interlinks

As an open, linkable ontology, SNIK can be connected to other ontologies, especially in the field of Medical Informatics. Some concept pairs are homonyms: they have the same label but differ in meaning. For example, bb:InformationManagement has the same label as dbr:Information_management, but the definition of information management differs. There should thus be no interlink between homonyms. On the other hand, dbr:Picture_archiving_and_communication_system and bb:PictureArchivingAndCommunicationSystem have a very similar meaning, so that an interlink is generated. Label-based interlinks are generated using LIMES in previous work [4].

Discussion

SNIK is based on the Meta Model from Figure 3. It can be considered an archetype for ontologies describing a given domain (here "HIS management"), which functions in a certain role to be carried out, with information a person with this role needs and provides while carrying out those functions. This is valuable knowledge, for instance as a basis for systems analysis projects [1]. Different user groups can use the interfaces of SNIK that are most suitable to them: Semantic Web experts can use the RDF dump and the SPARQL endpoint to integrate applications with the SNIK ontology. They can also use SNIK as a vocabulary to integrate data from different formats that result from applications from different vendors. Those experts can also use the SPARQL query editor directly to answer specific questions. Students and teachers use the RDF browser and SNIK Graph, which intuitively presents knowledge without requiring knowledge of the query language and the Meta Model.

Conclusions

We publicize knowledge on the management of information systems in medicine and health care using open standards over interfaces with different compromises between expressivity and accessibility. It can be combined with other knowledge in biomedical and health informatics and in other disciplines.

Acknowledgements

This work is supported by the DFG (German Research Foundation) under the project SNIK, grant no. 1605/7-1.

References

[1] E. Ammenwerth, R. Haux, P. Knaup-Gregori, and A. Winter, IT-Projektmanagement im Gesundheitswesen: Lehrbuch und Projektleitfaden Taktisches Management von Informationssystemen. Schattauer Verlag, 2014.

[2] T. Berners-Lee, Design issues: Linked data. Available at http://www.w3.org/DesignIssues/LinkedData.html. Accessed November 16, 2018.

[3] P. Hitzler, M. Krotzsch, and S. Rudolph, Foundations of Semantic Web Technologies. Chapman and Hall/CRC, 2009.

[4] K. Höffner, F. Jahn, C. Kücherer, B. Paech, B. Schneider, M. Schöbel, S. Stäubert, and A. Winter, Technical Environment for Developing the SNIK Ontology of Information Management in Hospitals, in: Building capacity for health informatics Vol. 243, pp. 122–126.

Address for Correspondence

Konrad Höffner, Leipzig University, IMISE, Härtelstraße 16-18, 04107 Leipzig, konrad.hoeffner@imise.uni-leipzig.de.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190594

Healthcare Professionals' Expectations of a Diabetes Care Performance Management System

Iiris Hörhammer[a], Juulia Jäppinen[a], Miika Linna[a]

[a] *The Institute of Healthcare Engineering, Management and Architecture, Department of Industrial Engineering and Management Aalto University, Espoo, Finland*

Abstract

The objective of this survey study was to identify health professionals' attitudes towards performance evaluation and their expectations of a diabetes care performance monitoring system. On average, the professionals supported the implementation of the system, and were interested in using cost-effectiveness information. Usability and willingness to compare performance results between units were most strongly associated with support. Clinicians' involvement in developing and using performance monitoring systems is encouraged.

Keywords:

Diabetes care, Healthcare professionals attitudes, Monitoring systems

Introduction

Performance monitoring systems can provide important information on the quality, effectiveness and cost-effectiveness of care. Observed deficiencies in practices require changes at managerial and clinician levels of the organization. In implementing new organization and management innovations, the role of the personnel at the clinical and operational level has, indeed, been found even more important than that of managers [1]. Despite healthcare professionals' crucial role in operational change and therefore as users of performance information, there is limited knowledge on how performance monitoring systems are accepted and used by this group. This study uses the extant literature on technology acceptance to shed light on what affects the healthcare professionals' acceptance of, and therefore intention to use a performance monitoring system in a pre-implementation phase.

In line with Davis' Technology Acceptance Model [2], Gagnon et al. [3] literature review found that the perceived benefits of the technology (*usefulness*) are the most common facilitating factor of ICT adoption among health care professionals, followed by *ease of use*. Professionals' perception of system usefulness is related to whether they find the information contained by the system interesting and whether they find their role in using this information as a developer as important. Therefore, professionals' *interest towards cost-effectiveness* information and perception of the importance of their *role in developing* care practices may affect their support for a performance monitoring system.

Healthcare professionals' typically work under tight schedules and therefore *time constraints and work load* related to the use of the new technology is a notable barrier to technology acceptance [3,4]. In the implementation phase, *sufficient resources* offered by the management and *support in the use of the new system*, especially in terms of a *super-user* in the clinical unit, contribute to the success [3].

In this survey study, we tested whether diabetes care professionals' interest towards and perceived role in performance management, their expectations of the performance management system, and perceptions of resource availability are associated with their support for the new reporting system.

Methods

Data

The study was conducted in a Finnish social and health care district serving a population of 170 000people. We conducted a web-survey to all potential future users of the system, that is, the health professionals treating patients with diabetes in the district, and their supervisors, altogether 170 professionals.

The questionnaire included questions on the present diabetes care practices in the region, care quality and cost-effectiveness improvement, and the implementation of the new monitoring system. Most of these questions were statements with a five-point Likert scale.

The invitation letter included a description of the new performance management system. The monitoring system allows evaluation of the quality of diabetes care, and the follow-up of the health outcomes, diabetes-related treatments and contacts of the entire diabetic patient population in the region. The system will be developed to also provide information on the association of cost and effectiveness of diabetes care (cost-effectiveness).

Analysis

Statistical analysis was performed with descriptive and linear regression analyses to assess different factors association with professionals' support for the monitoring system to be implemented.

In the linear regression analyses, professionals' support for the system was used as the dependent variable (DV). Professionals' perception of their role in quality improvement, interest in cost-effectiveness information, willingness to benchmark results with other units, expected usefulness, usability and willingness to develop the new system, time available for quality improvement at present, expected support from an expert user and having used and having tested the system were used as independent variables (IV). We first tested the associations of the IVs with the DV separately using univariate regression. Secondly, to test the relative contribution of each of the IVs to the total variance of DV explained, we used multiple regression analysis. Predictors that were significant in univariate regression were included

in the multiple regression analysis. All statistical analyses were performed using Stata version 15.0 (StataCorp LP, College Station, TX, USA).

Results

Responses from 86 professionals (response rate =51%) were received. Majority of the respondents were doctors (44%) and nurses (29%) who treat patients with diabetes, and a minority were doctors' (10%) and nurses' (10%) supervisors. Majority of the respondents work in primary care (70%). Little over half (53%) of the respondents had heard about the new system to be implemented and only 7% had tried the demo version. Respondents' support for the system was rather high and 92% reported that they are interested in the cost of care and the association between cost and health outcomes. Only 29% expected that it would be easy for them to use the system (Usability). At present, the respondents did not agree to have enough time for quality monitoring and 41% expect that there will not be enough resources for the implementation of the new system (see Table 1).

Table 1. Descriptive Information (n=86)

Characteristic or Factor		
Profession, N(%)		
	Doctor	38 (44%)
	Nurse	25 (29%)
	Doctors' supervisor	9 (10%)
	Nurses' supervisor	10 (12%)
	NA	4 (5%)
Support, Mean (SD)		4.4 (0.73)
Perceived role, Mean (SD)		4.0 (1.13)
Interest in cost-effectiveness, Mean (SD)		4.6 (0.70)
Willingness to benchmark, Mean (SD)		4.1 (0.95)
Expected usefulness, Mean (SD)		4.1 (0.74)
Expected usability, Mean (SD)		3.3 (0.98)
Willingness to develop, Mean (SD)		3.7 (0.96)
Available time, Mean (SD)		2.3 (1.09)
Expected resources, Mean (SD)		2.8 (1.03)
Expected support from expert user, Mean (SD)		3.7 (0.98)

To assess which factors would best predict support, we fitted a multiple regression model with support as the DV (see Table 2). In model A, all variables with significant association in the univariate regression were included as IVs. Available time and expected resources had not been associated with support in univariate analysis and were thus left out from this model. The results showed that only willingness to benchmark, expected usefulness and expected usability have significant association in the multiple regression. Model B shows that these variables alone explain 65% of the variation in the DV. The magnitude of the positive effect on support is moderate for expected usefulness and willingness to benchmark. In the multiple regression analysis, the association of expected usability with support changed signs.

Table 2. Multiple Linear Regression Analysis for System Support

	Model A	Model B
Independent variables		
Perceived role	-0.02 (0.10)	
Interest in cost-effectiveness	-0.02 (0.11)	
Willingness to benchmark	0.26 (0.08)**	0.29 (0.07)***
Expected usefulness	0.52 (0.09)***	0.50 (0.06)***
Expected usability	-0.14 (0.06)*	-0.13 (0.06)*
Willingness to develop	-0.01 (0.09)	
Expected support from expert user	0.05 (0.08)	
Has heard of the system	0.03 (0.13)	
R^2	0.68	0.65

Conclusions

In this survey study, we identified health professionals' attitudes towards performance evaluation and their expectations of a diabetes care performance monitoring system. The findings are representative of the diabetes care professionals in a Finnish integrated health and social care district. We found that clinicians who treat patients are interested in using performance management systems with information on cost-effectiveness at population level. Usefulness of the system is the most important aspect in professionals' support for a performance monitoring system. To encourage professionals' support towards performance monitoring, information provided by the systems should be used for constructive benchmarking.

Acknowledgements

This study is a part of the Service Solutions to Support Social Welfare and Health Care Reform (IMPRO) project funded by the Academy of Finland. The authors also thank Siun site personnel for their efforts in performing the survey in the district.

References

[1] J. Øvretveit, M. Andreen-Sachs, J. Carlsson, H. Gustafsson, J. Hansson, C. Keller, S. Lofgren, P. Mazzocato, S. Tolf, and M. Brommels, Implementing organisation and management innovations in Swedish healthcare: lessons from a comparison of 12 cases, *J Health Organ Manag*. **26** (2012) 237–257. doi:10.1108/14777261211230790.

[2] F.D. Davis, Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology, *MIS Quarterly*. **13** (1989) 319–340. doi:10.2307/249008.

[3] M.-P. Gagnon, M. Desmartis, M. Labrecque, J. Car, C. Pagliari, P. Pluye, P. Frémont, J. Gagnon, N. Tremblay, and F. Légaré, Systematic Review of Factors Influencing the Adoption of Information and Communication Technologies by Healthcare Professionals, *J Med Syst*. **36** (2012) 241–277. doi:10.1007/s10916-010-9473-4.

[4] A.K. Yarbrough, and T.B. Smith, Technology Acceptance among Physicians: A New Take on TAM, *Med Care Res Rev*. **64** (2007) 650–672. doi:10.1177/1077558707305942.

Address for correspondence

Iiris Hörhammer (email: iiris.horhammer@aalto.fi)

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190595

Can Solo Practitioners Survive in Value-Based Healthcare? Validating a Predicative Model for ED Utilization

Pamella Howell[a], Peter L. Elkin[a]

[a] *Department of Biomedical Informatics, State University of New York at Buffalo, Buffalo, New York, United States,*

Abstract

The health industry will see increased implementations of value-based models. This study validates a predictive model for determining emergency room utilization. Data from 2991 records are used for the analysis. To validate the model we used Poisson and random forest models. The results indicate that patients with one of six chronic conditions, who missed scheduled appointments or had higher body mass indexes were more likely to utilize the emergency department.

Keywords:

Value-based Healthcare; Machine Learning; Resource Allocation

Introduction

The value-based healthcare model is poised to change the practice of medicine in the United States. Value-based care addresses two concerns: are the available resources distributed to those who need it and how clinicians use the resources optimally [1]. Participating insurers and providers have aligned valued based programs with broader strategic objectives on quality. Value-based care serves three primary functions to reduce cost, improve outcomes and improve quality. As participation expands, patients, payers, and providers will experience a wide range of changes due to its effect. The most significant impact for patients is the improvement in the quality of care and outcomes. Both public and private sector insurers will benefit from any reductions in the cost of care. Insurers are yet to propose methods for decreasing premiums, which is a salient topic; however, it is beyond the scope of this study.

Clinicians serve dual roles in the value-based model. Firstly, they are primarily responsible for improving the quality of care and outcomes of patients. Secondly, they must assist payors by lowering the cost to care for patients. This study aims to provide a predictive model for identifying patients associated with high cost. Using the Health Cost Guidelines (HCG) as a reference, we have identified five categories of cost that a physician may need to reduce, including inpatient, outpatient, prescription drugs, physician, and other. Historical estimates show that inpatient, outpatient and pharmacy cost are highest representing 62 % of the national expenditure for 2016 [2]. The emphasis here will be outpatient cost specifically those incurred in the emergency department. Approximately 3.3% of all emergency department (ED) visits are avoidable [3]; however, statistics also show increased utilization for the treatment of chronic diseases [4].

How does the clinician identify patients who are likely to incur high ED cost? Can the factors that cause high ED usage be eliminated? In answering these question, we attempt to address two of many challenges. One, individuals might argue that the insurers will provide the information needed to monitor cost. The information provided by insurers is generally delayed due to claims processing; this significant time lag reduces a provider's ability to act. Delayed responses when attempting to improve patient adherence can have a negative impact. In this research, we utilize secondary data from the electronic health record (EHR) to reduce data lags and the resource intensity associated with value-based analytics.

The second challenge is that providers need technology and analytics infrastructure to explore cost. Though organizations like hospitals and large health systems have met some of these requirements, the outlay of capital has been significant. For solo and medium-size practices, the ability to participate effectively in the value-based ecosystem will be limited by finite human and financial resources. Clinicians need resources to implement value-based programs such as advanced analytic platforms. Given the current interoperability issues between EHRs and the limited efforts by some vendors to provide analytics support providers must find a way to supplement the need for these services. By implementing our proposed model physicians can reallocate current health information technology resources including team members without significant expenditure.

Methods

Using predictive analytics and machine learning models, we validate a model for predicting patient utilization of the outpatient services. To evaluate the research objective, we utilize a two-stage combination model. The first stage uses the outcomes of a model developed with regression-based machine learning techniques and standardized statistical techniques [5]. The second stage also used standardized regression techniques in the form a Poisson regression and a classification machine learning technique in the form of a random forest model. The feedback loop uses data from the EHR to create a predictive model, then in this study we validate the model using a new patient cohort covering a different period and with more observations. The dataset contains a total of 2991 records with six de-identified variables collected between 2017 and 2018.

Variable Selection and Description

The model variables include age, a polynomial age variable, body mass index (BMI), depression screen indicator, number of no-show appointments, gender and a chronic disease indicator – diabetes, heart failure, COPD, Hypertension, Hyperlipidemia and asthma. The variables can be subdivided into process variables including no-show visits and depression screens indicator and clinical outcomes such as chronic conditions and BMI. All the variables were established as significant in a prior study. The dependent variable is emergency department visits.

Statistical Methodology

We use two approaches to analyzing the dependent variable. For one method the dependent variable is a count of the discrete non-negative integer ED visits. The Poisson or negative binomial regression models are typically the first recommended

models. For the second method we dichotomize the count variable and use the random forest for the following benefits: it gives estimates of what variables are important, offers a technique for detecting variable interactions and has methods for balancing error in unbalanced datasets. Classification and Regression Trees (CART) like random forest are powerful predictive tools, they allow for the combining of predictors [6] [7]. The random forest model generates many CART models in which it chooses that model classification with the most votes.

Results

Using SAS 9.4 and R, we validate the model using Poisson and random forest models. We analyzed the response variable first as a count, then as a dichotomous indicator for the random forest model. We examine the association of ED count with age, gender, no show visits, depression screening, six chronic diseases, and body mass index. The model results suggest that four variables have a significant impact on the frequency of emergency room visits at $p < 0.05$.

The relationship between the response variable and age is convex when compared to the previous study; this likely due to the increased distribution of age. For a given level of all x variables, the expected number of emergency visits will decrease by a factor of exp(-0.055)=0.946 to its lowest, at this point the expected number of emergency department (ED) visits would increase by a factor of exp(0.004)=1.004. The number of ED visits for patients one of 6 chronic diseases is exp(1.362) = 3.90 times more than patients without one of the chronic diseases. No-show visits positively impact the expected number of ER visits for patients by a factor of exp(0.351)=1.420. Body mass index positively impacts the expected number of ER visits for patients by a factor of exp(0.099)=1.104. In this cohort, depression screen indicator and gender did not impact the utilization of the ED.

The preliminary model validation results using random forest, show that the area under the curve (AUC) 0.825 indication good discrimination. The accuracy of the test model is 0.983, sensitivity - 0.884 and specificity - 0.996. The variable importance results showed that the chronic disease indicator held the highest rank followed by no-show visits, BMI, gender, depression, and age.

Discussion

The movement of medical practice away from fee for service toward a value-based system will take time. Insurers have not changed their current products instead — for example, the Center for Medicare and Medicaid Services (CMS) as piloted some value-based programs for diseases such as end-stage renal disease. Private insurers have taken a different approach by contracting with health care providers to implement value-based services for population subsets. Solo and medium-sized practices will face many challenges when making decisions on how best to utilize limited human and financial resources to meet value-based objectives. Physicians who intend to benefit from participating in value-based quality initiatives must leverage their limited resources by using predictive modeling to define parameters that enable the reduction of patient cost like emergency rooms utilization.

The regression model indicates that chronic diseases indicator, no-show visits, BMI, and age can be used to predict the count of ED visits. By combining these finding with the CART prioritization estimates, we can provide suggestions for physicians on how to allocate resources. We suggest creating electronic patient registries that capture parameters outlined in the model. Providers should limit registries to patients with the six diseases indicated in the model due to the higher likelihood of predicting ED use. With a focus on a subset of the population, the practice can assign specific team members to work with the registry. The random forest importance also indicated that physicians should monitor patients with high no-show rates and body mass index this can also be achieved with a registry in the electronic health record. An automated alert and calling feature is available on most EHR; this feature permits the increases monitoring of patients who no-show with little human intervention. Clinical decision support systems parameters can be developed to encompass all the predicated variables for assisting physicians on which patients to counsel for increased risk of ED use. Otherwise, patients who do not meet these criteria should be monitored lightly freeing up human and technological resources.

Conclusions

Utilizing minimal data from the electronic medical record physicians can use existing resources to meet cost related value-based metrics. This method ensures that providers can implement the model by merely using registries in the EHR. Partnering with the insurer and utilizing our predictive model will delay a significant outlay of capital for specialized analytic technology and team members. To better understand the research question, future research will use additional machine learning techniques to improve the prediction capability of the model.

Acknowledgements

None.

References

[1] M. Gray, Value based healthcare, in, British Medical Journal Publishing Group, 2017.

[2] National Health Expenditure Projections 2013-2023, in: C.f.M.M. Services, ed., Centers for Medicare & Medicaid Services

[3] R.Y. Hsia and M. Niedzwiecki, Avoidable emergency department visits: a starting point, *International Journal for Quality in Health Care* **29** (2017), 642-645.

[4] J.G. Trogdon, L.B. Murphy, O.A. Khavjou, R. Li, C.M. Maylahn, F.K. Tangka, T.A. Nurmagambetov, D.U. Ekwueme, I. Nwaise, D.P. Chapman, and D. Orenstein, Costs of Chronic Diseases at the State Level: The Chronic Disease Cost Calculator, *Preventing chronic disease U6 - Journal Article* **12** (2015), E140.

[5] P.C. Howell, *Clinical Communication and Collaboration: Three Essays Examining the Impact of IT Interventions on At-Risk Populations Using Healthcare Analytics*, State University of New York at Buffalo, 2018.

[6] L. Breiman and P. Spector, Submodel selection and evaluation in regression. The X-random case, *International statistical review/revue internationale de Statistique* (1992), 291-319.

[7] L. Breiman, *Classification and regression trees*, Routledge, 2017.

Address for correspondence

Pamella Howel; pamellah@buffalo.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190596

Electronic Progress Note Reading Patterns: An Eye Tracking Analysis

Gretchen M. Hultman[a], Jenna L. Marquard[b], Swaminathan Kandaswamy[b], Elizabeth A. Lindemann[c], Serguei Pakhomov[a,d], Genevieve B. Melton[a,c]

[a]Institute for Health Informatics, University of Minnesota, Minneapolis, MN, USA
[b]College of Engineering, University of Massachusetts, Amherst, MA, USA
[c]Department of Surgery, University of Minnesota, Minneapolis, MN, USA
[d]College of Pharmacy, University of Minnesota, Minneapolis, MN, USA

Abstract

This study used eye-tracking to understand how the order of note sections influences the way physicians read electronic progress notes. Participants (n=7) wore an eye-tracking device while reviewing progress notes for four patient cases and then provided a verbal summary. We reviewed and analyzed verbal summaries and eye tracking recordings. Wide variation in reading behaviors existed. There was no relationship between time spent reading a section and section origin of verbal summaries.

Keywords:
Electronic health records, health communication

Introduction

The process of reviewing electronic progress notes is important for patient care but can be complex due to several factors. Practices by clinicians such as copy-paste and auto-population of structured data (e.g., medications, allergies, problem list), often result in clinical notes that are difficult to read with large amounts of redundant information. Moreover, users often use customized note templates resulting in inconsistent note formatting [1-3]. This points to a need to improve the experience of clinicians in reviewing progress notes.

Previous studies have shown that clinicians find the Assessment and Plan sections of progress notes to be the most important. However, this section traditionally appears later in the notes and clinicians need to scroll or click through lengthy supporting (often irrelevant) data to get to this section [1]. As a result, some have suggested that the top sections of progress notes should include Assessment and Plan sections [3,4], which would simplify the process of information retrieval in electronic progress notes [1,5]. Some clinicians are currently writing notes in this order.

Objective

The purpose of this study was to use eye-tracking to understand how progress note section order influences the way clinicians read electronic progress notes.

Methods

We conducted this IRB-approved study at the University of Minnesota. An Electronic Health Record (EHR) system prototype designed to mimic CPRS/VistA was populated with four de-identified patient cases [6]. Each case contained nine notes documenting visits to primary care physicians and various specialists. To test the impact of section order on note reading patterns, each patient case used a different section order within the notes. The four section orders were: SOAP (Subjective, Objective, Assessment and Plan), APSO (Assessment and Plan, Subjective, Objective), SAPO (Subjective, Assessment and Plan, Objective) and Mixed (the section orders were inconsistent across the nine notes in the case, with three notes of each order). Participants wore an eye-tracking device (Applied Science Laboratory; Bedford, MA) during the experiment and a screen capture software was used. Participants sat at a desktop computer with the EHR opened to the notes section of the first patient case. They were asked to review the notes for that patient as they normally would, and then give a brief verbal summary of the case. Participants repeated this process for the remaining three patient cases.

Visual Attention Analysis

We calculated three metrics for each patient case related to how they read the electronic progress notes: 1. the amount of time into the patient case when participants first glanced at each section (time to first fixation); 2. the total amount of time participants spent looking at each section over the course of the patient case (duration of glances); and 3. the total number of times participants looked at each section over the course of the patient case (number of glances). To determine whether there were significant differences in these measures across section orders, we used a mixed effects model to account for the repeated measures design of this experiment.

Verbal Summary Analysis

Verbal summaries were transcribed and parsed into discrete statements; an approach used in a previous study [7]. Each statement was mapped to note section(s) with that information. If a statement was found in more than one section, it mapped to all sections. A clinician with experience in qualitative analysis provided guidance and feedback on the coding process.

Results

Visual Attention

Glance Duration and Frequency

There was high variability across participants, sections, and patient cases. Within a single patient case, the number of glances at a specific section ranged from 1-86 separate glances, and the total duration of the glances at a specific section ranged from 1-356 seconds. The variability in total durations of glances

appeared to be lowest for SOAP ordered notes, and highest for the Mixed section ordering. The mixed order trials had especially high variation in numbers and durations of glances with respect to the subjective and objective sections.

Time to First Fixation

Time to first fixation (TFF) measures the first time an individual looked at a specific section in a note. We calculated two measures with regard to TFF: the actual time the participants first looked at that section, and the first time they looked at that section for longer than one second. While there was wide variability in TFF, participants seem to have glanced at the Subjective and Objective sections earlier, regardless of the note section order. It also appears that, for APSO section orders, participants were more likely to start reading the sections immediately or shortly after first looking at them.

Verbal summary Analysis

Figure 1 compares the average total duration of the glances at a specific section to the average number of verbal statements that correspond to that section. The section character count is represented by the size of the mark. Circles are drawn around data points corresponding to each information type. The subjective section does not have a circle around the data points due to the large variability in these data points. Although the Objective section had relatively higher character counts, relatively fewer verbal statements referred to information from this section. Regardless of the time spent looking at the Assessment section, relatively more verbal statements referred to information from this section. The highest variability was observed for the Subjective section, where there were inconsistent patterns related to the duration of glances and number of statements referring to information in that section.

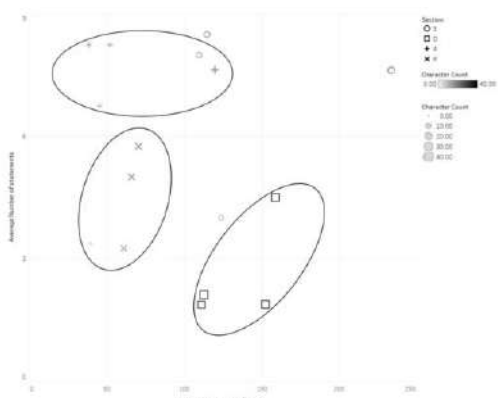


Figure 1-Verbal summary analysis

Conclusions

This study used eye tracking methodology to examine the influence of section order on clinician note reading patterns. Our study revealed several key insights about note reading patterns including high variability in reading patterns. Limitations of this study include having a small sample size and only including resident physicians. Future work should expand on the current study and examine the role of training, experience, and speciality in reading patterns. Work should also focus on identifying strategies to standardize note structure and present clinicians with the most relevant information in a way that is easy to locate.

Acknowledgements

We would like to acknowledge Osadebamwen Ighile, MBBS, MS and Oladimeji Farri MBBS, PhD for contributions to this study. This work was supported by the Agency For Healthcare research and Quality Award #R01HS022085 (GM) and National Science Foundation Award #CMMI-1150057 (JM).

References

[1] J. Shoolin, L. Ozeran, C. Hamann, and W.B. II, Association of Medical Directors of Information Systems consensus on inpatient electronic health record documentation, *Applied Clinical Informatics*. **4** (2013) 293–303.

[2] J.W. Beasley, T.B. Wetterneck, J. Temte, J.A. Lapin, P. Smith, A.J. Rivera-Rodriguez, and B.-T. Karsh, Information Chaos in Primary Care: Implications for Physician Performance and Patient Safety, *J Am Board Fam Med*. **24** (2011) 745–751.

[3] J.L. Belden, R.J. Koopman, S.J. Patil, N.J. Lowrance, G.F. Petroski, and J.B. Smith, Dynamic Electronic Health Record Note Prototype: Seeing More by Showing Less, *J Am Board Fam Med*. **30** (2017) 691–700.

[4] R.J. Koopman, L.M. Steege, J.L. Moore, M.A. Clarke, S.M. Canfield, M.S. Kim, and J.L. Belden, Physician Information Needs and Electronic Health Records (EHRs): Time to Reengineer the Clinic Note, *Journal of the American Board of Family Medicine : JABFM*. **28** (2015) 316–323.

[5] C.-T. Lin, M. McKenzie, J. Pell, and L. Caplan, Health care provider satisfaction with a new electronic progress note format: SOAP vs APSO format, *JAMA Internal Medicine*. **173** (2013) 160–162.

[6] O. Farri, K.A. Monsen, S.V. Pakhomov, D.S. Pieczkiewicz, S.M. Speedie, and G.B. Melton, Effects of time constraints on clinician–computer interaction: A study on information synthesis from EHR clinical notes, *Journal of Biomedical Informatics*. **46** (2013) 1136–1144.

[7] P.J. Brown, J.L. Marquard, B. Amster, M. Romoser, J. Friderici, S. Goff, and D. Fisher, What do physicians read (and ignore) in electronic progress notes?, *Applied Clinical Informatics*. **5** (2014) 430–444.

Address for correspondence

Genevieve B. Melton (Principal Investigator)
Institute for Health Informatics and Department of Surgery
420 Delaware Street SE, Mayo Mail Code 450
Minneapolis, MN 55455
Tel: 612-625-7992
E-mail: gmelton@umn.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190597

Realizing the Benefits of Managing Health Appointments via Mobile Application: Start of the Journey

Veronica Hung[a], Vicky Fung[a], Erica Lau[a], Stephen Lau[b], Dick Lam[b], Mavis Pow[a]

[a] *Health Informatics Section, Hospital Authority, Hong Kong Special Administrative Region*
[b] *Information Technology Department, Hospital Authority, Hong Kong Special Administrative Region*

Abstract

The Hospital Authority of Hong Kong ("HA") launched an interactive mobile application "BookHA" intending to use innovative technology to support healthcare delivery. "BookHA" provides an 'easy and convenient' alternative for patients to book specialist clinics along with the current methods – in-person or via fascmile. Patients, clinical staff, and the corporates benefit from this initiative as evidenced by the utilization figures, and the findings from patients surveys and staff surveys.

Keywords:

Ambulatory Care, Appointments, Mobile Applications.

Introduction

Hospital Authority (HA) is responsible for managing Hong Kong's public hospitals' services including 43 hospitals and institutions, 48 specialist out-patient clinics (SOPCs), and 74 general out-patient clinics. The hospitals, institutions, and clinics are organized into seven hospital clusters based on locations.

SOPC clinics serve more than 7 million attendances per year, of which around 10% of them were new cases. In 2015, a comprehensive review of the HA operation was conducted. The review recommended providing an alternative and more convenient means to empower patients to book SOPC new case appointments.

"BookHA" is the first interactive mobile application launched by HA for patients to request new SOPC appointments with smartphones. Users and patients can download the application from the Apple App Store or Google Play for free and is available in traditional Chinese and English languages, which are the official languages in Hong Kong.

"BookHA" empowered patients to manage their care process and enjoy the privileges of (i) submitting appointment requests anytime and anywhere, (ii) completing the booking request easily with three simple steps in a few minutes instead of going to the clinic in person or doing via fascimile, (iii) being notified of the appointment date via Short Message Service (SMS).

After the pilot launch of "BookHA" at Gynecology clinics in March 2016, HA rolled out the application to 10 specialties within two years (see Table 1). In parallel to the rollout, the agile development cycle of "BookHA" continued. HA upgraded "BookHA" with advanced features to enhance user experience including (i) appointments cancelation which is no longer be required, e.g., patients sought care from private doctors, (ii) "appointment reminders" that patients receive before their appointment dates, and (iii) notification if patients or providers reschedule the appointments.

Table 1 – Implementation Schedule of "BookHA" at Speicalist Out-patient Clinics

Specialtist Out-patient clinic	Rollout date
Gyneacology	March 2016
Ear, nose & throat	September 2016
Eye	September 2016
Neurosurgery	September 2016
Orthopaedics & traumatology	September 2016
Cardiothoracic surgery	March 2017
Medicine	March 2017
Surgery	March 2017
Obstetics	January 2018
Paediatric	January 2018

Mobile apps for booking health appointments are not rare. However, a large scale implementation with a usability survey in public settings with long term result would be valuable.

Methods

To identify the usability of "BookHA", we reviewed utilization figures. We conducted patients survey and staff survey to determine if the application requires any improvement.

Review on Utilization Figures

Leveraging on the utilization of "BookHA", the following figures were retrieved for analysis from the launch of "BookHA" till the end of 2017.

1. The number "BookHA" downloads
2. The number of requests for SOPC new appointment submitted via "BookHA"
3. The usage in office hour and non-office hours
4. The usage in weekdays and at weekends

Patient Survey

Complementary to the utilization figures, we designed a patient survey on the user experience with "BookHA". The application directed the persons who submitted appointment requests via "BookHA" to the survey within the app. We retrieved data from December 2016- December 2017 in this study. The survey included six questions:

1. "BookHA" makes submission of booking request easy and convenient
2. I am satisfied with my experience in using "BookHA"
3. I can complete booking request in a single attempt
4. I submitted booking request for myself/other people
5. I know about 'BookHA' from poster, pamphlet/hospital, or clinic staff/HA electronic

platform/HA referral letter/media, newspaper or magazine/others

6. Other comments

Staff survey

To prove that SOPCs staff benefit from "BookHA" in streamlining the workflow, we conducted a staff survey at two time points, i.e., June and December 2016. The survey included three questions:

1. The handling on "BookHA" request is simple
2. The instruction on handling BookHA request is clear
3. Comparing with fax booking, BookHA request is easy to handle

Results

Since the launch of BookHA in March 2016 till the end of 2017, the application had more than 178,000 downloads. Appointments were arranged for over 50,000 patients out of some 70,000 appointment requests. Patients and staff submitted around 60% of requests during non-working hours and around 20% of the requests during weekend and holidays.

Patient Survey

We collected a total of 25,500 surveys out of a total of 66,340 appointment requests via "BookHA", with a response rate of 38%, over 13 months (December 2016 – December 2017). Respondents found "BookHA" was easy and convenient, with an average score of 4.54 out of 5. In the surveys, 88% of patients were satisfied with the experience of using "BookHA". For the online appointment questions, 92% of respondents said they could complete the booking request in one go and 35% of respondents requested the SOPC new appointment for another person. Finally, comments collected via the open-ended question were overwhelmingly superb, for examples, "Good and user friendly", "Very convenient. Save lots of my time!".

Staff Survey

We distributed the survey in June and December 2016 respectively to ensure those SOPCs with BookHA rolled out were running effectively even with workflow changes. A total of 133 respondents filled out surveys and 90% of staff agreed that handling "BookHA" request was simple, the instruction on handling BookHA request was clear, BookHA request was easy to handle compared to fax booking.

Discussion

Patient empowerment in managing healthcare appointments is one of the most wanted features in mobile hospital applications [1]. The launch of "BookHA" is a win-win-win approach from patients, staff and corporate perspectives.

Patients' access to public health services is greatly enhanced as proven by the high download rate and utilization rate. The result from the patient survey supported "Simple and Easy" and "Anytime and Anywhere" features since the majority of the respondents (92%) could complete the booking requests in one go, over half of the appointments (60%) were being requested during non-office hours, and around one-third of the appointments (35%) were requested by care providers.

SOPC staff find the application appealing as proven by the staff survey findings. With "BookHA", the clinic operation was streamlined. Clinic staff could manage the booking requests by batch more efficiently. The online booking shortened the patient queue for in-person new booking in the clinics which would also reduce staff stress.

From the corporate perspective, "BookHA" is in line with the corporate strategies in optimizing clinic operations, e.g., providing an alert on duplicated appointment requests, reusing the quotas released from canceled appointments via the application. Finally, "BookHA" demonstrated the value of using new technology to engage patients and has lived out as modernized corporate version.

Our study can explore further directions for possible improvement in future studies on health appointment management via mobile applications. A possible direction is analyzing the age of users and the number of canceled appointments as well as no-show cases.

Conclusions

The adoption of mobile health apps in health care settings is encouraged with a view to improving patient experience [2]. The experience of planning, development, and the implementation of "BookHA" set a good reference for upcoming corporate patient mobile application projects. HA is now developing an integrated, personalized mobile application for patients to manage their own health records. One of the modules will provide an integrated appointment management service for patients to book, cancel, request re-scheduling. Moreover, the module will allow patients to mark 'arrival to clinic', make payment etc.

"BookHA" is the first step of HA to realize the benefits of managing health appointments via mobile application. We envisage envisaged that the benefits will further be attained in the integrated personalize mobile application platform in the future.

References

[1] 8 Types Of Mobile Hospital Apps And The 3 Features Patients Want Most - Merraine, (n.d.). https://www.merraine.com/8-types-mobile-hospital-apps-3-features-patients-want/ (accessed March 27, 2019).

[2] C.T. Lu, Y.H. Hu, J.Z. Xie, Q. Fu, I. Leigh, S. Governor, and G.P. Wang, The Use of Mobile Health Applications to Improve Patient Experience: Cross-Sectional Study in Chinese Public Hospitals, *Jmir Mhealth and Uhealth* **6** (2018).

Address for correspondence

Veronica HUNG Yuk Sai
Unit 601, 6/F, Kowloonbay International Trade & Exhbition Centre,
1 Trademart Drive,
Kowloon Bay
Hong Kong
Email: veronicahung@ha.org.hk

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190598

Technology to Assist Aging in Place: The Perspective of Health Organizations

Inga Hunter[a], Phoebe Elers[a], Caroline Lockhart[a], Dick Whiddett[a], Hans Guesgen[b], Amardeep Singh[b]

[a] *School of Management, Manawatu campus, Massey University, Palmerston North, New Zealand*
[b] *School of Engineering and Advanced Technology, Manawatu campus, Massey University, Palmerston North, New Zealand*

Abstract

This paper presents findings from a series of focus groups which is exploring the implications of, and stakeholder requirements for, integrating social media technologies and 'smart home' technologies to connect older adults with their formal support networks (i.e. to healthcare and social service providers) thus enabling them to live independently at home.

Keywords:
Technology, Delivery of Health Care, Social Networking

Introduction

Like other developed nations, life expectancy in New Zealand is increasing, which is resulting in an aging population. As older adults have higher rates of chronic conditions and disabilities that require regular support [1], this can place increased demand on health services. An option favoured by policy-makers and older adults [2] is to support individuals to remain in their own homes for longer and avoid residential care, known as 'aging in place'. Research has shown how aging in place can be assisted by smart home technologies, such as purpose built smart homes and home monitoring devices.

However, smart home technologies are underutilized [3], and there are challenges which these technologies frequently fail to address. In addition to barriers identified by technology acceptance models, such as the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT) [3], cost, psychosocial factors related to aging [3,4] and the needs of their informal carers who provide 40-90% of the care for older adults need to be addressed [5].

The present study is part of an exploratory project which takes a socio-technical system approach [5], which considers the adoption of new technologies to involve interaction between society's complex infrastructures and human behaviour.

This paper presents the findings from the health provider focus group phase of this project, to understand their perspective on the use of technology to help them provide services to older adults.

The findings, together with those from the older adults and their informal carer networks [6], identified barriers and facilitators for technology to support aging in place, which informed the development of a prototype system connecting older adults to both their formal (health provider) support and informal support networks tested in older adults' homes and currently under evaluation.

Methods

Seven face-to-face focus groups, comprising 44 participants, were undertaken between May and August 2017. Participants were recruited through convenience sampling through health organizations providing support for older adults in the Manawatū region of New Zealand. Four focus groups were held at the local hospital with nurses, social workers, physiotherapist and occupational therapists who worked with older adults in the community. Three focus groups were held at three general health practices, each involved a range of participants including doctors and nurses. Interested participants chose to attend and participate. They did not receive financial incentives for participation. Each focus group lasted approximately one hour and began with an introduction about the project and the use of home monitoring and digital technologies. Questions (at a macro level) covered their information needs, what information should be collected and transferred for what purpose, who should receive information, and ethical concerns. Participants were predominantly female, and their average age was 46.6 years. Study procedures were approved by the Massey University Human Ethics Committee (SAO 16/65).

The focus groups were transcribed and thematically analyzed inductively in NVivo Version 11.0. Due to the nature of focus groups, individual participants were not identified. Coding was an iterative process that involved rereading datasets to establish initial codes covering key ideas. Similar codes were combined into themes, which were reviewed alongside the original dataset. Codes and themes were discussed within the research team until saturation, stability, and logical organisation were attained and all data was accounted for.

Results

Two main themes emerged, relating to information collection and barriers.

Information Collection

Participants wanted more information about their patients, and were enthusiastic about technologies retrieving information, providing that informed consent is obtained. This theme encompasses three subthemes: information systems' disconnect, participants' information needs, and what information should be collected through technologies. A reoccurring issue identified was a disconnect between current information systems within and across organizations and the need for any new systems to be properly integrated with existing systems.

Participants' information needs varied. Compared to participants from general practices, community practitioners wanted more information about patients' living conditions and their social and physical wellbeing, such as whether they are socializing and leaving the house, their diet, home temperature and physical activity. However, the need to obtain informed consent was emphasized.

Barriers

Participants were optimistic about technology retrieving information but were clear about potential uptake barriers. This theme encompasses three subthemes: usability, cost, and privacy and data security.

Participants were concerned about software complexity and usability issues associated with aging, such as dexterity and cognitive decline. Many participants indicated that the training required for older adults would be substantial and that their desire to learn to use technology would be mixed.

Participants considered cost to be another barrier and discussed how it is already a barrier for their uptake of health products.

Ensuring privacy and data security was also considered a requirement to protect patients. Some participants felt that older adults would have concerns about privacy and data security, although this barrier did not present as strongly compared to usability and cost.

Discussion

The key findings identify several user-requirements as outlined in Table 1, below.

Table 1– User Requiremnts

Theme	Requirement
Information Collection	The technology should save users' time. Informed consent is required for collecting, accessing and sharing information.
Barriers	The interface should be stable and easy to navigate, with large text. The technology should be adaptable to meet older adults' changing needs. The technology needs to be low cost, and, when possible, utilize objects that already present within the home. The information needs to be stored securely and this needs to be clearly communicated to users. The technology should save users' time. Information should integrate with existing systems used by health providers

Some issues addressed by these requirements correspond to barriers already identified in health informatics scholarship. For example, cost is an issue for many older adults [3], and one way to address this is to utilize technologies that already exist within the home. Usability is another barrier [3,4], and so the technology should be designed for ease-of-use. In fact, a direct user-interface may not be suitable for some older adults, instead the technology could monitor the environment so that an older adult does not actively engage with the system.

This study has some limitations. Participant recruitment was not drawn from a random sample of subjects but was from organizations that volunteered to be involved in the study. This resulted in participants only being employed at general health practices and hospital departments, although they represented a range of professionals, including general practitioners, physiotherapists, social workers, and nurses.

Conclusions

The 44 health provider participants working in aged care want more information about their patients' wellbeing within their homes, demonstrating a potential for home monitoring and ICTs to connect older adults to their formal support networks. This would not only assist health practitioners, but it would also support aging in place, which is socially and economically beneficial. While the findings should not be generalized beyond the organizations involved in this study, the user requirements identified, regarding consent, cost, information access, data security and usability, have informed the development of a system that is currently being prototyped in older adults' homes.

Acknowledgements

Funded by Health Research Council, New Zealand, ref 16/679 and School of Management, Massey University.

We would like to thank our participants.

References

[1] Ministry of Health. Healthy ageing strategy. https://www.health.govt.nz/system/files/documents/publications/healthy-ageing-strategy_june_2017.pdf [accessed April 9, 2019]

[2] J.L. Wiles, A. Leibing, N. Guberman, J. Reeve, and R.E. Allen, The meaning of "aging in place" to older people, *Gerontologist* **52** (2012), 357-366.

[3] S.T. Peek, E.J. Wouters, J. van Hoof, K.G. Luijkx, H.R. Boeije, and H.J. Vrijhoef, Factors influencing acceptance of technology for aging in place: a systematic review, *Int J Med Inform* **83** (2014), 235-248.

[4] S. Yusif, J. Soar, and A. Hafeez-Baig, Older people, assistive technologies, and the barriers to adoption: A systematic review, *Int J Med Inform* **94** (2016), 112-116.

[5] S. Carretero, J. Stewart, C. Centeno, F. Barbabella, A. Schmidt A, F. Lamontagne-Godwin, G. Lamura. Can technology-based services support long-term care challenges in home care? http://ftp.jrc.es/EURdoc/JRC77709.pdf [accessed April 9, 2019]

[6] E. Coiera, Putting the technical back into socio-technical systems research, *Int J Med Inform* **76 Suppl 1** (2007), S98-103.

Address for Correspondence

Dr Inga Hunter, email: i.hunter@massey.ac.nz

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190599

FAIR Principles for Clinical Practice Guidelines in a Learning Health System

Tiffany I. Leung[a], Michel Dumontier[b]

[a] *Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands*
[b] *Institute of Data Science, Maastricht University, Maastricht, The Netherlands*

Abstract

The learning health system depends on a cycle of evidence generation, translation to practice, and continuous practice-based data collection. Clinical practice guidelines (CPGs) represent medical evidence, translated into recommendations on appropriate clinical care. The FAIR guiding principles offer a framework for publishing the extensive knowledge work of CPGs and their resources. In this narrative literature review, we propose that FAIR CPGs would lead to more efficient production and dissemination of CPG knowledge to practice.

Keywords:

Guidelines as Topic; Delivery of Health Care; Medical Informatics Applications.

Introduction

The *learning health system* (LHS) is a concept for an integrated social and technological system that seamlessly embeds continuous improvement and innovation for the effective delivery of healthcare [1,2]. Key parts of the LHS are collecting, integrating, and analyzing data from different sources. Evidence generated should be rapidly translated to practice, with additional data collected from practice to enable further evidence generation or modification. Ideally, a clinician could draw quickly and precisely from such knowledge to match patients to relevant evidence and data generated. Since the 1990s, encoding CPG knowledge into computer-interpretable guidelines (CIGs) has been one approach to implementing CPGs in practice [3], although this has remained a slow, difficult-to-scale process. Reusable software components will enable faster, cheaper protocol development.

The 2011 Institute of Medicine report on Digital Infrastructure for the Learning Health System envisions infrastructure that enables a variety of health data sources to improve population health. The 2010 President's Council of Advisors on Science and Technology report tasked U.S. government agencies to "establish initial minimum standards for the metadata associated with tagged data elements" towards realizing the potential of health information technology. More recently, a 2018 commentary highlighted the need for sharable digital knowledge objects, representing syntheses, systematic reviews, and scientific data from original studies [4] in an LHS, yet did not describe the principles needed to achieve this goal.

The *FAIR guiding principles* [5] offer an opportunity to address such thorny issues. Initially conceived to address the fundamental challenges associated with finding and reusing research data, the FAIR guiding principles outline 15 points to enhance the *F*indability, *A*ccessibility, *I*nteroperability, and *R*eusability of digital resources [5,6], from data to analytical pipelines. These recommendations include: unique and persistent identifiers for all entities (from documents to individuals, from concepts to medications); machine understandable data and metadata to facilitate structured search and query; the development and use of pragmatic community standards that reduce effort in reuse; use of shared vocabularies to enable aggregation and enhance data mining; improved provenance to support reproducibility; social and technological commitments to realize long term preservation of digital resources; and simpler terms of use to elucidate expectations and intensify innovation. These principles offer a sensible basis for the design of LHS digital infrastructure, towards meeting the data and information needs of a variety of healthcare stakeholders. This paper describes the FAIR guiding principles and their role in LHS digital infrastructure, using CPGs as a use case.

Methods

Citing key literature, we discuss the role of FAIR in reinventing CPGs, and how this would further enable the LHS digital infrastructure. Included literature was identified based on the authors' shared expertise, supplemented with focused literature searches using MEDLINE and Web of Science.

Discussion

Table 1 provides a detailed checklist for making CPGs FAIR. Fulfillment of this checklist can be verified using tools for performing FAIRness assessment [7].

FAIR principles can facilitate the management and stewardship of CPG knowledge.

The FAIRness of digital objects can be assessed, enabling accountability and transparency of data quality on which CPGs are based. Digital resources for clinical knowledge and healthcare systems include CPGs, as well as other forms of evidence and data generatio, such as EHRs, predictive algorithms, etc. In the 6S Hierarchy of Evidence Based Resources model of medical knowledge and information, CPGs are positioned second from the top [8]. At the model's base are original single studies, such as randomized clinical trials (RCTs), which often have highly selective inclusion criteria; this potentially marginalizes subpopulations who are excluded from such trials (e.g. patients with multimorbidity or cognitive disorders, women and minorities, etc.). As a result, a CPG recommendation may be inappropriately applied to a specific patient, for whom the original RCT results did not apply.

The FAIR principles also aim to address issues in data discovery and reuse. For example, consider a simple case where a clinician wishes to recommend evidence-based therapies for a Chinese American woman with diabetes. With findable and accessible data on patient characteristics, pharmacologic treatments, and clinical outcomes in single studies, systematic reviews, and original studies (from the 6S model), these data

can be readily reused in personalizing care for *this* patient – a desirable goal in building an LHS. Moreover, FAIR principles would empower further research involving social determinants of health and vulnerable populations.

The FAIR principles strengthen the connection between CPGs and the best evidence that support them. Multiple relevant CPGs may apply for single conditions, and vary in the quality and strength of their supporting evidence [9]. Inconsistencies between CPGs for the same condition and conflicting recommendations between different conditions can occur. A single guideline can also contain an overwhelming amount of evidence; one guideline update on ischemic stroke management included 217 recommendations and 421 references [10].

Table 1 – Checklist for developing FAIR CPGs

- CPGs are identified with a unique, persistent identifier, such as a Digital Object Identifier (DOI), which can be resolved to the latest version of a CPG.
- CPGs are versioned with an identifier generated through systematic assignment or from CPG content or its metadata, using identifier systems (e.g. TrustyURIs, DataGUIDs).
- CPGs are described with rich, standardized metadata. The NGC had 54 guideline attributes, but the list must be further discussed and adopted by the community.
- CPGs must be represented in a machine processable language that enables automatic document validation.
- CPGs and their metadata must each be released with clear, standard licenses so users understand and agree to the rights and responsibilities expected of them.
- CPGs and their metadata must include detailed provenance regarding how and when they came to be including details of authors, contributors, methods and resources (e.g. 6S components) used to generate them, and the publishers that make them available.
- CPGs should include standard identifiers for key components (e.g. drugs or diseases should use specific terminologies, such as RxNorm or ICD).
- CPG metadata should always be available, even if CPGs are behind a paywall.
- CPGs and their metadata should be searchable and accessible through a variety of human and machine interfaces.

New research that adhere to FAIR principles can more rapidly be identified and incorporated into the CPG literature review process. Each citation, including original and processed study data, would be FAIR and serve as evidence in CPG development and be accessible for the wider community. This can also ensure the inclusion and findability of important but potentially neglected information, such as conflict of interest information for the experts on CPG development committees.

FAIR principles can enable redesign of guideline repositories to facility CPG implementation.

The FAIR principles can guide the construction of a next-generation guideline repository. The National Guideline Clearinghouse (NGC), formerly a publicly accessible online resource for CPGs, performed quality assessments of published guidelines, summarized high-quality guidelines, and curated and annotated key information into guideline summaries [11,12]; each summary was structured using 54 guideline attributes, including tags with UMLS Metathesaurus concepts [13]. The NGC had performed an important public service in curating and annotating CPGs, although the process was resource-intensive. The sunset of NGC in July 2018 signals an opportunity to redesign guideline repository infrastructure in a manner that adheres to the FAIR principles, thereby enabling easier access to CPG knowledge and their digital objects needed for implementation in an LHS.

FAIR CPGs can also pave a way towards their implementation at a systems level. While almost all CPGs publish recommendations in narrative text, some also produce flow diagrams, schemata, and care pathways, which serve as visual summarizations of CPG recommendations. These can be outlines for translating CPG evidence into CIG formats and clinical decision support. Producing these supplementary CPG materials as FAIR digital objects would facilitate their rapid integration and updates into an LHS and clinical care.

Conclusions

In the aftermath of the NGC repository, CPGs are now further away from FAIR than before. We described key features of FAIR and detailed opportunities for next-generation CPG development, which are the next steps for further investigation.

References

[1] L.M. Etheredge, Rapid learning: a breakthrough agenda, *Health Aff.* **33** (2014) 1155–1162.
[2] C.P. Friedman, A.K. Wong, and D. Blumenthal, Achieving a Nationwide Learning Health System, *Sci. Transl. Med.* **2** (2010) 57cm29.
[3] M. Peleg, Computer-interpretable clinical guidelines: A methodological review, *J. Biomed. Inform.* **46** (2013) 744–63.
[4] J.-M. Guise, L.A. Savitz, and C.P. Friedman, Mind the Gap: Putting Evidence into Practice in the Era of Learning Health Systems, *J. Gen. Intern. Med.* (2018).
[5] M.D. Wilkinson, M. Dumontier, et al, The FAIR Guiding Principles for scientific data management and stewardship, *Sci. Data*. **3** (2016) 160018.
[6] B. Mons, C. Neylon, J. Velterop, M. Dumontier, L.O.B. Da Silva Santos, and M.D. Wilkinson, Cloudy, increasingly FAIR; Revisiting the FAIR Data guiding principles for the European Open Science Cloud, *Inf. Serv. Use*. **37** (2017) 49–56.
[7] M.D. Wilkinson, S.-A. Sansone, E. Schultes, et al, A design framework and exemplar metrics for FAIRness, *Sci. Data*. **5** (2018) 180118.
[8] A. DiCenso, L. Bayley, and R.B. Haynes, ACP Journal Club. Editorial: Accessing preappraised evidence: fine-tuning the 5S model into a 6S model, *Ann. Intern. Med.* **151** (2009)
[9] K.D. Wyatt, L.M. Stuart, et al, Out of context: clinical practice guidelines and patients with multiple chronic conditions: a systematic review, *Med. Care*. **52** (2014) S92–100.
[10] L.S. Sulmasy, A.M. López, C.A. Horwitch, American College of Physicians Ethics, Professionalism and Human Rights Committee, Ethical Implications of the Electronic Health Record: In the Service of the Patient, *J. Gen. Intern. Med.* **32** (2017) 935–939.
[11] P.G. Shekelle, Clinical Practice Guidelines: What's Next?, *JAMA*. **320** (2018) 757-75.
[12] Z. Munn and A. Qaseem, Disappearance of the National Guideline Clearinghouse: A Huge Loss for Evidence-Based Health Care, *Ann. Intern. Med.* **169** (2018) 648-649.
[13] T.I. Leung, H. Jalal, D.M. Zulman, et al, Automating Identification of Multiple Chronic Conditions in Clinical Practice Guidelines, *AMIA Jt Summits Transl Sci Proc*. 2015 (2015) 456–460.

Address for correspondence

Tiffany I. Leung, MD, MPH. t.leung@maastrichtuniversity.nl

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190600

Evaluation of the Fast Healthcare Interoperability Resources (FHIR) Standard for Representation of Knowledge Bases Encoded in the Arden Syntax

Robert A Jenders

Charles Drew University & University of California, Los Angeles, California, USA

Abstract

Context: Arden Syntax is a standard that encodes knowledge as Medical Logic Modules (MLMs) but lacks a standard query data model and terminology. Objective: Assess to what extent FHIR can represent MLMs. Method: A convenience sample of 340 MLMs were examined and tabulated for representation using the PlanDefinition resource. Result: While alignment between structured subsections is uneven, the PlanDefinition resource adequately represents procedural medical knowledge encoded as MLMs Conclusion: FHIR adequately represents Arden Syntax MLMs.

Keywords:
Clinical decision support systems, knowledge representation.

Introduction

Arden Syntax is a formalism supervised by Health Level Seven International (HL7) for representation of procedural medical knowledge to facilitate sharing units of knowledge known as MLMs that are executed in clinical decision support systems (CDSS) [1]. Each MLM consists of four categories - Maintenance, Library, Knowledge and Resources - each of which consists of one or more structured attribute value pairs that contain descriptions of the knowledge and executable statements. Key to interoperability of knowledge representation for CDS is the use of standard formalisms for knowledge representation itself as well as for the query syntax and data elements retrieved from local data stores in order to execute units of CDS knowledge. One increasingly popular formalism for facilitating such interoperability is the HL7 Fast Healthcare Interoperability Resouces (FHIR). FHIR consists of a library of resources, each of which is a standardized structured representation of a collection of closely related data elements [2]. Prior work has established the utility of standards, such as FHIR, to serve as the standard data model for queries within the Arden Syntax, thereby reducing barriers to sharing MLMs and enhancing their interoperability by decreasing the need for local customization when MLMs are moved from one health care organization to another [1].

Beyond standardizing just references to clinical data, a key FHIR resource that offers specific utility in the CDS space is PlanDefinition. A plan definition is a pre-defined group of actions to be taken in particular circumstances, often including conditional elements, options, and other decision points. The resource is flexible enough to be used to represent a variety of workflows, as well as clinical decision support and quality improvement assets, including order sets, protocols, and decision support rules [3]. The present work was undertaken to expand the prior work of standardizing just the queries of MLMs to assess the utility of using FHIR to represent and thereby store, transmit and reuse clinical knowledge encoded in the Arden Syntax in order to better integrate this latter standard within the increasingly popular FHIR framework.

Methods

A previously assembled convenience sample of MLMs was examined. MLMs were examined and tabulated to assess their alignment with and the extent to which they could be represented by FHIR Release 3, version 3.0.1 [3]. This extraction and tabulation process aligned with a similar effort to identify the utility of FHIR to represent the data variables inside non-query data mappings in MLMs [4]. In the present work, however, this was applied toward representation of the standard parts of the MLM outside of the curly braces that encode the details of queries and other mappings and are meant to be institution-specific. In particular, the text of the MLM corpus was concatenated in a single file. Using a text editor guided by Backus-Naur form of Arden Syntax v2.10 [5] and the format of the various kinds of statements, regular expressions were created in order to extract the instances of each kind of statement. Each of these batches of statements then were compared to the corresponding representation in the FHIR PlanDefinition resource by manual inspection. In particular, alignment between the subparts and attributes of the PlanDefinnition resource and the MLM structure were examined and issues related to the encoding of knowledge in the two standards were reviewed.

Results

A total of 340 MLMs were pooled from 5 source CDS systems, including 24 from 2 vendor knowledge bases and 316 from 3 academic medical centers. MLMs concerned with lab tests were the most common (138/340 = 41%), followed by clinical assessment (75 = 22%) and medication (45 = 13%). The remainder addressed administrative and miscellaneous topics. FHIR contains a rich collection of meta-knowledge attributes, so representation of Maintenance and Library category attributes such as author, title, purpose and so on in FHIR is adequate. While FHIR could represent all MLMs in the corpus, the challenge was the level of specificity. PlanDefinition has attributes for condition (or trigger), expression (for Boolean logic) and action (for defnining the result of knowledge execution) but does not support representation of some specific operators (e.g., substring extraction) or some other elements of Arden Syntax knowledge expressions such as fuzzy logic. On the other hand, all MLMs in the corpus contain at least one instance of temporal reasoning, and temporal attributes of the

PlanDefinition resource, including "timingAge" and "timingPeriod" could represent these Arden Syntax constructs.

Consequently, while all the MLMs could be represented in shells defined by PlanDefinition, the actual knowledge expressions and their components in at least some instances were not directly representable by specific FHIR constructs.

While the analysis highlighted constructs in the Arden Syntax that lacked explicit counterparts in the PlanDefinition resource, it also identified constructs in the latter that could be used to represent elements in MLMs that are represented only in narrative format currently. These include the attributes "code" and "reason" of an "action," which are codeable concepts representing the meaning of an action and its justification, providing additional structure that is lacking in the latest version of the Arden Syntax.

Discussion

Despite agreement that a common formalism for representation of computable clinical knowledge would be beneficial in fostering the deployment of CDS systems and reducing the cost of knowledge engineering therein, no widespread agreement has been reached on a specific formalism for doing so [6]. Workers have demonstrated the utility of the Arden Syntax in this space, and this solution for knowledge representation has been adopted by several software vendors, with resulting implementations in a number of countries. Nevertheless, lack of a standard data model and the resulting need to translate references to local data stores at each implementation site has hindered uptake, as this has with other formalisms [7]. Further, with an emphasis on the use of CDS for health care quality improvement and the consequent need to measure quality in reproducible ways, some workers have concluded that Arden Syntax is inadequate for this purpose. This is turn has led to the development of other standards such as the HL7 Clinical Quality Language, and additional attempts to synthesize an acceptable data model [6]. Paralleling these efforts, other work has led to formalisms, such as CDS Hooks, for accessing CDS systems and their underlying knowledge in ways that are agnostic regarding the specifics of the exact knowledge representation formalism [6]. Additional work has demonstrated the need for improved explicit representation of business processes within CDS knowledge bases, and meeting this need will introduce further complexity and utility in CDS standards. With the increasing popularity of FHIR as a data and information exchange formalism, the present work and related prior studies demonstrate that Arden Syntax complements this new milieu of standards through mechanisms for knowledge access and data representation, improving its utility for its original goals of standardizing knowledge representation and reducing its cost.

Conclusions

FHIR is adequate for encoding Arden Syntax MLMs. This would facilitate interoperability of MLMs as well as their integration in the FHIR framework. The FHIR PlanDefinition lacks specific constructs to support the details of knowledge expressions, so while it is a useful framework for storage, transmission and manipulation of knowledge artifacts, a specific expression lanaguage or standard beyond FHIR is required to use the knowledge artifacts within a CDS system. However, some features of the FHIR PlanDefinition provide additional structure for knowledge in Arden Syntax MLMs that currently may be represented only in narrative format.

Acknowledgements

This work was supported by NIMHD grants U54MD007598 and S21MD000103 and NCATS grant UL1TR001881 from the National Institutes of Health (USA).

References

[1] R.A. Jenders, K.-P. Adlassnig, K. Fehre, and P. Haug, Evolution of the Arden Syntax: key technical issues from the standards development organization perspective, *Artif Intell Med* **92** (2018), 10-14.

[2] J. Aerts, Towards a single data exchange standard for use in healthcare and in clinical research, *Stud Health Technol Inform* **248** (2018), 55-63.

[3] Health Level Seven International, *Fast Healthcare Interoperability Resources Release 3*, Health Level Seven International, Ann Arbor, MI, 2017.

[4] R.A. Jenders, Utility of the Fast Healthcare Interoperability Resources (FHIR) standard for representation of non-query data mappings in the Arden Syntax, *AMIA Annu Symp Proc* (2018), 1449.

[5] Health Level Seven International, *Arden Syntax for Medical Logic Systems Version 2.10*, Health Level Seven International, Ann Arbor, MI, 2014.

[6] K.-P. Adlassnig, P. Haug, and R.A. Jenders, Arden Syntax: Then, now, and in the future, *Artif Intell Med* **92** (2018), 1-6.

[7] R.A. Jenders, W. Sujansky, C.A. Broverman, and M. Chadwick, Towards improved knowledge sharing: assessment of the HL7 reference information model to support medical logic module queries, *AMIA Annu Symp Proc* (1997), 308–312.

Address for correspondence

Robert A. Jenders, MD, MS, FACP, FACMI. Center for Biomedical Informatics, Charles Drew University. Department of Medicine, University of California, Los Angeles. 1748 E 118th St, Los Angeles, CA 90059 USA. jenders@ucla.edu.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190601

Development of ICT-Based Comprehensive Health and Social-Needs Assessment System to Enhance Person-Centered Community Care

Eun Jeong Choi[a], Eunjung Lim[b], Minjung Kwak[a], Miri Jeong[a], Nayoung Lee[b], Il Bum Kwon[c], Wonpyo Lee[d], Hanwool Ku[e], Dongil Kim[e], Haesung Nam[f], Junsik Na[c], Myonghwa Park[a]

[a] College of Nursing, Chungnam, National University, Daejeon, Republic of Korea,
[b] Research Institute of Nursing Science Chungnam, National University, Daejeon, Republic of Korea,
[c] Mindle Health & Care Social Coop, Daejeon, Republic of Korea,
[d] Social Economy Institute, Daejeon, Republic of Korea,
[e] Department of Computer Science and Engineering, Chungnam National University, Daejeon, Republic of Korea,
[f] Department of Preventive Medicine, School of Medicine, Chungnam National University, Daejeon, Republic of Korea

Abstract

We developed a comprehensive health and social-need assessment system to evaluate the diverse needs of elders with chronic illnesses in the community and to enhance the connection of their needs to health and social services. A comprehensive needs-assessment tool and profiles were integrated into the ICT system. We found that care managers could assess elders' needs comprehensively and connect those needs to suitable health and social services systematically.

Keywords:
community health services, needs assessment, information technology

Introduction

Community health problems are increasing due to the aging of communities and chronic diseases. However, the current healthcare and social services system provides a disease-centered healthcare service and does not consider the comprehensive characteristics and social factors of individuals [1]. This format can risk providing redundant or deficient services that are not well suited to the individual [2]. Another problem is that services are provider-oriented rather than user-oriented, so services cannot meet individual need sensitively [3]. Therefore, in this study, we developed a comprehensive health and social-service need-assessment system based on ICT. The comprehensive health and social-needs assessment system not only enables evaluators to assess the comprehensive needs of elders with chronic illness and share the summarized profile with infographics but also enables them to logically connect individual needs to health and social services.

Methods

The assessment tool for comprehensive health and social-needs assessment system was developed based on the International Classification of Functioning, Disability and Health (ICF) [4], community case management, an elderly situation survey, a community-needs survey, a community health survey, and the WHO Disability Assessment Schedule. The tool was validated by the expert Delphi method. We visualized the assessment tools and evaluation result profiles with ICT, represented by various icons and pictures. We tried to reflect target users' opinions after pilot test using the system.

Results

The comprehensive health and social-needs assessment system tool and profiles with user-friendly design were integrated in the ICT system. The structured assessment tools of the comprehensive health and social-needs assessment system consisted of three areas: initial assessment; daily and social-life activity checklist; comprehensive assessment including physical health assessment, cognition, mobility, self-care, getting along with people, life activities, participation in society assessment, and environmental assessment. This tool is implemented using ICT, so targeted users can mark the answers directly on the Web, and the results appear with familiar images and graphs to easily recognize and store Web data for users. We applied a pilot test to the system to see users' experiences and user interfaces. As a result of the pilot test, first, time was much shorter than when using paper-based assessment forms. Second, need-assessment results were presented as an infographic rather than simply in a narrative format, so they could be easily understood. Third, the system is very useful for storing data so it is possible to compare present data with past data. Fourth, users could share decision making during the evaluation process and connect to appropriate health and social services.

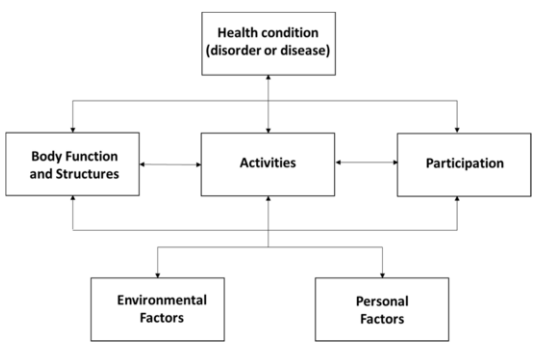


Figure 1– The Model of the International Classification of Functioning, Disability and Health (ICF) [4]

Conclusions

In this study, we developed a comprehensive health and social-need assessment system based on ICT for various needs-assessments of elders with chronic illnesses. An infographic assessment tool and result profiles, based on ICT, helped users to understand and provide information easily and logically. In future studies, we will continue to upgrade the system by applying more cases to the system. Based on the accumulated data, we will conduct further research to evaluate elders with chronic illnesses through a decision support system and present the results sensitively.

Acknowledgements

This work was supported by Community Business Development Program through the Korea Institute for Advanced of Technology (KIAT) funded by the Ministry of Trade, Industry, and Energy (Project No. P0004335).

References

[1] F. H. Barrow, The International Classification of Functioning, Disability, and Health (ICF), a new tool for social workers. *Journal of Social Work in Disability & Rehabilitation,* **5** (2006), 65-73.

[2] B. Dernek, S. Esmaeilzadeh, A. Oral, The utility of the International Classification of Functioning, Disability and Health checklist for evaluating disability in a community-dwelling geriatric population sample. *International Journal of Rehabilitation Research*, **38**(2015), 144-155.

[3] S. Spoorenberg, et al. Development of the geriatrics, an ICF-based and person-centred assessment tool for evaluation of health-related problems in community-living older adults, *International Journal of Integrated Care*, 2018, **18**(2018), s2.

[4] World Health Organization, *International Classification of Functioning, Disability and Health*, World Health Organization, Geneva, Switzerland, 2001.

Address for correspondence

Eun Jeong Choi

College of Nursing, Chungnam National University

266 Munhwa-ro, Jung-gu, College of Nursing, Chungnam National University, Daejeon

35015, Republic of Korea

Telephone number: +82-42-580-8407

Fax number; +82-42-580-8309

E-mail address: non337577@hanmail.net

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190602

Introduction of a Pathophysiology-Based Diagnostic Decision Support System and Its Potential Impact on the Use of AI in Healthcare

Sandra Kühnel, Milan Jovanović, Henry Hoffmann, Lennert Schneider, Sabrina Golde, Martin C. Hirsch

Ada Health GmbH, Adalbertstraße 20, 10997, Berlin, Germany

Abstract

Many currently available Diagnostic Decision Support Systems (DDSS) are based on causal condition-symptom relations that exhibit certain shortcomings. Ada's new approach explores the capabilities of DDSS based on pathophysiology, describing a disease as a dynamically evolving process. We generated a pathophysiology model for 8 conditions and 68 findings suitable to assess this approach. Preliminary results meet our expectations while leaving space for further improvement.

Keywords:

Decision Support Systems, Clinical; Artificial Intelligence; Population Health

Introduction

Artificial Intelligence (AI) in Medicine, first suggested by Ledley and Lusted in 1959 [1], aimed to improve diagnostic performance, treatment and clinical decision making. Today, these AI supported systems are typically referred to as *Clinical Decision Support Systems* (CDSS). [2] Many current CDSS that are assisting clinical diagnosis, more commonly referred to as *Diagnostic Decision Support Systems* (DDSS), are based on a causal condition-symptom relations. [3] This approach assumes the independence between symptoms and conditions. [4] Consequentially, symptoms that are part of the same pathophysiological pathway add redundant information. These so-called *additive effects* can reduce diagnostic precision of DDSS. Furthermore, DDSS based on causal condition-symptom relations do not address a disease as a dynamically evolving process.

Ada's Deep Reasoning (ADR), a new approach on DDSS development, extends current DDSS capabilities by utilizing pathophysiology to describe the dynamic components of disease pathogenesis.

Methods

The core of ADR is a graphical model based on concepts and relations. Concepts used to describe pathophysiological pathways of the model are either a pathological state (*patho-state*) or a pathological process (*patho-process*) (Fig. 1). Patho-states are defined by an *anatomical location* within the human body and an *attribute* describing the underlying pathology. The descriptions of anatomical locations are based on the Foundational Model of Anatomy (FMA). [5] The attributes represent different ordinally scaled dimensions such as structural integrity, functional integrity, or physical and biochemical measures. For example, myocardial ischemia as a patho-state can be described in the following way: the attribute "partial pressure of oxygen" in the state "reduced" within the location "myocardial tissue" (FMA ID: 14068). Furthermore, patho-processes describe the potential transition from one patho-state to another (Fig. 1). They are characterized by the *probability* of a process taking place and its *average duration*. In addition to patho-states and patho-processes, the model contains nodes that represent *conditions* and *findings* (Fig. 1), concepts commonly used in medicine. These are mapped to the underlying pathophysiology by *definition relations* (Fig. 1).

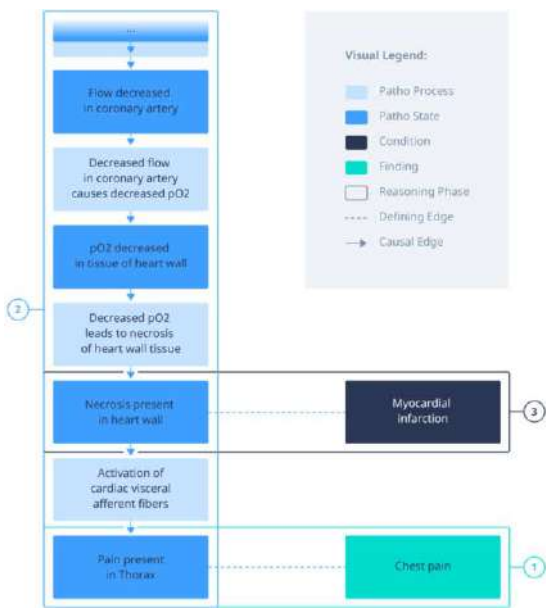


Figure 1 – Excerpt from Ada's Deep Reasoning graphical model for the finding chest pain in myocardial infarction

The pathophysiology model (*patho-model*) contains carefully chosen overall level of detail. It is able to describe the interaction between the components of the human body relevant for disease, while avoiding a granular level used for cell simulation. To create this model within a distributed team of doctors, we created a dedicated editor tool. Finally, we developed a model in order to test whether our approach

actively addresses the previously observed shortcomings of DDSS based on causal condition-symptom relations.

The main purpose of the described patho-model is to enable the inference of the underlying pathophysiological causes from observed findings. The current reasoning approach falls into three phases (Fig. 1). In the first phase, the evidence on different findings is projected through definition relations to the corresponding patho-state constellations. In the second phase, the evidence from these patho-states is propagated through the patho-processes using a probabilistic-deductive approach combining concept form logic and Bayesian inference. In the last phase, the posterior probabilities of all patho-states are finally projected through the definition relations to the condition concept layer.

An assessment of accuracy is performed for the total number of conditions modelled (*global score*). The model is tested against 71 condition-specific reference cases. These reference cases were previously used for the validation of other Ada engines. [6]

Results

The medical knowledge space for testing ADR consists of 8 conditions and 68 findings. The above mentioned assessment on accuracy, matching quality (MQ), was performed for all modelled conditions. The MQ results for ADR were calculated for the top suggestion (M1), the top 3 (M3) and the top 10 (M10) suggestions. The global MQ score for ADR was 0.70 for M1. The global MQ score for M3 and M10 were 0.95 and 0.96, respectively. Results are prior corrected.

Discussion

ADR explores the possibilities of a pathophysiology-driven DDSS. After successfully creating a proof-of-concept engine during the first phase of the ADR development, the current focus lies on reasoning refinement.

ADR introduces a level of detail that is missing in currently available DDSS. We observed that additive effects could be successfully addressed by explicit modelling of pathophysiological pathways that lead to observable symptoms. Additionally, the created editing tool supports the scaling of our existing medical knowledge space. It allows the inclusion of many medical experts to work together on a pathophysiological model for humans.

At this stage, the preliminary results are promising and meet our expectations. ADR is currently not addressing risk factors, e.g. sex, age, or BMI. Additionally, the patho-model neglects information that closely describe findings, e.g., quality, intensity, or position of abdominal pain. Based on our experience with other DDSS built in Ada, we expect these features to increase the M1 score by an additional 8-15%. Finally, our medical knowledge space needs to be extended to allow the testing of new features and support diagnoses of a wider range of conditions.

Conclusions

With ADR, Ada Health successfully demonstrates the feasibility of a pathophysiology-based DDSS. Through this innovative approach, ADR overcomes additive effects problem and considers more information on the underlying biology. This way, a broad spectrum of health-related data can be depicted and an individualized, high resolution model for personal healthcare can be created. Facilitating the integration of this kind of health-related data will significantly advance personalised medicine.

Acknowledgements

This work was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 674459 (SME Instrument).

References

[1] R.S. Ledley and L.B. Lusted, Reasoning Foundations of Medical Diagnosis, Science (1959) Jul 3;130(3366).

[2] J.C. Kunz et. al., Computer-assisted decision making in medicine, J Med Philos. (1984) May;9(2):135-60. J. Yoon et. al., Discovery and Clinical Decision Support for Personalized Healthcare, IEEE Journal of Biomedical and Health Informatics, Volume: 21, Issue: 4.

[3] Sox, H., M. Blatt, M. Higgins and K. Marton (1988). Medical Decision Making. Stoneham, MA, Butterworth-Heinemann.

[4] J. Yoon et. al., Discovery and Clinical Decision Support for Personalized Healthcare, IEEE Journal of Biomedical and Health Informatics, Volume: 21, Issue: 4.

[5] Rosse et. al., Anatomy Ontologies for Bioinformatics, Springer, New York, 2008.

[6] Ronicke S, Hirsch MC, Türk E, Larionov K, Tientcheu D, Wagner AD. The Accuracy and Potential Impact of a Diagnostic Decision Support System in Rare Disease Cases [abstract]. Arthritis Rheumatol. 2018; 70 (suppl 10). https://acrabstracts.org/abstract/the-accuracy-and-potential-impact-of-a-diagnostic-decision-support-system-in-rare-disease-cases/. Accessed on November 19, 2018.

Address for correspondence

Project lead Milan Jovanovic, mail to milan.jovanovic@ada.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190603

Designed Strategies and Adaptation of a Master Patient Index for Transgender Patients in a Tertiary Care Hospital

Julia Frangella, Melanie Cassarino, Fernando Plazzotta, Fernando Gassino, Carlos Otero, Daniel Luna

Department of Health Informatics, Hospital Italiano de Buenos Aires, city of Buenos Aires, Argentina

Abstract

Transgender people experience their gender identity as different from the sex assigned to them at birth and/or those listed on their legal identification. A Master Patient Index (MPI) is a centralized index of all patients in a health care system. The objective of this work was to describe the designed strategies and adapting of a MPI that contemplates transgender patient registration needs as regards as health and legal context.

Keywords:

Transgender Persons; Gender Identity, Health Information Systems.

Introduction

Transgender people experience their gender identity (GI) as different from the sex which was assigned to them at birth and/or those listed on their current legal identification [1]. "Gender identity data" can be defined as GI, birth-assigned sex, legal sex, chosen name and legal name [2; 3]. Current best practices for the collection of gender identity data recommend collection of both GI and birth-assigned sex [2]. Transgender patients have particular needs in what concerns to demographic information and electronic health records (EHR). Specifically, they may have chosen a name and GI that differs from their current legal name and gender. Transgender people face intense health disparities and lack of access to health care; failure to accurately document transgender identities increase these disparities through negative implications [4]. On the other hand accurate reporting of legal identity is necessary for appropriate insurance billing.

To date, at Hospital Italiano de Buenos Aires, we have a population of 30 transgender patients registered, including adult and pediatric. The lack of a registration method at the moment may underestimate the real number of patients.

The main objective of this work was to describe designed strategies to respond to transgender patient registration needs in a MPI, contemplating patients' GI, chosen name and the possible changes it might have in a health and legal context.

Methods

Settings

Hospital Italiano de Buenos Aires (HIBA) is a non-profit healthcare academic center founded in 1853. HIBA has a network of two hospitals with 750 beds (200 for intensive care), 41 operating rooms, 800 home care beds, 25 outpatient clinics and 150 associated private practices located in Buenos Aires city and its suburban area. Since 1998, the HIBA has run an in-house-developed health information system, which includes clinical and administrative data. It has been recently certified by the HIMSS as level 7 in the Electronic Medical Record Adoption Model [5].

Institutional MPI

Interoperability and Standards

The MPI was developed according to profiles described by the Integrating the Healthcare Enterprise model (IHE) [6]. Fast Healthcare Interoperability Resources (FHIR) was chosen as the health information technology interoperability standard. The administrative gender was taken into account, and the data was modeled from that profile [7].

Matching algorithms

In order to improve the safety and quality of data, patients registration process starts with a candidate search. The search of candidates based on the following data entered: First name; Other Names (Optional); Surname; Other surnames (Optional); Birthdate; Type and number of document; Sex. Different kind of matching algorithms, such as phonetic encoding systems utilized to counter misspelled names, had been used to determine potential matching candidate. At the end of the search process, the user visualized a match weight score, resulting from field match weights assigned to patient-identifying attributes such as last name, first name, date of birth and ID.

Strategies design process

A multidisciplinary group assembled to discuss the topic was formed. The first step was to analyze all possible use-cases, according to the law, literature and hospital previous experience. Then we planned the most adequate solution for each one. For this, tests were carried out with all possible modifications and potential errors that GI and/or name changed could cause in a subsequent search for candidates. According to results, adjustments were performed for the strategy design.

Results

During the registration process the system always takes the data of the National Identity Document (DNI) which has the official data at the time of birth. So that GI and chosen name could be registered in the MPI, two new data fields were added, one for each. If a patient requested to be called with a name or GI different than the one registered on his/her DNI, those fields will be used. In this way, this information is stored as demographic data. Legal name and sex would not be changed at all. After analyzing the data, five use cases (UC) were defined (Table 1).

Discussion

A proposal as an answer to satisfy transgender patients MPI registration needs was developed along the paper. For cases of transgender patients with legal documentation that do not reflect GI or chosen name, we add two new fields in the MPI for registration of GI and chosen named. The system allows to load this data in separate fields from legal name and birth assigned sex. During a candidate search or billing process, system will take the data according to legal data fields. EHR and administrative application will show patient's chosen name and gender, except for the billings system. After MPI new fields deployment, audit for searching duplicating patient will allow to evaluate the effectiveness of our strategy.

Implementation of the new MPI will require special training for employes performing registration process. Patients also need to understand why the issue is important and what they are being asked to do.

Conclusions

The adaptation of MPI to the registration needs of transgender patients is a necessity. Throughout this work the possible use cases were described and proposals were developed to solve each one of them. The design and development of strategies to respond to transgender patient registration needs, its implementation and subsequent audit are the beginning of the fundamental change in information systems.

Table 1 – Use Cases

Patient	ID	Patient's Request	Solution Proposed
New Patient	Original DNI	To be called by a name and GI different from DNI	Patients' register according to legal documentation data. GI and chosen name will be saved in separate fields from those previously mentioned. In a candidate search, the system will take the data from legal identificatory fields. EHR and administrative applications will show patient's chosen name and gender, except for the billings system.
New Patient	New DNI	To be register for the first time	Registration process will be performed according to DNI. If patient expresses gender and/ or name changes, this information will be registered exclusively in the EHR, conforming to the law.
Patient registered	Original DNI	To change Name and GI	Patient's previous data will not be modified. The system allows to load GI data and the chosen new name, in separate fields. During a candidate search, system will take the data from legal identificatory fields. EHR and administrative application will show patient's chosen name and gender, except for the billings system.
Patient registered	New DNI	To change Name and GI	Data modification for name and gender fields will be made, according to new DNI. The log of all changes made will be saved.
Patient registered	New DNI	To be register for the first time without clarifying that it's been already registered before	Registration will begin with a search for candidates. The patient will have a new name and gender, their surname and original identification number. A search for candidates of these characteristics will result in the patient already registered with an estimated weight of 87%. This is enough for the user to notice that the patient already exists in the registry.

References

[1] E. Coleman, W. Bockting, M. Botzer, P. Cohen-Kettenis, G. DeCuypere, J. Feldman, L. Fraser, J. Green, G. Knudson, W.J. Meyer, S. Monstrey, R.K. Adler, G.R. Brown, A.H. Devor, R. Ehrbar, R. Ettner, E. Eyler, R. Garofalo, D.H. Karasic, A.I. Lev, G. Mayer, H. Meyer-Bahlburg, B.P. Hall, F. Pfaefflin, K. Rachlin, B. Robinson, L.S. Schechter, V. Tangpricha, M. van Trotsenburg, A. Vitale, S. Winter, S. Whittle, K.R. Wylie, and K. Zucker, Standards of Care for the Health of Transsexual, Transgender, and Gender-Nonconforming People, Version 7, *International Journal of Transgenderism* **13** (2012), 165-232.

[2] M.B. Deutsch and D. Buchholz, Electronic health records and transgender patients--practical recommendations for the collection of gender identity data, *Journal of general internal medicine* **30** (2015), 843-847.

[3] Centers for Disease Control and Prevention, Health Resources and Services Administration, National Institutes of Health, HIV Medicine Association of the Infectious Diseases Society of America, and HIV Prevention in Clinical Care Working Group, Recommendations for Incorporating Human Immunodeficiency Virus (HIV) Prevention into the Medical Care of Persons Living with HIV, *Clinical Infectious Diseases* **38** (2004), 104-121.

[4] J. Frangella, C. Otero, and D. Luna, Strategies for Effectively Documenting Sexual Orientation and Gender Identity in Electronic Health Record, *Studies in Health Technology and Informatics* **247** (2018), 66-70.

[5] F. Plazzotta, D. Luna, and F.B. de Quirós González, Health information systems: integrating clinical data in different scenarios and users, *Revista peruana de medicina experimental y salud publica* **32** (2015), 343-351.

[6] IHE, Main Page, https://wiki.ihe.net/index.php/Main_Page, 2018. Accessed November 22, 2018.

[7] HL7.org, Patient - FHIR v3.0.1, http://www.hl7.org/fhir/patient.html, 2018. Accessed November 22, 2018.

Address for correspondence

Corresponding author: Maria Julia Frangella. Hospital Italiano de Buenos Aires. Juan D. Perón 4190 (C1181ACH) Buenos Aires, Argentina Tel. (5411) 4959-0200. maria.frangella@hospitalitaliano.org.ar

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190604

Development and Evaluation of a Prototype CDSS for Fall Prevention

Hyesil Jung[a], Hyeoun-Ae Park[a]

[a] College of Nursing, Seoul National University, Seoul, South Korea,

Abstract

We developed a prototype CDSS that 1) provides tailored recommendations by combining a fall-risk prediction model, patients data, and evidence from CPGs, and 2) helps nurses to plan nursing care and document their activities for fall prevention. The accuracy of rules in knowledge base and inference engine was verified using ten scenarios and heuristics of user interface evaluated by four experts. We are currently evaluating the effects of the system on nurses' workflow and patient outcomes.

Keywords:

Accidental falls; Decision support systems, Clinical; Evidence-based nursing

Introduction

According to the AHRQ, the learning health system is a system that integrates internal data and experience with external evidence, and puts that knowledge into practice. As a result, patients receive higher quality, safer, and more efficient care.

We have been working to develop a small scale of learning health system to prevent falls, which is the most common adverse event threatening patient safety in the acute care settings. In our learning health system, EHR data as 'internal data' is integrated with nurses' clinical experience and clinical practice guidelines (CPGs) for fall prevention as 'experience' and 'external evidence', respectively. A fall prevention clinical decision support system (CDSS) is a tool that enables knowledge to be put into practice. The fall prevention CDSS might allow nurses to provide a high quality of care for fall prevention and improve patient safety and efficiency of workflow.

Existing fall prevention CDSSs could not provide tailored actionable recommendations based on evidence at the point of care. Furthermore, CDSSs could not automatically extract patient data which was already inputted into EHR system, so users had to input data manually. To be able to 'learn' in the learning health system, the results of practice must be recorded.

Our final goal is to develop a near real-time fall prevention CDSS that provides tailored nursing recommendations by combining fall-risk model, EHR data, and evidence from CPGs, and supports nursing care plan and documentation. On that journey, we have developed a prototype CDSS.

Methods

Figure 1 represents three components of the fall prevention CDSS which we are developing as a final goal. A fall-risk prediction model (component A.) was developed and validated. In this article, we focused on how we developed a knowledge base, an inference engine (component B), and an user interface (component C).

The scope of development of a prototype CDSS is identical with the fall prevention CDSS except for automatic extraction of EHR data. Finally, we evaluated an accuracy of rules in knowledge base and inference engine, and heuristics of user interface.

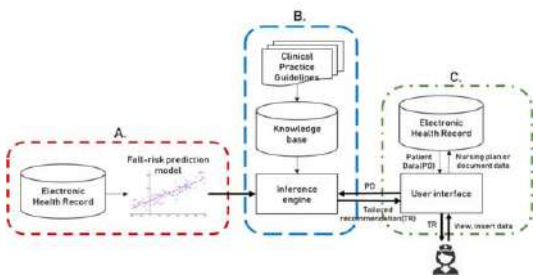


Figure 1-Components of the fall prevention CDSS

We developed the prototype CDSS in accordance with the following four stages of the system development life cycle.

Plan - functional requirements analysis

We examined a variety of tools CDSS encompassed from the literature [1] and specified functions of this system in detail through interviews with three nurses.

Design

Knowledge base

Published CPGs on fall prevention were retrieved from the National Guideline Clearinghouse and Google scholar. Knowledge was extracted from the selected 7 CPGs using MS Excel.

The knowledge was represented by decision rules using the IF-THEN rules.

User interface

Reflecting the results of the functional requirements, we designed a user interface. With our user interface, a nurse can modify information of a patient's fall-related variables and can plan nursing care and document their activities for fall prevention. The user interface was developed using Visual Studio 2017/2017 C#, NetFramework 4.0, and DevExpress 17.2 tools.

We have aligned recommendations with nursing statements the study hospital uses in order to document nursing activities performed according to the CDSS's recommendations. For recommendations not mapped to existing nursing statements, we have added new nursing statements.

Implementation

The fall-risk model, patient data, and knowledge extracted from CPGs were combined in order to provide tailored nursing recommendations by risk group and risk factors a patient has. For example, the process of inferring a patient with 'a high risk of falling' and providing recommendations for them is as

follow: First, patient data is entered into the fall-risk prediction model and the risk of falling is calculated. If the risk of falling is greater than or equal to the cutoff, the patient is identified as 'a high risk of falling'. Then, knowledge on nursing intervention for patient with 'a high risk of falling' is extracted from knowledge base and printed on the screen.

Evaluation

Rules in knowledge base and inference engine

We tested an accuracy of rules that calculate the risk of falling and provide the tailored nursing recommendations using ten scenarios. Of 15,450 patients' data, we randomly selected ten patients and compared their risk of falling calculated by the system to the risk of falling that we manually calculated. In addition, we manually modified values of the fall-related variables of ten patients and verified whether CDSS printed out nursing recommendations accurately based on modified risk factors.

Heuristics of user interface

In accordance with Nielsen's heurisitcs principles [2], four evaluators who majored in nursing informatics were asked to try the prototype and to rate the severity of each usability problem with a scale from 1 to 4. For principles with an average score higher than 3, the system has been revised to fix that issue.

Results

Plan- functional requirements analysis

We defined 'computerized alerts and reminders to care provider' and 'clinical guidelines' as key functionalities of a fall prevention CDSS. Specified functional requirements of the CDSS were 1) the CDSS is able to calculate the risk of falling using patient data, 2) the CDSS is able to alert nurses to identify patients with a high risk of falling, 3) the CDSS is able to provide tailored nursing recommendations according to risk factors that a patient has, 4) the CDSS allows nurses to manually modify a state of patient if necessary, and 5) the CDSS enables nurses to plan and document nursing care for fall prevention according to the recommendations.

Design

Knowledge base

We extracted 75 specific knowledge related to risk factors and 69 general knowledge related to environment. An algorithm was developed by linking 42 rules defining the value of each variable and 19 tailored recommendations.

User interface

When a patient's condition changes, a nurse can modify values of fall-related variables manually (Figure 2). Then, the system recalculates the risk of falling and displays tailored recommendations. Based on the recommendations, the system enables a nurse to plan nursing care and to document nursing activities for fall prevention.

Implementation

Figure 3 shows a screenshot of the prototype CDSS providing tailored recommendations according to a risk of falling (a high risk) and risk factors a patient has. Nurses can track the trends in a risk of falling and get tailored recommendations by risk factor when they click on the risk factors.

Evaluation

Rules in knowledge base and inference engine

There was no inconsistency between the risk of falling the CDSS calculated and the risk of falling we manually calculated. The CDSS printed out nursing recommendations without errors according to the modified risk factors.

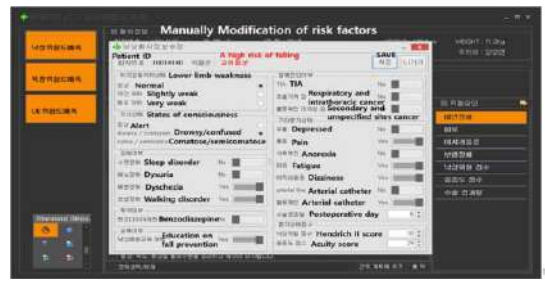


Figure 2-A modification of values of fall-related variables

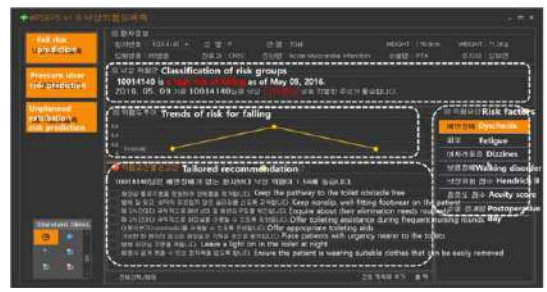


Figure 3-A risk of falling, risk factors, and tailored recommendations,

Heuristics of user interface

Only the 'consistency and standards' principle, which indicates compliance with platform conventions, were lower than 3.

Conclusions

The prototype CDSS has several limitations. First, it cannot extract EHR data automatically. Second, knowledge extracted from CPGs was not tested for its applicability to clinical settings. Thus, a validation of the nursing recommendations by the experts is needed. Nevertheless, this system can help nurses to provide evidence-based tailored nursing for fall prevention without any extra workload.

Acknowledgement

This work was supported by the Natioanl Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2018R1A2A2A05022021).

References

[1] J. E. Tcheng, S. Bakken, D. W. Bates, H. Bonner III, T. K. Gandhi, M. Josephs, K. Kawamoto, E. A. Lomotan, E. Mackay, B. Middleton, J. M. Teich, S. Weingarten, and M. Hamilton Lopez, *Optimizing Strategies for Clinical Decision Support: Summary of a Meeting Series*, National Academy of Medicine, Washington, DC, 2017.

[2] J. Nielsen. Heuristic evaluation. *Usability Inspection Methods*, John Wiley & Sons, NY, 1994.

Address for correspondence

Hyeoun-Ae Park, PhD, FAAN, FACMI, FIAHSI

College of Nursing, Seoul National University, 103 Daehak-ro, Jongno-gu, Seoul 03080, South Korea. Tel: +82-2-7408827, Fax: +82-2-765-4103, E-mail: hapark@snu.ac.kr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190605

Designing Archetype Models for Each Step of Workflow in Medication

Shinji Kobayashi[a,b], Naoto Kume[a,b], Hiroyuki Yoshihara[a,b]

[a] *Department of Electroenic Health Record, Kyoto University, Kyoto, Japan*
[b] *NPO openEHR Japan*

Abstract

Medication processes are consisted with multiple steps by professionals and consumers. Physicians prescribe drugs to patient with conformance, but sometimes they were changed to other drugs.

We designed concept models to capture medication workflow process records by openEHR archetype models, and four templates were determined with each step of the medication process.

We will show the detail of clinical modeling about medication workflow in this article.

Keywords:
Medication, openEHR, concept models

Introduction

Medication has four steps (shown below) of process with complex workflow.

1. **Prescribe**: Physicians prescribe drugs to patients with instruction to take.
2. **Dispense**: Pharmacists dispense drugs to patients or professionals by the instruction on prescription with packaging. Drugs might be altered to generics of original or others for some reason, such as allergy or shortage of drugs.
3. **Deliver**: If drugs should be delivered by professional in case of injection or other, dispensed drug are delivered by nurses, or other professionals with instruction by the physicians or pharmacists
4. **Consume**: Patients consume drugs by the instruction on the prescription in general, but sometimes abandoned or changed.

We have developed nation-wide Electronic Health Record (EHR) system based on ISO 13606/openEHR archetype technology to capture health related data from care providers [1; 2]. More than 30 hospitals were connected, and two groups of pharmacies will be connected in 2019. At first, we designed a single openEHR template to capture prescription information, but we found dispense information in pharmacies should be designed separately to involve tasks of pharmacies.

And the next, delivery and consumption were identified as other concepts.

In this article, we describe the concept models for the medication process.

Methods

We figured each concept model, prescribe, dispense, deliver, and consume by Mindmap with existing form of items by XMind8 [3] (Figure 1). 16 concept models were identified and compared with openEHR Clinical Knowledge Manager (CKM) [4].

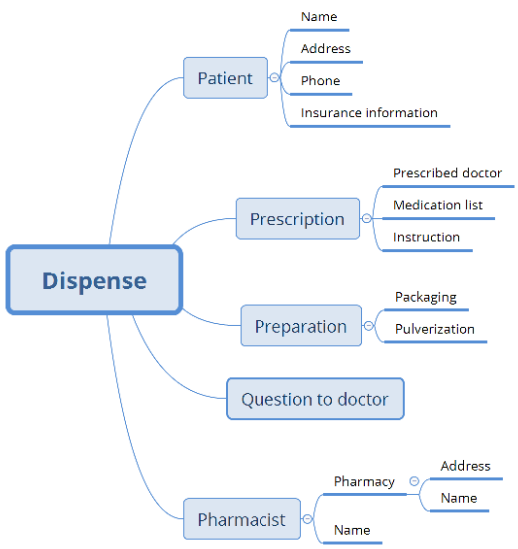


Figure1– Mindmap of drug dispense

Concept models were composed in composite archetypes to conform four openEHR templates by Ocean Template Designer [5].

Results

Table 1 shows the templates and major archetypes that conform medication steps.

Each step of process contains specific information, but existing archetypes on CKM were not well-designed for dispense, deliver, and consume. Therefore, we designed new archetypes, and specialized action-medication archetypes. Because there is specific manners for medication in Japan, specialization for prescription was also requied

Table 1– Major archetypes to construct medication workflow

Template	Archetype
Prescribe	composition-prescription
	instruction-prescription
	cluster-medication-details
Dispense	composition-dispense
	action-dispense
	cluster-packaging
Deliver	composition-delivery
	action-deliver
	cluster-deliver-details
Consume	composition-consumption
	action-consume

Discussion

We idetified clinical concepts in medication process, and designed four templates with archetype models. Prescription data were well-disigned already, but it needed specialization. Because each country or aria has each own manner of medication, universal design of archetypes could not be applicable.

Dispense, deliver, and consumption archetypes were still not well-designed, because the data were not able to capture from pharmacy and patients. However, IoT devices will be able to collect drug consumption data form patients, EHR should have concept model accoring to them.

Even though there are diversity in medication process, standardized process for medication should be desigined for the reference for each derived works.

Conclusions

We designed four steps of medication workflow. Each step required new archetypes and specialization, because existing archetypes were not well designed for dispense, deliver, and consume.

le if printed in black & white. Figures and graphs may span both columns.

Acknowledgements

This research is supported by "Clinical Study Oriented ICT Infrastructure Development Project - Sustainable Massive Health and Clinical Data Repository for Secondary Use" from Japan Agency for Medical Research and Development.

References

[1] Kobayashi S, Kume N, Yoshihara H. Restructuring an EHR system and the Medical Markup Language (MML) standard to improve interoperability by archetype technology. Stud Health Technol Inform. 2015;216:881.

[2] Kobayashi S, Kume N, Nakahara T, Yoshihara H. Designing Clinical Concept Models for a Nationwide Electronic Health Records System For Japan. European Journal of Biomedical Informatics. 2018 Jan 10;

[3] XMind - Mind Mapping Software [Internet]. [cited 2018 Nov 26]. Available from: https://www.xmind.net/

[4] openEHR Clinical Knowledge Manager [Internet]. Available from: http://www.openehr.org/ckm/

[5] Informatics O. Ocean Template Designer [Internet]. Available from: http://www.openehr.org/downloads/modellingtools

Address for correspondence

Shinji KOBAYASHI, MD, PhD

e-mail: kobayashi.shinji.5s@kyoto-u.ac.jp

Address: Room 506, 5th floor, Bldg. #9 Kyoto Research Park, 91 Awata-cho, Chudoji, Shimogyo-ku, Kyoto, Japan 600-8813

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190606

A New Approach for Ageing at Home: The CAPTAIN System

Evdokimos I. Konstantinidis[a,b], Despoina Petsani[a], Giuseppe Conti[b], Antonis Billis[b], Valentina Conotter[c], Guillaume Chican[d], Tim Llewellynn Lorenzo[e], Alejandro Rivero Rodríguez[f], Santiago Hors Fraile[f], Andoni Beristain[g], Gorka Epelde[g], Unai Diaz-Orueta[h], Louise Hopper[i], Maxim Kostin[j], Rosa Almeida[k], Raquel Losada[k], Wolfgang Kniejski[l], Giandomenico Nollo[m], Francesco Tessarolo[m], Panagiotis D. Bamidis[a]

[a] Lab of Medical Physics, Aristotle University of Thessaloniki, Thessaloniki,Greece,
[b] Nively SAS, Nice, France
[c] Social IT, Trento, Italy
[d] HOLOLAMP, France
[e] NVISO SA, Lausanne, Switzerland
[f] Salumedia, Seville, Spain
[g] Vicomtech, Donostia/San Sebastián, Spain
[h] Deparment of Psychology, Maynooth University, Maynooth, Ireland
[i] School of Nursing and Human Sciences, Dublin City University, Dublin, Ireland
[j] DIGIFLAK, Tallinn, Estonia
[k] INTRAS, Valladolid, Spain
[l] INI-Novation, Hessen, Germany
[m] Department of Industrial Engineering, University of Trento, Trento, Italy

Abstract

Our work exhibits how previous projects on the Active and Healthy Ageing field have advanced to the conception of CAPTAIN, a radically new approach towards increased end-user acceptance. The goal is to create intuitive technology that does not require specific skills for interaction and blends in with real life. CAPTAIN will be co-designed by all types of stakeholders, including older adults, involved in all stages, from the initial design to delivery of the final system.

Keywords:

Healthy Aging, Home Care Services, Needs Assessment

Introduction

Ageing is directly related to several issues that may greatly affect the level of independence of an older adult, often requiring institutionalization [1]. However, as a matter of fact, adults typically prefer living at their homes as long as possible retaining their independence. Furthermore, institutionalization has been linked with negative outcomes such as mortality and low quality of life [2].

A wide range of technologies for homecare scenarios have been developed over the last few years including the use of smartphones, tablets, sensors, etc. However, most of them intervene and renovate the user's home or require them to use wearable devices, unable to seamlessly "blend" within the person's daily life and living environment. Moreover, most technologies focus on the patronization of user activity without empowering personalized interventions.

CAPTAIN, which is the acronym of "Personalised coaching for well-being and care of people as they age", proposes an innovative technology that blends in the real-home environments, turning it into an intelligent and ubiquitous coach. The project, which has been funded by the Horizon 2020 work programme proposes to leverage a range of different innovative technologies such as projected augmented reality, 3D sensing technologies, tangible user interaction, and physiological and emotional data analysis to create a virtual coach capable of supporting daily tasks and activities, suggesting a healthy and active lifestyle, and encouraging social participation under the context of a long term coaching strategy.

CAPTAIN System Description

To deliver an intuitive, transparent, and yet effective interaction paradigm, CAPTAIN transforms the older adults' home into a tangible interface where instructions from the virtual coach are projected onto the real context, whenever and wherever needed. Interaction with the environment occurs by touching real life objects, according to the so-called tangible interfaces, and by voice commands using everyday language.

Non-invasive user and environment sensing

User and environmental sensing is done by 3D sensors, RGB cameras, and microphones which are used for:

- User identification and authentication automatically made using face recognition and human body pose recognition (skeleton/silhouette features).
- Creating a profile for each user, not only including data concerning their biological and demographic information, but also their interests and behavioral habits.
- Gait analysis that has been widely used as an indicator of different cognitive impairents [3]. With regards to analysis of movements, CAPTAIN will build upon previous work on indoor analytics [8] where density based clustering was applied to indoor (location) transitions recorded in real older adults' homes for a duration of approximately a year.
- User interaction with so-called "exergames" and cognitive games, engaging the user in physical and

cognitive training tasks, respectively. CAPTAIN will employ a serious games platform, called web-FitForAll [4], incorporating standard physical exercises protocols. The in-game metrics analysis have exhibited a classification accuracy greater than 73% when discriminating between cognitively normal from mild cognitive impaired seniors as well as promising results on creating user profiles using scores created from in-game metrics [5,6].

Coach behaviour design and Artificial Inteligence (AI) algorithms

The data coming from the cameras, projectors, sensors, etc. analyzed using Artificial Intelligence algorithms will permit information to be obtained about the person's behavior and preferences, as well as the environment. Then, the CAPTAIN coach will use algorithms from the fields of Artificial Intelligence and data mining to integrate the higher-level information about user and environment sensing, combined with behavior change models and domain specific knowledge, to produce personalized motivational guidance embodied through a virtual coach persona specifically designed to support physical and cognitive training, improve nutrition, lifestyle habits and social participation, and reduce risk.

Human Computer Interaction (HCI) system

The CAPTAIN's HCI will be based on a new prototype device that will create a living environment capable of turning any surface into a projected (augmented) information source. To do so, a spatial augmented reality pipeline will be developed based on two rendering passes, where the first renders the 3D content from the perspective of the senior while, during the second, the generated image is warped into an image to be projected that looks correct from the perspective of the user's eyes and for the environment's geometry.

Design Approach

While the CAPTAIN's consortium consists of a multidisciplinary team, the main actors of CAPTAIN will be a large group of stakeholders who will be involved throughout the project, providing ideas and insights to which the entire team will be constantly responding and adapting. One of the greatest challenges of introducing new forms of human computer interaction to older adults is to achieve high usability, usefulness, and effectiveness levels, making older adults adopt the solution and not adapt to it.

Agile methodology

CAPTAIN has been entirely conceived around the idea of co-creation, participatory, and user-centered design which have been proposed to mitigate user acceptance failure frequently observed in older populations. To do so, a hybrid approach leveraging on concepts from Design Thinking, Lean, Startup [7], and SCRUM agile framework [8] approaches are followed by the project. Design Thinking allows the consortium to identify unmet needs and create value from these insights while the Lean approach enables delivering a partially functional prototype frequently enough to the stakeholders in order to collect feedback, validate assumptions, and inform readjustment. The use of SCRUM helps organize work across technical development to collaborate towards delivering high value. This hybrid approach facilitates CAPTAIN to solve effectively, and with high flexibility, the complex project's developments required to achieve its goals.

CAPTAIN Stakeholders community

The CAPTAIN Stakeholder community becomes the basic source of requirements throughout the project's lifecycle. The CAPTAIN consortium aims for the creation and maintance of a community with strong support bonds based on two clusters of stakeholders: the first one will include those who will actively use CAPTAIN (older adults and their caregivers), and the second one will include those who can give suggestions or influence (service providers, nursing home management, and patient associations).

Discussion

CAPTAIN's continuous and unobtrusive sensing of physical, cognitive, and social functioning has the significant potential to support lifelong health management. CAPTAIN uses an innovative projective environment to provide a useful and effective contextualized (i.e. direct projection onto the real world) virtual coach to older adults living at home.

The advantages brought by CAPTAIN will be evaluated during an impact assessment phase which will be tailored to measure the effectiveness in terms of quality of life, societal, economic, and psychological terms The information and indicators that will be extracted during the assessment phase will be essential to fine-tune the final value proposition in light of later commercial exploitation of the project's results.

References

[1] M. Luppa, T. Luck, S. Weyerer, et al., Prediction of institutionalization in the elderly. A systematic review, *Age Ageing* **39** (2009) 31-38.

[2] N. Yang and K. Ornstein, The impact of symptom burden on hospitalization, nursing home placement and death among homebound older adults, *J Am Geriatr Soc* **63** (2015) S134.

[3] Q. Lin, D. Zhang, X. Huang, H. Ni, and X. Zhou, Deteecting wandering behavior based on GPS traces for elders with dementia, In *2012 12th International Conference on ICARCV*, Guangzhou, China, 2012, 672-677.

[4] E.I. Konstantinidis, A.S. Billis, C.A. Mouzakidis, et al., Design, implementation, and wide pilot deployment of FitForAll: An easy to use exergaming platform improving physical fitness and life quality of senior citizens, *IEEE J Biomed Heal Inform* **20** (2016) 189-200.

[5] D. Petsani, E.I. Konstantinidis, V. Zilidou, and P.D. Bamidis, Creating health profiles from physical and cognitive serious game analytics, In *2018 2nd International Conference on TSIHW*, Thessaloniki, Greece, 2018, 1-6.

[6] E.I. Konstantinidis, P.D. Bamidis, A.S. Billis, P. Kartsidis, S.G. Papageorgiou, Physical training in-game metrics for cognitive assessment: Evidence from extended trials with the FitForAll exergame platform, *Front Aging Neurosci* (2016).

[7] E. Ries, The Lean Startup, Crown Business, New York, NY, 2011.

[8] K. Schwaber and J. Sutherland, The Scrum Guide The Definitive Guide to Scrum: The Rules of the Game, Creative Commons, Mountain View, CA, 2013.

Address for Correspondence

Panagiotis D. Bamidis, Assoc. Prof., Medical Education Informatics

Lab of Medical Physics, School of Medicine, Aristotle University of Thessaloniki, P.O. Box 376, 54124, Thessaloniki, Greece

email: bamidis@med.auth.gr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190607

Information System Implementation Optimizes Medical Coding

Hong-Ling Lin[a], Shu-Meng Cheng[b], Dai-Fang Hsu[a], Chang-Chuan Huang[c], Ding-Chung Wu[a]

[a] *Medical Records Office, Tri-Service General Hospital, Taipei City, Taiwan (ROC)*
[b] *Head office, Tri-Service General Hospital, Taipei City, Taiwan (ROC)*
[c] *General Manager office, HISMAX CO., LTD., Kaohsiung City, Taiwan (ROC)*

Abstract

Diagnosis Related Groups (DRGs) and the Tenth Revision of the International Statistical Classification of Disease and Related Health Problems (ICD-10) were implemented to Taiwan in 2010 and 2016 respectively. New rules related to the medical costs reimbursement were great challenges facing medical institutions. One of the medical centers in north Taiwan introduced an ICD e-dictionary, DRGs cloud computing system, and integrated them into the hospital information system. Further, developing a medical coder specialization work model optimized the workflow, coding quality, and efficiency, which defeated the adverse effects of DRGs and ICD-10 implementation successfully.

Keywords:

Diagnosis-Related Groups, Hospital Information Systems

Introduction

In 2010, the National Health Insurance Administration (NHIA) of Taiwan applied Diagnosis Related Groups (DRGs) gradually to replace the fee-for-service payment policy, expected it could be comprehensive before 2019 and suppresses medical costs. Moreover, in order to describe disease exactly and conform to the statistical data format worldwide, NHIA announced International Statistical Classification of Disease and Related Health Problems, Tenth Revision (ICD-10) should be used as the basis for medical expenses application since 2016.

DRGs and ICD-10 implementations were great challenges facing the medical institutions. The most direct impact is the complexity of disease coding, which increases after the introduction of ICD-10. According to pioneering research from NHIA, the manpower of medical coders needs to be increased by 2.3 times when ICD-10 implementation [1]. Other impacts may be triggered by financial factors, such as medical ethical problems. Healthcare providers worried about DRGs increased financial burden to medical institutions, so that critically ill patients whose expected costs are higher than the associated reimbursement will be dumped, an experience seen with DRGs in other countries [2-3], even though some supporting measures have been executed by NHIA [4].

To reduce the adverse effects of DRGs and ICD-10 implementations, a medical center in north Taiwan introduced ICD-10 e-dictionary in 2015 and DRGs cloud computing system in 2017, which were integrated into a self-developed DRGs monitoring system. Moreover, they developed s specialized medical coder work model (*Figure 1*) with the systems in 2018. In the model, medical cost management shifts to the period of patients' hospitalization, when doctors could classify patients to appropriate DRGs, controlling the supplemental cost of specialized medical coder. Our hypothesis is that the doctors and medical coders can reach a consensus and propose the disease code that is closest to the disease, which can accurately control the consumption of medical resources and obtain reasonable payment. Also, the work model forms a positive feedback cycle.

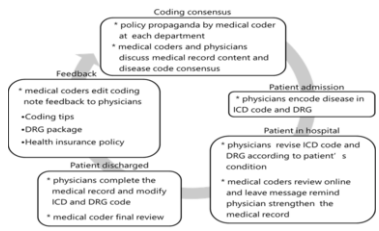


Figure 1–specialized medical coder work model

Methods

To evaluate the effects of the information system and work model, we made some tests and reviewed related indexes:

1. Difference in time distribution of disease coding between using the paper dictionary and e-dictionary

   300 DRGs cases were randomly selected from January 2017 to April 2018. 10 medical coders appointed by a medical center in the north of Taiwan used an ICD-10-CM/PCS paper dictionary (2014 version) and e-dictionary (HISMAX 2014 version) to find the code separately, and recorded the time based on the HIS system. We compared the coding time in different tools, classified by DRGs payments types (full reimbursement, quota, exceed quota). Matched paired *t*-tests were used to test for changes in time.

2. DRGs consistency rate of doctors and medical coders

   We collected all DRGs cases (918 cases in average each month) from January 2017 to August 2018 and compared doctors' last edited DRGs record with the final results of the medical coders, calculating consistency rate as a percentage. Mann–Whitney U test was used to test for change of the consistency rate before and after the specialized medical coder work model was put into practice.

3. Satisfaction with DRGs cloud computing and monitoring system and the specialized medical coder work model

We compiled a structured questionnaire based on the business content of the medical coder and the information system success model proposed by DeLone & Mclean [5]. A 5-point Likert scale was applied, where 1 was very dissatisfied, and 5 was very satisfied. The questionnaire consisted of 19 questions and was distributed to the departments of Infectious Diseases, Neurology, Neurosurgery, and Nephrology in a medical center in north Taiwan between March 2018 to April 2018. A convenience sample of 55 doctors was included. The recovery rate of the questionnaire was 98.2% and Cronbach α=0.967, indicating that the measurement results were highly reliable. Questionnaire scores were analyzed by descriptive statistical methods (mean and standard deviation).

4. Related indexes were also reviewed from the 2016-2018 annual reports of the hospital, including internal and external medical coding experts' review consistency rate (Review sample size: 25 cases; Internal/ External review frequency: monthly/ yearly), manpower of medical coder, and DRGs surplus monthly average ratio.

Statistical analysis was performed with SPSS 20.0 program.

Results

Encoding full reimbursement, quota, and exceed quota case in a paper dictionary took medical coders 11.63 minutes, 7.36 minutes, and 13.06 minutes in average; in the e- dictionary, it took them 9.16 minutes, 5.03 minutes, and 9.28 minutes, respectively. Encoding in e-dictionary required less time (p < 0.001) than paper-based work (*Table 1*).

Table 1 —Comparison of coding time between paper and electronic dictionary

Form of dictionary	DRG reimbursement type								
	full reimbursement			quota			exceed quota		
	n	mean (SD)	p	n	mean (SD)	p	n	mean (SD)	p
paper	40	11.63 (10.61)	<0.001	189	7.36 (7.56)	<0.001	71	13.06 (11.40)	<0.001
electronic		9.16 (8.94)			5.03 (4.65)			9.28 (7.83)	

In 2017, the average DRGs consistency rate of doctors and medical coders was 58.68% and it was a slightly downward trend. In 2018, the average rate was 63.20%. It showed an uptrend after the specialized medical coder work model put into practice. Compared with the average rate in 2017, 2018 was higher significantly (p = 0.003) (*Table 2* and *Figure 2*).

Table 2 —2017 and 2018 DRGs consistency rate of doctors and medical coders（%）

	Mena	Min	Max	SD	p
2017	58.68	54.89	64.58	3.11	0.003
2018	63.20	57.09	65.21	2.74	

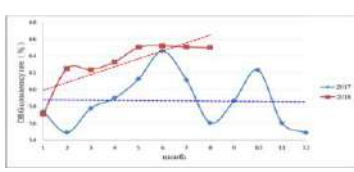


Figure 2 —the trend of DRGs consistency rate (%) of doctors and medical coders (2017-2018)

The satisfaction survey results showed that the average scores of system quality, information quality and benefits with the DRGs cloud computing and monitoring system was 4.28, 4.29, and 4.27 respectively. In the specialized medical coder services, the average scores of policy advocacy and coding services were 4.22 and 4.21 respectively. Overall, the average score of satisfaction was a high rating of 4.25, (*Table 3*).

Table 3 —Questionnaire and scores of satisfaction

Facet	Item	Question	Mean	SD
DRG computing and monitoring system	system quality	Computing speed	4.26	0.42
		Stability of the system	4.26	0.4
		Easy to use	4.31	0.38
	information quality	System could compute the correct DRG and fit the condition	4.28	0.4
		System could be revised instantly whenrule (such as ICD or DRG) are changed	4.3	0.42
		Advanced review could help you choose reasonable DRG	4.31	0.44
		Light can remind you to manage the abnormal medical expenses	4.25	0.35
	benefits	Using the system improves the quality of medical records	4.24	0.45
		Functions of the system can meet clinical needs	4.26	0.4
		Using the system helps increase efficiency	4.31	0.36
Specialized medical coder service	policy advocacy	Whether the content of the promotion is satisfactory	4.26	0.44
		Is the time properly arranged	4.23	0.42
		Policy advocacy and coding discussions are helpful for medical record writing	4.18	0.39
		Policy advocacy and coding discussions are helpful for coding	4.2	0.4
	coding services	Does the medical coder specialization improveyour coding accuracy	4.18	0.4
		The reminder service for the abnormal DRG case	4.2	0.4
		Medical coders review medical record and correct the wrong DRG	4.2	0.36
		Coding note is helpful for coding	4.27	0.45
		Specialized medical coder in charge for discussion on coding	4.22	0.42

By reviewing 2016-2018 annual reports of the medical center, we found out that in the face of ICD revision and DRGs implementation, the manpower of the medical coder was maintained at 10. The external and internal medical coding experts review consistency rate was above 90% each month. In summary, that means the medical coding quality could be maintained without increasing personnel. Moreover, the DRGs surplus rate maintained at 17% to 19% which means the medical expenses were effective controlled.

Conclusions

The information systems do help medical coders work more efficiently, and optimize the workflow and coding quality. The results also confirm our hypothesis. As the DRGs consistency rate of doctors and medical coders was 63.2%, which is higher than before, and the doctors' satisfaction level was rated high, that means doctors are willing to do it and give more suitable disease codes by the system with specialized medical coders' supports so that every DRGs case could be monitored more precisely and reimburse reasonable payment. In this way, it successfully reduced medical resource wasting and potential financial risk which may lead to medical ethics issues.

Acknowledgements

No potential conflicts of interest relevant to this article were reported.

References

[1] H.M. Lin, National Health Insurance Administration designated hospital ICD-10-CM/ PCS Pioneer Project, National Health Insurance Administration, Taipei, Taiwan (ROC), 2013.

[2] B. Reinhard, J. Schreyögg, and P.C. Smith, Hospital case payment systems in Europe, *Health care management science* 9.3 (2006), 211-213.

[3] Q. Kevin, After the revolution: DRGs at age 30, *Annals of internal medicine* 160.6 (2014), 426-429.

[4] National Health Insurance Administration, What are the supporting measures for DRGs payment system, National Health Insurance Administration, Taipei, Taiwan(ROC), 2016.

[5] W.H. Delone and E.R. McLean, The DeLone and McLean model of information systems success: a ten-year update, *Journal of management information systems* 19 (2003), 9-30.

Address for correspondence

Hong-Ling Lin, Ding-Chung Wu
No.325, Sec.2, Chenggong Rd., Neihu District, Taipei City 11490, Taiwan (R.O.C)
lynnrainbow106@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190608

Implementation of a REDCap-Based Research Data Collection System in Cameroon

Lionel Ngamani[a], Rogers Ajeh[a], Akindeh Mbuh[a], Anastase Dzudie[a], Stephany N. Duda[b]

[a] *Clinical Research Education, Networking and Consultancy (CRENC), Yaoundé, Cameroon*
[b] *Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, Tennessee, USA*

Abstract

Implementing electronic data collection for health research can be challenging in resource-limited settings, where electricity, Internet access, and study staff with computer training may be limited. Our team has established a successful research data infrastructure using the REDCap software at three HIV clinics and one coordinating center in Cameroon. We describe our recommended network architecture and guidance for study data teams working in similar settings.

Keywords:

Data Collection, HIV infections, Cameroon

Introduction

Cameroon is a member country of the Central Africa International epidemiology Databases to Evaluate AIDS (CA-IeDEA), one of seven regional consortia of HIV cohorts encompassing 45 countries globally [1]. Clinics participating in IeDEA contribute longitudinal patient care data for observational HIV research and the development of care guidelines for groups like the World Health Organization. The participating Cameroon HIV clinics used paper charts, so the Cameroon IeDEA research team needed to build a low-cost electronic cohort database for three clinics plus an in-country data center in order to send data to the US-based Regional Data Center. The solution selected should not interfere with the government's long-term planned rollout of electronic health records. Furthermore, the Cameroon National Ethics Committee advised that all data should be housed in-country before sharing with outside collaborators.

Methods

We implemented our research data collection system using REDCap [2], a free, secure and flexible web-based clinical research data capture platform. Unlike the single webserver solution typically seen in high-resource settings, we implemented three tiers of REDCap servers (Figure 1). To accommodate electricity and Internet outages, we set up a Windows desktop at each clinic as a "REDCap server" on a local area network (LAN) with battery backup. Patient enrollment and routine visit data were transcribed from paper forms into REDCap by data entry personnel working on laptops. This enabled data entry work to continue despite recurrent Internet and power interruptions. We configured a REDCap server (with fixed IP address) at the Data Center in Yaounde to receive data from all sites over Internet. Data officers at each clinic pushed data to the central server weekly ensuring timely data availability at the in-country Data Center. We implemented REDCap forms with integrated quality checks (e.g. range and valid values, date formats, skip logic) and conducted a monthly data quality control at the central level. At the end of every data quality review, a report was sent to the sites for verification and correction. The data quality control procedure was repeated to ensure all corrections were implemented. Cleaned data were uploaded quarterly to a regional data center. Variable names and codes were the same across REDCap projects on all server tiers, allowing easy transfer and harmonization of data.

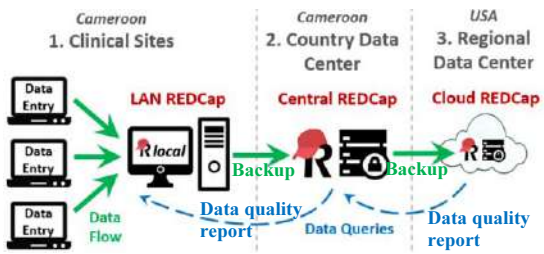


Figure 1 – Diagram of the 3-step data flow from HIV clinics in Cameroon to the regional data coordinating center

Results

This multi-tier REDCap architecture has enabled us to capture high quality, longitudinal data on over 6,000 unique patients despite weeks of Internet interuptions at both the Cameroon data center and participating clinics. Challenges with data collection and entry and REDCap data uploads have been resolved collaboratively with increased site engagement and training [3]. The Cameroon-based server meets requirements for in-country data storage and secure data sharing.

Conclusions

A multi-tier REDCap infrastructure can be a workable data capture solution for research in resource-limited settings.

Acknowledgements

This work was supported by U.S. NIH U01AI096299.

References

[1] IeDEA Central Africa Regional Collaboration https://iedeaca.org/

[2] P.A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez and J.G. Conde., Research Electronic Data Capture (RED-Cap) - A Metadata-driven Methodology and Workflow Process for Providing Translational Research Informatics Support, *J Biomed Inform* 42 (2009) 377-381.

[3] Lang TA, White NJ, H.T. Tran, J.J. Farrar, N.P. Day, R. Fitzpatrick, B.J. Angus, E. Denis, L. Merson, P.Y. Cheah, R. Chilengi, R. Kimutai and K. Marsh,Clinical Research in Resource-Limited Settings, *PLoS Negl Trop Dis* 4(6) : e619 (2010)

Address for Correspondence

Lionel Ngamani, lionel.ngamani@crenc.org

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190609

Analyzing the Demographics of Virtual Care Users

Songzi Liu[a], Tanzila Zaman[a], Barbara Edson[b], Robert Gianforcaro[b], Saif Khairat[a]

[a] *Carolina Health Informatics Program, University of North Carolina, Chapel Hill, NC, USA*
[b] *UNC Health Care, University of North Carolina, Chapel Hill, NC, USA*

Abstract

Using telemedicine to provide care is an attractive alternative for patients. However, few studies have examined the demographics of the patients using virtual care. In this paper, we investigate the demographic features of the Virtual Care (VC) users; Age, gender, roles, and preferred methods of communication are studied.

Keywords:

Demographic features, Patients, Telemedicine

Introduction

Telemedicine is the use of technology to provide care to patients that are at a distance from health care providers [1]. Studies show virtual care is an attractive alternative for in-person urgent care for patients mostly living in the rural communities [2]. Some of the reasons for using virtual consultation are to save cost of travel and time, ease of accessibility, and to reduce pollution [3]. Although there has not been many studies done to understand the demographics of the population using telemedicine, one study shows women over 55 years of age are less willing to use telemedicine for ear and hearing appointments [4].

Methods

The subject of this study is a virtual care, web-portal created to host telemedicine consultation sessions between patients and board-certified providers from multiple disciplines. The virtual care service functions 24 hours a day, seven days a week. Through the portal, patients have the independence to choose to be seen the same day or schedule a visit at a later day. Based on the patient's complaint, a list of expert physicians and their availabilities are presented. Patients have the ability to talk to the preferred physician by phone or video call. This feature is particularly important to accommodate personal preferences with regards to access to technology, privacy, security, and convenience.

Data was collected through the web-portal and stored in a secure, HIPAA compliant server. Institutional Review Board (IRB) approval was obtained to conduct this research. The datasets were cleaned and preprocessed using OpenRefine and Microsoft Excel. All data points were kept because they have complete information. The descriptive statistical analysis was carried out using R. In further examining of the relationship between different variables, Tableau 10 was used in generating data visualizations.

Results

Since Go-Live, a total 2,467 patients registered to the portal; 996 patients completed a visit in Virtual Care over a period of 9 months.

Age

The self-reported age of users ranged from 1 to 118 years old, with a median of 39 and a mode of 34, as seen in Figure 1. The age data is further categorized into four age groups representing different population characteristics: with age 0-18 as *Minors*, age 19-34 as *Young Adults*, age 35-64 as *Older Adults*, and over 65 as the *Elders*. Users aging from 35 to 64 years old composed 57.68% of the entire user population.

Gender

A total of 1,763 female users constituted 71.46% of the user body, significantly outnumbering the 704 male counterparts (28.54%). A closer examination of the relationship between *Gender* and *Age* reveals that the male user population has a more spread-out distribution of users from different age groups. The female users, on the other hand, come heavily from the *35-64* group (63.36%), Figure 1.

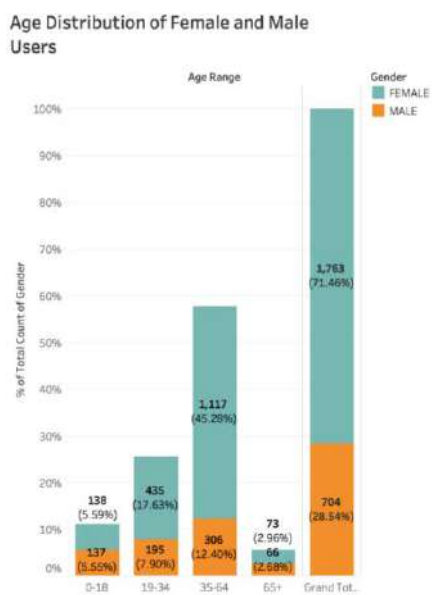


Figure 1. Age Distribution of Female and Male Users

Encounters

Among the 996 instances of patient encounters, 915 (93.56%) were completed and successfully paid. The remaining 6.44% of cases were either *Non-billable due to Missed Session* (46 instances) or either *Non-billable due to Cancellation* (17 instances), where encounters were initiated but never took place. The duration of patient encounters ranges from under 1 minute to 15 minutes, with an arithmetic mean value of 4.55 minutes and a median of 4.51 minutes, as seen in Figure 2.

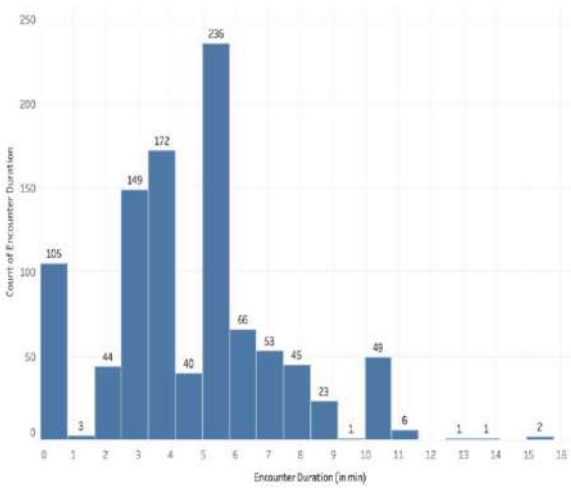


Figure 2. Distribution of the Duration of Patient Encounters

Primary and Dependent Users

2,016 users (81.72%) fell under the *Primary* roles; the remaining 451(12.28%) are classified as *Dependent*. A further look into user roles with respect to the age parameter indicates that 85.82% of Minors are *Dependent*. The other three age groups are predominantly *Primary*, each with over 80% of users being Primary as seen in Figure 3 below.

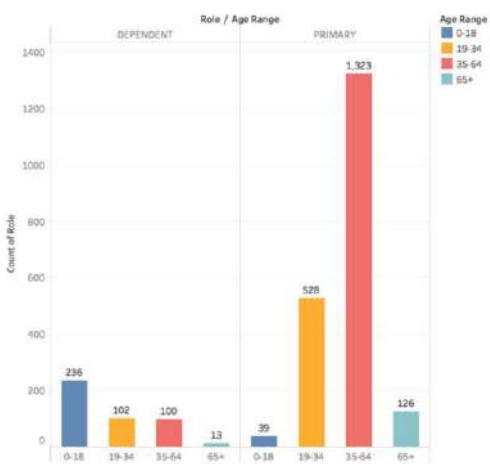


Figure 3. Age Composition of Primary and Dependent Users

Preferred Methods of Encounter

827(83.03%) of instances took place via *Phone* call; the other 167 (16.97%) encounters were facilitated by *Video* call, as seen in Figure 4. A closer look at the preferred methods by gender indicates that a greater percentage of male users prefer *Video* than the female users.

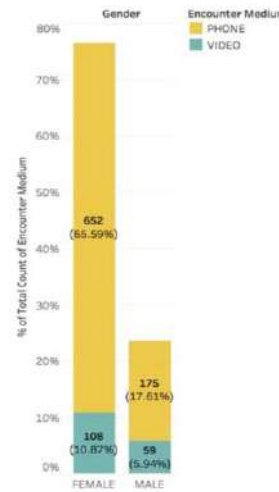


Figure 4. Preferred Methods of Encounters (Video vs. Phone) by Gender

Discussion

The major finding of this study showed 70% of the users were females. A likely explanation of this trend is women tend to seek care in general or are more open to integrating technology into their care seeking efforts. Another rational is that females may use VC to seek consultation or care for their children.

Furthermore, the choice of video or phone call varied by gender. Most female participants preferred a phone call over video, while males had a higher rate of video calls.

More research is needed to understand the reasons behind the choice of encounter. Future work will include the collection and analysis of broader data, such as chief complaint and diagnosis, to assess the efficacy of VC from clinical, technical, and operational perspectives.

References

[1] R. Roine, A. Ohinmaa, and D. Hailey, Assessing telemedicine: a systematic review of the literature, Cmaj 165 (2001), 765-771.

[2] A.C. Smith, M. Bensink, N. Armfield, J. Stillman, and L. Caffery, Telemedicine and rural health care applications, Journal of Postgraduate Medicine 51 (2005), 286.

[3] S. Khairat, T. Zaman,, S. Liu,, B. Edson,, R. Gianforcaro, Evaluating the Feasibility and Acceptability of Virtual Care Delivery (2018).

[4] R.H. Eikelboom, M.D. Atlas, Attitude to telemedicine, and willingness to use it, in audiology patients, Journal of telemedicine and telecare 11 (2005), 22-25.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190610

Smartphone-Based Self-Empowerment App on Secondary Prevention of Patients with Cardiovascular Disease

Yisi Liu[a], Jingyi Chen[b], Karen V Lamb[c], Peirong Wu[d], Polun Chang[d], Yanyan Cui[d], Ying Wu[a]

[a]*School of Nursing, Capital Medical University, Beijing, China*
[b]*Cardiovascular Unit, Peking First Hospital, Beijing, China*
[c]*Rush University College of Nursing, Chicago, IL, USA*
[d]*Institute of Biomedical Informatics, National Yang-Ming University, Taiwan*

Abstract

We developed a HeartGuardian app and explored its effects supporting people with CVD on lipid control and medication adherence. Fifty-seven patients were enrolled, 29 in the intervention group and 28 in the control group. The 12-week intervention resulted in a moderate improvement in lipid level and greater improvement in medication adherence (82.14% vs 37.93%, P=0.001). These outcomes translate into significant differences in occurrence of major adverse cardiac events (28.75% vs 72.43%, P =0.001).

Keywords:
Cardiovascular Diseases, Secondary Prevention, Medication Adherence.

Introduction

Cardiovascular disease (CVD) is a leading cause of death. High lipid level and poor adherence to CVD secondary prevention medications have been reported as the main risk factors of mortality in CVD [1]. Yet, lipid control rate and medication adherence are below ideal [1]. Mobile technology, such as smartphone application, might be part of a solution, as increasing evidence indicates that these apps support behavior change [2]. The aim of this study was to explore the short-term effectiveness of a HeartGuardian app in supporting people with CVD on lipid control and medication adherence.

Methods

In this twelve-week, two-armed, parallel, randomised control trial, fifty-seven patients with CVD were randomly assigned to either the intervention group (use of HeartGuardian app on smartphones and receive weekly text messages on health education) or the control group (receive weekly text messages on health education). The inclusion criteria were (1) hospitalized for CVD; (2) prescribed at least one CVD secondary prevention medication; (3) age 18 to 70; (4) owning a mobile phone with Android operating system; (5) competent to receive and read text messages. Patients who were critically ill, at an unstable or acute stage of illness, with mental illness, or unable to communicate were excluded. Approval of the study was obtained from the Capital Medical University Review Board.

The HeartGuardian app is designed based on behavior change theory [3]. Interventions integrated in the app include: (1) health education (tailored health plan and daily updates on lifestyle, diet, and treatment); (2) behavior interventions (medication reminders); (3) follow-up (medication recording and daily feedback. The app will remind the participants when it is time to take their medications and mark the medication as a "taken" or "missed" dose.); and (4) self-empowerment via automatic intelligent, real-time video feedback based on the subjects' adherence to taking medication. The videos show either improved or worsening vessel pathology based on the subjects' specific actions. The app is compatible with the Android operating system.

The primary outcome was the lipid level, including low-density lipoprotein cholesterol (LDL), total cholesterol (TC), triglyceride (TG), and high-density lipoprotein cholesterol (HDL). Secondary outcomes included medication adherence, and the occurrence of major adverse cardiac events (MACE) which are defined as the occurrence of unstable angina, ST-segment evaluation myocardial infarction, non-ST-segment evaluation myocardial infarction, and coronary revascularization. Primary and secondary outcomes were measured at the baseline and twelfth week by a research staff member who was blind to participants' random assignment. The baseline data regarding participants' demographics, medical and medication details were collected from participants' charts before discharge. Participants were required to visit the clinic at the twelfth week after discharge for routine health review and collection of follow-up data. The fasting blood samples of the participants were collected at their clinic visit for measurement of lipid levels. Medication adherence was measured by the Morisky Scale [4]. The occurrence of MACE was collected from participants medical charts.

Statistical analyses were conducted using SPSS 19.0 for Windows (SPSS Inc, Chicago, Illinois). The Chi-square test or Students t-test was used to test the difference between the two groups. A matched paired t-test and McNemar test were applied to test the differences within groups between the baseline and at the twelfth month follow up. Statistical significance was set at $P < 0.05$.

Results

Fifty-seven patients were enrolled with 29 in the intervention group and 28 in the control group. There was no significant difference between the two groups on the baseline data (data not shown here due to word limits). The mean (SD) age was 59.05 (7.23) years. 82.46% were male. 63.16% were retired.

71.70% of the participants were diagnosed with unstable angina and 88.30% had comorbidities.

As shown in Figure 1, no significant difference was observed in the baseline lipid levels between two groups. After the 12 weeks of intervention, both triglyceride and total cholesterol were significantly lower in the intervention group, compared with the control group (P=0.020, 0.014, respectively). Pre-post tests showed triglyceride level at the twelfth week was significantly lower than the baseline in the intervention group with a mean difference in change of -0.263 mmol/l ([95%CI, -0.040 to 0.487], P=0. 024), while it was significantly higher than the baseline in the control group with a mean difference in change of 0.378 mmol/l ([95%CI, 0.044 to 0.712], P=0. 029). No significant changes were observed with respect to low-density lipoprotein (LDL) and high-density lipoprotein cholesterol (HDL).

The baseline medication adherence was similar between the intervention group and the control group (43.33% vs 30.00%, P =0.284). After 12-week follow up, adherence increased significantly in intervention group (43.33% vs 82.14%, P =0.002), while the increase was minor in the control group (30.00% vs 37.93%, P =0.520). The difference was statistically significant, with the intervention group exhibiting higher medication adherence at the twelfth week than the control group (82.14% vs 37.93%, P=0.001).

Regarding the terminal clinical outcomes, the HeartGuardian app achieved a positive effect in lowering the occurrence of MACE in the intervention group, compared to the control group (28.75% vs 72.43%, P =0.001).

Discussion

Recent studies support the effectiveness of applications in improving lipid management and medication adherence [5,6]. But few studies assessed the effects on terminal clinical outcomes, such as MACE [5]. In the current study, a 12-week smartphone-based application intervention resulted in a moderate improvement in lipid level and greater improvement in medication adherence. These outcomes translate into significant differences in the occurrence of MACE. The possible reason that smartphone-based application improves the management of CVD risk factors might be that it enhances patients' self-efficacy and provides a form of social support [7]. This app features multiple integrated interventions and customized health education, which are reported to be a more effective approach to achieve behavior change than single interventions and generic health education [2]. This study has several limitations. One of the main limitations is the Hawthorne bias. Patients in the intervention group were aware that they were enrolled in a study and the improvements in lipid management and medication adherence could just be the result of better motivation in terms of a healthy lifestyle. We sent weekly text messages on health education to patients in both groups to minimize this bias.

Conclusions

A smartphone-based self-empowerment app could help patients with CVD better manage their dyslipidemia and improve medication adherence, and reduce the rate of MACE. This SmartPhone application appears valuable in CVD secondary prevention. Effects longer than twelve weeks are worth further investigations.

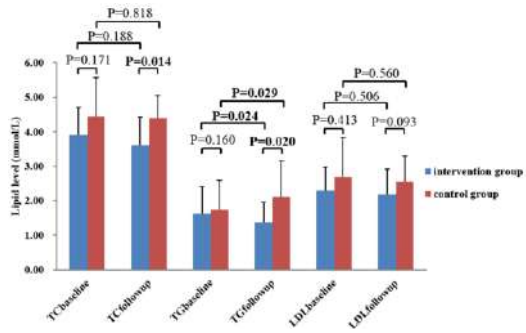


Figure 1. Effects of HeartGuardian Application on Lipid Levels. TC: total cholesterol; TG: triglyceride; LDL: low density lipoprotein; followup: at twelfth week after discharge.

Acknowledgements

This study was supported by research program 15HL06 from the Capital Nursing Foundation. The first two authors contribute the same to the study.

References

[1] M. F. Piepoli, A. W. Hoes, S. Agewall, C. Albus, C. Brotons, A. L. Catapano, M. T. Cooney, U. Corra, B. Cosyns, C. Deaton, and I. Graham, 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice, *European Heart Journal* 37(2016), 2315-2381.

[2] H. Eyles, R. McLean, B. Neal, Y. Jiang, R. N. Doughty, R. McLean, and C.Ni Mhurchu, A salt-reduction smartphone app supports lower-salt food purchases for people with cardiovascular disease: findings from the SaltSwitch randomised controlled trial, *European Journal of Preventive Cardiology* **24**(2017) , 1435-1444.

[3] M. Al-Durra, M. B. Torio, and J. A. Cafazzo, The use of behavior change theory in Internet-based asthma self-management interventions: a systematic review, *Journal of Medical Internet Research* **17**(2015), e89.

[4] D. E. Morisky, L. W. Green, and D. M. Levine, Concurrent and predictive validity of a self-reported measure of medication adherence. *Medical Care* **24**(1986),67-74.

[5] K. Santo, A. Singleton, K. Rogers, A. Thiagalingam, J. Chalmers, C. K. Chow, and J. Redfern, Medication reminder applications to improve adherence in coronary heart disease: a randomised clinical trial, *Heart* **105**(2018), 323-329.

[6] Y. Gandapur, S. Kianoush, H. M. Kelli, S. Misra, B. Urrea, M. J. Blaha, G. Graham, F. A. Marvel, and S. S. Martin, The role of mHealth for improving medication adherence in patients with cardiovascular disease: a systematic review. *European Heart Journal Quality of Care & Clinical Outcomes* **2**(2016), 237-244.

[7] de T. Jongh, I. Gurol-Urganci, V. Vodopivec-Jamsek, J. Car, R. Atun, Mobile phone messaging for facilitating self-management of long-term illnesses. *The Cochrane Database of Systematic Reviews* **12**(2012), Cd007459.

Address for correspondence

Ying Wu, School of Nursing, Capital Medical University, Beijing, China, 100069. helenywu@vip.163.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190611

Using openEHR's Guideline Definition Language for Representing Percutaneous Coronary Intervention Patient Safety Rules in a Dynamic Checklist System

Leixing Lu[a,b], Shan Nan[a,b,c], Sicui Zhang[a,b,c], Xudong Lu[a,b,c], Huilong Duan[a,b]

[a] *Department of Biomedical Engineering, Zhejiang University, Hangzhou, P.R. China*
[b] *Key Laboratory for Biomedical Engineering, Ministry of Education, P.R. China*
[c] *School of Industrial Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands*

Abstract

openEHR's Guideline Definition Language is designed for standardizing clinical decision support systems. In this study, we use Guideline Definition Language to represent patient safety rules in pre-operation of Percutaneous Coronary Intervention for the dynamic checklist system. After using Guideline Definition Language in this case, we had some results about its expression adaptability to requirements of patient safety rules.

Keywords:

Electronic Health Records, Patient Safety, Checklist

Introduction

The dynamic checklist system is a new type of guideline-based decision support system [1]. It can make judgments on actual data of patients with an execution engine and dynamically generates a personalized checklist accordingly. The core function of such system is based on implementation of rules. However, rules are closely associated with clinical data which makes the systems difficult to implement.

openEHR is an open standard maintained by openEHR Foundation. Its core design is to separate medical domain knowledge from specific clinical information by dividing models into two levels: archetype model and reference model. Its organizers nowadays pay attention to integration problem between clinical decision support systems (CDSS) and electronic health records (EHR). They also found that expressing and sharing computerized clinical decision support content across languages and technical platforms are a long term goal. Lack of commonly shared clinical information models and flexible support for various terminology resources have been identified as two main challenges of sharing decision logic across sites. Therefore, they proposed a new guideline definition language (GDL) and added it to the openEHR Foundation.

We carried out a case study for constructing clinical rules of the dynamic checklist system in preoperative stage of Percutaneous Coronary Intervention (PCI). We used GDL to express decision logic and build knowledge base for the dynamic checklist system. Also, we used openEHR's information models as data model to represent clinical data concepts used in this case. The data source used in this case is a clinical data platform, which is also based on openEHR foundation. Then, we combined them with an execution engine. Based on these steps, we also summarized some general methods about rule editing with GDL. Finally, we discussed advantages, limitations, and directions of its future development.

Methods

Data Models for Medical Concepts

First, we chose several archetypes based on openEHR corresponding to data requirements in the preoperative check procedure of PCI. Archetypes can be searched in the official platform named Clinical Knowledge Manager (CKM) by using keywords corresponding to domain medical concepts [2].

Edit GDL Rules

A complete GDL rule mainly consists of four parts, as shown in Figure 1:

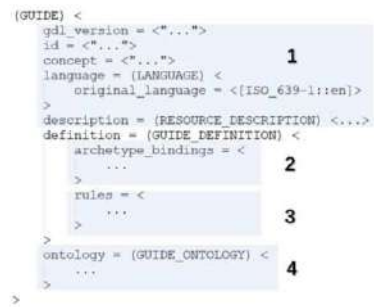


Figure 1– Basic structure of GDL

After editing a large-scale number of GDL rules, we concluded a general method for rule editing based on GDL syntax, which is mainly divided into four steps:

1. **Defining Metadata**: The metadata part is based on metadata description in the ADL, which includes mostly version information, language information, author information, keywords, rule usage and using scopes.
2. **Binding Archetypes and Elements**: We used local variables to associate with unique path of archetypes or elements.
3. **Authoring Rules**: The definition part of the rule is part of the GUIDE_definition module. Keywords WHEN and THEN are used to guide the condition and action of each rule. Priority is used for setting execution order of the rules.
4. **Defining Ontology**: The concept represented by **gt** code used in rule definition is indexed with clinical concept and can also support bindings of external term sets like SNOMED CT and ICD10.

Data Source and Engine for Rule Execution

In this study, our data source was a clinical data center platform called CDR based on openEHR. The CDR uses archetypes as the primary form for storage and access of clinical data. Data stored in CDR can be queried by the interface and returned in JSON format with mapping of path and value. The path corresponds to what we defined in the rules and the value is what we need related to the element in archetypes. By using the CDS Workbench developed by Cambio Healthcare [3], we translated the rule format from GDL to Drools. Then, we used Drools as the execution engine.

Results

The core requirement for constructing rule base is expression of a single rule. Rules were classified into six types and we made an analysis about the percentage of expressing result. The result is shown in Table 1.

Table 1– Results of rule construction

Type	Quantity	Examples	Results
Hospital procedures	5	Name, sex, age, deposit for surgery, medical insurance type.	100%
Surgical preparation	4	Medical orders and drug-related verification.	100%
General risk control	6	Blood potassium, left ventricular ejection fraction, allergies.	100%
Visual signs	8	Blood pressure, body temperature, pulse.	75%
Physical exams	6	Ultrasound, radiation.	67%
Laboratory tests	8	Blood, urine, stool, and blood coagulation routines.	50%

By executing these rules with Drools engine and running with data source from CDR platform [4], a personalized checklist was produced. Figure 2 shows one of the execution results.

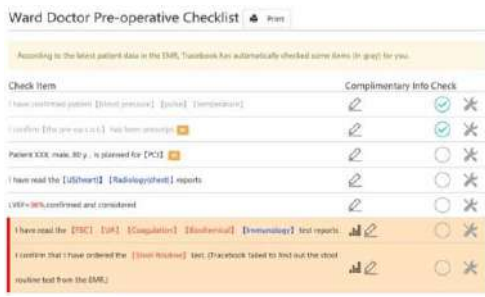


Figure 2– Execution result of the System

Discussion

According to requirements of rule expression of PCI preoperative surgery, we concluded several basic types we needed, mainly boolean expression, relational expression, logical expression, arithmetic expression, collection expression, interval expression, and data and time expression. After practicing rule construction, we found that GDL is designed to be user-friendly in expressing boolean, relation, logical, and arithmetic types; and lacking in expressing collection, interval, date and time. Collection and interval operations are important to dynamic checklist systems; keywords like foreach, where, size, contains and so on are urgently needed in cases of construction of patient safety rules. As for interval type, it can combine with data and time to express accurate date/time frame and interval without complex calculation.

The above results mainly focus on the condition part of a rule. Using GDL to fit the content of the action specification part has some shortcomings. It can hardly support our requirements totally. For example, we needed to add basic check items into a checklist template through execution of rules. As such, we needed some functions like "addTask()". Presently, we cannot find scalable functions in the action part of GDL.

Conclusions

In this study, we mainly discussed a case study about using openEHR's guideline definition language to express patient safety rules in pre-operation of PCI and its practical results. We conclude that GDL has better adaptability when constructing rules of simple expression form. At the same time, it can effectively solve the problem of data acquisition and interoperability between systems. However, the expression ability of GDL has certain limitations. Thus, it cannot completely meet requirements of dynamic checklist systems.

Future research will focus on design and implementation of a new type of guideline language to meet requirements of the dynamic checklist systems. Based on existing efforts, this new guideline language is expected to be robust, standards-based, open source tools supported, and clinically user-friendly.

Acknowledgements

Research leading to these results was funded by Chinese National Science and Technology Major Project (grant number 2016YFC0901703). The authors would like to thank their research group for their contribution to implementing the reported platforms.

References

[1] S. Nan, P.V. Gorp, X. Lu, U. Kaymak, H. Korsten, R. Vdovjak, and H. Duan, A meta-model for computer executable dynamic clinical safety checklists, *BMC Med Inform Decis Mak* **17** (2017), 170.

[2] L. Min, Q. Tian, X. Lu, and H. Duan, Modeling EHR with the openEHR approach: an exploratory study in China, *BMC Med Inform Decis Mak* **18** (2018), 75.

[3] N. Anani, M.V. Mazya, C. Rong, T.P. Moreira, O. Bill, N. Ahmed, N. Wahlgren, and S. Koch, Applying openEHR's Guideline Definition Language to the SITS international stroke treatment registry: a European retrospective observational study, *BMC Med Inform Decis Mak* **17** (2017), 7.

[4] L. Min, Q. Tian, X. Lu, J. An, and H. Duan, An openEHR based approach to improve the semantic interoperability of clinical data registry, *BMC Med Inform Decis Mak* **18** (2018), 15.

Address for correspondence

Corresponding author: Xudong Lu
Email: lvxd@zju.edu.cn
Department of Biomedical Engineering, Zhejiang University, Room 512, Zhouyiqing Building, Zhejiang University, 38 Zheda Road, Hangzhou, Zhejiang, China

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190612

Mindup: A Platform for Monitoring and Cognitive Enhancement for Patients with Alzheimer's Disease

Matheus Moreira Luna[a], Diogo Dantas Moreira[a], Fabio Abrantes Diniz[a]

[a] *Curso de Tecnologia em Análise e Desenvolvimento de Sistemas, Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, Cajazeiras, Paraíba, Brasil*

Abstract

Alzheimer's Disease (AD) is an illness that degenerates an individual's cognitive functions, leaving them unable to take care of themselves. Even without a definitive cure, AD should be treated with remedies and cognitive enhancement. This article presents an application that assists in the cognitive reinforcement of AD patients through games, supports the medical follow-up of patients, and facilitates the daily exchange of information between the caregiver and the doctor.

Keywords:
Smartphone; Alzheimer Disease

Introduction

Alzheimer's disease (AD) is the most common form of dementia, characterized by being a disease that destroys nerve cells, disrupting synaptic connections, causing severe memory loss in addition to other symptoms that worsen over time [1]. Currently, AD has no cure but can be treated through medication and cognitive enhancement activities to prevent disease progression [2].

In the literature, we have found several applications that may help in cognitive enhancement in patients diagnosed with AD through games, however, none of the applications presented a dedicated platform to specialized observation of patient's performance in the games by the doctor, that is, it does not provide a form of communication between them and the caregivers.

As a way to help doctors, caregivers, and patients with the initial level of Alzheimer's disease to have better interactions, this work presents an application called "MindUp" for mobile devices and web, where it's possible for patients to have access to specific games of cognitive exercise. These exercises serve as daily reinforcement of their functions in a controlled way and the caregiver can make out daily reports of the patient and assist the doctor monitoring the evolution of the disease.

Methods

This work began with a bibliographical review of studies done by neurology experts on Alzheimer's Disease and its predictions for the next years and generations, as well as informal interviews with neurology professionals to try to validate the idea and collect system requirements. In this context, the stages of the software development process, based on the ideas of the Scrum process [8] and adapted to the text of the work, were started. The requisites collected for the system were: A) Medical account management; B) Caregiver account management; C) Patient account management; D) Cognitive reinforcement exercises (games); E) Results; F) Exercise performance monitor (for doctors); G) Caregiver's Reports history and H) Daily Report Fill.

A comparative analysis was made with 4 applications that serve the same purpose: "Fit Brain Trainer" [3]; "Lumosity" [4]; "Elevate" [5] and "Peak" [6]. These applications are the ones closest to the approach proposed by this work. The comparison can be seen in Table 1, where C indicates that the application has the associated requirement and N/C where it doesn't. When observing the collected data, the proposed system has as a differential the functions related to A, B, F, G and H requirements in relation to the related applications.

Table 1 – Applications comparison

(R)	[3]	[4]	[5]	[6]	M
A	N/C	N/C	N/C	N/C	C
B	N/C	N/C	N/C	N/C	C
C	C	C	C	C	C
D	C	C	C	C	C
E	C	C	C	C	C
F	N/C	N/C	N/C	N/C	C
G	N/C	N/C	N/C	N/C	C
H	N/C	N/C	N/C	N/C	C

Based on these requirements, the "MindUp" system, composed of three dedicated interfaces for each user role (Doctor, Patient, and Caregiver), is being implemented. Figure 1 shows the activity flow of the actors in the proposed system. Subsequent sections explain in detail each step of this flow.

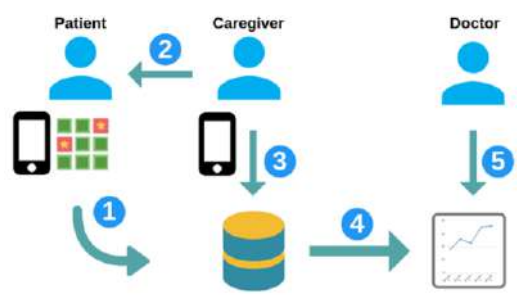


Figure 1 – Activity flow on MindUp

For the patient, the application has a simple and intuitive interface inspired by the guide to serious games for cognitive treatment [9], where the patient will have access to the game of cognitive reinforcement. In "MindUp", a memory game is used and can only be accessed at a determined time, which is

configured by the doctor, so that there is no overload of exercise and stress to the patient.

In step (1), the patient must access the platform at the time specified by the doctor and perform a round of play. Each round consists of 3 levels of difficulty, the first with 3 pairs of cards, the second with 6 pairs and the third level with 8 pairs.

For the caregivers, we used the test to catalog the daily behavior of the patient, which is observed by the caregiver in step (2), and capture details of the patient's daily life.

In step (3), the caregiver will be presented to 10 multiple choice questions that were developed by neurologist James E. Galvin, whose main purpose is to detect the level of the disease in patients, applying the questions seasonally [7].

The doctor can register their patients on the platform, giving each of them a credential, and configure the patient's times of use of the application. Subsequently, after processing the data collected by the patient's use and by the caregiver's follow-up, the system will generate data on the patient's daily income, as demonstrated in step (4).

The daily yield (Dy) is measured by the average points of each session (S) marked by the patient on a given day. The score of each session (S) is made by the average score (Pt) of each level of the game, which is calculated by the number of pairs of cards marked (x) times 10 divided by time (t). Therefore, the calculation made to gauge the yield is given by the following formulas adopted by the author 1)Ap = (10 * x) / t; 2)S = (Ap1 + Ap2 + Ap3)/3 and 3)Dy = (S1 + S2 + S3)/3

In addition, the doctor can access reports produced by the caregivers and can access the last one produced or of a certain date. This process is represented in step (5)

Results

Currently, the project is undergoing validation, prototyping, implementation, collection, and analysis of new requirements, as well as refinement of existing ones, and improvements in system architecture.

Figure 2 shows the main functionalities of the system, already implemented. The interface (1) demonstrates how its user interface (UI) is proposed by the Caregiver, who will answer the multiple choice questions; the UI (2) already demonstrates a use of the game of memory that will be accessed by the patient; Finally, the UI (3) demonstrates the proposal of the physician interface, which can be used to monitor the patient's treatment.

MindUp attempts to bring greater interaction among doctors, caregivers, and patients, collecting information onthe daily and specific behavior that may occur on sporadic days, and may encourage further research.

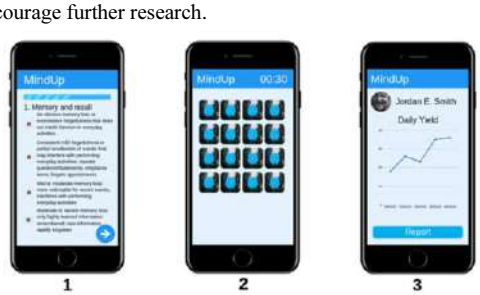


Figure 2 – UI for different roles in Mindup

Through informal interviews with neurologists, a great deal of acceptance was noted by practitioners. It was noticed that the application "MindUp" is useful those involved with Alzheimer's disease treatment, and therefore, we had a greater interaction of professionals in this project, where they were always giving ideas of what the proposed system could do.

The system proposed here differs from other related platforms in that "MindUp" proposes to be an open platform and most of the features that its competitors promise are paid. None of the related applications showed any interactivity with physicians and caregivers, showing the results of the training only for the end-user, and also does not give the caregiver the ability to record daily patient behaviors

Conclusions

The "MindUp" system still has the potential for growth, mainly because it has partnerships with professionals to understand, collect and validate requirements, which can bring new functionalities and integrating other types of cognitive reinforcement exercises.

Continuous use of the system, which has not yet been carried out, and a follow-up of a considerable number of patients can give new visions about the system and prove its relevance in the face to other available solutions.

Acknowledgments

This study was supported by Prof. Carlos Kennedy and Universidade Federal de Campina Grande (Cajazeiras, PB, Brazil).

References

[1] B. Dubois, Revising the definition of Alzheimer's disease: a new lexicon, 9 (2010). doi:10.1016/S1474-4422(10)70223-4.

[2] Alzheimer's Association, Treatments, 2019. https://www.alz.org/alzheimers-dementia/treatments (accessed April 1, 2019).

[3] Fit Brains Trainer, Fit Brain Trainer, 2019. https://fit-brains-trainer.br.uptodown.com/android (accessed April 1, 2019).

[4] Lumosity, Lumosity, 2019. https://www.lumosity.com/en/ (accessed April 1,2019).

[5] Elevate, About Elevate, 2019. https://www.elevateapp.com/about (accessed April 1, 2019).

[6] Peak, About Peak, 2019. https://www.peak.net/about (accessed April 1, 2019).

[7] J. Galvin, The Quick Dementia Rating System (QDRS): A rapid dementia staging tool, 1 (2015) 249-259. doi: 10.1016/j.dadm.2015.03.003

[8] C. Brod, Scrum: Guia Prático para Projetos Ágeis, 2nd ed., Novatec, Brazil, 2015.

[9] B. Bouchard, F. Imbeault, A. Bouzouane, and B. Menelas, Developing Serious Games Specifically Adapted to People Suffering from Alzheimer, (2012) 243-254. doi: 10.1007/978-3-642-33687-4_21.

Address for correspondence

Matheus M. Luna,
matheus.luna@academico.ifpb.edu.br
Engenheiro Coelho Sobrinho St., 292, Cajazeiras, Paraiba, Brazil

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190613

Role of Nursing Informatics in Implementation of SNOMED-CT in India

Sweety Rai [a], Owais M Siddiqui [a], Susil K. Meher [a], A Shariff [a], Shefali Banga [a]

[a] All India Institute of Medical sciences, Ansari Nagar, New Delhi, India-1

Abstract

SNOMED-CT project under the Ministry of Health and Family Welfare is operational at AIIMS since July 2016. A team of nurses were recruited under SNOMED project who actively works for integrating existing EHR with SNOMED-CT, monitoring, training of users auditing the data, resets creation and development of National Drug Database.

This paper emphasizes role of Nursing Informatics in implementation of SNOMED-CT project in India as well as in any other country.

Keywords:
SNOMED-CT, Nursing Informatics (NIS)

Introduction

SNOMED-CT is the foundation for standardizing clinical content in Electronic Health Records (EHR) that contains concepts with unique meanings along with specific codes and formal logic based definitions organized into hierarchies.

Nursing informatics (NIS) is a field of nursing that incorporates nursing, computer, and information sciences to maintain and develop medical data and systems to support the practice of nursing, and to improve patient care outcomes.The quality of care depends on effective communication among healthcare providers, since healthcare providers communicate primarily through the notes they write in a patient's health record/in-patient chart, nurse informaticist seek to continually improve the speed, efficiency, timeliness, accuracy of patient charting, reducing costs and thereby enchancing patient care.

Methods

SNOMED-CT project under the Ministry of Health and Family Welfare is operational at AIIMS since July 2016. The team of nurses and a programmer were recruited under an exsiting SNOMED project. The programer actively works on integrating existing EHR with SNOMED-CT, monitoring, training of users auditing the data, reference set creation and development of a national drug database .

AIIMS is the premier medical institute of India and it conducts teaching programs in medical and para-medical courses both at undergraduate and postgraduate levels and awards its own degrees.

The NIS team working for SNOMED-CT at AIIMS, NEW DELHI has accomplished the following task in order to enhance the implementation of SNOMED-CT.

- Twenty six reference sets were created and validated based on the NHP (National Health Program) requirement.
- LOCAL REFSETS: Apart from these, the department wise local Refsets are also prepared in the Refset Management Tool by the AIIMS SNOMED team
- Validation of some Refsets are in process (Future Targets)
- Creation of The National Drug Library Database has been done and it includes, name of the drug with SNOMED-CT code, Manufacturer code, Drug classification which includes a SNOMED-CT code and a Medicine brand name (1,69,000 drugs have been mapped with SNOMED-CT codes)
- SNOMED-CT coding done for 4000 nationalized lab investigations given from ministry
- SNOMED-CT coding for the diagnoses collected from clinical areas of the hospital
- SNOMED Integrated EHR : Various online applications running in AIIMS have now been fully integrated with SNOMED-CT which are as follows:
  - National death registry of India (NDRI)
  - e-birth module (Online birth certificate)
  - e-MLC (Online medico legal certificate)
  - e-vitals Entry for Nurses
  - e-Hospital-Admission Discharge Transfer
  - PDS (patient display system)
  - e-Blood request
  - OT module
  - Online Admission Requisition
  - Quality Assurance Module for academics
  - Heart Transplant Registry

Results

The NIS team working for SNOMED-CT at AIIMS, NEW DELHI has been successful in creation and validation of 26 Refsets, along with Local Refset. The creation of National Drug Library Database has been done and 1,69,000 drugs have been mapped with SNOMED-CT codes, 4000 nationalized lab investigations and diagnoses have been mapped. Along with this SNOMED integrated EHR are running in AIIMS. NIS have

been successfully carrying out the ongoing implementation of SNOMED-CT.

Discussion

To make electronic health record meaningful and useful we need a standard code of terminology. SNOMED-CT provides the same, however before implementing it we need to customize it as per national requirement.

For the customization of SNOMED-CT as per national requirement comes the role of NIS.They are a neccesity in the implementation of the same .

Conclusions

Nursing informatics (NIS) is an intersection between patient care and patient health information and is emerging as the backbone for bridging the gap to implement **SNOMED-CT**. As a result, this standardized reporting helps in reducing lot of inaccuracies and thereby ensures comprehensiveness in recording the patient's entire health information.

References

[1] Working Together: SNOMED International and the World Health Organization and the SNOMED-CT to ICD-11-MMS Map Position Statement.
[2] Pickering, B.W., Dong, Y., Ahmed, A., Giri, J., Kilickaya, O., Gupta, A., Gajic, O. and Herasevich, V., 2015. The implementation of clinician designed, human-centered electronic medical record viewer in the intensive care unit: a pilot step-wedge cluster randomized trial International journal of medical informatics, 84 (5), pp.299-307.
[3] Kate RJ. Towards Converting Clinical Phrases into SNOMED-CT Expressions. Biomedical Informatics Insights 2013 6(Suppl. 1):29-37.

Address for correspondence

Lt Col (Dr) Owais M Siiddiqui.
Room No 21.
Computer Facility,
All India Institute Of Medical Sciences
Ansari Nagar, New Delhi-1
drowaismsid@gmail.com
.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190614

Development of Augmented Reality in Learning for Nursing Skills

Yukie Majima[a], Seiko Masuda[a], Takeshi Matsuda[b]

[a] Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University, Sakai, Osaka, Japan,
[b] Department of Information Security, University of Nagasaki, Nagasaki, Japan

Abstract

This study was designed to develop a wearable learning support system that enables novices to learn the skill for blood withdrawal while imitating its images displayed in front of them. For them to master "tacit knowledge," such as "proficient art" and "knacks," in intravenous injection and blood drawing techniques. This paper outlines a learning support system incorporating augmented reality (AR), which we have developed based on earlier studies.

Keywords:

Augmented Reality, Nursing Skills, Tacit Knowledge

Introduction

Along with the increasing sophistication of medical care, it is important to foster "nursing practical knowledge" with a view of nursing that gives due consideration not only to nursing skill proficiency, but also to emotional aspects of work, such as intentions of patients and their families in individual and diverse situations. Nevertheless, such practical knowledge of nursing is highly tacit in nature. It is therefore difficult to verbalize [1]. Furthermore, because nursing practice at clinical sites is conducted in a closed environment, no opportunities exist for it to be disseminated or for the information to be shared; few opportunities exist for the practices to be refined.

Other studies investigating the area of video reflexive ethnography have indicated a study method with which structural changes to clinical communication were achieved for reshaping ICU ward round practices [2]. Another study explored how clinicians used the built environment to provide safe communication in an ICU [3].

In the area of leadership education, Gordon et al. [4] reported that being able to "see" oneself at work gave participants the opportunity to discuss and analyze their everyday leadership practices and challenge some of their sometimes deeply entrenched values, beliefs, practices, and assumptions related to healthcare leadership.For those reasons, the authors have conducted research continually since 2008, specifically examining the tacit nature of nursing skills, for which efforts have been undertaken to visualize manuals for nursing skills at hospitals and to share them using an e-learning system.

Into some specific types of blood vessels, novices have difficulty inserting a needle. This study was undertaken to develop a wearable learning support system that enables novices to learn the "art" in blood vessels while imitating its images displayed in front of them. For them to master "tacit knowledge," such as "proficient art" and "knacks," in intravenous injection and blood drawing techniques, we specifically examine fingers holding no tool, those which play a supportive role, which has been a blind spot as a subject for study. To that end, we are carrying out this study using the following three approaches.

(1) Wearing sensors on supportive-side fingers for intravenous injection techniques to analyze the contact force and facilitate the insertion of a needle into a blood vessel.

(2) Consideration of a method to visualize learning contents to ease understanding of techniques.

(3) Consideration and development of a system to support real-time learning using wearable glasses.

We are now analyzing contact pressure and skin extension pressure of supportive fingers and comparing the characteristics of experts and novices in relation to (1), concurrently with the pursuit of (2). Regarding (3), we have developed a learning support system prototype using augmented reality (AR). This paper presents an outline of the system.

Earlier Studies

The authors have conducted the following studies to date, devoting particular attention to the tacit knowledge of nursing skills.

(1) We conducted an interview survey and obtained the following findings: a) Many nurses believe that, in most cases, once they have identified a vein, they are able to perform an intravenous injection. However, they just verbally express the moment when they have done it well as "feeling like entering a blood vessel 'smoothly ('kukutto' or 'sutto' in Japanese)'". It is difficult to express exactly how they feel [5]. b) Novice nursing students consider that the knack for techniques is to remember procedures [6]. It is therefore important for the process of learning support to compel novices to master procedures before their advancement to sharing the performance characteristics of experts.

(2) Having analyzed lines of sight in practicing intravenous injection techniques, we found that line-of-sight flows differed between novices (nursing students) and experts (nursing personnel) and that experts move their lines of sight to the next practice (prior treatment) [7].

(3) We developed a reflective learning support system that supports to review skills based on the procedures and the line-of-sight flows [8] and a system that can reproduce finger movements in practicing skills in computer graphics using motion capture for fingers [1].

(4) To extract tacit characterization data of nursing skills from different perspectives, we analyzed "brain waves" and "heart rate" data of practitioners and clarified the state of relaxation and that of tension in the process of practicing nursing skills. Subsequently, the characteristics of "techniques" were compared between experts and novices [9].

(5) We analyzed tuning information (pull-in effect) related to biological rhythms (brain waves and heart rate) of nurses and patients on the hypothesis that the good or bad mutual relationship between a nurse and a patient might be related to the proficiency of techniques. Results suggest the possibility that nurses' states of relaxation might cause patients' state of relaxation [10].

In injection skills education, both instructors and learners are typically highly interested in how to handle a syringe. In fact, the know-how to support the learning has been accumulated. However, only purposes such as "fixation and extension of a blood vessel" are presented in terms of the techniques of left fingers matched to the characteristics of blood vessels of individual patients for whom giving an injection is difficult. The characteristics include a rolling over-and-over ('korokoro' in Japanese) of a moving blood vessel. How to fix and extend a blood vessel has not been fully taught.

In blood drawing and intravenous injection techniques, fingers holding no tool that play a supportive role facilitate the insertion of a needle into a blood vessel. For this reason, we aim to develop a learning support system that particularly addresses those factors which play a supportive role (fingers that facilitate inserting a needle into a blood vessel) other than the main factors and processes. We compared the characteristics of experts and novices to observe the contact pressure of supportive fingers and found that experts practice skills with a more stable contact pressure [11] and skin extension pressure [12] than novices do.

Wearable AR System for Learning

In conventional education for nursing skills, novices typically first learn the purpose and procedures of the skills in lectures. Then they actually undergo training individually or as a group after having seen demonstrations in technical seminars. Because novice nursing students consider that the knack for techniques is to remember procedures, instructors have been required to try to provide easy-to-understand demonstrations. In recent years, when ICT has developed, video learning and eLearning that can be learned anywhere, anytime has become possible. Nonetheless, in actual nursing skills training, learners often practice while watching a monitor that displays video images. Real-time learning is difficult because of line-of-sight flows and other problems. No image clues are found in terms of fingers holding no tool, which might play a supportive role in a blind spot, but which might be a subject for study. It is particularly problematic to learn the "art" for blood vessels into which it is difficult for novices to insert a needle.

For that reason, we first developed a learning support system using AR (Argument Reality) in which learners can learn experts' nursing skills without moving their lines of sight.

Figure 1 portrays a specific screen for learning. When practicing skills training, learners can learn skills by following and imitating (tracing) the images of experts' techniques that are displayed transparently in front of them in real time. The prototype system verified that training is possible by overlaying images on a simulation arm model.

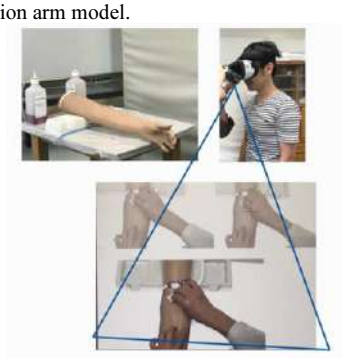


Figure 1- Wearable learning system using AR: Learners obtain skills by watching and imitating images of experts' skills projected on an actual simulation model.

Conclusions

We described the development of a wearable system for the learning of tacit knowledge of injection skills, following our research conducted to date. Future challenges include the following. Because adjustment of overlay images is necessary before training, additional efforts must be undertaken in overlaying experts' (others') images and beginners' images of skills, including the determination of object positions and how to imitate objects. Furthermore, we will compare practice performance of students using AR and those not using AR to be clearly seen whether AR makes a difference, for example, whether the time to achieve proficiency of techniques reduces by using AR.

Acknowledgements

This work was supported by JSPS KAKENHI (JP15K15805).

References

[1] Y. Majima, Y. Maekawa, and M. Soga, Learning Support System Reproducing Finger Movements in Practicing Nursing Techniques, *Proc. 11th International Congress on NI* (2012), 278-282.

[2] K. Carroll, R. Iedema, and R. Kerridge, Reshaping ICU ward round practices using video-reflexive ethnography, *Qualitative Health Research*, 18.3 (2008), 380-390.

[3] S.Y. Hor, R. Iedema, and E. Manias, Creating spaces in intensive care for safe communication: a video-reflexive ethnographic study, *BMJ Quality & Safety*, 23.12 (2014), 1007-1013.

[4] L. Gordon, C. Rees, and J. Ker, Using video-reflexive ethnography to capture the complexity of leadership enactment in the healthcare workplace, *Advances in Health Sciences Education: Theory and Practice*, 22.5 (2017), 1101-1121.

[5] Y. Majima, and Y. Maekawa, Analysis of nurses' knacks in injection skills: An interview survey. *Proc. 28th Academic Conference Japan Academic of Nursing Science,* (2008), *p.299*.

[6] Y. Maekawa and Y. Majima, Comparison of consciousness between expert nurses and nursing students for knacks in practicing intravenous injection skills: An interview survey. *Proc. 29th Academic Conference Japan Academic of Nursing Science,* (2009), *p.315*.

[7] Y. Maekawa and Y. Majima, Comparison of vein skills between nurses and nursing students: Focusing on lines of sight and nursing skill procedures. *Proc. 41st Annual Conference of Japanese Society for Information and Systems in Education,* (2009), *pp.28-29.*

[8] Y. Majima, M. Sakoda, Y. Maekawa, and M. Soga, Evaluation of Nursing Skills Acquisition of Reflective e-Learning System for Nursing Students by Different Learning Methods, *Proc. 20th International Conference on Computers in Education*, (2012) pp. 460-467.

[9] Y. Maekawa, Y. Majima, T. Kawano, M. Katagiri, and M. Kawasaki, Analysis of practical nursing knowledge from biological data heart rate variability and drawing blood, *Proc. 32nd Joint Conference on Medical Informatics*, (2012), *pp. 604-607*.

[10] T. Kawano, Y. Majima, Y. Maekawa, M. Katagiri, and A. Ishigame, Inter-brain Synchronization between Nurse and Patient During Drawing Blood, *HEALTHINF*, 5, (2016), pp. 507-511.

[11] T. Matsuda, Y. Majima, and K. Ootani, Pressure data analysis for vascular fixation in venipuncture, *IEICE Technical Report, MICT2017-41*, (2018), pp. 1-3.

[12] N. Ueda, M. Izumi, Y. Majima, T. Matsuda, and Y. Maekawa, Correlation between left hand contact force and skin development in injection technique, *MICT2017-41*, (2018), pp. 11-16.

Address for correspondence
Yukie Majima, Email: majima@kis.osakafu-u.ac.jp

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190615

Adding Instructiveness to IHE-XDS Based Electronic Health Records

Patrick Mangesius[a,c], Dmytro Rud[e], Samrend Saboor[b], Thomas Schabetsberger[a,c,d]

[a] *ITH icoserve GmbH, Innsbruck, Austria,*
[b] *Siemens Healthineers GmbH, Erlangen, Germany*
[c] *University of Health Sciences, Medical Informatics and Technology, Hall, Austria*
[d] *Scientific society for innovative health information systems, Mils, Austria*
[e] *Swiss Post, Bern, Switzerland*

Abstract

eHealth system users need adequate, timely access to health-related information; including the functionality of alerting and notifications on events. This was addressed through requirements engineering techniques. The analysis resulted in a requirements' list and architecture based on IHE standards. Patients, professionals and administrators were identified as main actors, core notification attributes were type and transport channel, and notifications always event-triggered. One challenge is system-based data access events, potentially leading to an invalid triggering of events.

Keywords:
Electronic health records, access to information, health personnel

Introduction

Social services and health care, are areas of welfare expected to benefit extensively from digitalization [1,2]. Both sectors require a large public investment and are areas that most citizens come in contact with over their lifetime. For eHealth system users it is crucial to be able to access health related information in an adequate and timely manner. Patients are increasingly engaged in their treatment process, transforming healthcare IT into a personalized, largely connected and patient empowering infrastructure [3]. Improving the continuity of care and establishing an effective cooperative health care is expected to reduce costs. To address this challenge, electronic health record (EHR) systems are introduced to close the gap between institution-specific patient data and a comprehensive, longitudinal collection of health related patient data [4]. Patient access to the EHR and interactive messaging functions may improve communication, patient empowerment, engagement, and self-management [5]. Those challenges and promises are drivers of many eHealth projects around the globe focusing on introducing regionally/nationally wide and connected, standards-based EHRs [6]. The underlying standards have also been endorsed and promoted by the European Union [1]. However, a recently published study on the status of EHR implementations within the European Union [7] shows that introducing the infrastructure for a standards-based EHR is not enough for a successful implementation and acceptance by the major key players in the health care system. Especially the usage by medical practitioners, and therefore the primary and secondary usage of EHR data, is in several countries an aspect receiving a low rating. One important aspect for physicians regarding the usage and acceptance of eHealth systems is the functionality of alerting and notifications on events on or within data of the EHR system [8]. Focusing on this aspect, this paper has the following goals: 1. To identify requirements for an alerting and notification system expected to be provided by a state-of-the-art EHR system; 2. To elaborate on a solution on how to incorporate such a system in an already existing IHE based EHR system.

Methods

Our approach was to address this topic by using requirements engineering techniques focusing on the business analysis and project definition phases [9]. We did not address the project execution and maintenance phase, since implementing a prototype was not within the scope of this paper. This will be a part of future research. Therefore, our research evolved through three phases leading to a list of requirements and an architecture based on IHE standards: **conception** (business analysis), **architecture and workflows** (project definition) and **validation** of the architecture against requirements.

Results

Requirements for Notification Systems in EHRs

Based on the gathered material and information, three rough types of actors of an eHealth system were identified:

- Patient: a person with a medical history. Related PHI data for the identified person is stored within the EHR. This Actor has no or limited medical knowledge and background.
- Professional: Professional user of the eHealth system connected with zero or several patients. The usage of the system is within the scope of the treatment process of associated patients.
- Administrator: Technical user with no medical background. This person is responsible for maintaining and operating the eHealth system and its functionality.

Since this paper targets the acceptance and requirements of an eHealth or EHR system for its end users, the actor "Administrator" was not considered. In contrast to our initial knowledge, the conception phase showed that the requirements on the level of which notifications are defined and received are identical for the Actors "Professional" and "Patient". For both

actors the requirements can be arranged into two core functionalities: "Define Notification" and "Receive Notification", where the latter is a consequence from the attributes defined for the former. Therefore the core characteristics of a notification are the type and transport channel.

Notification Types:

- Instant Notification: Notification of this type is submitted immediately, without non-technical delay, to the intended recipient.
- Aggregated Notification: This type of notification describes a summary of the occurrences of a certain event type. The timeframe of the summary is expected to be configured when defining the Notification.

Notification Channels:

1. In-Application: The information of the event is displayed within the application of the user. This could be a Primary System of the physician, a portal or web page. The notification is delivered as part of the application's user workflow in the user interface.
2. E-Mail: The notification is delivered in an E-Mail inbox. The transport media as well as the receiving party is expected to be insecure or not reliable.
3. Secure E-Mail: This channel is similar to the E-Mail channel, however, transport and storage are secure through state of the art security mechanisms such as encryption and receiving confirmation.
4. SMS: Delivery of the information to a mobile phone using SMS technology. Similarly to the E-Mail notification, this is not regarded as a secure channel.
5. Letter: The notification is delivered in paper form using postal deliveries. Depending on the service, this can be a secure or insecure channel.

Notification types and channels are closely related to each other since, for obvious reasons, not each Notification type can be transported through each notification channel, e.g. a Letter-based notification channel cannot transport instant notifications. Additionally, the channel has a severe impact on the notification itself. Less secure channels (e.g. E-Mail, SMS) can only transport general information events, whereas secure channels (Secure E-Mail, In-Application) are capable of transporting more detailed information such as patient names or information about the data that caused the event. This also has a direct impact on the usefulness of the notification.

Notifications are always triggered by events. Two event categories could be identified:

- User Interaction Events: Events triggered by users of the EHR (e.g. A physician accesses data of a patient)
- Data Processing Events: Events triggered by data within the EHR (e.g. a document was submitted containing critical information such as an important laboratory parameter outside its ranges)

Events are the only driver of notifications. The events are bound to the functionality of the eHealth system. In this paper, we used Siemens Healthineers eHealth Solution as an exemplary eHealth and EHR system.

Architecture for Notifications in IHE based EHR systems

As discussed previously two event categories could be identified: 1) user interaction events and 2) data processing events. In IHE based EHR systems, events are audited using the IHE ATNA (Audit Trail and Node Authentication) profile. The audit events are stored by an IHE Actor called Audit Record Repository (ARR). Our architecture uses this as the main source of triggering events. In order to generate notifications triggered by the IHE ATNA audited events, we introduce a Service called Notification Service. The ARR, acting as an Audit Record Forwarder, submits relevant events as ATNA based messages to this newly introduced service. Therefore the Notification Service is required to be capable of receiving ITI-20 transactions.

Conclusions

Based on the requirements, transaction flows and nature of IHE based eHealth systems we conclude that it is possible to enrich existing IHE based eHealth infrastructures with a notification system that is based on IHE ATNA messages and automated fetching of additionally needed information through standards-based transactions.

References

[1] The European Commission, COMMISSION DECISION (EU) 2015/1302 of 28 July 2015 on the identification of 'Integrating the Healthcare Enterprise' profiles for referencing in public procurement, *Off. J. Eur. Union*. **L 199/43** (2015).

[2] Vision for eHealth 2025, 2016. https://www.government.se/4a3e02/contentassets/b0fd09051c6c4af59c8e33a3e71fff24/vision-for-ehealth-2025.pdf.

[3] W.H. Frist, Connected health and the rise of the patient-consumer, *Health Aff.* **33** (2014) 191–193.

[4] F. Burns, Information for Health: An Information Strategy for the Modern NHS 1998–2005. A national strategy for local implementation, *Dep. Heal. Publ.* (1998). doi:10.1007/s13244-014-0339-z.

[5] C. Rathert, J.N. Mittler, S. Banerjee, and J. McDaniel, Patient-centered communication in the era of electronic health records: What does the evidence say?, *Patient Educ. Couns.* (2017). doi:10.1016/j.pec.2016.07.031.

[6] BGBl., Bundesgesetz: Elektronische Gesundheitsakte-Gesetz ELGA-G, **BGBl I 111** (2012).

[7] A.S.O. Gonçalves, N. Bertram, and V.E. Amelung, European Scorecard Zum Stand Der Implementierung Der Elektronischen Patientenakte Auf Nationaler Ebene, 2018. https://www.stiftung-muench.org/wp-content/uploads/2018/09/Scorecard-final.pdf.

[8] C. de Grood, A. Raissi, Y. Kwon, and M.J. Santana, Adoption of e-health technology by physicians: A scoping review, *J. Multidiscip. Healthc.* (2016). doi:10.2147/JMDH.S103881.

[9] C. Ebert, Understanding the product life cycle: Four key requirements engineering techniques, *IEEE Softw.* (2006). doi:10.1021/pr050271e.

Address for Correspondence

Patrick Mangesius
patrick.mangesius@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190616

An Innovative Platform to Analyze Heart Failure Biomarkers in Saliva

A. Martínez-García[a], P. García-Ocaña[b], S. Rodríguez-Suárez[b], C.L. Parra-Calderón[a]

[a] *Group of Research and Innovation in Biomedical Informatics, Biomedical Engineer and Health Economy, Institute of Biomedicine of Seville, IBiS/"Virgen del Rocío" University Hospital /CSIC/University of Seville, Spain*
[b] *Integral Healthcare Unit, "Virgen del Rocío" University Hospital, Seville, Spain*

Abstract

This paper explains a methodology to improve patient safety through early detecting of HF complications and analyzing HF biomarkers in saliva obtained using a biosensor developed in HEARTEN project. A pilot study was performed in two hospitals in Spain and Italy respectively. A direct correlation was identified between TNF-alpha levels in saliva and weight. The weight gain in HF patients could predict a HF decompensation, consequently TNF-alpha could be a new biomarker of these decompensations.

Keywords:

Heart Failure, Biomarkers, Telemedicine

Introduction

Heart Failure (HF), a cardiovascular chronic disease, has been rapidly increasing, positioning this disease as a main cause of mortality and poor quality of life in western societies [1].

For this reason, in the beginning of 2015, the consortium of HEARTEN project [2] started working to develop and validate a co-operative ecosystem of tools (Clinical Decision Support - CDS- tools directed to healthcare professionals, and empowerment tools directed to patients and caregivers), with the aim to improve the HF self-management for patients, and the HF management for all the actors that participating in HF care, both regarding the adherence and compliance of drugs, diet, and physical exercise recommendations. As part of this ecosystem of tools, one of the CDS tools developed by the consortium, is a saliva biosensor to obtain cortisol and TNF-alpha HF biomarkers [3]. This couple of biomarkers are very interesting to know the HF status, and currently could be obtained taking a blood sample from the patient. HEARTEN proposal is to obtain the same biomarkers only taking a saliva sample from the patient, technique much less invasive. Our hypothesis on that subject is that a correlation between TNF-alpha and/or cortisol and HF decompensation prediction could be identified obtaining these biomarkers from saliva.

The work presented in this paper is related to the patient safety, one of the project outcomes, because the aim is to early detect HF complications analyzing saliva samples.

Methods

To validate the ecosystem of tools, a pilot study has been performed, including the collection and analysis of saliva samples. Two clinical sites participated in this study: 'Virgen del Rocío' University Hospital (VRUH) of Andalusian Health Service in Seville, Spain, and Cardiology Unit of the University Hospital (CUUH) in Pisa, Italy. Before the pilot study started, a clinical protocol was agreed between the involved partners, and approved by the ethical committees.

The pilot study started in July 2017, and finalized in March 2018. In this period, VRUH and CUUH recruited 29 and 38 patients respectively in the intervention group of the study. A 3-months follow-up was performed for each patient, validating the interactive ecosystem, and monitoring the specific saliva biomarkers. Patient used the ecosystem at home. A nurse visited patients one/two times a week, to solve doubts in the use of the ecosystem, and to collect saliva samples in a Salivette® container. The nurse carried the saliva samples to a laboratory, where a technician analyzed the samples. The laboratory technician used HEARTEN biosensors to monitor cortisol and TNF-alpha in saliva samples.

Results

There were 30 cortisol saliva samples and 28 TNF-alpha saliva samples of 15 patients (4 from Italy, 11 from Spain) analyzed with HEARTEN biosensors to find relation between cortisol and TNF-alpha samples at the beginning and at the end of the follow-up. Table 1 shows the detail of these observations.

Table 1– Cortisol and TNF-alpha Data

Biomarker	Follow-up	Nº of obs	Mean	St. dev	Min	Max
Cortisol	Begin	15	16.35	14.48	1	46.6
Cortisol	Begin	15	15.76	12.99	0.2	34.2
TNF-alpha	End	15	18.43	16.84	2.8	49.8
TNF-alpha	End	13	7.91	5.24	2	18

In Table 2, a decrease in the levels of cortisol, TNF-alpha, diastolic blood pressure (DBP) and weight can be identified, and a small increase in heart rate and systolic blood pressure (SBP). However, the only significant change can be read in the TNF-alpha levels.

Table 2– Analyzed Variables

Variable	Mean Begin	Obs Begin	Mean End	Obs End	p-value
Cortisol	16.35	15	15.76	15	0.91
TNF-alpha	18.43	15	7.91	13	0.04
SBP	129.3	15	133.13	15	0.55
DBP	71	15	69.41	15	0.74
Heart rate	71.93	15	73.27	15	0.83
Weight	81.51	15	80.64	14	0.86

The boxplot in Figure 1, a reduction in TNF-alpha at the end of the period can be graphically shown. The significance of this conclusion may be seen in the Wilcoxon signed-rank test, which shows a p-value of 0.55, allowing rejecting the null at a significance level of 10%.

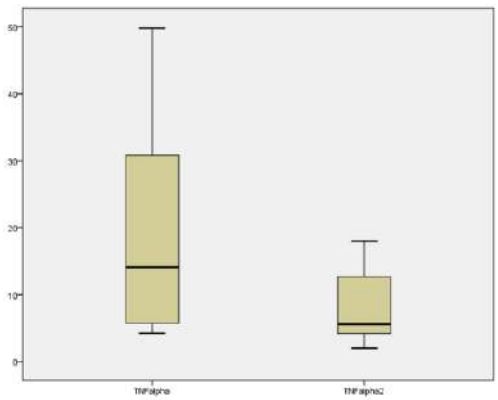


Figure 1– Reduction in TNF-alpha.

In Table 3, we can see the Spearman correlations between the two indicators and the measures of SBP, DBP, heart rate, weight, age, NYHA class and NT-proBNP at the beginning.

Table 3– Spearman Correlation Coefficient of Saliva Cortisol and TNF-alpha Samples with HEARTEN Biosensors and Biological Data at the Beginning

Variable	CORTISOL (n=15)	TNF-alpha (n=15)
SBP	-0.06	0.24
DBP	0.20	0.07
Heart rate	-0.06	0.19
Weight	0.03	0.05
NT-proBNP	0.26	-0.39
NYHA class	0.02	0.28
Age	-0.13	0.07

We cannot find any significant correlation between clinical information, although it seems to be a inverse relation between NT-proBNP and TNF-alpha, with lower levels of TNF-alpha in patiens with higher levels of NT-proBNP. This correlation is low and there is no signifficant differences (Spearman correlation coefficient 25,7% p 0,3). If this correlation is demonstrated in bigger studies patient prognostic with a non invasive sample could be predicted. Table 4 reported the same analysis considering the measures at the end of the period.

Table 4– Spearman Correlations of Saliva Cortisol and TNF-alpha Samples with HEARTEN Biosensors and Biological Data at the End

Variable	CORTISOL (n=15)	TNF-alpha (n=13)
SBP	-0.01	-0.29
DBP	0.32	0.19
Heart rate	-0.23	-0.63
Weight	0.21	0.64
NYHA class	0.12	-0.12
Age	-0.52	0.51

There is a significant negative correlation (at 5%) between TNF-alpha and heart rate. We also find a significant positive correlation between TNF-alpha and weight. We cannot analyze differences in samples at the end of the follow up because NT-proBNP has been measured only at the beginning.

Discussion

Initially in the project, the samples analysis was performed directly at patient's home. Finally, this can not be possible because the samples needed a preprocessing (for example, the saliva must be centrifuged) in a laboratory.

Although 67 patients were recruited in the pilot study, to analyze the TNF-alpha and cortisol correlation, only data from 15 patients could be used, because in some cases the patient refused to collect the sample, or the sample was corrupted. It was not possible to use all analyzed samples obtained from TNF-alpha/cortisol, since we missed data from the patient's clinical situation due to failures of HEARTEN ecosystem.

No significant correlation between cortisol and biological data has been detected, mainly due to the reduced sample size. However, it seems to have a direct correlation between TNF-alpha levels in saliva and the weight. This correlation can be used to predice HF decompensation, if the same results in bigger studies will be obtained, with a higher sample size. With higher sample size, we would like to look for a correlation between NT-proBNP and TNF-alpha, predicting patient prognostic with a non invasive sample.

Conclusions

HEARTEN ecosystem enables HF patients to achieve self-manage of their disease. The saliva samples analysis has been performed in a satisfactory way, however the sample size has been too little. Authors are interested in performing a following pilot study to obtain a higher sample size that allows to analyze in depth the correlation between TNF-alpha and cortisol in saliva and the clinical situation of patients.

Acknowledgements

HEARTEN project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643694. And many thanks to the clinical team of the Cardiology Unit of the University Hospital in Pisa, Italy, for their contributions related to this work.

References

[1] D. Mozaffarian, E.J. Benjamin, A.S. Go, et al, Executive summary: heart disease and stroke statistics-2016 update: a report from the American Heart Association, *Circulation* **133** (4) (2016), 447-454.

[2] http://www.hearten.eu/ [accessed March 23, 2019]

[3] F.G. Bellagambi, A. Baraket, A. Longo, et al, Electro-chemical biosensor platform for TNF-α cytokines detection in both artificial and human saliva: Heart failure, *Sensors and Actuators B: Chemical*, **251** (2017), 1026-1033.

Address for Correspondence

Corresponding author name: Alicia Martínez García. Full mailing address: Av. Manuel Siurot s/n 41013 Seville Spain. Email: alicia.martinez.exts@juntadeandalucia.es. Phone: +34 955013616.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190617

The Impact for Medical Management of the Health Information Exchange Through Changes of the Number of the First Visit Patients and Admission Patients in Japan

Takehiro Matsumoto, Naota Taura, Masayuki Honda

Department of Medical Informatics, Nagasaki University Graduate School of Biomedical Sciences, Japan

Abstract

Recently Health Information Exchange (HIE) is gradually spreading in Japan. But there are few reports about the effect of the hospital management. So we examined the effect for hospital management through the number of the newly admitted patients from 2013 to 2017. In this study both the total number of the admitted patients who was reserved as first visit patients was higher than not reserved and the number of the admitted patients who visited with introduction letters was higher than without introduction letters. (p<0.01) The total number of the admitted patients who had been registered to the Ajisai-net was 624 (41.2%) for the same 5 years, and was revealed to be significantly higher than the number of the admitted patients who was reserved first visit with introduction letters. (p<0.01) In conclusion, use of the Japanese type of the HIE has a potential to increase the newly admitted patients finally.

Keywords:
Health Information Exchange, hospital administration, Electronic Health Records

Introduction

The continuous increase of the Medical costs is one of the most important national problem all over the world. It is an important problem in Japan as well. Because the income of the hospitals had been supported with the public universal insurance has been reduced, many hospital incomes are getting worse in Japan. On the other hand, Health Information Exchange (HIE) has been gradually spreading in Japan. [1,2] The Japanese Medical Association reported the number of these networks was at least 309 in 2017 all over Japan, and is increasing even now. But there are few reports about the economic and management effects of these networks in Japan[3]. The number of the newly admitted patients is one of the most influential factor of the hospital income, and this is greatly affected from the number of introduced patients from clinics or the other hospitals on the Japanese medical system. The Ajisai-net, one of the most popular HIE of Japan, has been organized and used since 2004 at the Nagasaki Prefecture located at the western end of Japan. Over 90,000 patient's medical records of 36 regional big hospitals of Nagasaki Prefecture are shared among over 370 medical institutions such as hospitals, clinics, pharmacies and other medical institutions. In this study, we evaluated the economic and management effect of HIE through the changes of the patient introduction.

Methods

The Ajisai-net is one of the HIE of Japan. The feature of this network is the type in which the whole medical information of the Electronic Medical Record of the regional big hospitals was shared among many regional clinics, hospitals, pharmacies, nursing care stations and other several medical institutions. And this type of the HIE is the most popular in Japan. Nagasaki University Hospital has joined the Ajisai-net since 2009, and almost all the medical information of the Electronic Medical record of our hospitals is shared at the clinics, pharmacies, and other medical institutions based under the patient consents. We investigated the number and rate of the newly admitted patients and first visit patients using the statistical open data and the first visit management system.And the number of registered patients to Ajisai-net and additional data were obtained from the first visit management system and the Ajisai-net registration management system of the Nagasaki University Hospital. Emergency visit cases and dental visits were excluded, and the remaining first visit patients were

Table 1 The Number and Rate of the First Visit and Admission to Nagasaki University Hosipital for Recent 5 Years

			2013	2014	2015	2016	2017	total	
with introduction leter	reservation	First visit	5,106	5,652	5,485	7,353	8,802	32,398	
		Hospitalization	1,905	2,106	2,045	2,623	3,558	12,237	※
		rate	37.3%	37.3%	37.3%	35.7%	40.4%	37.8%	
	no reservation	First visit	5,884	4,592	4,849	4,631	4,593	24,549	
		Hospitalization	1,891	1,580	1,709	1,663	1,921	8,764	※
		rate	32.1%	34.4%	35.2%	35.9%	41.8%	35.7%	※
without introduction leter		First visit	1,538	1,050	1,746	1,499	1,334	7,167	
		Hospitalization	143	110	342	321	361	1,277	
		rate	9.3%	10.5%	19.6%	21.4%	27.1%	17.8%	
total		First visit	12,528	11,294	12,080	13,483	14,729	64,114	
		Hospitalization	3,939	3,796	4,096	4,607	5,840	22,278	
		rate	31.4%	33.6%	33.9%	34.2%	39.6%	34.7%	

※ p<0.01

Table 2 The Number and Rate of the Registered Patients to Ajisai-net for 5 Years

	2013	2014	2015	2016	2017	total	
total number of Ajisai-net registered	841	935	956	1,231	1,787	5,750	
first visit patient Ajisai-net registered	652	596	618	749	779	3,394	
rate	77.5%	63.7%	64.6%	60.8%	43.6%	59.0%	
Introduced under Ajisai-net registered	265	266	265	334	384	1,514	※
effective usage for the first visit	162	177	179	241	294	1,053	
rate	61.1%	66.5%	67.5%	72.2%	76.6%	69.6%	

※ p<0.01

Table 3 The Comparison of the Number of the Admitted Patients between First Visit Patients Registered to the Ajisai-net and Reserved First Visit Patients with the Introduction Letters

	2013	2014	2015	2016	2017	total		
Reserved first visit with introduction letter	5,106	5,662	5,485	7,353	8,802	32,408		
the number of admission	1,906	2,106	2,045	2,623	3,558	12,238		
rate	37.3%	37.2%	37.3%	35.7%	40.4%	37.8%		
Introduced under Ajisai-net registered	265	266	265	334	384	1,514	※	
the number of admission	111	107	101	125	180	624		※※
rate	41.9%	40.2%	38.1%	37.4%	46.9%	41.2%		
effective usage for the first visit	162	177	179	241	294	1,053		
the number of admission	78	65	62	93	133	431		
rate	48.1%	36.7%	34.6%	38.6%	45.2%	40.9%		

※ p<0.01 ※※ p=0.03

divided to 3 types which are reserved visit bringing the introduction letters, non-reserved visit bringing the introduction letters and non-reserved visit not bringing the introduction letters. And 'the effective usage' in the table 2 has been defined as the case the doctor shared medical information through the Ajisai-net by themselves before introduction for this patient. Statistical analysis adopted the chi-square test.

Results

The total number of the first visit patients from 2013 to 2017 was 64,120 at the Nagasaki University Hospital, and the average admission rate was 34.7% for 5 years. (Table 1) And the admission rate had increased every year. The admission rate of the first visit patients who had been non-reserved not bringing the introduction letters was the worst of the 3 types, 17.8% in average. Conversely the highest admission rate was the type which was reserved visit bringing the introduction letters. There was a statistically significant differences between the number of the admissions of the reserved first visit and of the non-reserved first visit, ($p<0.01$) and also between the number of admission of the first visit with and without the introduction letters ($p<0.01$). On the other hand, the number of total registered to the Ajisai-net with patient consent was 5,750 for the same 5 years (Table 2). The total number of the first visit patient who had been registered to the the Ajisai-net was 3,394, and registration rate was 59.0% for the same 5 years. And the total number of the registered patients to the Ajisai-net who was introduced from clinics was 1,514. And the total number of 'the effective usage' whose Electronic Medical Records of Nagasaki University had been shared through the Ajisai-net before introduction was 1,053 (69.6%), and significantly higher than the former ($p<0.01$). The total number of the admitted patients who had been registered to the Ajisai-net was 624 (41.2%) for the same 5 years (Table 3), and was significantly higher than the number of the admitted patients of the reserved first visit patients bringing the introduction letters ($p<0.01$). And the number of the admitted patients who had been registered to the Ajisai-net as the 'effective usage' was 431 (40.9%), and also higher than the number of the admitted patients of the reserved first visit patients bringing the introduction letters ($p=0.03$).

Discussion

Recently hospital management in Japan is getting more difficult year by year because of the governmental policy of the suppression of the medical social cost. So many hospital has made much efforts to hospital management. Because in general the income of admission fee of Acute hospitals is far more than the outpatient department of the same hospital in Japan, the one of the most important factor of the hospital economic management is pointed out to be the increase of the number of the newly admitted patients. Although there are some reports about the effect to the economic and hospital management of HIE in North America[4,5], there are few in Japan. In this study the number of the admitted patients was significantly higher on the patients who had been introduced bringing the introduction letter than those who had not bringing the introduction letter ($p<0.01$) (Table 2). And the number of the admitted patients whose first visit to the outpatient sections had been reserved with FAX was also significantly more than not reserved ($p<0.01$). These results seemed to indicate the more severe patients who had been selected and introduced by the clinic doctors by means of the introduction letters and first visit reservations. On the other hand, the number of the admitted patients who had been registered to the Ajisai-net was also significantly higher than no use of the Ajisai-net. The total number and the number of the 'effective usage' of the admitted patients regsitered to the Ajisai-net was 624 (41.2%) ($p<0.01$) and 431 (40.9%) ($p=0.03$) respectively. Though the rate of the admission of the 'effective usage' patients who had registerd to the Ajisai-net before their firist visits was less than the rate of admission of the total number of the first visit who had registered to the Ajisai-net use (41.2% VS 40.9%) , these results seems to indicate the sharing medical information through the Ajisai-net which is the most popular type of HIE in Japan has clear incentive for the doctor who introduced thier patients to our hospital. In this study, although the direct increase effect of the number of hospitalized patients who had been introduced as the first visit patient after registration to the Ajisai-net is not clearly proved, at least the number of hospitalized patients who had been introduced as the "effective usage" was significantly higher than the total number of hospitalized patients who had been introduced as the first visit patients. There is 2 typical use cases on the Ajisai-net. One is for making use of the past or old medical information for the diagnosis or past history taking, etc, and the other is for the observation and evaluation of the updated patient condition or obtaining the latest medical information for the lifelong self-learing for doctors through the Ajisai-net. These results seems to indicate that there is the clearly incentives for the doctor to use the Ajisai-net through these 2 use cases when they introduce their patients to the big or professional hospitals. By the above, because the increase of the number of the hospitalized patients causes to increase the hospital income in Japan, the Japanese type of HIE is proved to be effective for the economic hospital management indirectly in this study.

Conclusions

The Japanese type of HIE like the Ajisai-net has much potential of increasing the number of the introduced first visit patients and admitted patients because of its high value in sharing medical information.

References

[1] A. Winter, K. Takabayashi, M. Honda, T. Matsumoto, et al, Quality Requirements for Electronic Health Record Systems A Japanese-German Information Management Perspective, *Methods of Inf Med* **56**(2017), e92-e104.

[2] T. Matsumoto, M. Honda, The Evaluation of the Health Information Exchange with the Number of Usage and the introduction of Outpatient to the Hospitals at Nagasaki Japan, *Stud Health Technol Inform.* **245**(2017), 1373.

[3] H.Nakamura, N.Iwamoto, T.Matsumoto, A.Kawakami, An effective medical partnership in Nagasaki, Japan for patients with rheumatoid arthritis, *Mod Rheumatol.* **22**(2016), 878-884.

[4] Frisse ME, Johnson KB, Nian H, et al, The financial impact of health information exchange on emergency de partment care, *J Am Med Inform Assoc*. **19**(3)(2012), 328-33.

[5] Welk B, Liu K, Al-Jaishi A, McArthur E, Jain AK, Ordon M, Repeated Diagnostic Imaging Studies in Ontario and the Impact of Health Information Exchange Systems, *Healthc Q* **19**(1)(2016), 24-28.

Address for correspondence

Takehiro Matsumoto, Department of Medical Informatics,
Graduate School of Biomedical Sciences Nagasaki University,
E-mail: tmatsumo@nagasaki-u.ac.jp

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190618

The Use of Culturally-Tailored Telehealth Interventions in Managing Anxiety and Depression in African American Adults: A Systematic Review

Terika McCall[a], Clinton S. Bolton III[b], Rebecca McCall[c], Saif Khairat[a]

[a] *Carolina Health Informatics Program, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA*
[b] *School of Education, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA*
[c] *Health Sciences Library, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA*

Abstract

This systematic review examined studies aimed at reducing anxiety or depression in African American adults through use of telehealth interventions. Three small independent studies were identified. The findings showed significant reduction of depressive symptoms post-intervention (all p values < .05). However, effectiveness of telehealth intervention compared to face-to-face was not determined. The results highlight the need for additional research into the effectiveness of using telehealth to manage anxiety and depression in this population.

Keywords:

Telemedicine; African Americans; Mental Health

Introduction

In 2017, the U.S. Substance Abuse and Mental Health Services Administration (SAMHSA) estimated that 46.6 million adults in the U.S. experienced mental illness in the last year [1]. The prevalence rate of mental disorders, excluding serious mental illness, has remained constant since 2008 [1]. A decade later, technology has allowed us to increase the options for delivery of mental health services; however, the percentage of adults with a mental illness who received mental health services in the past year (42.6%) was comparable to the rate for most years from 2008 to 2016 [1]. The expanding adoption of telehealth in the U.S. can increase access to mental health services, improving health outcomes and reducing health care costs. Although telehealth may not work for everyone, it has the potential to help many people included in the estimated 26.7 million U.S. adults (57.4%) who have a mental illness but did not receive mental health services in the last year [1].

Anxiety and depressive disorders are among the most common mental illnesses. A report by the Centers for Disease Control and Prevention (CDC) noted that 12.9% of non-Hispanic whites and 9.3% of non-Hispanic blacks reported having received a diagnosis of anxiety in their lifetime [2]. Moreover, the estimated prevalence of depression was higher for African Americans (12.7%) than their white counterparts (7.5%) [2].

Due to underreporting of mental illness, especially in underserved populations, it is unclear if the prevalence estimates depict the true rates, and if the differences observed between racial groups are accurate; however, there is significant unmet need and racial disparity in the use of mental health services [1]. Although there is less than a 4% difference in the prevalence of mental illness among white adults compared to black adults (20.4% vs. 16.5%), African Americans utilize mental health services at less than half the rate of their white counterparts (8.9% compared to 18.1%) [3]. Barriers to African Americans receiving health services are well documented in literature.

Methods

The methods outlined in the *PRISMA Statement* for systematic reviews were followed (see Figure 1). First, a comprehensive literature search was conducted using the PubMed, PsycINFO, Scopus, and Web of Science electronic databases for relevant articles published from 1970 to September 2018. A combination of keywords relating to African Americans, depression, anxiety, and telehealth were used. Second, reference lists of the included primary articles and retrieved systematic reviews were examined to identify any relevant publications.

The criteria used for inclusion in this review were: (1) intervention targeted Black/African American adults (≥ 18 years old); (2) primary outcome(s) include either diagnosis or symptom severity of depressive or anxiety disorder; (3) telehealth-based psychological intervention; (4) intervention effectiveness evaluated using one or more standardized measures of anxiety or depression administered pre- and post-intervention; and (5) feasibility of using telehealth modality to receive psychological help, acceptability of using telehealth modality to receive mental health services, or self-management through participation in the intervention assessed post-intervention.

Covidence software was used for article screening and data extraction. Two independent reviewers (T.M. and C.B.) independently analyzed each title and abstract of articles retrieved to determine their relevance. The full text of potentially eligible studies was retrieved and similarly analyzed by T.M. and C.B. to exclude papers that did not meet inclusion criteria. Any disagreement over the eligibility of particular studies was resolved by an adjudicator (S.K.). The *Critical Appraisal Skills Programme* (CASP) checklists were used to assess the quality of the identified studies.

Significant themes and trends of the studies were identified and discussed. A quantitative analysis and comparison of treatment effectiveness across studies was considered. However, due to the limited number of randomized controlled trials (RCTs) included, comparison was not feasible.

Results

A detailed search in all databases identified a total of 622 articles. Twenty-one additional articles were identified through examining other sources. After the removal of duplicates, 529 articles were included in the title and abstract screening process. The majority of the articles that were excluded did not involve psychological interventions that targeted black adults. Fifty papers were identified as relevant to the research question and

included in the full-text screening process. Following the independent full-text screening process, three articles were identified that met the inclusion criteria and were included in the analysis [4–6].

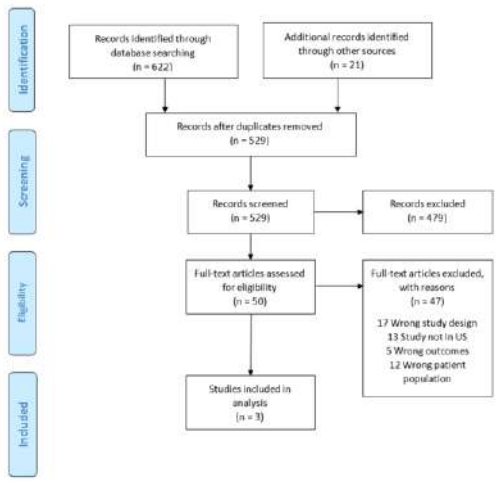


Figure 1– PRISMA Flow Diagram of Search Strategy

The three independent studies had a pooled total of 32 participants (range 6-15 per study). Telephone-based cognitive behavioral therapy (CBT) [4,6], or a mobile phone-optimized online intervention [5] was used. Only one of the studies was a randomized controlled trial (RCT) [4]. The remaining two studies were prospective cohort studies assessing the feasibility of interventions [5,6]. Two of the studies reported a majority (> 80%) of their sample were female [4,6]. All participants reported to be HIV-positive in one study [6], and another reported 60% of participants were HIV-positive [5]. All of the studies used validated assessments (the Center for Epidemiological Studies Depression Scale (CES-D) [4,5] or Hamilton Depression Rating Scale (HAM-D) [6]) to assess the primary outcome of depression pre and post intervention.

All of the studies [4–6] met a majority of the criteria used for consideration in the assessment of study quality in the CASP appraisal checklists (i.e. reviewers T.M. and C.B. answered 'yes' for > 80% of the questions). Two studies lacked control groups [5,6], and all had small sample sizes (6-15 participants) and short follow-up periods (< 6-months) [4–6]. This resulted in the reviewers rating all of the studies to be of 'fair' quality; however, all studies were published in peer reviewed journals and deemed appropriate for inclusion.

Statistically significant reduction of depressive symptoms (measured by improvement in mean CES-D or HAM-D score) was observed post-intervention in all of the studies ($p < .05$) [4–6]. In addition, one study showed a significant reduction in depression severity post-intervention ($p = .02$, measured by improvement in mean Quick Inventory of Depressive Symptomology (QIDS-SR) score) [6]. All studies reported patient satisfaction with the interventions [4–6]. Hightow-Weidman et al. also reported improvements for social support ($p = .012$) and social isolation ($p = .050$) for study participants post-intervention [5]. Furthermore, the RCT conducted by Glueckauf et al. found significant within-subjects effects for time across caregiver subjective burden ($p < .02$, measured by improvement in the subjective burden subscale of the Caregiver Appraisal Inventory (CAI)) and assistance support ($p < .03$, measured by improvement the Assistance subscale of Interpersonal Support Evaluation List (ISEL)) post-treatment [4]; however, no statistically significant effects were observed for group (telephone vs. face-to-face CBT) and the group x time interaction (all $p > .05$) on any of the measures [4].

Conclusions

Given the burden of unmet need and disparity in mental health utilization among African American adults, there is great potential to use telehealth to deliver services to this population. Telehealth is proving to be acceptable for some mental health services, such as CBT; however, to increase the likelihood of adoption among African American adults, telehealth interventions should be culturally-tailored. A "one size fits all" approach to designing telehealth interventions to help African American adults manage anxiety or depression may lead to more options but continued disparity in receiving care.

Acknowledgements

The first author is supported by funding from the National Library of Medicine's Institutional Training Grant for Research Training in Biomedical Informatics and Data Science (T15-LM012500).

References

[1] Substance Abuse and Mental Health Services Administration, Key substance use and mental health indicators in the United States: Results from the 2017 National Survey on Drug Use and Health (HHS Publication No. SMA 18-5068, NSDUH Series H-53), Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration, Rockville, MD, Retrieved from https://www.samhsa.gov/data/ [accessed November 13, 2018].

[2] Centers for Disease Control and Prevention, Mental Illness Surveillance Among Adults in the United States, MMWR 60 (2011).

[3] Substance Abuse and Mental Health Services Administration, 2017 National Survey on Drug Use and Health [data file], RTI International [distributor], Research Triangle Park, NC, Retrieved from https://pdas.samhsa.gov/#/survey/NSDUH-2017-DS0001 [accessed November 18, 2018].

[4] R. L. Glueckauf, W.S. Davis, F. Willis, et al., Telephone-based, cognitive-behavioral therapy for African American dementia caregivers with depression: initial findings, Rehabilitation psychology **57** (2012), 124–139.

[5] L.B. Hightow-Weidman, K.E. Muessig, E.C. Pike, et al., HealthMpowerment.org: Building Community Through a Mobile-Optimized, Online Health Promotion Intervention, Health education & behavior: the official publication of the Society for Public Health Education **42** (2015), 493–499.

[6] S. Himelhoch, D. Mohr, J. Maxfield, et al., Feasibility of telephone-based cognitive behavioral therapy targeting major depression among urban dwelling African-American people with co-occurring HIV, Psychology, health & medicine **16** (2011), 156–165.

Address for correspondence

Corresponding Author: Saif Khairat, PhD, MPH, FAMIA, 428 Carrington Hall, Chapel Hill, NC, USA 27514; Email: saif@unc.edu.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190619

Multi-Institutional, Large-Scale, International Applied Clinical Informatics Research Through the Clinical Informatics Research Collaborative (CIRCLE)

Allison B. McCoy[a], Adam Wright[b], Dean F. Sittig[c]

[a] Vanderbilt University Medical Center, Nashville, TN, USA,
[b]The University of Texas School of Biomedical Informatics at Houston, Houston, TX, USA,
[c]Harvard Medical School, Brigham and Women's Hospital, Boston, MA, USA

Abstract

There is a critical need need for multi-institutional, large-scale, international applied clinical informatics research, given the global, widespread use of commercially-available electronic health records with different designs, capabilities, configurations, and implementation strategies. The Clinical Informatics Research Collaborative (CIRCLE) aims to identify and develop best practices for safe and effective health information technology design, development, implementation, use, and evaluation.

Keywords:
Medical Informatics, Research, Health Facilities

Introduction

The last decade witnessed dramatic growth in electronic health record (EHR) adoption, from less than 10% to nearly universal. Until recently, only large, academic healthcare organizations implemented locally developed EHRs, which enabled extensive customization and novel research but had limited generalizability [1]. Currently, 95% of EHRs are commercially-developed systems [2]. Commercial EHRs partially addressed problems related to system maintenance, long-term reliability, and variable annual costs. Nevertheless, commercial systems significantly curtail the ability of informatics researchers at a single organization to experiment with different designs, configurations, and implementation strategies, and satisfaction with current EHR performance remains low [3]. In the absence of purposeful studies, natural experiments nevertheless happen. Organizations adopting various commercially-developed EHRs choose different configurations, use systems in distinct ways, and learn valuable lessons based on mistakes during such processes [4]. These insights are rarely captured formally or disseminated widely. There is a critical need to conduct multi-institutional, large-scale, international applied clinical informatics research that can improve EHR usage and outcomes.

The Clinical Informatics Research Collaborative

The mission of the Clinical Informatics Research Collaborative (CIRCLE) is to identify and develop best practices for safe and effective health information technology design, development, implementation, use, and evaluation. CIRCLE currently has 41 collaborators representing 30 organizations across five countries. CIRCLE collaborators use many EHRs, including Epic, Cerner, MEDITECH, McKesson, Allscripts, CPRS/Vista, GE Centricity, eClinicalWorks, and others. Primary roles filled by CIRCLE collaborators include researcher, educator, physician, CMIO/CMO, nurse, data analyst, CCIO/CIO, and pharmacist. Most collaborators work in academic medical centers, though four work in community hospital settings. To date, CIRCLE collaborators have worked together to complete numerous multi-institutional research projects and have published 29 peer-reviewed manuscripts of varying types, including cross-institutional interviews and observations [5,6], surveys [7,8], system demonstrations [9,10], distributed queries [11,12], and experimental interventions [13]. Some results of these projects made available through publication in peer-reviewed journals include success factors for people involved in CDS [6], recommended practices for CDS governance [8], and guidelines for future development of user-defined CDS [10].

Future Work

To facilitate the continued efforts by CIRCLE and other research consortia, there is a need to develop standard procedures for carrying out multi-institutinal, applied clinical informatics research. Reporting guidelines improve the quality, transparency, and reproducibility of research by defining the necessary components for describing a study [14]. Some guidelines exist for health informatics research, including including STARE-HI (Statement on reporting of evaluation studies in Health Informatics) [15] and CONSORT-EHEALTH [16], but none specifically address the unique challenges posed by use of different commercially-available EHRs. Most EHRs have different local configurations across different healthcare organizations. Institutional decisions about specific implementations and configuration decisions regarding the same vendor's EHR can substantially alter their local performance [17]. External rules and regulations that can impact EHR requirements and implementations may also differ across countries [18]. Many organizations and commercial EHR vendors restrict what can be publicly shared or published regarding their EHRs [19]; thus, special considerations apply to interpreting reported results unique to the site or EHR. Improved guidance for conducting and reporting collaborative applied informatics research across varied implementations and sites may help researchers overcome such challenges. Improved quality and reproducibility of the studies would follow, ultimately increasing the effectiveness of EHRs.

Conclusions

Efforts by CIRCLE to date suggest that collaborative projects can effectively identify best practices for EHR

implementations. Future success of CIRCLE depends on continued expansion and inclusion of members from diverse healthcare settings and from sites with access to currently unrepresented EHRs, as well as established guidelines for carrying out new multi-institutional, collaborative research. New informatics research carried out through CIRCLE that is high quality, transparent, and reproducible will enable healthcare organizations to better evaluate and improve their EHR capabilities.

References

[1] K.G. Shojania, A. Jennings, A. Mayhew, C.R. Ramsay, M.P. Eccles, and J. Grimshaw, The effects of on-screen, point of care computer reminders on processes and outcomes of care, *Cochrane Database Syst. Rev.* (2009) CD001096. doi:10.1002/14651858.CD001096.pub2.

[2] Hospital Health IT Developers, (n.d.). /quickstats/pages/FIG-Vendors-of-EHRs-to-Participating-Hospitals.php (accessed August 22, 2018).

[3] D.F. Sittig, and H. Singh, Which electronic health record is better: A or B? Realities of comparing the effectiveness of electronic health records, *J. Comp. Eff. Res.* **3** (2014) 447–450. doi:10.2217/cer.14.46.

[4] D.F. Sittig, J.S. Ash, J. Zhang, J.A. Osheroff, and M.M. Shabot, Lessons from "Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system," *Pediatrics*. **118** (2006) 797–801. doi:10.1542/peds.2005-3132.

[5] J.S. Ash, D.F. Sittig, C.K. McMullen, A. Wright, A. Bunce, V. Mohan, D.J. Cohen, and B. Middleton, Multiple perspectives on clinical decision support: a qualitative study of fifteen clinical and vendor organizations, *BMC Med. Inform. Decis. Mak.* **15** (2015) 35. doi:10.1186/s12911-015-0156-4.

[6] A. Wright, J.S. Ash, J.L. Erickson, J. Wasserman, A. Bunce, A. Stanescu, D. St Hilaire, M. Panzenhagen, E. Gebhardt, C. McMullen, B. Middleton, and D.F. Sittig, A qualitative study of the activities performed by people involved in clinical decision support: recommended practices for success, *J. Am. Med. Inform. Assoc. JAMIA*. **21** (2014) 464–472. doi:10.1136/amiajnl-2013-001771.

[7] D.F. Sittig, B.L. Hazlehurst, J. Brown, S. Murphy, M. Rosenman, P. Tarczy-Hornoch, and A.B. Wilcox, A survey of informatics platforms that enable distributed comparative effectiveness research using multi-institutional heterogenous clinical data, *Med. Care*. **50 Suppl** (2012) S49-59. doi:10.1097/MLR.0b013e318259c02b.

[8] A. Wright, D.F. Sittig, J.S. Ash, D.W. Bates, J. Feblowitz, G. Fraser, S.M. Maviglia, C. McMullen, W.P. Nichol, J.E. Pang, J. Starmer, and B. Middleton, Governance for clinical decision support: case studies and recommended practices from leading institutions, *J. Am. Med. Inform. Assoc. JAMIA*. **18** (2011) 187–194. doi:10.1136/jamia.2009.002030.

[9] D.F. Sittig, D.R. Murphy, M.W. Smith, E. Russo, A. Wright, and H. Singh, Graphical display of diagnostic test results in electronic health records: a comparison of 8 systems, *J. Am. Med. Inform. Assoc. JAMIA*. **22** (2015) 900–904. doi:10.1093/jamia/ocv013.

[10] A.B. McCoy, A. Wright, and D.F. Sittig, Cross-vendor evaluation of key user-defined clinical decision support capabilities: a scenario-based assessment of certified electronic health records with guidelines for future development, *J. Am. Med. Inform. Assoc. JAMIA*. **22** (2015) 1081–1088. doi:10.1093/jamia/ocv073.

[11] A.B. McCoy, A. Wright, M.G. Kahn, J.S. Shapiro, E.V. Bernstam, and D.F. Sittig, Matching identifiers in electronic health records: implications for duplicate records and patient safety, *BMJ Qual. Saf.* **22** (2013) 219–224. doi:10.1136/bmjqs-2012-001419.

[12] A. Wright, A.B. McCoy, T.-T.T. Hickman, D.S. Hilaire, D. Borbolla, W.A. Bowes, W.G. Dixon, D.A. Dorr, M. Krall, S. Malholtra, D.W. Bates, and D.F. Sittig, Problem list completeness in electronic health records: A multi-site study and assessment of success factors, *Int. J. Med. Inf.* **84** (2015) 784–790. doi:10.1016/j.ijmedinf.2015.06.011.

[13] A. Wright, J. Pang, J.C. Feblowitz, F.L. Maloney, A.R. Wilcox, K.S. McLoughlin, H. Ramelson, L. Schneider, and D.W. Bates, Improving completeness of electronic problem lists through clinical decision support: a randomized, controlled trial, *J. Am. Med. Inform. Assoc.* **19** (2012) 555–561. doi:10.1136/amiajnl-2011-000521.

[14] D. Moher, D. Altman, K. Schulz, I. Simera, and E. Wager, Guidelines for reporting health research: a user's manual, John Wiley & Sons, 2014.

[15] J. Talmon, E. Ammenwerth, J. Brender, N. de Keizer, P. Nykänen, and M. Rigby, STARE-HI--Statement on reporting of evaluation studies in Health Informatics, *Int. J. Med. Inf.* **78** (2009) 1–9. doi:10.1016/j.ijmedinf.2008.09.002.

[16] G. Eysenbach, CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions, *J. Med. Internet Res.* **13** (2011) e126. doi:10.2196/jmir.1923.

[17] J. Metzger, E. Welebob, D.W. Bates, S. Lipsitz, and D.C. Classen, Mixed results in the safety performance of computerized physician order entry, *Health Aff. Proj. Hope*. **29** (2010) 655–663. doi:10.1377/hlthaff.2010.0160.

[18] D.F. Sittig, and H. Singh, A new sociotechnical model for studying health information technology in complex adaptive healthcare systems, *Qual. Saf. Health Care*. **19** (2010) i68–i74. doi:10.1136/qshc.2010.042085.

[19] R. Koppel, and D. Kreda, Health care information technology vendors' "hold harmless" clause: implications for patients and clinicians, *JAMA*. **301** (2009) 1276–1278. doi:10.1001/jama.2009.398.

Address for correspondence

Allison B. McCoy, PhD

2525 West End Ave., Suite 1475

Nashville, TN, 37203, USA

allison.b.mccoy@vumc.org

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190620

Secondary Data Use in Rwanda: Leveraging OpenMRS for Global HIV Research

Benjamin Muhoza[a], Eric Remera[b], Qiuhu Shi[c], Jules Kabahizi[a], Ellen Brazier[d], Jean d'Amour Sinayobye[a], Stephany N. Duda[e]

[a] Rwanda Military Hospital, Kigali, Rwanda; [b] Rwanda Biomedical Center, Kigali, Rwanda
[c] Department of Epidemiology and Community Health, New York Medical College, Valhalla, USA
[d] Institute for Implementation Science in Population Health, City University of New York, New York, USA
[e] Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, USA

Abstract

The Rwandan Ministry of Health supports a countrywide installation of the Open Medical Record System (OpenMRS) to improve clinical recordkeeping and patient care. However, electronic medical records also can be a valuable source of data for observational and experimental studies. We describe the challenges and lessons learned when reusing OpenMRS data in Rwanda for global HIV epidemiology research.

Keywords:

Electronic Medical Records, Observational Study, Rwanda

Introduction

The International Epidemiology Databases to Evaluate AIDS (IeDEA) consortium includes over 400 HIV care and treatment clinics in 45 countries that pool routine patient care data to answer research questions of global scope (iedea.org). Ten HIV care and treatment clinics in Rwanda are participating in IeDEA. Clinic researchers contribute scientific expertise and their patient data extracted from OpenMRS, an open source, customizable electronic medical record system (EMR) implemented in Rwanda since 2006 to support patient care [1]. At each participating clinic, patient data are recorded on paper charts during clinic visits and subsequently entered into OpenMRS by data entry personnel. Each clinic has an Ubuntu Linux server running OpenMRS on a local area network to ensure availability without internet. The EMR infrastructure is maintained by the government of Rwanda, which supports secondary use of the data for research.

Methods

The local IeDEA data manager visits each clinic bi-annually, extracts patient data from the clinic's OpenMRS database, and runs a custom C# program that removes patient identifiers. With Rwandan government and ethics board permission, de-identified copies of the database are securely uploaded to the central Africa IeDEA coordinating center in the United States (US), which runs data quality checks and merges datasets from the entire region. Data queries are resolved by the Rwandan team. OpenMRS data are compared to paper source documents where necessary and query responses are passed back to the US data center, which then maps local data to the IeDEA data exchange standard (iedeades.org) for global IeDEA collaborations.

Results

Using this model of data preparation, the Rwanda IeDEA team has provided 60 data uploads since late 2015 and contributed 31 high quality, merged datasets for global and regional IeDEA publications. We encountered two noteworthy challenges with this approach to reusing OpenMRS data for research:

Transforming data in the OpenMRS star data model and concept dictionary to a functional research dataset is complex, with 100-130 database tables per HIV clinic. Customized data de-identification and data quality review can slow the data sharing process. However, due to the standardized OpenMRS implementation in Rwanda, we had the same concept dictionary and nearly the same database tables available at all participating care sites, which simplified the data extraction and merging.

Involving local OpenMRS experts was critical when reusing routine clinical care data. We initially used a custom OpenMRS module that generated encrypted data export files for each clinic, but this approach prevented the Rwanda IeDEA team from de-identifying the data and participating in data quality assurance. When US-based support for the module ended and there was no sustainability plan, we developed the current model. Although our current approach may be more labor intensive, it has proven secure and resilient to changes in the OpenMRS data model (due to software upgrades), and relies on a more egalitarian paradigm of shared Rwandan-US involvement with research data processes.

Conclusions

Clinical care data from OpenMRS can be utilized successfully for HIV observational and experimental research but it requires custom data de-identification, quality checks, and engaged local data teams.

Acknowledgements

This work was supported by the U.S. NIH U01AI096299.

References

[1] B.W. Mamlin, P.G. Biondich, B.A. Wolfe, et al., Cooking Up An Open Source EMR For Developing Countries: OpenMRS – A Recipe For Successful Collaboration, *AMIA Annu Symp Proc*. **2006** (2006) 529–533.

Address for correspondence

Benjamin Muhoza, benmuhoza@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190621

Developing a Biomedical Information System for Clinical Assessments and Interventions of Mental Health and Substance Abuse Patients Belonging to YMSM of Color with HIV Using Multidimensional Scaling Analysis and Paired Comparisons Techniques

Sukrit Mukherjee[a], Eric Houston[b]

[a] *Center for Biomedical Informatics, Charles R. Drew University of Medicine and Science, Los Angeles, CA, USA,*
[b] *Department of Psychiatry and Human Behavior, Charles R. Drew University of Medicine and Science, Los Angeles, CA, USA*

Abstract

In mental health and substance abuse treatment, individualized assessments provide information on the specific thoughts and cognitive processes influencing a person's behavior, emotional responses, and psychological functioning. Given the lack of automated assessment procedures or individualized clinical interventions in the growing health disparities in the South Los Angeles of USA, we developed a novel system using idiographic techniques to automatically and quickly generate individualized patient assessment data for use in clinical interventions.

Keywords:
Mental Health, Information Systems, HIV

Introduction

Depression affects people disproportionately living with HIV and other chronic health conditions in a widesspread manner [1]. A study in the United States that was based on a nationally representative sample of people living with human immunodeficiency virus (HIV) found that 36% of participants had a probable diagnosis of major depressive disorder. This rate was nearly five times greater than that of the general population [2]. Therefore, receiving treatment for depression becomes very critical for the HIV patients. At the same time among many people living with HIV (PLH), depression continues to remain largely untreated [3][4]. This low percentage of of treatment for depression among HIV patient population have given rise to several stigmas that may include tedency of individuals to avoid engagement in treatment for depression as well as desire of non participation in depression intervention research [4][5][6][7].

In mental health and substance abuse treatment, individualized assessments provide information on the specific thoughts and cognitive processes that influence a person's behaviors, emotional responses, and psychological functioning [8]. Having information about the individual's thoughts and cognitive processes facilitates the development of clinical interventions tailored specifically to an individual patient. Research suggests that interventions based on information from such individualized assessments are likely to have greater effectiveness than those that do not [9]. Interventions that incorporate these individual-specific thoughts and cognitive processes are likely to resonate more strongly with patients than interventions lacking this focus. In addition, using individualized assessments to inform and implement treatment also heightens patient motivation and may strengthen efforts to promote their retention in treatment.

Little research has focused on how to reach and engage patients with depression and HIV who may be at the intersection of multiple social identities faced with stigma. Due to the negative impact of depression on the health status of people living with HIV, there is a need for individualized clinical interventions especially for the growing health disparities in the South Los Angeles of USA. Paired comparisons techniques and MDS analyses are two data analytic techniques that have traditionally been used in statistics and research [10]. But to date, paired-comparisons techniques and MDS have not been combined for use in either automated assessment procedures or individualized clinical interventions. We intended to fill this gap by utilizing novel use of such idiographic approaches that may be particularly valuable in clinical settings that lack staff trained in clinical interviewing or individualized assessment techniques.

Methods

We followed a mixed method research design in ths study. The study called, Project STEP ("Steps Toward Embodying Positivity") was targetted to conceive a new cognitive training intervention system to address depressive symptoms and suboptimal HIV treatment adherence among African American and Latino YMSM (ages 18-29) in the Los Angeles metropolitan area. The study received guidance and support from a community advisory board consisting of eight members for participant recruitment procedures and strategies. The Institutional Review Board of Charles R. Drew University of Medicine and Science (CDU) approved the study. Participants were recruited for a focus group session to obtain feedback on an initial version of the web-based intervention system and for a preliminary brief pilot in which individuals from the target population used the web-based intervention. Participants in the pilot were asked to use the system on a daily basis at home as well as in a clinical setting for up to four weeks. Pilot study participants also received information related to behavioral strategies for HIV treatment adherence, mood management, and alcohol and substance abuse prevention based on modules from Centers for Disease Control (CDC) of USA's evidence-based intervention. The system was accessible as web based modules that were delivered to participants via laptop computers and iPad devices during weekly visits with a study counselor. Study participants were asked to complete questionnaires to assess depressive symptoms, adherence, treatment motivation, self-

efficacy, HIV stigma, internalized homophobia, and challenges they face with regard to medication management. Pilot study participants were also asked to complete interviews regarding their views of the intervention upon completing their enrollment [11].

The system consits of 2 stages. The first stage was designed using Java Server Pages (JSP) as frontend, Java as middle tier and MySQL as backend hosted on Tomcat Web server on a Linux platform with statement session of the input data (paired-comparisons ratings) from a participant and and it processes the data using MDS analysis and transforms the data into a matrix. The combined use of paired-comparisons techniques and MDS through automated, cross-platform processes allowed for the collection of critical information from patients. We used idiographic assessment procedure to uniquely allow for this information to be obtained and processed with significantly less time and then subsequently making it available for immediate use for a range of clinical interventions and approaches, including those involving attention bias modification training, cognitive mapping techniques, and motivational interviewing which were implemented in the second stage of the system that was developed using .Net as frontend, C# as middle tier and MS SQL Server as backend hosted on Internet Information Server on a Windows platform. The primary purpose of this stage was to generate stimuli for the users in Attention Bias Modification (ABM) training interventions, particularly individualized ABM interventions. The automated nature of our assessment approach and the relative speed with which the information can be applied is novel and has much potential in addressing gaps in the delivery of care.

Results

The first stage of our 'Project STEP' system receives input data (paired-comparisons ratings) from a participant and transforms the data into a matrix followed by processing of the data in MDS analysis through a very unique and specialized algorithm that results in the automatic generation and display of a n-dimensional spatial configuration or mapping and provides stress values, a least distance measure and R values of the individual user. The second stage of this unique system that receives statements, ratings, dot location and emotion values from first stage is designed with idiographic assessment program that generates stimuli for use in ABM training interventions, particularly individualized ABM interventions. The idiographic assessment program embedded in the system could operate cross-platform using data from an elicitation task and paired-comparisons techniques. The first stage generates and displays n-dimensional configurations automatically providing data to inform and implement clinical interventions using attention bias modification training implemented in the second stage. Thus, the resulting system design is likely to have strong market potential. The customer base for such an assessment procedure would include hospitals, clinics and substance use treatment centers as well as private practitioners and agencies and organizations involved in health care delivery in the United States and abroad.

Conclusions

We have attemped to solve the problem of clinical assessment for mental health and substance abuse patients belonging to YMSM of color with HIV across the continuum of care by creating a novel system that automatically and quickly generates individualized patient assessment data for use in clinical interventions like attention bias modification training with strong market potential.

Acknowledgements

The work was supported in part by Grant U54MD007598 from US National Institute on Minority Health and Health Disparities.

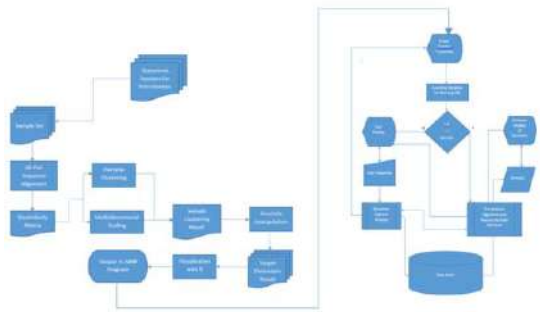


Figure 1– Process Flow in 'Project STEP' System

References

[1] Ali, S., Stone, M., Peters, J., Davies, M., & Khunti, K. (2006). The prevalence of co-morbid depression in adults with Type 2 diabetes: A systematic review and meta-analysis. Diabetic Medicine, 23, 1165–73.

[2] Bing, E.G., Burnam, M.A., Longshore, D., Fleishman, J.A., (2001). Psychiatric disorders and drug use among human immunodeficiency virus-infected adults in the United States. Archives of General Psychiatry, 58, 721–728.

[3] Asch, S. M., Kilbourne, A. M., Gifford, A. L., Burnam, M. A., Turner, B., Shapiro, M. F., & Bozzette, S. A. (2003). Underdiagnosis of depression in HIV. Journal of General Internal Medicine, 18, 450-460.

[4] Shippy, R. A., Mendez, D., Jones, K., Cergnul, I., & Karpiak, S. E. (2004). S-adenosylmethionine (SAM-e) for the treatment of depression in people living with HIV/AIDS. BMC Psychiatry, 4, 38.

[5] Earnshaw, V. A., Bogart, L. M., Dovidio, J. F., & Williams, D. R. (2013). Stigma and Racial/Ethnic HIV Disparities: Moving Toward Resilience. *The American Psychologist*, *68*(4), 225–236.

[6] Nichols, J.E., Speer, D., Watson, B. (2002). Aging with HIV: Psychological, Social, and Health Issues. Academic Press; San Diego.

[7] Reece, M. (2003). HIV-related mental health care: Factors influencing dropout among low-income, HIV-positive individuals. AIDS Care, 15, 707-716.

[8] Rockville (MD): Substance Abuse and Mental Health Services Administration (US); 2004. (Treatment Improvement Protocol (TIP) Series, No. 39.) Chapter 3 Approaches to Therapy. Available from: https://www.ncbi.nlm.nih.gov/books/NBK64259/

[9] Anderson et al, editors. Critical Perspectives on Racial and Ethnic Differences in Health in Late Life. Washington (DC): National Academies Press (US); 2004. 17, Behavioral Health Interventions: What Works and Why? Available from: https://www.ncbi.nlm.nih.gov/books/NBK25527/

[10] Zhu C, Yu J. Nonmetric multidimensional scaling corrects for population structure in association mapping with different sample types. Genetics. 2009;182(3):875-88.

[11] Houston, E., Argueta, C., Lacey, J., Mukherjee, S., & Breda-Hill, K. (2018). Reaching Mental Health Research Participants with Multiple Stigmas; A Description of Strategies Used in a Depression Intervention Study for YMSM of Color with HIV. Advanced Journal of Social Science, 3(1), 1-7. https://doi.org/10.21467/ajss.3.1.1-7

Address for correspondence

Sukrit Mukherjee, PhD, MSE.

Email: sukritmukherjee@cdrewu.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190622

Development of an Guideline-Based Decision Support System for Effective Diagnostic Workflow for Oncologic Pathologists

Yoko Nakanishi[a], Ryo Takahashi[b], Takuya Haga[c], Noriyuki Inoue[d], Yoshiaki Kondo[e], Shinobu Masuda[a], Yuichiro Gomi[f]

[a]Division of Oncologic Pathology, Department of Pathology and Microbiology, Nihon University School of Medicine, 30-1, Ohyagushi-kamimachi, Itabashi-ku, Tokyo, Japan,
[b]Graduate School of Science and Technology, Nihon University, 7-24-1, Narashinodai, Funabashi City, Chiba, Japan,
[c]Sakura Finetek Jappan Co.,Ltd., 2-31-1, Nihonbashi-Hamacho, Chuo-ku, Tokyo, Japan,
[d]Nihon University Business, Research and Intellectual Property Center, 4-8-24, Kudan-Minami, Chiyoda-ku, Tokyo, Japan,
[e]Division of Health Care Service Management, Department of Social Medicine, Nihon University School of Medicine, 30-1, Ohyagushi-kamimachi, Itabashi-ku, Tokyo, Japan,
[f]College of Science and Technology, Nihon University, 7-24-1, Narashinodai, Funabashi City, Chiba, Japan

Abstract

Accurate and rapid differential diagnosis are required for personalized cancer treatment. However, owing to the numerous molecular tests used for establishing a diagnosis, pathologists need time to investigate and confirm necessary test items. We present a guideline-based decision support system for effective workflow with regards to the molecular tests for pathological differential diagnosis.

Keywords:

Pathology, Clinical Decision Support System, Practice Guideline

Introduction

A guideline-based Clinical Decision Support System (CDSS) enables medical personnel to provide improved medical care. Many CDSSs have been studied in the field of pathology. Most of these systems extract or classify free-text from pathology reports using natural language processing[3] or electronic patient data using machine learning approaches based on dictionaries and coded data[1] and International Classification of Diseases for Oncology (ICD-O)[2]. Some systems used in pathology utilize image analysis approaches that are based on morphology[4,5]. For providing personalized medicine, another type of system uses graphical models for biomarkers [6]. These are the elements of data mining, but regulatory obstacles are encountered in terms of personal data, lack of standardization, and difficulties in versatility[4]. Appropriate pathological diagnosis is important in cancer treatment, particularly in personalized medicine. Recently, the diagnostic criteria and treatment strategies for various cancer types have undergone frequent revisions. Therefore, in cancer diagnosis, a pathologist spends a large amount of time and effort to review literature such as guidelines, authorized classification books, and journals. Reducing this time and effort will enable clinicians and patients make faster clinical decisions. There is also need for a system to compensate for the chronic shortage of pathologists.The aim of this study was to construct a prototype of a decision support system that can provide effective guideline-based pathological diagnostic workflow for oncology pathologists, regardless of the pathologists' experience and knowledge.

Methods

Position of the present system in the pathological workflow

An overview of the current system is shown in Figure 1. Consideration of personal information security is the most important aspect in medical system development. However, this application completely blocked patient data from the electronic medical records.

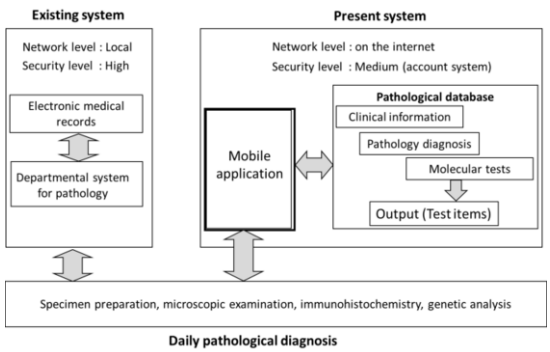


Figure 1. Overview of the current system in daily diagnostic work

Data

We collected data of each cancer type to create a comma separate value (CSV)-based master file on the local server. Data for each cancer type were extracted from "General Rule for Clinical and Pathological Record," "Japanese Clinical Guideline," "Pathology Atlas," and "WHO classification of Tumours" and coded sequentially. The following data were extracted: organ; site; method of tissue collection (biopsy or surgery); primary cancer – metastasis or un-clear; epithelial, non-epithelial, or unclear; and malignant, benign, or unclear. Diagnosis (superordinate), diagnosis (sub-ordinate), diagnosis (subordinate, English), ICD-O, target molecule for establishing the diagnosis, detection methods, expected results, and evidence (references). The expected results of molecular tests were displayed as microscopic images, because many patterns are observed in the results of immunohistochemical target molecule detection. Thus, it is difficult to understand the images based only on the text. Each photo was linked to the

corresponding text in the result. Actual immunohistochemical examination was performed for each stained sample, and the images were retrieved by the microscopic imaging software cellSense®V1.7 (Olympus Co., Tokyo, Japan).

User interface

Consideration of personal information security is the most important aspect in medical systems. This application was developed such that it could be used without inputting the IDs of patients or doctors. Therefore, this application can be used without direct access to electronic medical records of the hospital. Furthermore, considering the users and their locations, we believe that the usability of the application must not be restricted to personal computers but must also be extended to smart phones and tablets.

Decision making process

The current system was built using a web-form design with access to sequential files for making decisions regarding the molecular tests to be selected for pathological differential diagnosis. Data types were boolean nodes and guideline-based standard text nodes without free text entering. However, in addition to alternative Boolean nodes, the current system has a choice named “unidentified”. Pathology samples may not clearly indicate primary cancer or recurrence or the malignant or benign status. Sample related data, including the “unidentified” choice, are sequentially referred from the master .CSV file to help with extraction the result of candidate diagnosis. If the user selects “unidentified”, all the possible differential diagnostic nodes are extracted and displayed.

Results

On performing a demonstration test using the prototype CDSS for modeling differential diagnosis for lung cancer, we confirmed that our system operates properly in various web browser environments, including Internet Explorer, Google Chrome, Mozilla Firefox, Microsoft Edge, Safari, in each smart phone version, and in various terminals, including computers (Windows and Mac), and smart phones. Furthermore, the characteristics of the current system were studied. One of the potential benefits is that pathologists can obtain information of the latest guideline-based diagnostic terms and molecular tests, regardless of their experience and knowledge, with easy operations. The“unidentified” option will serve as an electronic dictionary and consultant to enable effective pathological diagnostic workflow. Therefore, we expect this system to promote workflow standardization, compared to the current workflow that relies on an individual's competence. Furthermore, portability to a mobile platform will allow oncology pathologists to use this state-of-the-art guideline-based CDSS in routine diagnostic workflow, without concerns about patients’ personal data breach. However, at the same time, the point of separation from personal data is a limitation of this system. If pathologists need personal patient data from electronic medical records, we will have to consider additional tools while simultaneously ensuring security of personal information. It is also necessary to increase the processing performance of this CDSS by creating a database to enable frequent revisions of the master file, in accordance to revisions of the guidelines. Furthermore, we plan to conduct a larger study with free distribution at pathology-related academic meetings in Japan to verify the utility and accuracy of this system. By increasing the number of pathologist users, accumulation of search logs will likely lead to new knowledge, particularly on rare cancer.

Conclusions

We present a guideline-based CDSS for effective diagnostic workflow for oncology pathologists. This system is expected to be more practical in terms of protecting personal data and mobile use. It is also expected to promote workflow standardization, regardless of the pathologists’ experience and knowledge.

Acknowledgements

This work was supported by the Nihon University Social Implementation Research Grant (2018-2019), MEXT KAKENHI Grant Number JP 18H03497, and Grant-in-Aid for Scientific Research (B). The authors thank staff and students at the Nihon University School of Medicine for their cooperation.

References

[1] S.N. Kasthurirathne, B.E. Dixon, and S.J. Grannis, Evaluating Methods for Identifying Cancer in Free-Text Pathology Reports Using Various Machine Learning and Data Preprocessing Approaches, Stud Health Technol Inform 216 (2015), 1070.
[2] M. Oleynik, M. Finger, and D.F. Patrão, Automated Classification of Pathology Reports, Stud Health Tech-nol Inform 216 (2015), 1040.
[3] N. Ashish, L. Dahm, and C. Boicey, University of California, Irvine-Pathology Extraction Pipeline: the pathology extraction pipeline for information extraction from pathology reports, Health Informatics J 20 (2014), 288-305.
[4] M.G. Hanna and L. Pantanowitz, The role of informatics in patient-centered care and personalized medicine, Cancer Cytopathol 125 (2017), 494-501.
[5] J. Arevalo, A. Cruz-Roa, V. Arias, E. Romero, and F.A. González, An unsupervised feature learning framework for basal cell carcinoma image analysis, Artif Intell Med 64 (2015), 131-45.
[6] P. Antal, J. Reiter, and P. Mátyus, From biomarkers to clinical decision support, Orv Hetil 156 (2015), 2077-81.

Address for correspondence

Correspondence to Yoko Nakanishi, PhD.
Division of Oncologic Pathology, Department of Pathology and Microbiology, Nihon University School of Medicine, 30-1, Ohyagushi-kamimachi, Itabashi-ku, Tokyo, 173-8610 Japan.
E-mail:nakanishi.youko@nihon-u.ac.jp
Phone: +81-3-3972-8111EXT2256
FAX: +81-3-3972-8163

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190623

Identify Facilitators and Challenges in Computerized Checklist Implementation

Shan Nan[a,b], Ashley De Bie[c,b], Sicui Zhang[a,b], Hendrikus Korsten[c,b], Xudong Lu[a,b], Huilong Duan[a]

[a] *College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China,*
[b] *Faculty of Industry Engineering and Innovation Science, Technical University of Eindhoven, Eindhoven, The Netherlands,*
[c] *Department of Anesthesiology and Intensive Care, Catharina Hospital, Eindhoven, The Netherlands*

Abstract

Safety checklists have been considered as a promising tool for improving patient safety for decades. Computerized checklists have better performance compared with paper-based checklists, though there are barriers to their adoption. Given previous literature, it is still unclear what assists implementations and their challenges. To address this issue, this paper summarizes the implementation of two successful computerized checklist implementations in two countries for two different clinical scenarios and analyzes their facilitators and challenges.

Keywords:

Patient Safety; Checklist; Decision Support Systems, Clinical

Introduction

In recent years, several studies have proved that process-oriented and context-aware computerized checklists can improve adherence to safety checklists and consequently improve clinical outcomes [1]. Previous studies have addressed several technical barriers to the wide adoption of computerized checklists [2]; however, it is still unclear what the facilitators and challenges are in the implementation phase.

Due to the limited amount of implementation research, it is still not possible to summarize the facilitators and challenges out of literature. However, our research team has recently implemented computerized checklist systems in two countries for two different clinical scenarios [3]. It would be meaningful to summarize and communicate the facilitators and challenges gained from these two studies.

An intensive care unit (ICU) round computerized checklist and a supporting system were developed and implemented in a Dutch tertiary hospital in 2014. The computerized checklist was connected to the electronic medical record (EMR) system in the hospital. Each item is customized to specific patient conditions by executing clinical rules. Clinical rules help intensivists to double check items, highlight items that were critical for specific patients, and tailor items to become specific for the patient. A simulation-based study was carried out to validate the user acceptance and effectiveness. Compared to the paper-based checklist, the adherence to the checklist increased from 73.6% to 100% [3].

A computerized percutaneous coronary intervention (PCI) peri-operative checklist set was developed for a Chinese tier-three hospital from 2015 to 2017. The computerized checklist was also connected to the hospital EMR. Patient data were extracted out of the EMR database. A semi-structured interview was carried out while implementing this checklist. Cardiologists reported this checklist helped them to complete the safety checklist faster, since the checklist helped them by showing relevant patient data and check items automatically.

In the remainder of this paper, we identify facilitators and challenges by analyzing our own experiences gained from these two cases.

Methods

While developing the computerized checklists in these two studies, we followed a proof-by-demonstration approach [4]. With this approach, we divided the development into several iterations. In each iteration, we demonstrated the latest software to the clinical users, interviewed them while demonstrating, analyzed users' comments, and applied validated comments to the next version of the software that would be discussed in the next iteration. Change Management Principles [5] were applied while implementing the software. After several iterations, the software was stable and brought to daily use.

In the Dutch case, two engineers and one intensivist had worked on developing checklist items and related clinical rules for half a year based on the local protocols. Another large part of the clinical rules was derived from the Clinical Decision Support System (CDSS) knowledge base.

In the Chinese case, three engineers and two cardiologists had been working on the computerized checklist for two years. The related clinical rules were based on their existing paper-based checklist and narrative clinical guidelines.

Users' comments were collected and analyzed afterward. Those comments on what they liked about computerized checklists were categorized as facilitators. Those comments on what may hinder the acceptance were considered challenges. Additionally, engineers' own development experiences were summarized into these two categories.

Results

Facilitators

Use the Established Local Standard of Care

Both hospitals already had a great amount of established local standard of care, includes existing guidelines, pathways, safety checklists, CDSSs, and other approaches that aimed to

improve quality of care. These standards were derived from published literature and adapted to their hospitals.

All providers who participated in this study suggested that this knowledge was already accepted and followed in the department level. Using this knowledge made checklist items more acceptable and reasonable to clinicians.

Make Computerized Checklists Dynamic

Providers claimed they refused to use paper-based checklists due to its static nature that did not fit in their dynamic and demanding daily workflow.

A computerized checklist that could be adapted to each specific patient made much more sense to clinicians. During the interviews, clinicians mentioned that they liked the dynamic properties as it saved time to complete the checklists and it could detect more patient-specific problems that were worth noticing.

Adaptive to Clinicians' Requirements

In both of the studies, during our implementations, clinicians also mentioned new potential clinical rules sparking their interest in computerized checklists. Especially in the Dutch case, a visualized knowledge acquisition tool was used to build clinical rules. An intensivist implemented and updated the clinical rules independently.

Provided Users Additional Value

The length of time to complete computerized checklists was a major concern for checklist users. Users claimed that the exercise of ticking boxes was time-consuming, even though it was necessary. Nevertheless, they eventually liked the idea of the computerized checklist after several iterations, because it provided additional value. While computerized checklists could help clinicians complete the checklists and worked as cognitive aids resulting in better acceptance and compliance, the additional value of the digital checklist was also their capability to extract and display relevant patient data, to analyze data automatically, and to provide evidence-based literature or local protocol.

Reuse Existing Hospital Information Systems

In our practice, we found that by reusing existing hospital information systems and their components as much as we could, we could save a great amount of time and make the computerized checklist easier to be accepted.

One example was the CDSS used in the Dutch hospital. The clinical rules in the system had been developed and tested for decades. Intensivists had all agreed on them and were familiar with these rules. Reusing these rules saved not only time for the development phase, but also made the checklist items more convinceable to the intensivists

Challenges

Users' Perception and Medical Culture

Checklists could help deliver more transparent healthcare, that people could know who did what at what time; however, some providers had their concerns. Some clinicians worried about patients and their family, who could lack medical knowledge and could misinterpret the records that they made. This could be used as evidence against them if something wrong were to happen. Therefore, these clinicians refused to use the checklists, or they ticked everything to avoid trouble.

The Right Level of Variability

Healthcare processes were highly variable. It was yet difficult for computerized checklists to cover every path of the healthcare process. In some situations, there were no checklists which could cover them or a checklist could no longer validate for those cases.

Knowledge Acquisition

The cost of knowledge acquisition was high. Knowledge engineers and clinical experts had difficulties in understanding each other. The knowledge provided by experts did not always reflect the specific problem in a specific department. It sometimes took several iterations to finalize a clinical rule.

Conclusions

Safety checklists have been considered a promising tool for improving patient safety for a decade. Computerized checklists can help implement safety checklists in a process-oriented and patient context-aware way so that they fit better in medical practices; however, due to the lack of experiences, it remains unclear what the facilitators and challenges are when implementing them in hospitals. To help accelerate the widespread adoption of computerized checklists, this paper summarizes facilitators and challenges of two successful computerized checklist implementations. Suggestions for future computerized checklist implementations and other possible research directions are considered for future research.

Acknowledgments

The research leading to these results has received funding from Chinese National Science and Technology Major Project (grant number 2016YFC0901703) and the Brain Bridge Project sponsored by Philips Research. The authors would like to thank the intensivists and cardiologists involved in this study.

References

[1] H.S. Kramer, F.A. Drews, Checking the lists: A systematic review of electronic checklist use in health care, *J. Biomed. Inform.* **71** (2017) S6–S12.
[2] S. Nan, X.-D. Lu, P. Van Gorp, H.H.M. Korsten, R. Vdovjak, U. Kaymak, H.-I Duan, Design and implementation of a platform for configuring clinical dynamic safety checklist applications, *Front Inf. Technol Electron Eng*. **19** (2018) 937–946.
[3] A. de Bie Dekker, S. Nan, L. Vermeulen, *et al.*, Improving checklists: Making them dynamic and connected to the EHR, in: Proceedings of the 11th Annu. Int. Conf. Rapid Response Syst. Med. Emerg. Teams, (2015).
[4] A.S. Vinzé, D.G. Gregg, U.R. Kulkarni, Understanding the philosophical underpinnings of software engineering research in information systems, *Inf. Syst. Front.* **3** (2001) 169–183.
[5] R. Kandt, Organizational change management principles and practices, *Fourth Int. Conf. Prod. Focus.* (2002).

Address for correspondence

Correspondence to Xudong Lu, lvxd@zju.edu.cn

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190624

A Novel Platform to Define Chemotherapy Templates and Their Prescriptions

Hernán Navas[a], María Eugenia Liva[b], Federico Rossi[b]

[a] *Health Informatics Consultant at Cloodie S.A., Buenos Aires, Argentina*
[b] *Software Engineer at Cloodie S.A., Buenos Aires, Argentina*

Abstract

Medication errors have been identified as a major type of medical errors. Chemotherapy medication errors that occur in the prescription phase appear to be related to more significant adverse outcomes. The use of pre-printed templates increases patient safety. The functionalities required for the prescription of chemotherapy are not usually part of Clinical Physician Order Entry. The implementation of electronic chemotherapy templates will reduce the prescription errors.

Keywords:
Electronic Health Records, Antineoplastic Agents, Patient Safety.

Introduction

Medication errors have been identified as a major type of medical errors. It is one of the most risky for patients. The highest percentage occurs at prescription time and its cause is multifactorial. Chemotherapy medications pose an additional risk due to the limited safety margin existing in its dosage and its high toxicity. The prescription phase medication errors appear to be related to more significant adverse outcomes.

The use of pre-printed templates on paper with complete treatment schemes increases patient safety [1]. The use of the Clinical Physician Order Entry (CPOE) also improves patient safety by generating clear, complete and accurate orders.

The implementation of CPOE in chemotherapy is related to a reduction in the risk of errors [2,3], mainly in the dose calculation, and in an improvement in administration times.

The precision of the preparation of the dilution to ensure adequate concentration is essential to optimize the efficacy and safety of the medications administered.

The aim of this paper is to describe the development of a system to create and keep electronic chemotherapy templates.

Methods

Cloodie HIS Electronic Medical Record is unique per patient. In inpatient areas, the system allows the request and visualization of complementary tests through the integration with ancillary services (laboratory, imaging, etc.), write clinical notes, and medication prescription, for general drugs.

As a requirement for a specialized Institution were conducted a joint evaluation of the prescription system for chemotherapy orders. The existent functionalities were not enough to allow the safe prescription of these medications.

They were collected and analyzed the existing paper templates. The templates filled out by the physicians were also included. In addition to the information contained in each field, all handwritten notes beside the structure of the templates were taken into account (such as marginal notes, footnotes or notes on the back). Such analysis sought to identify:

- Prescription categories, according to their structure.
- Order and relation between administrations.
- Forms of preparation/ dilution of drugs and mixtures.
- Additional variables related to administration

Results

35 empty templates and 55 complete templates were analyzed. It was possible to identify:

- Prescription categories: hydration, general and alternatives.
- Order and relation between administrations: absolute indications (as regards the time of administration), and relative to other existing ones.
- Forms of preparation/ dilution of drugs and mixtures: total and partial dilutions. Variables were also detected in terms of the dependence of the amount of solute and/ or solvent to be used.
- Additional variables related to administration: infusion rates variable in time.

A. Template Management System

The information was distributed in 4 tabs:

- General information.
- Medications.
- Preview, in calendar format.
- Signatures.

B. Parameterization of indications

The definition of the medication has been structured using an association scheme linked to two fundamental concepts. On the one hand, the conformation of a grouping entity of one or more drugs (forming a mixture), called “Drug Group”, and, on the other hand, the definition of how this group is administered over time, allowed by the template, called “Planning”. Each group of medications must have at least one plan, but it may have more than one (see Figure 1).

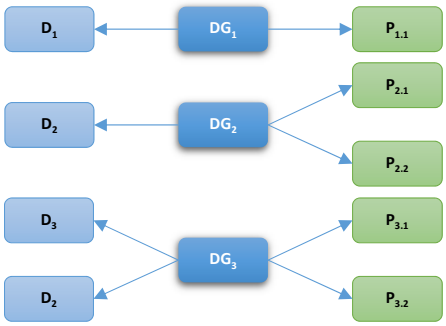


Figure 1 – Relation between "Drug", "Drug Group" and "Planning"

B1. Drug Group

It allows creating sets of medication that will be prepared by means of a mixture, if their presentation permitting, and if administered by the same route. Otherwise, only one medication per group is allowed (if the dosage form is a tablet).

B2. Definition of each plan

Each "Drug Group" may contain multiple plans. Each plan shall have as main data the dose and the time of application. In this way, the same group may be administered with a dose on certain days and with a different dose on other days.

Each plan must include:

- Basic information: the administration route must be defined (according to the selected medication), the days on which it will be administered and the temporal relation with other medications, if applicable.
- Dose: it can be defined relatively according to some anthropometric parameter of the patient; or absolutely. When expressed in a relative manner, the system will make the calculation at the prescription time.
- Method: to determine the administration frequency.
- Dilution scheme:
  - Fixed: i.e. regardless of the amount of solute (drug), it will always be diluted by the same volume of solvent.
  - Relative concentration: When a single final mixture is prepared with variable quantity of both the solute (drug) and the solvent according to the anthropometric parameters.
  - Absolute concentration: the same preparation is repeated in "n" number of bags, keeping the concentration in each of them.
- Flow: It allows specifying different rates of administration of the diluted medication.
- Observations: Free text for any relevant information.

C. Safety and control mechanisms

C1. Internal Validation

Some of the blocking validations are as follows:

- Administration days: the system will not allow values out of the range comprised by the duration in days to be entered on the template.
- Modifications from pharmacy: each time data are entered on the template, the system verifies if a change was made from the pharmacy master system on any of the attributes of each of entered medications. If so, it will alert the physician and request them to carry out the homologation.
- Maximum dose: At the prescription time, the attending physician must not exceed this value under any circumstances. If the automatic calculation resulted in a value higher than the allowed one, the physician shall adjust the parameter below the maximum established.

There are other warning that may be ignored, for example:

- Percentage of dose adjustment: for each drug on each template a percentage may be defined, within which the adjustment of the dose is recommended. If the attending physician modifies the dose beyond such percentage, a limit warning will be displayed.

C2. External Validation

A circuit was defined for the safe use of templates.

1. Creation.
2. Signature: The signing physician must evaluate the completeness and accuracy of the template.
3. Available for use by the attending physicians.
4. Template without signature: If the template was modified after the signature, it will be left out of the healthcare process, requiring the pertaining signature again. This does not affect the templates in progress.

Conclusions

The implementation of electronic chemotherapy templates will make it possible to reduce errors related to the prescription, while speeding up the prescription process by the attending physicians. It will simplify the management of such templates every time it is necessary to carry out any modification to update the treatment or removal due to obsolescence, guaranteeing the reliability of the available templates.

References

[1] A. Saad, C.A. Der-Nigoghossian, R. Njeim, R. Sakr, P. Salameh, and M. Massoud, Prescription Errors with Chemotherapy: Quality Improvement through Standardized Order Templates, *Asian Pac. J. Cancer Prev.* **17** (2016) 2329–2336.

[2] J. Bubalo, B.A. Warden, J.J. Wiegel, T. Nishida, E. Handel, L.M. Svoboda, L. Nguyen, and P.N. Edillo, Does applying technology throughout the medication use process improve patient safety with antineoplastics?, *J Oncol Pharm Pract*. **20** (2014) 445–460. doi:10.1177/1078155213514469.

[3] B.R. Meisenberg, R.R. Wright, and C.J. Brady-Copertino, Reduction in chemotherapy order errors with computerized physician order entry, *J Oncol Pract*. **10** (2014) e5-9. doi:10.1200/JOP.2013.000903.

Address for correspondence

Dr. Hernán Navas - hnavas@cloodie.com

Health Informatics Consultant. Cloodie S.A. Juana Manso 670, Floor 34, Appartment 2, C1107CBN, Ciudad Autónoma de Buenos Aires.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190625

Incidence of Falls in a General Hospital in Southern Brazil

Aline Tsuma Gaedke Nomura[a], Murilo dos Santos Graeff[a], Lisiane Pruinelli[b], Miriam de Abreu Almeida[a]

[a] *School of Nursing, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil*
[b] *School of Nursing, University of Minnesota, Minneapolis, Minnesota, USA*

Abstract

Understanding the multifactorial dimension of fall risk contributes to the search for better preventive interventions. The objective of this study was to identify the events types and the most frequent causes related to falls in adults and newborns. In the included period, 378 falls occurred. Understanding the predictive factors for the occurrence of falls has the potential to guarantee better safety for the patient.

Keywords:

Nursing Informatics, Accidental Falls, Electronic Health Records

Introduction

Falls are the second leading cause of accidental or unintentional injury deaths worldwide [1]. A fall is defined by the World Health Organization (WHO) as "an event which results in a person coming to rest inadvertently on the ground or floor or other lower level" [2].

A study pointed to a drop rate of patients in hospitals in developed countries that ranged from 3 to 5 falls per 1,000 patient-days [3]. In addition, it pointed out that these falls resulted in damages for 30% to 50% of these cases, such as fractures, subdural hematomas, and bleeding, which can lead to death. This creates a negative impact on the mobility of patients, as well as anxiety, depression, and fear of falling again, which increases the risk of new falls.

In another study that sought to evaluate the risk and incidence of falls in hospitalized adult patients [4], authors have identified that if the patient has a high-risk score for falls on admission, this risk tends to remain high until hospital discharge. The incidence rate was 1.68%, with a higher percentage of patients classified as having a high risk for falls.

In this context, a study suggested there was an improvement in patient care practices associated with better nursing records after the implementation of specific protocols in electronic health records (EHR), and the prevention of falls is one of the greatest improvements [5].

Risks of falling is multifactorial. Among the most frequent are: disorientation/confusion, frequent urination, walking limitations, absence of caregiver, postoperative period, and number of medications administered within 72 hours before the fall (last dose of the classes: benzodiazepines, opioids, barbiturates, antipsychotics, antidepressants, antihypertensive, laxatives, diuretics, antihistamines, anticonvulsants, and sedatives) [6]. Thus, understanding the dimension of these events, through EHR, supports nursing clinical decision-making, which contributes to the search for the best preventive interventions and positively impacts patient safety [6].

The use of clinical decision support (CDS) associated with EHR has been widely used as a tool to provide the best evidence assisting physicians, patients, and other team members to achieve better care outcomes and to reduce costs in health systems [7]. According to the Office of the National Coordinator for Health Information Technology (ONC), these tools include computerized alerts and reminders for professionals, clinical guidelines, focused reports and summaries of patient data, documentation models, diagnostic support, and contextually relevant reference information, among other tools [8].

Thus, the questions are: How are the incidence of falls? What are the most common types and causes? The objective of this study was to identify the events related to, the types of, and the most frequent causes of falls in adults and newborns.

Methods

This was a quantitative, descriptive, and documental study, conducted in a private hospital in Porto Alegre in southern Brazil. The population is composed of all medical records of patients admitted to the institution from January 2011 to October 2018.

The EHR was implemented at the institution in 2008, and includes all items related to patient care, as well as decision support tools such as event management, quality, and other administrative functions. Through computerized record entries, nurses reported the falls of patients hospitalized in adult and neonatal units.

These units account for a total of 189 beds. For each notification, the Patient Safety Center (NSP) opens an event analysis form to identify possible causes, thereby generating actions to mitigate event damage. In addition, it promotes improvements targeting the prevention of other falls for the same reasons. Since the implementation of the NSP, the team has already promoted several actions to reduce the risk of falls and minimize damages.

Among the actions are: identify the patient at risk by the adoption of a bracelet, mandatory use of safety in wheelchairs, seat belts, transport of the newborn in crib or litter, placement of support bars in corridors and restrooms, and implementation of bedridden patient routines such as low beds with grids, among others. In addition, system alerts were implemented into the EHR, as a form of communication among the multi-professional health team.

Data was obtained through the EHR software database of the institution and descriptive statistics were used for analysis.

Results

During the study period, there were 378 falls, which represented 1.07 falls per 1,000 patient-days. Of these, 252 (66.66%) were falls from their own height and 126 (33.34%) were falls from a stretcher, chair, or armchair. The most frequent causes were: sensitivity deficit (24.94%), loss of balance and gait (14.40%), mental state alteration (12.08%), drugs that affect the central nervous system (6.68%), wet floor (6.43%), furniture and equipment interference (4.88%), inadequate footwear (3.60%), and bed without grid (3.34%). These numbers represented 76.35% of all events.

Conclusions

The study allowed the identification of the characteristics of events related to falls using a large EHR system with the aid of a clinical-decision tool specifically designed for the identification of the risk of and the event of falls. With the adoption of electronic systems and the systematization of care, it was possible to understand the predictive factors for the occurrence of falls in this field of study. From the data generated, it is possible to obtain resources to evaluate, plan, and implement actions aiming to reduce falls in the hospital environment in order to promote greater patient safety.

Although multifactorial, the findings determined the main factors that led to falls. Through critical thinking and clinical judgment, it is possible to implement a preventive care plan that is more attentive to related causes. A preventive care plan includes: to plan a safer environment with the physical structure and furniture, to evaluate the need for greater human resources, to implement patient, family, and professional educational programs, and to make use of specific instruments and protocols to prevent falls. These are some strategies to be implemented and/or optimized.

It is important to highlight the important role of the nurse in management, risk control, and hospital safety. It should also be emphasized that prevention processes should be systematically analyzed and reviewed, guided by best care practices. EHR and the standardization of nursing records, combined with information systems improvement, can contribute to the development and management of quality indicators, promoting decision-making and patient safety.

Acknowledgements

We would especially like to thank colleagues at the Patient Safety Center for their support in planning and conducting this study. We also give special thanks to the technicians of the information technology who provided technical support during the data collection.

The authors have been developing research projects on Nursing Processes and electronic health records, based on Computer Science.

References

[1] Brazilian Society of Geriatrics and Gerontology, S.R.M. Pereira, S. Buksman, M. Perracini, L. Py, K.M.L. Barreto, and V.M.M. Leite, Project Guidelines: Falls in the Elderly, Federal Council of Medicine, SP, 2008.

[2] Word Health Organization (CH), WHO Global Report on Falls Prevention in Older Age, (2017). http://www.who.int/ageing/projects/falls_prevention_older_age/en/.

[3] D. Oliver, F. Healey, and T.P. Haines, Preventing falls and fall-related injuries in hospitals, *Clinics in Geriatric Medicine* **26** (2010) 645-692.

[4] T.S. Pasa, T.S.B.S. Magnago, J.S. Urbanetto, M.A.M. Baratto, B.X. Morais, and J.B. Carollo, Risk assessment and incidence of falls in adult hospitalized patients, *Revista Latino-Americana de Enfermagem* **25** (2017) e2862.

[5] A.T.G. Nomura, L. Pruinelli, M.B. Silva, A.F. Lucena, and M.A. Almeida, Quality of electronic nursing records: The impact of educational interventions during a hospital accreditation process, *Computers, Informatics, Nursing* **36** (2018) 127-138.

[6] I.M. Severo, R.S. Kuchenbecker, D.F.V.B. Vieira, A.F. Lucena, and M.A. Almeida, Risk factors for fall occurrence in hospitalized adult patients: A case-control study, *Revista Latino-Americana Enfermagem* **26** (2018) e3016.

[7] J.E. Tcheng, S. Bakken, D.W. Bates, H. Bonner III, T.K. Gandhi, M. Josephs, K. Kawamoto, E.A. Lomotan, E. Mackay, B. Middleton, J.M. Teich, S. Weingarten, and M.H. Lopez, Optimizing Strategies for Clinical Decision Support: Summary of a Meeting Series, National Academy of Medicine, Washington, DC, 2017.

[8] USA HealthIT.gov, Clinical Decision Support, (2018). https://www.healthit.gov/policy-researchers-implementers/clinical-decision-support-cds.

Address for Correspondence

Aline Tsuma Gaedke Nomura

963 São Manoel St, Porto Alegre, RS, 90620-110, Brazil

alinenomura@hotmail.com

Phone: 55 51 996677978

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190626

Midwives' Perception of Using a Knowledge Base on Fetal Impact of Drugs

Ulrika Nörby[a,b], Tero Shemeikka[a], Birger Winbladh[b,c]

[a] *Health and Medical Care Administration, Stockholm County Council, Sweden*
[b] *Center of Reproduction Epidemiology, Department of Clinical Sciences, Lund University, Sweden*
[c] *Department of Clinical Sciences and Education, South Hospital, Karolinska Institutet, Stockholm,* Sweden

Abstract

A non-commercial knowledge base providing assessments of fetal risks of medicinal drugs is a useful tool in the everyday work of midwives. The information is freely available on the internet, and according to a questionnaire study, nearly 95% of the midwives are familiar with the database, 30% use the information weekly, and 80% express that it affects their medical decisions. A vast majority of the midwives also state that it is time-saving.

Keywords:
Knowledge Bases, Risk assessment, Midwifery

Introduction

The midwife has a central position in perinatal care in Sweden with its own responsibility for normal pregnancy, delivery, and care after birth [1]. As a result, the midwife is usually the primary contact person for the pregnant woman. The midwife must assess whether there are any complications that have to be referred to a physician.

Over the years, drug treatment during pregnancy has become more and more common [2; 3], probably due to the fact that pregnant women are older nowadays and because effective therapy of chronic diseases allows more women to give birth. Hence, most women use drugs at some point during their pregnancies [4-6]. A common question is thereby whether exposure to different medicinal drugs might harm the unborn baby. Often, both pregnant women and health care professionals believe that the fetal risk of drug exposure is higher than the actual risk [7; 8].

To support health care professionals, the non-commercial Swedish knowledge base *Janusmed Drugs and Birth Defects* provides assessments of fetal risks of the approximately 1450 drug medicinal substances on the Swedish market. It is available at www.janusinfo.se/fosterpaverkan, as well as integrated into several electronic health records [9].

The contents are based on substances and in most cases one document deals with one substance. Each document is after that linked to a medical products register via a combination of unique substance identifiers and ATC codes (Anatomical Therapeutic Chemical classification codes), which allows the users to search for both substances and drug products. The texts are written in a structured way with short risk assessments, background, and references. They are also classified according to the 3-tier system dependent on the risk level.

The medical contents are better adapted to clinical decision making than the official summaries of product characteristics for the drug products, which are not seldom inconsistent and sometimes even incorrect [10; 11].

The knowledge base has previously been shown to be useful for pregnant women, even though it is not adapted for lay people [12]. In this survey, we have evaluated how midwives as key persons in antenatal care, use and perceive the web-based version of the database.

Methods

An electronic questionnaire focusing on the use and value of the database was sent by e-mail to 960 midwives in Sweden. The e-mail addresses were provided by coordinating midwives who have regional responsibility for training, routines and quality control of antenatal care. The questionnaire consisted of multiple-choice, scaled and open-ended questions. We used the survey tool Easyresearch provided by Questback to set up and collect the data.

Results

In total, 408 midwives answered the questionnaire which corresponds to a response rate of 48% among respondents with a valid e-mail address. 93% of the midwives were familiar with the database: among them, 4% used it daily, 26% weekly, 35% at least once per month and 35% more seldom. They used the information both during the pregnant woman's visit at the clinic (80%) and before/after the consultation (74%) and to a lesser extent for teaching and own studies (17%).

The most common situation was when a pregnant woman already had used a drug, in which case 96% of the midwives consulted the database. However, 46% used it to choose suitable drug treatment for a pregnant woman. Overall, the midwives perceived that the knowledge base provided answers to their questions and that it had an impact on their medical decisions (Table 1). They also often read the texts together with the pregnant woman and perceived that the information is useful to decrease the woman's anxiety of teratogenic effects, see Table 1. Finally, 76% of the midwives answered that the knowledge base saves time.

The following answers to open-ended questions illustrate the opinions of the midwives:

"Typically, I am able to calm down a worried patient who used a medication during her pregnancy"

"I use the database to collect facts concerning a drug and provide the woman with proper information. Sometimes, an appointment with a physician is made."

"Saves a lot of time. It is tremendously useful!"

Table 1– Midwives' use of the knowledge base – key questions

Question	Response option	n	%
To what extent does *Janusmed Drugs and Birth Defects* provide answers to your questions?	Very high extent	258	54
	Rather high extent	178	38
	Neither high nor low	15	3
	Rather low extent	3	1
	Very low extent	3	1
	Don't know	17	4
How often do the risk assessments in *Janusmed Drugs and Birth Defects* affect your medical decisions?	Very often	96	27
	Often	99	28
	Sometimes	83	24
	Seldom	12	3
	Never	3	1
	Not relevant	58	16
How does the information in *Janusmed Drugs and Birth Defects* affect the pregnant woman's anxiety for negative fetal impact?	Decreases to a high extent	120	34
	Decreases to some extent	184	52
	Neither high nor low	2	1
	No impact	2	1
	Increases to a high extent	2	1
	Not relevant/can not judge	81	17
Do you read the texts together with the pregnant woman or advice her to use the knowledge base?	Yes, often	84	24
	Yes, sometimes	178	51
	No	88	25

Discussion

By using the knowledge base *Janusmed Drugs and Birth Defects,* midwives are able to handle many questions independently, which unburdens the work-load of the physicians and provides the pregnant women with evaluated advice in an efficient way. It is an advantage that the database benefits both health care professionals and pregnant women. Using the same primary information resource likely facilitates communication and prevents misunderstanding.

Conclusions

The knowledge base *Janusmed Drugs and Birth Defects* is extensively used and appreciated by midwives in Sweden. Further, it has a direct impact on their medical decisions and advice at the point of care.

Acknowledgements

The authors would like to thank the coordinating midwives for providing the
e-mail addresses to the respondents.

References

[1] The Swedish Society of Obstetricians and Gynecologists (SFOG): *Mödrahälsovård, Sexuell och Reproduktiv hälsa [Antenatal care, sexual and reproductive health]. Report nr 59.* 2008. Updated 2016.

[2] A.A. Mitchell, S.M. Gilboa, M.M. Werler, K.E. Kelley, C. Louik, S. Hernandez-Diaz, and S. National Birth Defects Prevention, Medication use during pregnancy, with particular focus on prescription drugs: 1976-2008, *Am J Obstet Gynecol* **205** (2011), 51 e51-58.

[3] S.H.L. Thomas and L.M. Yates, Prescribing without evidence - pregnancy, *British Journal of Clinical Pharmacology* **74** (2012), 691-697.

[4] A. Lupattelli, O. Spigset, M.J. Twigg, K. Zagorodnikova, A.C. Mardby, M.E. Moretti, M. Drozd, A. Panchaud, K. Hameen-Anttila, A. Rieutord, R. Gjergja Juraski, M. Odalovic, D. Kennedy, G. Rudolf, H. Juch, A. Passier, I. Bjornsdottir, and H. Nordeng, Medication use in pregnancy: a cross-sectional, multinational web-based study, *BMJ Open* **4** (2014), e004365.

[5] K. Palmsten, S. Hernandez-Diaz, C.D. Chambers, H. Mogun, S. Lai, T.P. Gilmer, and K.F. Huybrechts, The Most Commonly Dispensed Prescription Medications Among Pregnant Women Enrolled in the U.S. Medicaid Program, *Obstet Gynecol* (2015).

[6] O. Stephansson, F. Granath, T. Svensson, B. Haglund, A. Ekbom, and H. Kieler, Drug use during pregnancy in Sweden – assessed by the Prescribed Drug Register and the Medical Birth Register, *Clinical Epidemiology* **3** (2011), 43-50.

[7] C. Csajka, A. Jaquet, U. Winterfeld, Y. Meyer, A. Einarson, and A. Panchaud, Risk perception by healthcare professionals related to drug use during pregnancy: a Swiss survey, *Swiss Med Wkly* **144** (2014), w13936.

[8] H. Nordeng, E. Ystrom, and A. Einarson, Perception of risk regarding the use of medications and other exposures during pregnancy, *Eur J Clin Pharmacol* **66** (2010), 207-214.

[9] U. Norby, K. Kallen, B. Eiermann, S. Korkmaz, B. Winbladh, and L.L. Gustafsson, Drugs and Birth Defects: a knowledge database providing risk assessments based on national health registers, *Eur J Clin Pharmacol* **69** (2013), 889-899.

[10] I. Fusier, C. Tollier, and M.C. Husson, Infovigilance: reporting errors in official drug information sources, *Pharm World Sci* **27** (2005), 166-169.

[11] P. Warrer, L. Aagaard, and E.H. Hansen, Comparison of pregnancy and lactation labeling for attention-deficit hyperactivity disorder drugs marketed in Australia, the USA, Denmark, and the UK, *Drug Saf* **37** (2014), 805-813.

[12] U. Norby, K. Kallen, T. Shemeikka, S. Korkmaz, and B. Winbladh, Pregnant women's view on the Swedish internet resource Drugs and Birth Defects intended for health care professionals, *Acta Obstet Gynecol Scand* **94** (2015), 960-968.

Address for correspondence

Ulrika Nörby, Health and Medical Care Administration, Stockholm County Council, Stockholm, Sweden,
e-mail: ulrika.norby@sll.se

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190627

Development of Albumin Analyzer with Whole Blood and Application to Telemedicine for Patient Nutrition Management in Home Health Nursing

H. Nozaka[a], H. Sazawa[b], S. Shiroto[a], M. Nakano[a], H. Takami[a]

[a] Hirosaki University Graduate school of Health Sciences, Hirosaki city, Aomori, Japan
[b] Department of Clinical Labolatory, Hokkaido Cancer Center, Sapporo city, Hokkaido, Japan

Abstract

Albumin level is q significant indicator of patient nutritional status. However, Point of Care Testing (POCT) devices and telemedicine system that nurses can operate easily in-home medical care is not developed. The aim of this work is the development of a POCT device for Albumin level and application to a telemedicine support system. The operability of our system was simple and easy for the nurse or patient. We believe our method is useful for Nutrition Support Team activities in-home medical care.

Keywords:
Telemedicine, Home Health Nursing, Nutrition Status

Introduction

Albumin (ALB) level is a significant indicator of a patient's long-term nutritional status[1,2]. Nutritional management is reported as strongly related to patient prognosis, recovery period, and survival rate. Nutrition Support Team (NST) activities become more important not only in hospital but also in home health nursing. Biochemical analysis with Point of Care Testing (POCT) in home health nursing is useful for the understanding of patient nutritional status, but measuring principle of albumin level with whole blood specimen has not been developed[3]. The aim of this work is the development of a POCT device for albumin test with whole blood specimen and application to a telemedicine support system for patient nutrition management in home health nursing.

Methods

We designed a POCT device that can measure hemoglobin and albumin level with whole blood from the fingertip. The hardware consisted of two photo-transistor units, LED units, and "NI-USB DAQ" (Figures 1, 2). The SLS-hemoglobin method was adopted for hemoglobin testing, and the BCG method was adopted for albumin testing. We improved the BCG method for whole blood test specimens and devised the data convert algorithms from whole blood result to serum result (Figure 3).

Tele-medicine system for patient nutrition management

We used the iPad and telepresence robot "Kubi" as the telemedicine support system. Web conference system "Zoom" was used for face-to-face coaching between nurse and patient. Also, we developed a chatbot system, "e-nutrition" as an automatic response system in non-face-to-face (Figures 4, 5).

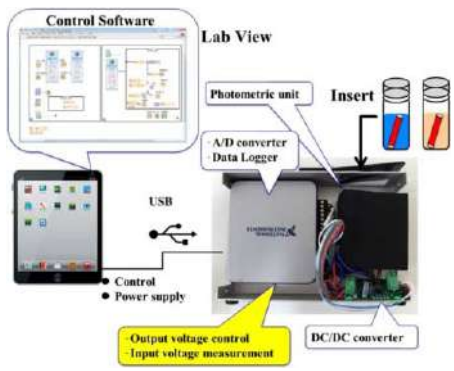


Figure 1– Whole blood albumin analyzer

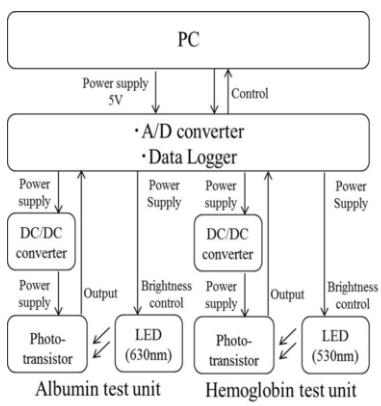


Figure 2 – System Block Diagram

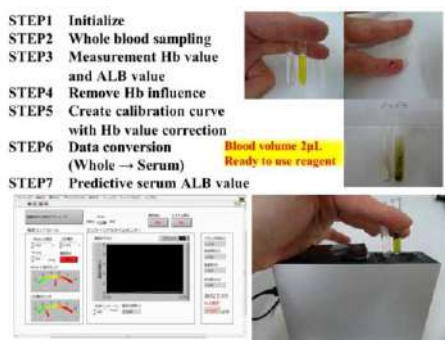


Figure 3– Prototype system of ALB POCT device

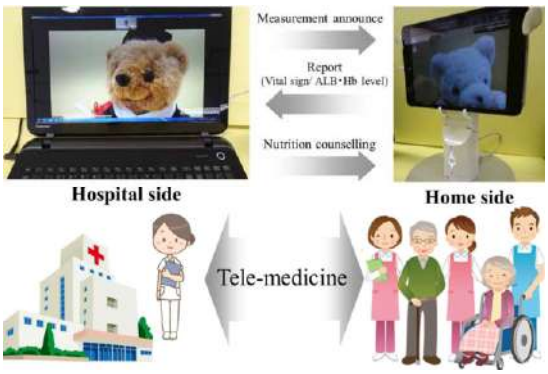


Figure 4– Face-to-face coarching between nurse and patient

Figure 5– Automatic response system by chat bot system

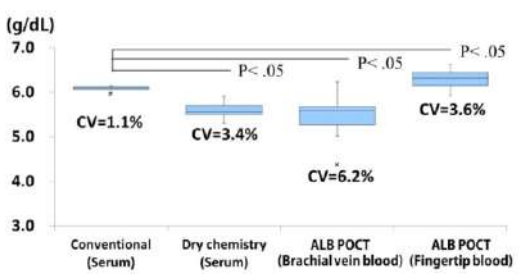


Figure 6– Repeatability and Accuracy

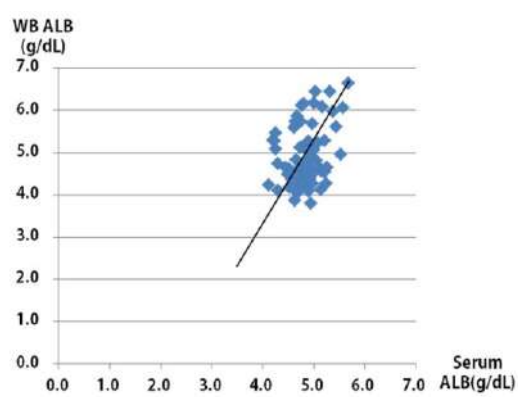


Figure 7– Correlation between actual measurementserum ALB level and estimated ALB level with whole blood

Clinical trial

We evaluated the correlation and accuracy of ALB level between the hospital and the home. Evaluation of the telemedicine system was carried out by patients survey. 1) Study subjects: fifty normal healthy persons. 2) Home side: Patients draw fingertip blood with lancet by themselves, and collected into a microcapillary. ALB level was measured with whole blood by our POCT system. 3) Hospital side: We draw peripheral blood from a brachial vein and collected into a blood collection tube. Serum was recovered after centrifuge. Albumin level was measured with serum by biochemical analysis.

Results

Analytical repeatability showed C.V. 6.2% or C.V. 3.6% with whole blood method, and it was sufficient in clinical fields use (Figure 6). Approximate levels were obtained between estimated ALB level and serum ALB level in about 70% of cases. Deviation between estimated ALB level and serum ALB level was caused in high Hb level cases (Figure 7). Automated response system by chatbot scored high in operability on input. It was also excellent that patients could respond at any time, and it was useful for the report when the nurse is absent.

Conclusions

Our POCT showed a good correlation between whole blood and serum. The correction with Hemoglobin level was effective for the albumin level data convert from whole blood to serum. Our system is possible to measure estimated ALB level with a small amount of blood collection, and operability is simple and easy for home health nursing. Our POCT method with liquid reagent is low cost and the chatbot system contributed to labor-saving of hospital nurses. We believe this system can contribute to home health nursing in telemedicine.

Acknowledgements

This work was supported by SCOPE of the Japan Ministry of Internal Affairs and Communications.

References

[1] Jin Y, Yong C, Ren K, Li D, Yuan H, Effects of Post-Surgical Parenteral Nutrition on Patients with Gastric Cancer, *Cell Physiol Biochem*, 49(4), 1320-1328.
[2] Zhang YH, Xie FY, Chen YW, Wang HX, Tian WX, Sun WG, Wu J, Evaluating the Nutritional Status of Oncology Patients and Its Association with Quality of Life, *Biomed Environ Sci*, 31(9), 637-644.
[3] Cho IH, Lee J, Kim J, Kang MS, Paik JK, Ku S, Cho HM, Irudayaraj J, Kim DH, Current Technologies of Electrochemical Immunosensors: Perspective on Signal Amplification, *Sensors*, 18(1), 207.

Address for correspondence

Hiroyuki Nozaka
Hirosaki University Graduate School of Health Sciences
66-1, Honcho, Hirosaki-shi, Aomori 036-8564, Japan
E-mail: hnozaka@hirosaki-u.ac.jp

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190628

From Fax to Blockchain: Sharing Health Information Democratically and Safely

David Parry

Department of Computer Science, Auckland University of Technology, New Zealand

Abstract

Information sharing in healthcare remains an unsolved problem despite a plethora of standards and architectures. Effective information sharing is difficult because of the heterogeneity of health information users and data sources, organisational, ethical and legislative constraints and the very demanding requirements of clinical practice. This paper argues that the key requirement of a viable sharing architecture is to support trust in the system and between stakeholders. It uses the concept of a "democratic " approach where citizens can control and verify the use and sharing of data about them and identify ways that some of the value extracted from the data could be assigned to the patient themselves. The reasons for the survival of obsolescent methods are used to inform the design of a proposed citizen-centric architecture using blockchain technology.

Keywords:
Health Information Exchange, Privacy, Communication

Introduction

Sharing medical information between healthcare providers is a key task in improving healthcare to avoid over testing, and misdiagnosis.. At the same time access to, and value extraction from data by citizens, governments and machine learning agencies is becoming ever-more prominent. Many current data sharing approaches are not universal and have led to the perhaps surprising survial of older technologies such as the fax machine.

Review

Clinical information is often described as residing in "silos"- separate and sealed repositories of data that are generally aligned around a clinical service rather than the needs of the patients. Before any information sharing can take place, the location of possibly relevant information must be identified and systems must have semantic interoperability.

A low-overhead, relatively flexible means of transferring time-sensitive information is attractive. Fax's continue to meet this need in many part of the health system [1].

A major reason why data is not routinely and effectively shared is a lack of trust between providers [2]. Centralized systems such as the *Myhealth record* in Australia [3] have had particular problems in getting both clinical and patient support, with privacy being seen, sometimes perhaps more than is warranted, as a major barrier. Health information exchanges are one approach to addressing this problem, and although often successful when implemented within regions, costs and boundary issues remain a major problem [4]

The middle-out approach [5] involves setting interface standards and allowing users and vendors to build systems that can interface with this. The effect of adding new systems to a standards-based middle-out system is to add value to the network without increasing complexity.

Key technologies and standards

Blockchain is a technology derived from the underlying principles of crypto currency. It is attracting increasing interest as a medium for health data exchange [6]. Issues of file-size have bedevilled early implementations – the blockchain gets larger with every interaction- but this is being addressed with newer approaches [7].

SNOMED CT can be used for storing vocabulary information and *open*EHR architypes can be used to incorporate metadata about clinical information. Semantic web approaches can be used as an universal directory, for both the transfer of data and the location of data stores. Blockchain-based messages using FHIR or other low-overhead interfaces may be added to health systems and devices. Practical approaches to doing this have been described in Peterson et al. [8].

Methods

A system must allow the patient to know who is sharing information and why There must be an accessible and non-repuduiatable audit trail showing use, change and viewing of the data. However the the patient may want to restrict access in some cases and both aspects must be supported. People are increasingly aware of such issues and developing models that go beyond simple privacy and control.

Ownership of financial and other benefits of use of the data must be clear and a process for negotiation around such income must be possible [9]. As much as possible, the data shared should be atomic – only including as much data as needed . Interfaces to the transfer system, directory and audit system must be relatively simple and based around standards that are scalable. Network value increase with the number of connections, and blockchain-based networks appear to be following Metcalfe's law [10] that the value of the network increases exponentially with the number of nodes. The payload of data transfer should be stored efficiently but with fully integrated metadata. Non-traditional forms of medical information – e.g personal electronic health records, data from wearables etc. must be able to be integrated into the system

Results

Architecture of a block chain-based health information exchange

An outline design for such an approach is shown in Figure

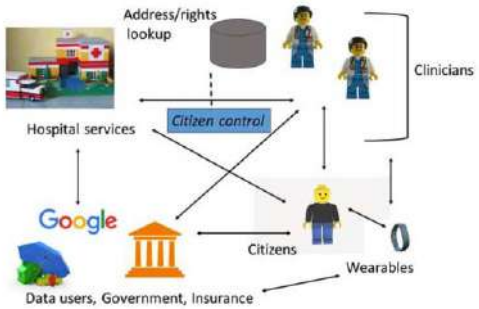


Figure 1: Architecture - each connection is made using block chain-based messaging

Use-case example.

Patient A has high blood pressure and prediabetes. She communicates her wellness data from here wearable devices via the blockchain architecture to her family doctor. The specialist in internal medicine is on a pre-agreed "whitelist" of data receivers so the family doctor adds clinical observations and the result of blood tests to this information as part of a consultation process. A pharmaceutical company has interest in patients with similar characteristics, so the specialist sends a request to the patient, including an estimated fee, to the patient to allow the pharmaceutical company to extract some data from the blockchain-based message. The original patient has complete visibility of both who has viewed and who has changed the data. If there is a delay in the specialist reviewing the data, this can be communicated to the family doctor and a reminder sent.

Discussion

There exists a growing gap between the expectations of citizens that they are to have control of their data – for example the recent general data protection regulation in the EU [11] and current practice. Verifiable control of personal data hsa become very important to large numbers of individuals. However onerous privacy policies, are unpopular and may be used as an excuse to not share data between health providers where appropriate [12]. Granular sharing of data – that is data being shared with certain entities for defined reasons- is popular in principle but difficult to implement in conventional models [13]. Attitudes to data sharing and use have become more citizen-centric – as demonstrated by the EU adoption of GDPR.

Conclusions

Blockchain approaches are at an early stage of development and are certainly suffering from a great deal of "hype" and unrealistic expectations [14]. The democratisation and simplification of health information exchange and value extraction is an important and continuing challenge for health professionals and citizens and other users of health data.

To replace the Fax, the healthcare system and citizens need to be able to understand and trust systems that transfer information safely and fairly apportion the value extracted from that data.

Acknowledgements

Many ideas in this paper have arisen through discussion with Dr. Mohsin Chaudhry, CEO & Co-founder of HealthDex.

References

[1] Vox. (2018). The fax of life. Available: https://www.vox.com/health-care/2017/10/30/16228054/american-medical-system-fax-machines-why

[2] A. Poor, "What are the obstacles to Health Information Systems interoperability in the Auckland region primary health care sector?," Auckland University of Technology, 2016.

[3] Australian Digital Health Agency. (2018). My health record. Available: https://www.myhealthrecord.gov.au/

[4] J. R. Vest and L. D. Gamm, "Health information exchange: persistent challenges and new strategies," Journal of the American Medical Informatics Association, vol. 17, no. 3, pp. 288-294, 2010.

[5] E. Coiera, "Building a national health IT system from the middle out," Journal of the American Medical Informatics Association, vol. 16, no. 3, pp. 271-273, 2009.

[6] T.-T. Kuo, H.-E. Kim, and L. Ohno-Machado, "Blockchain distributed ledger technologies for biomedical and health care applications," Journal of the American Medical Informatics Association, vol. 24, no. 6, pp. 1211-1220, 2017.

[7] X. Liang, S. Shetty, D. Tosh, C. Kamhoua, K. Kwiat, and L. Njilla, "Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability," in Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, 2017, pp. 468-477: IEEE Press.

[8] K. Peterson, R. Deeduvanu, P. Kanjamala, and K. Boles, "A blockchain-based approach to health information exchange networks," in Proc. NIST Workshop Blockchain Healthcare, 2016, vol. 1, pp. 1-10.

[9] A. M. Al-Khouri, "Data ownership: who owns "my data"," Int. J. Manag. Inf. Technol, vol. 2, no. 1, pp. 1-8, 2012.

[10] K. Alabi, "Digital blockchain networks appear to be following Metcalfe's Law," Electronic Commerce Research and Applications, vol. 24, pp. 23-29, 2017.

[11] J. M. M. Rumbold and B. Pierscionek, "The effect of the general data protection regulation on medical research," Journal of medical Internet research, vol. 19, no. 2, 2017.

[12] W. G. Van Panhuis et al., "A systematic review of barriers to data sharing in public health," BMC public health, vol. 14, no. 1, p. 1144, 2014.

[13] K. Caine and R. Hanania, "Patients want granular privacy control over health information in electronic medical records," Journal of the American Medical Informatics Association, vol. 20, no. 1, pp. 7-15, 2012.

[14] P. Michelman, "Seeing Beyond the Blockchain Hype," MIT Sloan Management Review, vol. 58, no. 4, p. 17, 2017.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190629

Application of Process Metrics to Compare Children's Asthma Diagnostic Pathways in 30 EU/EEA Countries

Fabrizio Pecoraro, Daniela Luzi, Oscar Tamburis

National Research Council. Institute for Research on Population and Social Policies, Rome, Italy

Abstract

The paper presents the results of the application of the business process approach to analyse and compare healthcare pathways performed in 30 EU/EEA countries focusing on children's asthma care. The adoption of process metrics allows the identification of different levels of fragmentation across countries resulting from the interactions among primary and secondary healthcare professionals as well as from parents' involvement in the process.

Keywords:

Business process modelling, Bronchial Asthma, Child health

Introduction

Asthma is the most common chronic disease among children [1], representing one of the top 20 conditions for global ranking of disability-adjusted life years (DALY) [2]. Although different international widespread guidelines [3] provide criteria to diagnose, manage and monitor asthma in children and young adults, this disease is still one of the most difficult disease to diagnose and control [4]. A correct, timely and early diagnosis of asthma is necessary to improve its control. This requires speeding up the provision of diagnostic tests, such as spirometry, to confirm asthma diagnosis and its severity, as symptoms alone are not enough to assess the presence and the severity of this respiratory disease [5]. In some countries this is achieved performing spirometry in the primary care setting. Moreover, communication procedures have a crucial role to improve the continuity of care in the management of patients with asthma [6].

This study is part of the MOCHA (Models of Child Health Appraised) project1 that aims to compare and appraise existing national models of primary care for children. Within this project the objective of this paper is to analyse the efficiency of the diagnostic pathway carried out in 30 EU/EEA countries focusing on the communication among professionals and parents as well as on the performance of the spirometry test.

Methods

The communication procedures between primary and secondary care as well as the provision of the spirometry test in each MOCHA country are captured analysing the answers provided by local experts in child health services, i.e. Country Agents (CAs), in an ad hoc questionnaire. In particular the analysis is focused on the following questions, as reported by 22 CAs: 1) is the spirometry usually performed by a primary care (PC) provider?; 2) how does the secondary care (SC) specialist provide the results of the spirometry test to the primary care provider?; 3) which are the usual procedures to refer the child to secondary care?. Moreover, the CAs were asked to report the availability of the medical record in PC setting and whether the record is shared among other professionals. The result of this analysis has been represented using the Unified Modelling Language (UML) diagrams that facilitate the comparison between the different business processes performed in the MOCHA countries in the delivery of child care [7]. In particular, the UML use case diagram has been developed to provide an overall picture of the activities performed as well as of the actors involved in each activity of the care process. This diagram represents the starting point to group countries that have similar procedures, used similar services, and are based on similar caregivers for the provision of care. Responses reported in each use case can help to cluster countries that have a similar behaviour to complete the process. A set of UML activity diagrams have been then developed to summarize the behaviour of each group of countries describing activities performed and messages exchanged by the different actors as well as triggering conditions taking also into account their location and timeline. The process flows adopted in each group of country is analysed to detect differences and similarities among countries in the performance of the different parts of the diagnostic pathway. This analysis requires the identification, quantification and application of business process metrics [8]. Considering the purpose of this paper and giving the high-level description captured from the questionnaire, in this paper we focused on the straightforwardness versus the fragmentation of the care process considering two main aspects: 1) the number of interactions among actors (i.e. professionals, child and parents); 2) the points of the process in which time (waiting for appointments, referral, etc.) has a role in getting the diagnosis and monitoring child health status. Finally, to carry out a performance analysis the activity diagrams were mapped using the NETIMIS process simulation tool. This tool supports the representation of possible bottle necks in a clinical pathway, describing the patients' journey through a series of nodes (e.g. activities) that can be associated with a time and/or a cost for their execution [9].

Results

For space reasons the UML diagrams are not reported in this paper. Starting from the UML use case diagram, countries have been classified and clustered as reported in Table 1. From the cluster analysis three countries have been selected

[1] MOCHA website available at http://www.childhealthservicemodels.eu/

and analyses to capture the level of complexity of the process in terms of its straightforwardness: 1) Spain; 2) Italy and 3) Lithuania. Each activity diagram was subsequently mapped in the NETIMIS simulation tool to represent the differences in the execution of the process, where the probability of the alternatives of the diagnostic pathways decision points is reported in the relevant branches. Figure 1 reports the NETIMIS representation of the activity diagram of Italy.

Table 1– Cluster of countries considering the referral (columns) and communication procedures (rows) and the availability of spirometry in PC (= available; # = partially available). The presence of a shared EHR is reported in brackets (E)*

		PC refers to SC		
		Direct	***Both***	***Parents***
SC communicates to PC	***Direct***	Austria Croatia * Finland (E) * Spain (E) *	Cyprus Germany # Ireland *	France # Iceland * Netherlands #
	Both	Portugal (E)#	Estonia (E)* Italy # Norway (E) *	Malta Greece *
	Parents		Belgium # Latvia #	Poland * Bulgaria Lithuania Romania *

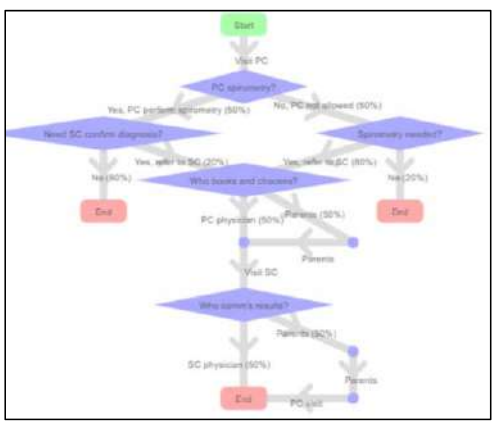


Figure 1– NETIMIS process diagram

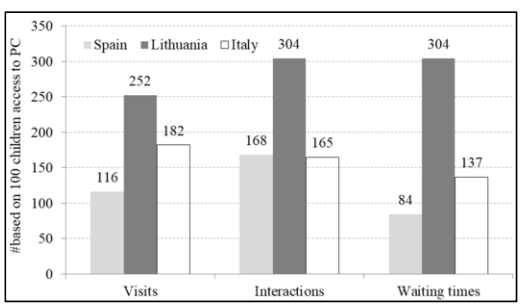


Figure 2– Simulation results reporting the number of visits, interactions and waiting times occurred in each scenario

Discussion and Conclusions

Our approach supports the analysis of the different diagnostic pathways that represent a mix of organisational solutions adopted at local level to address childcare needs [10]. In the present example of asthma, the analysis considered number of encounters, interaction transactions, presence and use of EHR and waiting times across countries. It contributes to outlining the efficiency of organizational models of asthma diagnosis for children on the basis of process metrics highlighting bottlenecks and - through cross-country comparison - possible solutions to improve the process. The need of an early and appropriate diagnosis speaks in favour of spirometry done at PC encounters, especially in cases of exacerbations, when the major focus is posed on the organisational efficacy. Under the business metrics' perspective this resulted in the reduction of the number of visits and interaction between PC and SC physicians and consequently waiting times. In the countries where this is not allowed, the presence of shared EHR and/or close collaboration between primary and secondary care can support both accurate and timely diagnosis (decrease of number of interactions) as well as provide relief of parents' burden in childcare with asthma. The mapping in the NETIMIS simulation tool provides further insights in the analysis of the organisational process, as time and costs of resources can be evaluated. The association of these indicators with children health outcomes could contribute to evaluate the efficacy in the adopted diagnostic process outlining optimal pathways of childcare.

Acknowledgements

This study is part of the Horizon 2020 MOCHA project. Grant Agreement no. 634201.

References

[1] World Health Organization (WHO) – Asthma. Available at www.who.int/respiratory/asthma/en/
[2] I. Asher, and N. Pearce, Global burden of asthma among children, *Int J Tuberc Lung Dis* **18** (2014), 1269-1278.
[3] Global Initiative for Asthma (GINA) report, Global Strategy for Asthma Management and Prevention. Available at: ginasthma.org/download/832/
[4] J.A. Castro-Rodriguez, The Asthma Predictive Index: early diagnosis of asthma, *Curr Opin Allergy Clin Immunol* **11** (2011), 157-161.
[5] C.L. Yang, E. Simons, R.G. Foty, P. Subbarao, T. To, and S.D. Dell, Misdiagnosis of asthma in schoolchildren, *Pediatr Pulmonol* **51** (2016), 1159-1167.
[6] A. Rebuge, and D.R. Ferreira, Business process analysis in healthcare environments: A methodology based on process mining, *Inf Syst* **37** (2012), 99-116.
[7] D. Luzi, F. Pecoraro, and O. Tamburis, Appraising Healthcare Delivery Provision: A Framework to Model Business Processes, *Stud Health Technol Inform* **235** (2017), 511-515.
[8] R. Dijkman, M. Dumas, B. Van Dongen, R. Käärik, and J. Mendling, Similarity of business process models: Metrics and evaluation, *Inf Syst* **36** (2011), 498-516.
[9] O.A. Johnson, P.S. Hall, and C. Hulme, NETIMIS: Dynamic simulation of health economics outcomes using big data, *PharmacoEconomics* **34** (2016), 107-114.
[10] D. Luzi, F. Pecoraro, O. Tamburis, M. O'Shea, P. Larkin, J. Berry, M. Brenner, Modelling collaboration of primary and secondary care for children with complex care needs: long-term ventilation as an example, *Eur J Pediat* (2019), 1-11.

Address for correspondence

Fabrizio Pecoraro, IRPPS-CNR. Via Palestro, 32, 00185 Roma, Italy. Phone: +39(0)6 492724278. E-mail: f.pecoraro@irpps.cnr.it

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190630

The Impact of Hybridisation on the Accuracy of Fluid Balance Documentation: A Retrospective Cross-Sectional Analysis of Intravenous Fluid Order and Administration Documentation Using a Partly-Computerized Medical Record in an Australian Tertiary Teaching Hospital

Samuel Perotti[a], Angus Ritchie[b]

[a] *M.D Student, The University of Sydney School of Medicine, Sydney, Australia*
[b] *Health Informatics Unit, Sydney Local Health District, Sydney, Australia*

Abstract

Inaccurate intravenous (IV) fluid documentation remains a major cause of fluid-related adverse health outcomes in hospitals. In this study, we characterise the integrity and documentation completeness of a hybrid paper-computer documentation system for IV fluids and explore how inconsistencies within it are associated with adverse patient health outcomes. We highlight key areas of weakness specific to IV fluid documentation that need to be addressed in moving to a fully computerised system in the future.

Keywords:

Infusions, Intravenous; Medical Order Entry Systems; Electronic Health Records

Introduction

IV fluids are the most commonly prescribed intervention in Australian hospitals and are responsible for more adverse reactions than any other single drug alone. It is estimated that up to 20% of patients experience some complication from their use [1]. A major contributor to this issue is incomplete and inaccurate documentation, with studies finding that 20% of fluid orders are incorrectly documented [2,3]. To tackle this, documentation is shifting towards electronic platforms. In 2014, Concord Repatriation General Hospital (CRGH) in Sydney, Australia implemented a "hybrid" model whereby IV fluid orders and administration were documented on paper, and the fluid balance charts were documented in the hospital's electronic medical record (eMR) database. To date, no study has explored the integrity of such a system on IV fluid documentation completeness and how it relates to patient health outcomes. This study sought to address this by: assessing documentation completeness of paper IV fluid orders at CRGH, exploring how well these match their corresponding electronic eMR entries in the hybrid documentation system, and establishing whether inconsistencies between the paper and electronic platforms are associated with poorer health outcomes.

Methods

This study consisted of a retrospective cross-sectional analysis of nearly all inpatients at CRGH between October and December 2016 who received more than 100 ml of crystalloid IV fluids during their admission (excluding paediatric, intensive care and emergency department admissions). Eligible patients were extracted from the hospital's eMR database, and a representative sample was randomly selected for inclusion in the study. For each patient, paper IV fluid orders were examined for completeness of fluid order and administration details (as listed in Table 1). Each paper documentation was then compared with its corresponding eMR fluid balance chart entry. To be assessed as being correct and consistent, the paper and eMR entries needed to record the same fluid type, volume, and infusion duration. The incidence of calls for Clinical Review (CR; i.e., an escalation system for deteriorating patients [4]) made for each patient during their hospital stay was used as a marker of poor patient health outcomes (Figure 1).

Results

A total of 1679 patients met eligibility criteria, of which a random sample of 91 was included in the analysis. A total of 360 IV fluid prescriptions were made for these 91 patients and were assessed for completeness of order and administration details (Table 1). Our findings demonstrate high completeness of IV fluid identifiers and start time, but low completeness for documentation of delivery site (35.6%), time ceased (24.7%) and total volume delivered (18.9%). When compared with their corresponding eMR documentation (Figure 2), it was found that 35.8% of orders were correctly documented, 38.1% were incompletely documented, and 26.1% had no electronic entry at all. The most common inconsistency between paper charts and fluid balance records was showing a different volume of IV fluids delivered on the eMR compared to the paper order (Figure 2). Twenty-two patients received at least one CR call during their admission, which was associated with twice as many documentation inconsistencies between the paper and electronic records (median of 3 vs 1.5 per patient; P=.0007; see Figure 3).

Discussion and conclusions

Paper documentation completeness at CRGH was generally higher than that the 60% completeness reported in the literature [5,6]. However, this study found a significant issue relating to inacurate documentation of total volumes delivered, which is essential information for guiding clinical decision-making and likely to contribute to IV fluid-related adverse events. This is the first study to assess the documentation consistency in a hydrid paper-electronic documentation workflow for IV fluid management. This analysis identified a

high rate of documentation inconsistencies. Possible contributors include poor staff adherence to recommended workflows, difficulties in managing the hydrid interface, time pressures and interruptions in a busy clinical environment, and failure to document changes in fluid orders contemporaneously. Our results also support a strong association between documentation inconsistencies and poorer patient outcomes, as measured by CR call incidence. Though it remains difficult to contextualize these results due to the lack of similar studies in the field, it is clear that hospitals must carefully evaluate the implementation of hybrid documentation platforms as they transition to fully computerized systems in the future.

Acknowledgements

I would like to thank the University of Sydney for its support.

References

[1] K.G. Callum, The 1999 Report of the National Confidential Enquiry into Perioperative Deaths, National Confidential Enquiry into Patient Outcome and Death, 1999.

[2] L. Roughead, S. Semple, and E. Rosenfeld, Literature Review: Medication Safety in Australia, *Australian Commission on Safety and Quality in Health Care*. (2013).

[3] M.M. Walton, R. Harrison, P. Kelly, J. Smith-Merry, E. Manias, C. Jorm, and R. Iedema, Patients' Reports of Adverse Events: A Data Linkage Study of Australian Adults Aged 45 Years and Over, *BMJ Qual Saf*. **26** (2017), 743–750.

[4] NSW Health, Standard Calling Criteria, (2016). www.cec.health.nsw.gov.au/adultpatientsafetybetweenthef-lags/standardcallingcriteria (accessed July 25, 2017).

[5] H.I. Asfour, Fluid Balance Monitoring Accuracy in Intensive Care Units, *J Nurs Health Sci*. **5** (2016), 53–62.

[6] K.D. Evans, S.W. Benham, and C.S. Garrard, A Comparison of Handwritten and Computer-Assisted Prescriptions in an Intensive Care Unit, *Critical Care*. **2** (1998), 73.

Address for Correspondance

Samuel Perotti, email sam-ppp@hotmail.com

Table 1 – Completeness of 360 paper IV fluid orders for 91 patients

Comparator	Percentage of IV fluid prescriptions correctly depicting the comparator
Type of fluid	98.9 %
Volume to be delivered	98.6 %
Flow rate	99.2 %
Site of delivery	35.6 %
Medical Officer signature	96.1 %
Start date	91.1 %
Start time	90.6 %
Time ceased	24.7 %
Total volume delivered	18.9 %
Signed by 2 nursing staff	91.7 %

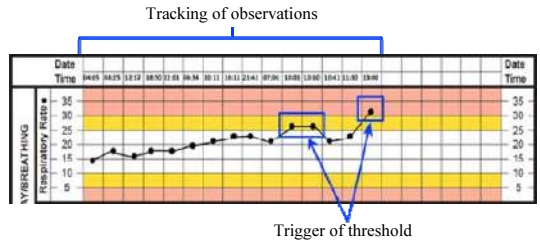


Figure 1 – Clinical review call criteria on a between the flags patient observation chart. Yellow zone: indication for clinical review by a medical officer. Vital signs incorporated include respiratory rate, heart rate, temperature and peripheral capillary oxygen saturations (SPO2) [4].

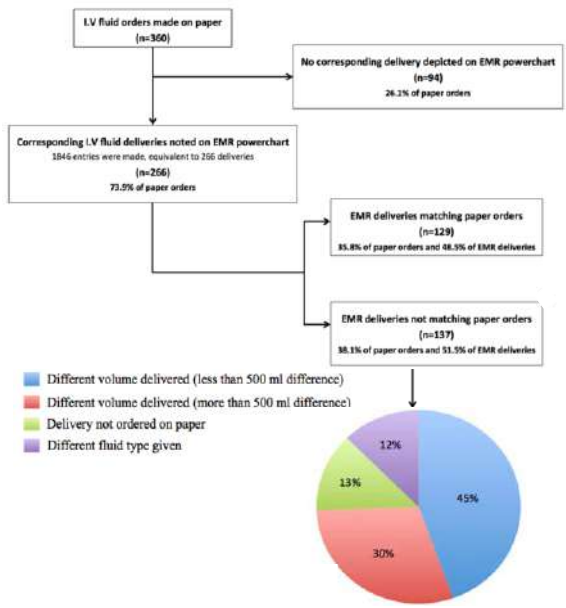


Figure 2 – Distribution of inconsistencies between paper IV fluid orders and their corresponding eMR

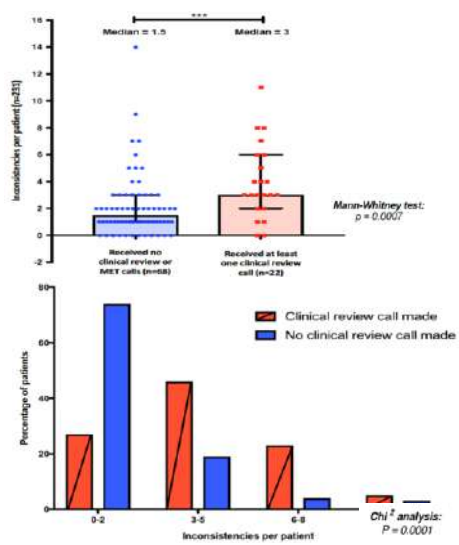


Figure 3 – Association between documentation inconsistencies and clinical review calls. Top: Inconsistencies made per patient stratified by clinical review call status. Bottom: Distribution of patients (%) based on the number of inconsistencies made.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190631

A Study Design to Model Clinical Management Process for the Health and Wellbeing of Patients with Alcohol Use Disorders

Siyu Qian, Ting Song, Ping Yu

Centre for IT Enabled Transformation, School of Computing and Information Technology, Faculty of Engineering and Information Sciences, University of Wollongong, NSW 2522, Australia

Abstract

Public hospitals are often plagued by patients with unexpected acute alcohol withdrawals because of failure of diagnosis at the time of their presentation. To address these problems, this study will develop models to predict the risk of alcohol withdrawal and recommend corresponding management processes to optimize patient safety using health data analytics.

Keywords:

Alcohol-induced disorders, public hospitals, patient safety

Introduction

Public healthcare systems face the challenge of providing safe and cost effective services to patients with alcohol use disorders. Internationally, alcohol is the most commonly abused substance and causes preventable morbidity, mortality and related behavioural risks [1, 2]. In North America, Europe and Asia there is high prevalence of alcohol use disorders in patients presenting to hospitals [3-5]. Likewise, in Australia, 27% of the patients presenting to hospitals have alcohol use disorders and require certain level of intervention [6]. It is estimated that the annual hospital cost for patients with alcohol use disorders is $19.8 billlion in the US [7]. These patients are often unidentified due to other seemingly unrelated presentation problems such as abdominal pain [8]. Consequently, public hospitals are often plagued by unexpected acute alcohol withdrawal developed during management processes [9]. Our preliminary study found that 73% of the patients with alcohol use disorders were not identified at the time of presentation and the clinical management for these patients was inconsistent and unsafe.

To address these challenges, this study aims to investigate the clinical identification and management processes for patients with alcohol use disorders in Australian hospitals. Multiple research methods, including clinical guidelines review, expert panel discussions with clinicians, machine learning and data mining of the routinely collected inpatient data in the electronic medical records (EMR), will be corroborated to maximize output. Specific research questions (RQs) are:

RQ1. How can we effectively identify patients with alcohol use disorders at the time of presentation?

RQ2. What are the recommended clinical management processes for these patients according to the clinical guidelines?

RQ3. What are the actual clinical management processes for these patients?

RQ4. Do these actual processes comply with the recommendation of the clinical guidelines?

RQ5. What are the optimal processes for safe and cost effective management for these patients?

Methods

This study is designed with two research components: 1. Development of a prediction model to identify patients with alcohol use disorders at the time of presentation (addressing RQ 1); 2. Development of optimal process models for clinical management for these patients (addressing RQs 2-5). Figure 1 illustrates the research components, tasks, methods and theoretical frameworks that guide this research.

The following research theory and frameworks will be adapted to guide this research:

Activity theory [10] will provide an overarching framework to guide our identification of clinical actors and actions from clinical guidelines review and panel discussions with clinicians. This theory provides a systematic approach to inspect a clinical activity from six aspects: who (e.g., clinicians) conducts the activity for what purpose (e.g., to manage alcohol withdrawal symptoms), using what tools (e.g., alcohol withdrawal scale), following what rules (e.g., clinical guidelines), supported by which community (e.g., healthcare team) and the roles of community members (e.g., clinician, nurse) [10]. The usefulness of this theory to our project is drawn from its successful application to study activities of patient discharge and maternal data collection [11-13]. We have successfully applied this theory in a previous study for medication management processes in a residential aged care facility [14].

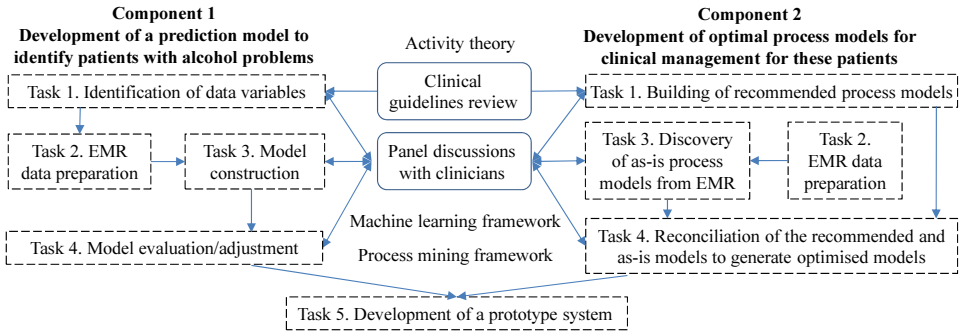


Figure 1– Research components, tasks, methods and theoretical frameworks.

Machine learning technique will be applied to deliver Research Components 1 and 2 [11]. Four steps will be followed to deliver Research Component 1: identification of data variables as model predictors, EMR data preparation, model construction, and model evaluation and adjustment. Whereas a process mining technique [15] will be applied to Research Component 2. The technique will be applied to discover as-is process models from the EMR. The key steps include — EMR data preparation, discovery of as-is process models, and model evaluation.

Conclusion

This paper presents a study design to model the clinical management process for the health and wellbeing of patients with alcohol problems in hospital environment. It addresses the practical challenges of identification and management of patients with alcohol problems in hospitals.

References

[1] S.S. Lim, T. Vos, A.D. Flaxman, G. Danaei, K. Shibuya, H. Adair-Rohani, M.A. AlMazroa, M. Amann, H.R. Anderson, and K.G. Andrews, A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010, *Lancet* **380** (2012), 2224-2260.

[2] N.B. Johnson, L.D. Hayes, K. Brown, E.C. Hoo, and K.A. Ethier, CDC National Health Report: leading causes of morbidity and mortality and associated behavioral risk and protective factors—United States, 2005–2013, (2014).

[3] C.J. Cherpitel and Y. Ye, Drug use and problem drinking associated with primary care and emergency room utilization in the US general population: data from the 2005 national alcohol survey, *Drug and Alcohol Dependence* **97** (2008), 226-230.

[4] S.P. Chou, S. Chun, S. Smith, J. Ruan, T.-K. Li, and B.F. Grant, Episodic heavy drinking, problem drinking and injuries–results of the WHO/NIAAA collaborative emergency room study in South Korea, *Alcohol* **46** (2012), 407-413.

[5] Q. Youmans, R.C. Merchant, J.L. Baird, T.J. Langan IV, and T. Nirenberg, Prevalence of alcohol, tobacco and drug misuse among Rhode Island hospital emergency department patients, *Medicine and Health, Rhode Island* **93** (2010), 44.

[6] K. Butler, R. Reeve, S. Arora, R. Viney, S. Goodall, K. Gool, and L. Burns, The hidden costs of drug and alcohol use in hospital emergency departments, *Drug and Alcohol Review* **35** (2016), 359-366.

[7] J. Gryczynski, R.P. Schwartz, K.E. O'Grady, L. Restivo, S.G. Mitchell, J.H. Jaffe, Understanding patterns of high-cost health care use across different substance user groups. *Health Affairs (Millwood)* **35** (2016), 12-19.

[8] D. Indig, J. Copeland, K.M. Conigrave, and I. Rotenko, Attitudes and beliefs of emergency department staff regarding alcohol-related presentations, *International Emergency Nursing* **17** (2009), 23-30.

[9] K. Butler, R. Reeve, R. Viney, and L. Burns, Estimating prevalence of drug and alcohol presentations to hospital emergency departments in NSW, Australia: impact of hospital consultation liaison services, *Public Health Res Pract* **26** (2016), e2641642-e2641642.

[10] L.C. Yamagata-Lynch, Understanding cultural historical activity theory, in: *Activity systems analysis methods*, Springer, 2010, pp. 13-26.

[11] F. Lin, W. Chaboyer, M. Wallis, and A. Miller, Factors contributing to the process of intensive care patient discharge: an ethnographic study informed by activity theory, *International Journal of Nursing Studies* **50** (2013), 1054-1066.

[12] P. Sadeghi, P. Andreev, M. Benyoucef, K. Momtahan, and C. Kuziemsky, Activity Theory Driven System Analysis of Complex Healthcare Processes, in, 2014, pp. 1-14.

[13] C. Ngoma and F.T. Igira, Empowering Community Health Workers to Collect and Record Maternal and Child Health Data by Resolving Contradictions, *Journal of Health Informatics in Africa* **3** (2015), 1-18.

[14] S. Qian, P. Yu, D. Hailey, N. Wang, and A. Bhattacherjee, Medication administration process in a residential aged care home: An observational study, *Journal of Nursing Management* **0** (2018).

[15] W.-S. Yang and S.-Y. Hwang, A process-mining framework for the detection of healthcare fraud and abuse, *Expert Systems with Applications* **31** (2006), 56-68.

Address for correspondence

Ping Yu

Associate Professor
Director, Centre for IT-enabled Transformation
School of Computing and Information Technology
Faculty of Engineering and Information Sciences
University of Wollongong, Wollongong, NSW 2522
Building 3 Room 204
Email: ping@uow.edu.au
T + 61 2 4221 5412 | F + 61 2 4221 4045

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190632

HomeCoRe: Bringing Cognitive Rehabilitation at Home

Silvana Quaglini[a], Silvia Panzarasa[a], Anna Alloni[b], Michele Sacchi[a], Elena Sinforiani[c], Sara Bottiroli[c], Sara Bernini[c]

[a] Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy;
[b]Biomeris srl, Pavia, Italy; [c] IRCCS C. Mondino, Pavia, Italy

Abstract

CoRe is a system for cognitive rehabilitation that has been successfully used for several years in hospital settings. Leveraging on the positive survey results from the potential final users (patients and their home caregivers), we developed HomeCoRe. This new version of the system will allow discharged patients to continue the rehabilitation treatment at home.

Keywords:

Home care services; neurological rehabilitation; therapy, computer-assisted therapy

Introduction

CoRe is a program for computer-supported cognitive rehabilitation. Since the development of first CoRe version over ten years ago, we have enhanced the through several iterations. It has been used in different clinical settings for face-to-face rehabilitation sessions between patients and their therapists [1]. A recent clinical study on hospitalized patients affected by Parkinson's disease [2] showed that patients undergoing one month of rehabilitation through CoRe outperformed control patients in the standard battery of neuropsychological tests administered at the clinical cognitive assessment before discharge. However, this advantage was less evident at the six months follow-up after discharge. This observation motivated the development of HomeCoRe, a version of the system that allowed patients to continue rehabilitation at home, after becoming familiar with the program during the hospital stay. We propose that a longer rehabilitation period will maintain the benefits achieved in the first month.

Methods

For the first stage, we interviewed and surveyed patients to investigate their willingness to continue rehabilitation at home. Home caregivers were also study participants, due to their role in both supporting and motivating patients.

The technical stage had the HomeCoRe architecture embedded into two main components, therapist-side and patient-side, and communication channels between them. Consideration was given to addressing the lack of availability of internet connection at home, and particularly for the elderly. The two components communicated through XML files that, from case to case, were either automatically sent using the internet (online mode) or manually shared (offline mode) through flash memories like usb drives, at the control visits.

From the functional point of view, to increase the patient motivation for the rehabilitation treatment, we exploited a unique characteristic of CoRe that was capable of generating "ever new" exercises, due to a stimuli ontology composed by thousands of images, words and sounds, and relationships among them. Moreover, HomeCoRe was able to generate patient-tailored exercises, by adding to the ontology a "private", patient-specific set of stimuli representative of his daily life (pictures from his house, his relatives, sentences related to his job, hobbies, etc.). To monitor a patient's progress, the system calculated an "overall weighted score (OWS)", taking into account the correctness of the answers, the execution time, and the difficulty of the exercises. A temporal abstraction algorithm informed the therapists about the OWS trend. All components were developed in Java.

Results

Out of 21 patient participants, 15 (71.5%) had definite positive responses, 2 (9.5%) had definite negative responses, and 4 (19%), although appreciating HomeCoRe, expressed a refusal to have a further home commitment. Even better results were obtained from caregivers. We were able to interview 16 of them, and 15 (94%) answered positively. The study results encouraged us to start the HomeCoRe project.

Discussion

The therapist's view, migrating from CoRe to HomeCoRe implied to add some functionalities to the therapist's interface. Figure 1 details the top (a) and lower (b) left panel of therapist's interface for creating/selecting patients, managing patients' data and therapeutic plans, and the system dashboard. Figure 2 illustrates the therapist's interface for visualizing the details of the active and expired therapeutic plans for the selected patient.

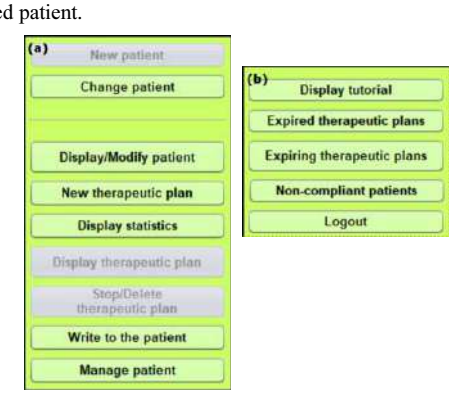


Fig. 1 – Details of the Therapist's Interface with all the System Functionalities

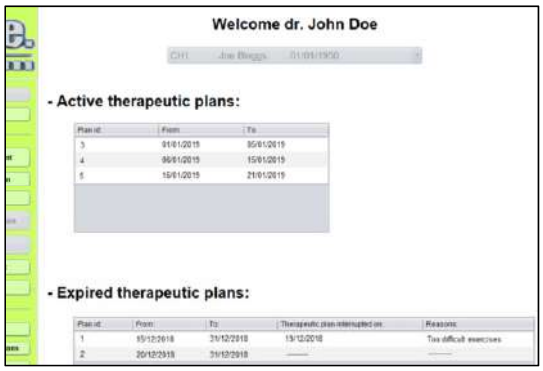


Figure 2 – The Therapist's Interface for Viewing the Active and Expired Therapeutic Patient Plans

The psychologist was able to prepare an entire rehabilitation plan, personalized to a specific patient (see Figure 3). As mentioned, the plan was represented through a XML file that, according to the clinicians' requirements, described (i) the type of exercises for each rehab session, (ii) the duration of the plan in days and daily frequency of execution, (iii) the difficulty level, and (iv) a set of parameters defining some other details, such as the automatic modification of the difficulty level according to the patient's performance, the feedback for the patient, the stimuli exposure time, etc.

The therapist was also able to set a *goal*, i.e. a score that the patient should achieve at the end of the plan execution. The idea was that, if a patient achieved OWS_{disch} at discharge, the goal could be OWS_{final}= OWS_{disch} +y at the end of the plan, with y>=0, decided according to the estimated potentialities of the individual patient. Similarly, goals could be defined for the neuropsychological tests scores (even if they were not continuously measured as the OWS, since they required a face-to-face visit).

Fig. 3 – The Therapist's Interface for Setting the Requirements for the Exercise Plan

At any time, the therapist could check a patient's progress, and the system showed him the percentage of goals achieved. Accordingly, the therapist could modify the rehabilitation plan or update (upgrading or downgrading) the goals.

On the patient's/caregiver's side, the interface was very simple to use (Figure 4). Patients were able to view the exercises of the day, read the message from their therapist and exit. The caregiver could access his private area with a username and a password. The caregiver had a restricted area, where he could upload the XML file and download the results (offline mode) and see all the messages sent by the therapist.

Figure 4 – Home Page of the Patient/Caregiver

All the exercises were performed through a touchscreen (Figure 5). The interface always showed information about the completion percentage of the tasks.

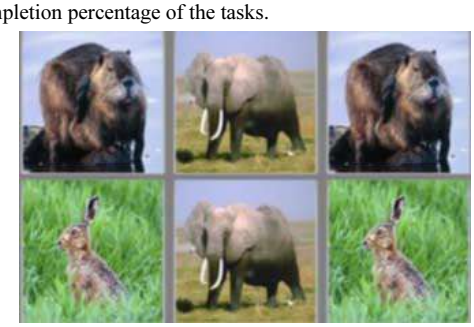


Figure 5 –Memory Exercise Example in the Therapeutic Plan

Conclusion

HomeCoRe is a tool for personalized cognitive rehabilitation that allows patients to continue a rehabilitation plan, which is often abandoned after discharge, due to scarcity of healthcare personnel for homecare. The limitation of the presented work is the lack of a clinical study to show the users' acceptance and efficacy of our proposal. We plan to start a pilot study to refine the system functionalities, in particular communication among patients, caregivers and therapists, and training modalities. Possible next steps include installation of in-hospital workstations to allow patients and caregivers to simulate future home sessions and assessment of the organizational changes that occur with the introduction of technological innovations [3].

References

[1] A. Alloni, E. Sinforiani, C. Zucchella, G. Sandrini, S. Bernini, B. Cattani, D.T. Pardell, S. Quaglini, C. Pistarini. Computer-based cognitive rehabilitation: The CoRE system. Disabil Rehabil, **39** (2017), 407-417

[2] A. Alloni, S. Quaglini, S. Panzarasa, E. Sinforiani, S. Bernini. Evaluation of an ontology-based system for computerized cognitive rehabilitation, *Int J Med Inform,* **115** (2018), 64-72

[3] A. Marino A., O. Tamburis. Technological innovation and organisational context information management in the high specialized hospital: the case of South Italy, *IJHCTM*, **6**(3) (2015), 276–287

Correspondence

Silvana Quaglini, PhD, e-mail address: silvana.quaglini@unipv.it

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190633

Polesat-Web-2018: A Simulation IT Tool with Immediate Prospective and Strategic Views of Hospital Spatial Planning

Anne Quesnel-Barbet[a,f], Julien Soula[b], Erik-André Sauleau[c,e], Pierre Parrend[d,e], Pierre Bazile[f], François Dufossez[g], Arnaud Hansske[h,i]

[a] *Regional Home of Clinical Research (MRRC) - Public Health Pole, the University Hospital Center (CHU), Lille, Nord, France*
[b] *CERIM-Medical Informatics Lab - EA 2694: Public Health, Medical Faculty of Lille University, Lille, Nord, France*
[c] *Public Health Department - Clinical Research Methods, the University Hospital Center (CHU), Strasbourg, Bas-Rhin, France.*
[d] *High School of ECAM Strasbourg-Europe, Schiltigheim, Bas-Rhin, France*
[e] *ICube-Bioinformatics & System complex – CNRS – CSDC-UNESCO-Unitwin, Strasbourg University, Strasbourg, Bas-Rhin, France.*
[f] *Geographic Information Association (AFIGEO), 73 avenue de Paris, Saint-Mandé, Val de Marne, France.*
[g] *Medical Information Department (DIM), Béthune Hospital Center (CH), Béthune, Pas-de-Calais, France.*
[h] *Medical Information & Hospital Information System Departments (DIM-SIH), Catholic Hospitals (GHICL), Lille, Nord, France.*
[i] *KASHMIR-DataReuse Lab, Catholic Lille University (UCL), Lille, Nord, France.*

Abstract

Medical geo-informatics allows the Health world to address major challenges thanks to attractive concepts, methods and user-friendly IT. PoleSat-web-2018 presents a decision support system – a modelling "variable geometry" IT tool for simulation of hospital spatial planning. The outputs enable quasi-instantaneous analytic visualization at several geographic levels. PoleSat-web-2018 provides prospective views of hospital catchments (by grouping, closing) and proves to be relevant for the French planners of the Ministry of Health.

Keywords:

Health Services Accessibility, Hospital Planning, Expert Systems

Introduction

Geo-informatics has become an important science and faces deep challenges to assist the medical policy makers through attractive concepts and methods. Among them, 1-spatial accessibility is paramount within the framework of a gravity model in order to provide the most valid measures [1-3]. 2- Decision Support Systems (DSS) as management tools through geospatial applications [3]. The main objective is to deliver a user-friendly Information Technology (IT) simulation environment that enables studies and test assumptions [4] and provides immediate strategic spatial views. PoleSat-web-2018 is based on a Graphical User Interface (GUI) with an embedded optimized geometric algorithm (an article was submitted to JOMS on March 2019) for simulation modelling hospital spatial planning. The algorithm, related to a refined gravity model [2; 5], has evolved in order to be completely automated inside a spatial DSS and is intended for the non-geomatician, i.e. non-Geographic Information System (GIS) specialists. An online demonstration (alpha version) is currently available for both specialties: traumatology – total hip prosthesis (PTH)) and Clinical Onco-Hematology (COH) on four geographical scales/levels including 1- former regions, 2- new regions, 3- Territory Hospital Grouping (GHT), 4- France.

This article presents the modelling results (by specialty) through the Advanced 1st Order Simulation **(A1OS)** and the Advanced 2nd Order Simulation **(A2OS)** examples, which are based firstly on "a grouping of hospitals" and secondly on "a suppression of hospitals followed by a conditional-based grouping of the remaining hospitals".

Materials and methods

PoleSat-web-2018 uses sources from the French Diagnosis-Related Group "DRG"-based information system (PMSI). This database (DB) is derived from aggregated data retrievals and represents private-public sectors in 2014. The specialties concern "rare and onerous diseases – OH", and "common and less expensive illnesses" – traumatology-PTH. Patient identification is rendered impossible. More details on the geographic DBs can be found inside the GUI aid.

Software access, architecture and algorithm process

An online PoleSat-web-2018 access is given by https://thymine.univ-lille2.fr/polesat2018/ with a log-in (ID: demo3/PW: polesat4). As of now, no scalability management has been made. Clients and server machines communicate in a request-response messaging pattern. All human-computer exchanges are made in a secured environment. The client can check the progress of the process. The calculation takes 2-3 min (per region) and 45 min (for France coverage). The algorithm workflow can be depicted through the succession of four main steps. 1- Keep all or delete hospital services (P5, default setting: keep all services = 0). 2- Group the hospitals with an expert distance (P8, default setting: 15,000 m). 3- Create referring hospital poles after relocating hospitals in "hospital and hospital-basket poles". 4- Deliver outputs detailing hospital catchment areas (CAs) in maps (.pdf, .shp) and spreadsheets (.csv).

Results

The modelling results derive first in "the A1OS" by grouping hospitals for the region: "Hauts-de-France". Next, the A2OS represents a removal of hospitals followed by grouping of the remaining hospitals. The GUI is intended either for a basic user (with a video guide and a pre-recorded default setup) or for a more advanced user (knowledge of the planner and setting adjustment).

The A1OS: grouping hospitals – COH DB

The A1OS represents (Figure 1) with 53 hospital services, a grouping of 12 services and a final number of 41 referring poles.

Figure 1– The A1OS: Grouping Hospitals (a CA map with black administrative census (IRIS) & white PMSI limits). The main settings: D5: COH, P1: New regions, P2: 32, ***P5: 0****, P8: 15,000 m, checked boxes of CHUs, label, ID, IRIS & PMSI*

The A2OS: removal & grouping hospitals – COH DB

The A2OS (Figure 2), with 53 hospital services, is based on A: removal of 26 hospitals (mass < 2) followed by B: grouping of the 3 remaining hospitals (inside a radius <= 15,000 m), into simple or basket poles, which are 24 final referring poles.

Figure 2– The A2OS by Closing & Grouping (a CA map with black IRIS & white PMSI limits). The main settings: D5: COH, P1: New regions, P2: 32, ***P5: 2****, P8: 15,000 m, checked boxes of CHUs, label, ID, IRIS & PMSI*

Discussion

The results provide both a quasi-instantaneous visualization thanks to ready-to-use maps and spreadsheets, and other format files suitable for geomatic use. To note that the related symbology, used in hospital CA maps, is distinguished by random and contrasting automatic colors. Also, a non-administrative numbered circle is assigned for each referring pole.

The advanced simulations: A1OS & A2OS

After launching the A1OS, maps and spreadsheets are available for verifications and adjusting the settings (Figure 1) and before launching A2OS if hospital closing is desired. This time, there are removals of services followed by a grouping of remaining hospital poles (Figure 2). However, the hospital closing insight (even by specialty) may be terrifying (for citizens), but that should not lead planners to give up this idea. It is why a tailored-made version could be focused to specifically start on hospital supply view and grouping modelling. Only after deeper simulation analyses, a new option proposal of service closing could be adopted by planners & public because of the objective, detailed and clearly analyzed proofs.

Besides, catchment area differences for specific border hospital pole could appear if we use a former or a new region. In consequence, the geographical analysis level choice must take place according to the initial problem of the planner and with full background knowledge.

Obviously, future versions should allow the processing of all PMSI and other health DBs over long periods. A tailored-made version has just begun for the French Health Ministry. The future updates are not yet specified since we are in a preliminary test period.

Conclusions

PoleSat-web-2018 is an effective and easy-to-use web-based GUI (alpha version), with an embedded optimized and automatized algorithm for prospective hospital spatial planning. PoleSat offers practical simulations for studying different scenario effects. It aims at helping planners in predicting the hospital provision reorganization. It proves to be a modelling "variable geometry" IT tool and appears to be relevant for planners with a productive and impartial foundation for strategic decisions. It can therefore be generalized.

Acknowledgements

The authors wish to thank the Strasbourg and Lille Universities, the Lille University Hospitals (CHU and GHICL) and the Altense Consulting SA.

References

[1] M.F. Guagliardo, Spatial accessibility of primary care: concepts, methods and challenges, *International Journal of Health Geographics* **3** (2004), 1-13.

[2] A. Quesnel-Barbet, F. Dufossez, M. Souris, D. Delerue, and A. Hansske, Geomatics in health and applications in the north of France, *Techniques de l'Ingénieur* ***MED4000*** **V1** (2016), 48.

[3] A.K. Lyseen, C. Nohr, E.M. Sorensen, O. Gudes, E.M. Geraghty, N.T. Shaw, and C. Bivona-Tellez, A review and framework for categorizing current research and development in health related geographical information systems (GIS) studies, *Yearbook of Medical Informatics* **9** (2014), 110-124.

[4] M. Debacker, F. Van Utterbeeck, C. Ullrich, E. Dhondt, and I. Hubloue, SIMEDIS: a discrete-event simulation model for testing responses to mass casualty incidents, *Journal of Medical Systems* **40** (2016), 1-15.

[5] A. Quesnel-Barbet, Geography of spatial utilization of the health services in the Nord - Pas-de-Calais region: a Newtonian modeling of hospital catchment areas, Th. doct. : Quantitative geography, USTL Lille I, 2002.

Address for correspondence

Anne Quesnel: aquesnel@univ-lille.fr Tel + 33 (0)658066808

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190634

Evaluation of the Belgian Guidelines Website *EBMPracticeNet* in French General Practice

Rambaud C[a,b], Fauquert B[c], Charbonnel P[b], Falcoff P[b], Letrilliart L[a,d]

[a] *Collège universitaire de médecine générale, Université Claude Bernard Lyon 1, France*
[b] *Collège de la médecine générale, Neuilly-sur-Seine, France*
[c] *Belgian Centre for Evidence Based Medicine (CEBAM), Leuven, Belgium*
[d] *HESPER EA 7425, Lyon, France*

Abstract

EBMPracticeNet is a Belgian website of guidelines translated and adapted from the Finnish EBM guidelines. During the experimentation of the EBMPracticeNet website in France, physicians globally got an accurate answer and found the information provided by the website reliable and useful for practice. They perceived its ergonomics as good and wished to continue using it. Improvements should focus on the indexation and adaptation of the guidelines, and on physicians' training.

Keywords:

Point-of-care system, practice guidelines, general practice, France

Introduction

Web-based medical compendia are point-of-care tools specifically designed to deliver pre-digested, rapidly accessible, comprehensive, periodically updated and evidence-based information to clinicians, which could be a solution to facilitate the use of clinical practice guidelines [1]. More of 20 systems have been developed worldwide, including BMJ Practice in the UK, Dynamed and Uptodate in the USA, which have been assessed in the top three according to their volume, editorial quality and methodology [2].

EBM Guidelines is a collection of 1000 clinical practice guidelines for primary care, produced by the Finnish medical society. It has been translated into French, adapted to the Belgium context and made available to all Belgian physicians on the website *EBMPracticeNet* [3].

As no such web-based point of care information summary was available in France, we have experimented this Belgian website. This aim of this study was to assess physician's participation and searches' effectiveness and users' global evaluation of *EBMPracticeNet* in French general practice.

Methods

A sample of French physicians experimented the Belgian website from March to September 2017. We contacted the approximately 2000 members of the French College of General practice and 50 post-graduate trainees from two medical schools. We collected data from three sources: 1) The website logbook where all searches were recorded with the following fields: user ID, timestamp(s), query(ies), guideline(s) opened; 2) a search-specific assessment questionnaire, evaluating searches effectiveness, periodically activated online; 3) a global assessment questionnaire, evaluating the website ergonomics (System Usability Scale, SUS) and content, and the participants' global satisfaction, sent by mail at midtime and at the end of the study. The frequency of use by each participant was defined as the number of clicks performed into the website, either for entering a query or for opening a guideline.

Results

Among 417 physicians registered, 262 (62.8%) performed at least one search on the website. The participants clicked into the website on average 5.9 folds per month (2.9 queries entered, 3.0 guidelines opened). The number of monthly users decreased during the study period from 168 to 53. In the meantime, the mean number of clicks per monthly user decreased from 17.0 to 11.3. We collected 194 search-specific questionnaires from 85 (32.4%) physicians. Most of them had performed their search out of the consultation (61.3%) and got an accurate answer (74.2%) (Table 1).

We collected a global assessment from 158 (60.3%) physicians at midtime and 103 (39.3%) physicians at the end of the study. At the end of the study, most physicians found the information provided by the website rather reliable (92.2%) and useful for practice (78.6%), were rather satisfied (63.1%) and wished to continue using the website (95.1%) (Table 2).

Table 1- Search-specific assessment (n=194)

	n (%)
Search time	
During the consultation	75 (38.7%)
Out of the consultation	119 (61.3%)
Easy use of the website	
Yes	184 (94.2%)
No	10 (5.2%)
Quick search	
Yes	183 (94.3%)
No	11 (5.7%)
Search answer	
Accurate	144 (74.2%)
Inaccurate	50 (25.8%)
Guideline(s) but no answer	35 (70.0%)
No guideline	8 (16.0%)
Uncertain	7 (14.0%)
Clear information	
Yes	162 (83.5%)
No	32 (16.5%)
Information useful for patient care	
Yes	119 (61.3%)
No	75 (38.7%)

Table 2 - Final global assessment (n=103)

		At the end n (%)	
Information reliability	Rather reliable 95 (92.2%)	Medium 5 (4.9%)	Rather unreliable 3 (2.9%)
Information adaptation to the French context	Rather adapted 82 (79.6%)	Medium 17 (16.5%)	Rather not adapted 4 (3.9%)
Website usefulness for practice	Rather useful 81 (78.6%)	Medium 14 (13.6%)	Rather useless 8 (7.8%)
Global satisfaction	Rather satisfied 65 (63.1%)	Medium 31 (30.1%)	Rather not satisfied 7 (6.8%)

The mean SUS score was 70.3 at the end of the study. No change was observed for these variables between midtime and the end of the study. On the 183 comments collected, the main barriers reported by respondents concerned the time and the effort required to find an accurate answer and the uneven relevance of the information retrieved.

Conclusion

Before implementing EBMPracticeNet at large scale in France, improvements should focus on the indexation of the guidelines and their adaptation to the French context, and on physicians' training in searching medical databases.

Acknowledgements

The French experimentation and assessment of *EBMPracticeNet* has been funded by the French public health insurance system (Assurance Maladie).

References

[1] Kwag KH, González-Lorenzo M, Banzi R, Bonovas S, Moja L. Providing doctors with high-quality information: an updated evaluation of web-based point-of-care information summaries. *J Med Internet Res* **18** (2016), e15.

[2] Banzi R, Liberati A, Moschetti I, Tagliabue L, Moja L. A review of online evidence-based practice point-of-care information summary providers. *J Med Internet Res* **12** (2010), e26.

[3] Delvaux N, Van de Velde S, Aertgeerts B, Goossens M, Fauquert B, Kunnamo I, Van Royen P. Adapting a large database of point of care summarized guidelines: a process description. *J Eval Clin Pract* **23** (2017), 21-28.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190635

Design and Evaluation of an Automatic Speech Recognition Model for Clinical Notes in Spanish in a Mobile Online Environment

Alejandro Renato, Hernan Berinsky, Mariana Daus, Miguel Fantin Dachery, Oscar Jauregui, Fernando Storani, María Laura Gambarte, Carlos Otero, Daniel Luna

Department of Health Informatics, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina

Abstract

Clinical documentation in healthcare institutions is one of the daily tasks that consumes most of the time for those involved. The adoption of mobile devices in medical practice increases efficiency among healthcare professionals. We describe the design and evaluation of an automatic speech recognition system that enables the transcription of audio to text of clinical notes in a mobile environment. Our system achieved 94.1% word accuracy when evaluated on pediatrics, internal medicine and surgery services.

Keywords:

Speech Recognition Software, Mobile Applications, Electronic Health Records.

Introduction

Clinical documentation is an important task, but also time consuming, often taking up 50% of the total time of health professionals [3]. This is where the need arises for new support technologies such as speech recognition, which could allow professionals to free up time and resources. Professionals could then focus more attention on the relationship with the patient. In addition, speech recognition technology maintains the quality of clinical documentation carried out and applied according to the workflow of the professional [4].

In this work carried out in the Hospital Italiano of Buenos Aires, we describe the development and implementation of an application of automatic speech recognition (ASR) oriented to medical domain in a mobile environment.

Methods

In 1998, the Hospital Italiano de Buenos Aires (HIBA) began implementation of its Health Information System (HIS), integrating the administrative and clinical workflows with the applications in use. This system handles all clinical and administrative information, from capture to analysis. Currently, the Electronic Health Record (EHR) is a software in web format, with problem oriented architecture and is focused at the patient level.

Automatic Speech Recognition System

The ASR system was developed internally by HIBA using the open source Kaldi system [2]. For the construction of the system, both the acoustic and the language models, and the acoustic dictionary were developed as part of our work, as well as a reverse text normalization module.

The acoustic model was trained with approximately 800 hours of audio from Spanish speakers from different locations in Argentina. At the end of the training period, we obtained a Neural Networks model (NNET3 in terms of Kaldi, based on sequential models with an I-Vectors module).

The language model used in this work was constructed from the clinical notes performed at HIBA for 5 years, a total of 12 million. The notes comprise 800 million tokens, and a vocabulary close to 80,000 terms. More than 60 medical specialties were represented in the notes. For the construction of the model, the text was standardized and normalized. With this textual database, we built a trigrams model consisting of 62,000 words.

The pronunciation of the 80,000 words of clinical notes was obtained automatically by the grapheme to phoneme converter and then supervised manually. Finally, a module was built to perform the inverse normalization.

The ASR System architecture is shown in Figure 1. ASR Service supports full-duplex communication based on websockets in a scalable architecture: HAProxy distributes incoming requests on multiple independent Kaldi GStreamer Servers [1]; each server supports concurrent processing running multiple workers. GStreamer plugins (GST plugin), installed on Kaldi, executes the decoding task. This task transforms input audio signals to text transcription output based on the ASR Model. The Mobile application for Android is built with Ionic 1 framework and AngularJS (using Cordova plugins to access to the microphone).

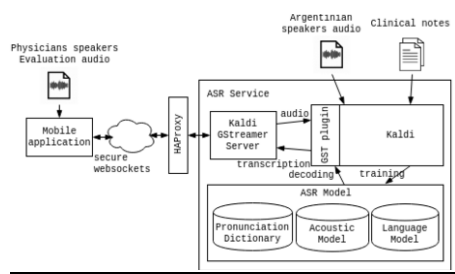


Figure 1– Automatic Speech Recognition (ASR) System architecture

Results

Over eight hours of audio recorded from the reading of clinical notes were used to assess the performance of the system.

The texts were selected from clinical notes of three specialties: internal medicine, pediatrics and surgery. The objective of the selection of texts was that several dimensions of the texts were represented: length of clinical notes measured in words, coverage of the vocabulary and less frequent phonemic combinations. To fulfill this objective, a random selection was made with the following requirements: length between 100 and 200 words, containing one of the 20,000 least frequent words, and containing at least one of the less frequent triptones found during acoustic model training. Following these criteria, 12 batches of data were obtained, each consisting at least 60 clinical notes, and were given to read to HIBA physicians who knew the vocabulary of each specialty.

The speakers were selected from the set of resident physicians at HIBA. Of the twelve residents, eight were women and four were men. One of the speakers was not a native Spanish speaker (Brazilian Portuguese), while another is a native Spanish speaker, but with a different dialect (Colombia). From the remaining speakers, nine speak Rio de la Plata Spanish as their native dialect.

The texts were read in a silent environment, using the mobile client designed for the final application. The logs of the system were collected, along with the original transcription and the audios of the system.

The results of the system without tuning were 91.7% Accuracy, 8.3% WER (Word Error Rate), and 93.4% Percent Correct.

Table 1 shows the results obtained after tuning the system. WER is determined by subtracting Accuracy from 100. Accuracy is different from Percent Correct Word, since it includes both inserts and deletions. The second column in Table 1 gives the percent values while the third column shows absolute values.

Among the serious errors, which are the most difficult to control, are the numerical quantities. Of the 5666 numerical expressions that appear throughout the recorded corpus, there were 307 errors in recognizing numbers (5.41%). Although it seems that the error rate is not high, the occurrence of one of these errors in some contexts can threaten the patient's safety.

Finally, a comparison was made with a set of typed clinical notes of similar characteristics extracted randomly from the set of notes. An analysis of the errors found was then made. The analysis shows that the degree of correct words is 89.2%, versus 95.8% obtained by the ASR system.

Table 1– Results Evaluation

Measurements	Percent	N Words
Word Error Rate	5.9%	2745
Percent Correct	95.8%	44406
Substitution	2.9%	1354
Deletions	1.3%	607
Insertions	1.7%	784
Word Accuracy	94.1%	

Conclusions

We conclude that our mobile ASR system for taking clinical notes has been shown to perform acceptably. Other studies are underway to verify parameters such as: tool safety, clinical applicability, performance compared to the traditional method and quality of the record.

References

[1] T. Alumäe, Full-duplex Speech-to-text System for Estonian, in: *Proceedings of the Sixth International Conference Baltic HLT 2014,* G.G. A. Utka, J. Kapočiūtė-Dzikienė, ed., IOS Press, 2014, pp. 3-10.
[2] A.G. Daniel Povey, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, Jan Silovsky, Georg Stemmer, Karel Vesely, The Kaldi speech recognition toolkit, *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding* (2011).
[3] N.H. Solutions, Accuracy & completeness of clinical documentation, in, 2015.
[4] W.D.A. T.H. Payne, J.A. Markiel, K. Lybarger, and A.A. White, Using voice to create hospital progress notes: Description of a mobile application and supporting system integrated with a commercial electronic health record, *Journal of Biomedical Informatics* **77** (2018), 91-96.

Address for correspondence

Alejandro Renato
alejandro.renato@hospitalitaliano.org.ar

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190636

Continuous Improvement of Clinical Decision Support via an Embedded Survey Tool

David Rubins[a,b,c], Sayon Dutta [a,b,d], Adam Wright[a,b,c], Gianna Zuccotti[a,b,c]

[a] *Brigham & Women's Hospital, Boston, MA, USA,*
[b] *Harvard Medical School, Boston, MA, USA,*
[c] *Partners HealthCare, Boston, MA, USA,*
[d] *Massachusetts General Hospital, Boston, MA, USA*

Abstract

Clinical decision support systems (CDSS) are widely used to improve patient care and guide workflow. End users can be valuable contributors to monitoring for CDSS malfunctions. However, they often have little means of providing direct feedback on the design and build of such systems. In this study, we describe an electronic survey tool deployed from within the electronic health record and coupled with a conversation with Clinical Informaticians as a method to manage CDSS design and lifecycle.

Keywords:
Clinical Decision Support, Feedback, Quality Improvement

Introduction

Clinical decisions support systems (CDSS) are widely used in electronic health records (EHR) and have been shown to improve patient safety, increase adherence to guideline-based care and decrease healthcare costs. However, increasing adoption and reliance on CDSS can have unintended consequences that can negatively impact patients [1-3] and lead to increased and inappropriate alerts for healthcare workers when CDSS malfunction.

In addition to automated monitoring systems to prevent CDSS from malfunctioning or quickly remedy errors, EHR users are a potential source of important feedback about the functioning and usefulness of the systems. Engaging the large number of end-users to comment on and report issues with CDSS related to differing workflows and errors in logic can allow malfunctions to be identified more quickly. Further, designing systems to directly alert Clinical Informaticians on an organization's Clinical Decision Support Team can accelerate the remediation of a malfunction. End-users may also have ideas about strategies to optimize a clinical workflow or changes to the EHR build to address a clinical problem without resorting to interruptive alerts.

We developed and implemented a user feedback tool embedded within our CDSS and assessed the impact of user-feedback on our ability to recognize malfunctions, improve CDSS design and manage the CDSS lifecycle.

Methods

In our EHR (Epic Systems v2018), there were no dedicated methods of eliciting user feedback. Healthcare workers were able to leave comments on alerts received at the point of care, but this feedback was limited because the end-user was unclear who read the comments and comments were generally directed towards concerns related to patient care, rather than system functioning. In addition, comments entered in this manner are stored in the patient's permanent medical record, which is not ideal for describing errors in the alerting logic or suggestions for EHR improvement.

To address these limitations, at our institution have we leveraged an external survey tool (REDCap) to collect feedback about how alerts were functioning. We placed a "provide feedback" link to the survey in every alert (Figure 1) and created a single question survey asking for user feedback about the alert they had just received (Figure 2). The link contained associated metadata about the alert such as time, patient medical record number, identity of alert displayed and alert ID. The survey encouraged users to leave any sort of feedback through the alert. Whenever possible, a member of the Clinical Informatics Team responded directly to the user with explanations about the current-state CDSS build and encouraged them to engage in direct discussion about what, if anything, would be changed based on their survey feedback or why it could not be changed.

Results

Over the past year, we have received 983 user feedback comments. Of those, 112 (11%) were blank or single character responses; 321 (33%) related to clinical care (i.e. not feedback on the alert) and 550 (56%) were feedback about the alerts. We received feedback on 89 different alerts (~19% of our CDSS inventory) from 415 unique users (highest user left 35 responses); 304 (55.3%) responses related to 10 alerts.

We responded to 152 (27.6%) user feedback comments that were not about patient care or were blank. We categorized the comments into 6 categories: Unclear Why Alert Fired (27%), Suggestion for Improvement (24%), Error in Alert (21%), Frustration (19%), Clinical Misunderstanding of Alert (6%), Positive Feedback (2%).

Based on the feedback received, we made changes to 43 alerts ranging from updating display text to adjusting errors in logic to creating entirely new version of the alerts. In addition, the user feedback system alerted us to 3 abnormal firing rates before detection by our automated daily monitoring system and we were able to correct those errors on the same day.

Discussion and Conclusions

Launching the REDCap survey tool from within the CDSS alert workflow provided important real-time feedback on the design and build of our CDSS. User adoption took place seemlessly and without specific training. We were able to identify malfunctions in our system, as well as improve workflows. Ad hoc feedback from users leads us to believe that direct conversation between the Clinical Informatics staff and users can provide both groups with new information to be used for system design and result in an improved sense of partnership. In the future, we plan to expand the feedback and user engagement to include a response to all comments. We also plan to incorporate routine review of aggregated comments regarding individual alerts as part of our routine CDSS review and life cycle maintenance.

Figure 1– Best Practice Alert showing link to surrvey (arrow)

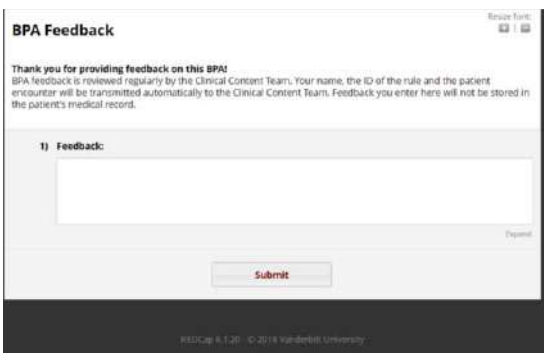


Figure 2– Survey Landing Page

Acknowledgements

The authors appreciate the assistance of the Partners eCare Extensions Team and the many EHR users who have provided real-time feedback to improve the system.

References

[1] T.J. Bright, A. Wong, R. Dhurjati, E. Bristow, L. Bastian, R.R. Coeytaux, G. Samsa, V. Hasselblad, J.W. Williams, and M.D.J.A.o.i.m. Musty, Effect of clinical decision-support systems: a systematic review, **157** (2012), 29-43.

[2] A.B. Landman, S.S. Takhar, S.L. Wang, A. Cardoso, J.M. Kosowsky, A.S. Raja, R. Khorasani, and E.G.J.J.o.t.A.M.I.A. Poon, The hazard of software updates to clinical workstations: a natural experiment, **20** (2013), e187-e190.

[3] E.G. Stone, Unintended adverse consequences of a clinical decision support system: two cases, *Journal of the American Medical Informatics Association* **25** (2017), 564-567.

Address for correspondence

Gianna Zuccotti, MD
Partners Healthcare System
399 Revolution Drive,Suite 1065
Somerville, MA 02145-1466
USA
gzuccotti@partners.org

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190637

Patient-Empowered Electronic Health Records

Tony Sahama[a], Andrew Stranieri [b], Kerryn Butler-Henderson[c]

[a] Federation University of Australia Brisbane, QLD, Australia
[b] Federation University of Australia Ballarat, VIC, Australia
[c] University of Tasmania, Tasmania, TAS, Australia

Abstract

Electronic Health Records (EHRs) constitute evidence of online health information management. Critical healthcare information technology (HIT) infrastructure facilitates health information exchange of 'modern' health systems. The growth and implementation of EHRs are progressing in many countries while the adoption rate is lagging and lacking momentum amidst privacy and security concerns.

This paper uses an interrupted time series (ITS) analysis of OECD data related to EHRs from many countries to make predictions about EHR adoption. The ITS model can be used to explore the impact of various HIT on adoption. Assumptions about the impact of Information Accountability are entered into the model to generate projections if information accountability technologies are developed. In this way, the OECD data and ITS analysis can be used to perform simulations for improving EHR adoption.

Keywords:

Electronic Health Records, Intervention Study, Information Accountability.

Introduction

The health sector plays a significant role in any country's productivity and economic development. Health sector expenses are astronomical, and factors associated with service deliveries and healthcare products, directly impact on quality of life. The health sector is also an information intensive sector in that information is generated and consumed in enormous quantities by healthcare professionals, patients, managers, and other stakeholders. In this information, intensive sector, the research challenges include discovery of ways that health information can be integration that linked and, measuring appropriate use of health information under the given privacy and security constraints (e.g., Information Accountability) [1, 2].

Electronic health records (EHR) capture healthcare information prior to the inception of an individual's birth, through the life span until after death. These records are either captured in paper format or are electronic in nature and, should and would empower the consumer and their families care. It is indeed the comprehensive, unified, longitudinal records that aid in improving 'quality of life'. With technological maturity, EHRs constitute evidence of online records from interactions between professionals (e.g., Physicians and GPs), the public (Consumers and Patients) and healthcare service providers (healthcare policy makers and funding agencies). Furthermore "EHRs consist of patient information such as demographics, medications, laboratory test results, diagnosis codes, and procedures" [3].

Realization of the health information flow and contextual awareness of EHR sharing capabilities are vital to have meaningful and quality healthcare decision making [4]. This socio-technical, information accountability driven, conceptual EHR knowledge-based information sharing model is considered for this study [4], illustrated in Figure 1 below.

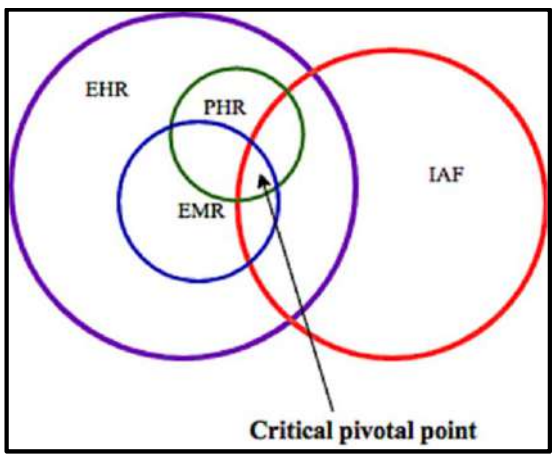


Figure 1– Information Accountability Driven EHR Model [4]

A brief description of Figure-1 elements are presented below:

1. **EMR** (electronic medical record)– information is amended; updated personal health records (**PHR**s) are managed by authorized clinicians and healthcare organizations (the professional view). While the information stored in this area is vital for the clinical decision-making process, at the same time the **PHR** will be managed to link the "Patient" for specific information sharing.
2. **PHR**– is recognizable; where individual information is stored, collected, shared and controlled by the individual (the public view). This set of information supports the individual "Patient" healthcare journey. The Healthcare Information Exchange (**HIE**) requires meaningful linkage of the **EHR** to complete the "Patient" journey.
3. **EHR**– is a key element of the model that is recognizable [1, 2]; and "*a comprehensive interconnected health information record that can capture and share a variety of information about people's health status, their history of encounters with the healthcare system, the results of all diagnostic and therapeutic interventions, and (ideally) their key social and demographic characteristics*".
4. **IAF**– is the Information Accountability Framework, the principle defined and implanted [1, 2] for the model. The governing principles of Information Accountability is "*that information usage should be transparent so it is possible to determine whether a use is appropriate under a given set of rules*".

5. Ω–this pivotal point is a significant balance [4] between healthcare services providers (policy makers and practitioners) and consumers ("Patient") to be maintained transparently when sharing the EHR.

The aims of this paper are to establish a socio-technical, information accountability driven, conceptual EHR knowledge based information sharing model demonstrated in preceding sections. The following sections present the data set retrieved based on infodemiological discoveries, organized and analyzed using an intervention study using Interrupted Time Series Analysis [6 & 7], as well as a discussion of the findings.

Methods

The OECD has assessed the Technological Readiness of countries (TOR) and the Data governance readiness through a series of surveys and assessments between 2015 and 2017. A Quasi–Poisson model was used to determine the variances. The statistical analysis was carried out using the **R** function ***glm*** and necessary adjustments were accommodated to estimate time seasonality.

Results

The dispersion parameter for the ***quasipoisson*** family is taken to be 0.1841. The ITS generated post 2015 projection illustrated that some countries are within the acceptable range while others are not. This illustrates that by using assumptions about the impact of Information Technology, some countries are ready to accommodate this HIT whereas others are not.

Discussion

An adoption of EHR associated technologies provides sustainability of healthcare systems and enables population-based outcomes. Furthermore, 'modern' healthcare systems affordability and equity depend on its performance. The proposed model (Figure 1) allows consumer-directed exchange of health information. With the patient–empowerment principle in the background, it is reasonable and practicable to search for the coexistence of HIT infrastructure that facilitates the model capabilities, possibilities and opportunities to establish the patient controlled EHR system.

This study describes an *infodemiological* exploration of patient empowerment measured through EHR adoption using OECD data analyzed through ITS models. This approach would help and "*could shed light on unmet medical needs and research priorities for the future, and provide guidance for the decision making in public policy*" [7]. An alternate approach would be to employ a randomized control trial (RCT), however this exercise would be impractical and difficult to generalize to "real world" settings.

Conclusions

Use of personal health data creates opportunities for health system improvement, research and disease surveillance, but requires the right governance frameworks to realize these benefits while managing risks. Patient–empowerment facilitates and provides HIT factors including sharing electronic health records providing functional data interoperability and meaningful use to minimize the security and privacy concerns. This paper investigated potential and opportunities for shared EHRs in Patient–empowerment from a sociotechnical perspective. The intervention study used an interrupted time series analysis (ITS) of the publicly available data, and its prediction diagnosis for opportunities and usefulness are based on recent developments of the EHR on two readiness factors: Technical and Operational (TOR), and Data governance (DGR), which are part and parcel of the socio-technical interactions.

This makes better use of health data in this exponential growth of digital technologies, which is warranted to establish the support of Patient–empowerment. Sharing data under a thoroughly functional data interoperability regime would support other community benefits like migrant inflow to the health watch of the communities. Establishing 'quality of care' among such community elements should also be supported. The Healthcare expenditure and health policy implementation is controversial and a troubling location. Healthcare spending might be managed and reduced by use of HIT however, long–term planning on studies like this and longitudinal experiments on the use of **shared EHR** should benefit most counties that require immediate healthcare support.

Based on the information available from OECD countries, employing and implementing an intervention study of interrupted time series regression analysis (ITS) on available and practicable HIT infrastructure information is a promising start.

Acknowledgements

The authors acknowledge that this paper neither analyses the situation per country nor provides predictions for a country. This analysis is out of the scope of this paper since the granularity of the data is complex and unknown.

Data sources for the work was obtained from the OECD iLibrary and we appreciate the support rendered by the iLibrary team. There is no financial transaction involved in this endeavor.

References

[1] R. Gajanayake, R. Iannella, and T. Sahama, Sharing with Care: An Information Accountability Perspective. IEEE Internet Computing 15:4 (2011), 31–38

[2] R. Gajanayake, R. Iannella, and T. Sahama, An Insight into the Adoption of Accountable-eHealth Systems. An Empirical Research Model Based on the Australian Context. IRBM 37:4 (2016), 219–231

[3]. P. Yadav, M. Steinbach, V. Kumar, and G. Simon, Mining Electronic Health Records (EHRs): A Survey. ACM Comput. Surv. 50, 6, Article 85 (2018), 40 pages.

[4]. T. Sahama, L. Simpson, and B. Lane, Security and Privacy in eHealth: Is it possible? In 2013 IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013). 249–253.

[5]. M. Huang, O. ElTayeby, L. Zolnoori and M. Yao, Public Opinions Toward Diseases: Infodemiological Study on News Media Data. J Med Internet Res 20:5 (2018), e10047.

[6]. B. J. Lopez, S. Cummins and A. Gasparrini, Interrupted time series regression for the evaluation of public health interventions: a tutorial, International Journal of Epidemiology 46:1 (2017), 348—355

[7] E. Kontopantelis, T. Doran, D. A Springate, I. Buchan, and D. Reeves, Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. BMJ (2015), 350—h2750.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190638

Analysis of the Stay Time of Patients in Gunma University Heavy Ion Medical Center (GHMC) Using RFID Technology

Partha Protim Hazarika, Kota Torikai, Rei Noguchi, Yuichiro Saito

Department of Medical Informatics, Gunma University Graduate School of Medicine
Maebashi City, Gunma, 371-8511 JAPAN

Abstract

We observed the stay time of patients and staff in Gunma University Heavy Ion Medical Center. The stay time of patients with the prostatic cancer and the facing time with radiotherapy technicians in treatment rooms were significantly reduced as times goes by. This decreasing in time has an implication in scheduling algorithm development: for patients. RFID technology can be a potential method to track both staff and patients and thereby to assess the resource utilization efficiency.

Keywords:

Radio Frequency Identification Device (RFID), heavy ion therapy, prostatic cancer.

Introduction

Radio Frequency Identification Device (RFID) technology is a short-range wireless electromagnetic field technology used to communicate digital information between a stationary location and a movable object or between movable objects. Use of RFID-based unique identification of objects is presently being explored in many sectors, including health. We have reported that RFID technology was found to provide feasible and accurate means for capturing and evaluating nursing time spent in patient rooms [1]. We checked feasibility of this system to evaluate resource utilization of our Gunma University Heavy Ion Medical Center (GHMC). It is available only in 12 centers across the globe, as its installation necessitates huge infrastructure cost and skilled manpower resources [2]. Appropriate treatment scheduling of patients at a medical office in a hospital using RFID system has the potential to reduce patient's waiting time [3]. It will have implications in improving patients' satisfaction and better compliance to treatment, in addition to optimizing utilization of precious heavy ion treatment facilities.

Methods

The RFID system at GHMC (by TOPPAN FORMS, Tokyo, Japan) used a frequency band of 93.75 KHz and 300 MHz. All patients and staff at GHMC were issued a specific semi-active integrated circuit tag (MXAT-MV-14, MATRIX, MATRIX, Osaka, Japan) which was activated by two trigger coils during movement to/from a room. To perform this retrospective study, we retrieved all the time stamp information stored on the server from September 2010 to November 2013. The data were initially sorted out and separated based on tag ID holders – patients, doctors, and radiology technicians. Clinical data (age, diagnosis, and performance status) of all the available patients is retrieved from the hospital information system. All the statistical analysis of this study was done using statistical data analysis package of Microsoft Excel 2016. Statistical significance indications were $p < 0.05$ by using unpaired, two-tailed t-tests. This study was approved by the Institutional Review Board (IRB) of Gunma University (reg. no. 2016-030).

Results

In patients with the prostatic cancer, we observed a significant decrease in the average stay time on 2nd, 4th, 5th, 8th, 9th, 11th, 12th, 13th, 14th and 15th days of treatment as compared with the first day of treatment during the early period. On the other hand, in case of patients with the liver cancer and the lung cancer, there was a decreasing trend in stay time on subsequent days, although the reduction was not statistically significant because of small number of patients (n=32 in 3.6 years). When the respective treatment days on the early period were compared to other periods in patients with the prostatic cancer, the stay time was significantly reduced in a later period (Figure 1). Next, we analyzed the stay time of radiology technicians to confirm if a reduction in patients' stay time is associated with parallel reduction of radiology technicians stay time in treatment unit. We noticed that there was statistically significant reduction in stay time of radiology technicians on 2nd, 3rd, 6th, 9th, 12th, and 14th days in the early period. Furthermore, the stay time of radiology technicians in the later period was significantly decreased as compared with the early period (Figure 2). On the contrary, the stay time of patients inside examination unit with doctors was similar throughout the time independent of the treatment unit installation status. Based on the analysis of tracking results, we developed a radiology information system to book a treatment schedule of our patients in GHMC. The number of patients treated in GHMC increased from 266 in 2013 to 574 in 2018

Discussion

We observed that the stay time of the prostatic cancer patients in treatment unit was considerably reduced in the last several days as compared to the 1st day of treatment. It was noticed that stay time of patients as well as the face time with radiotherapy technicians in the treatment unit was significantly decreased in most of the days of the later period. This suggests that the radiotherapy technicians took comparatively lesser time during pre-treatment setting up and post-treatment patient removal and it proves their increasing efficiency with time. Generating a transparent information flow and using it for coordination

between healthcare workers and patients is an effective way to minimize waiting times, which is possible by tracking patients as well as staff by using RFID technology. We had a few limitations in our study. In 3.51% of patients, there was no data recorded in server, due to missing signal capture from the tag readers. Vakili et al. found that an RFID system's event record rate was 83.7% in the rooms of an ambulatory clinic in Johns Hopkins University School of Medicine [4]. Our data were better than their rate, however we need efforts to improve accuracy of RFID systems in future. Long waiting time for treatment is a very important but overlooked infrastructural problem [5] and it can lead to poor adherence to prescribed treatment schedule especially in case of elderly cancer patients, thereby compromising possibility of cure [6]. Patient experience and/or outcome measure will be topics for future study.

Conclusions

Our study suggests that RFID technology can be a potential method to track both staff and patients and thereby to assess the resource utilization efficiency.

Figures and Graphs

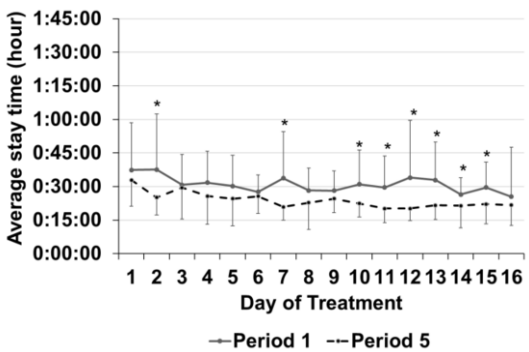


*Figure 1. Comparison of average stay time of patients with the prostatic cancer in treatment unit. Period 1 was September, 2010 – December, 2010 (n= 25), period 5 was April, 2012 – July, 2012 (n= 39). *p<0.05 vs. peiod5.*

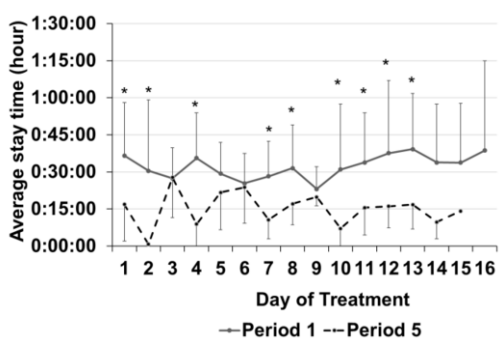


*Figure 2. Comparison of average facing time of radiology technicians to patients with prostatic cancer in treatment unit. Period 1 was September, 2010 – December, 2010 (n= 25), period 5 was April, 2012 – July, 2012 (n= 39). *p<0.05 vs. period 5.*

Acknowledgements

This study was supported by a grant (no. 26460854) from the Japan Society for the Promotion of Science.

References

[1] Y. Saito, R. Suzuki, K. Torikai, T. Hasegawa, T. Sakamaki, Efficiency and safety of new radiofrequency identification system in a hospital, Stud. Health Technol. Inform 192 (2013), 1032.

[2] Particle Therapy Co-Operative Group, Particle therapy facilities in clinical operation, (2019) https://www.ptcog.ch/index.php/facilities-in-operation.

[3] D.S. Kim, J. Kim, S.H. Kim, and S.K. Yoo, Design of RFID based the Patient Management and Tracking System in hospital, Conference Proceeding IEEE Eng Med Biol Soc, (2008), 1459-1461.

[4] S. Vakili, R Pandit, E. L. Singman, J. Appelbaum and M. V. Boland, A. comparison of commercial and custom-made electronic tracking systems to measure patient flow through an ambulatory clinic, Int. J. Health Geogr, 14 (2015), 32-38.

[5] J.C. Chen, T.J. Collins, Creation of a RFID based real time tracking (R-RTT) system for small healthcare clinics. J. Med. Syst. 36 (2012), 3851–3860.

[6] M.T Puts, HA Tu, A Tourangeau, D Howell, M Fitch, E Springall, SM Alibhai. Factors influencing adherence to cancer treatment in older adults with cancer: A systematic review. Ann. Oncol 25 (2014), 564–577.

Address for correspondence

Yuichiro Saito
Department of Medical Informatics,
Gunma University Graduate School of Medicine
3-39-22 Showa-machi, Maebashi City, Gunma, 371-8511 JAPAN
Preferred method of contact is e-mail.
E-mail; saitoyui@gunma-u.ac.jp
Phone; +81-27220-8771
Fax number; +81-27220-8770

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190639

Prototype of Care Application for Obstetric Telemonitoring of Hypertensive Syndromes in High Risk Pregnancy

Danielle Santos Alves[a,b], Érika Maria Alves da Silva[a], Mikellayne Barbora Honorato[a], Magdala de Araújo Novaes[a,c]

[a] *Telehealth Center, Clinics Hospital, University Federal of Pernambuco, Recife, Pernambuco, Brazil*
[b] *Department of Nursing, Post-Graduation Program in Computer Science, Federal University of Pernambuco, Recife, Pernambuco, Brazil*
[c] *Internal Medicine Department, University Federal of Pernambuco, Recife, Pernambuco, Brazil*

Abstract

This article describes the development of a prototype application for obstetric telemonitoring of hypertensive syndromes during pregnancy. A workflow was elaborated with the conduct for gestational hypertensive syndromes. Subsequently, prototyping was performed using Balsamiq. The prototype presents daily monitoring of blood pressure, signs and symptoms, displays accompanying charts, and generates alerts when there are changes to more or less the normal values, which are sent to the pregnant woman and the health professional.

Keywords: Mobile applications; Hypertension, Pregnancy; Obstetrics.

Introduction

Pregnancy is a physiological and dynamic process that occurs most often without complications. However, a considerable number of women develop gestational problems or some health impairments during the gestational period. Gestational hypertensive syndromes (GHS) represent a high risk for maternal and fetal health [1, 2]. In addition to diseases such as hypertensive encephalopathy, cardiac and renal failure, GHSs are associated with severe perinatal complications [2].

GHSs are considered the second cause of maternal mortality [3]. They can often be avoided by screening family, personal, habits and lifestyles, as well as controlling and monitoring blood pressure levels. Thus, early screening of risk factors for the development of SHGs becomes essential in assisting the pregnant woman so that early detection of injuries to the mother-baby binomial is possible.

There are several strategies that the health professional can use to monitor the possible gestational risk factors for GHS. With the increasing use of the internet by pregnant women and family members for gestational follow-up, the use of applications for this purpose has shown to be a great potential [4]. As the Internet has become a tool of support for pregnant women in seeking information about pregnancy, the websites and applications available do not guarantee to provided the basis of scientific evidence [4].

Considering the importance of the screening of risk factors, monitoring pressure levels and detecting warning signs of complications due to hypertensive syndromes, as well as the increasing and positive use of technology in health by pregnant women and health professionals for maternal, this study aims to describe the development of a prototype of care application for obstetric telemonitoring of hypertensive syndromes during pregnancy.

Methods

The pressent study was conducted in the period from January to October 2018. This study describes an application prototype related to gestational hypertensive syndromes, which will be incorporated into an obstetric telemonitoring system.

First, a literature review of international guidelines, manuals and protocols of the Brazilian Ministry of Health (MS), the World Health Organization (WHO), health institutions and scientific evidence was conducted, followed by a search in the Play Store of gestational applications that deal with blood pressure monitoring and hypertensive syndromes. Then, a workflow for SHGs using BPMN in the Bizagi Modeler software was created with the specification of a clinical protocol with all the information. A prototyping of the GHS service module from Balsamiq software was developed. In this preliminary study, no usability testing has been performed, limiting only to prototyping. Thus, no ethical appreciation was required.

Results

The data collected from the materials were unified in a single flowchart that encompasses different aspects including screening, diagnosis, prevention, and management of hypertensive syndromes during pregnancy.

The search in the Play Store showed that only one application presented a feature of monitoring blood pressure. Regarding health education, 6 applications presented general educational information about the changes in blood pressure in pregnant women. None of them evaluated aspects related to antecedents and risk factors or allowed to fulfill obstetric telemonitoring data by health professionals.

A protocol was created containing the signs and symptoms, risk factors, explanations, possible diagnoses, behaviors for pregnant women and health professionals, as well as warning messages for both, when necessary.

The prototyping of the application was performed based this protocol using Balsamiq, a medium fidelity prototyping software, which allows the elaboration of functional interfaces and gives the user an initial impression of the application's operation.

Regarding the main functionalities, the prototype presents space for insertion and daily monitoring of the pregnant woman's blood pressure (BP) (Figure 1). The pregnant woman not only can insert the BP from a external device at home, but

also can register the data recorded on her prenatal card or follow the data recorded by the health professional.

Figure 1- Prototype screens: data record of health and symptoms

The alerts are sent to both the pregnant woman and her health professional in prenatal care. The prototype also presents educational information on hypertensive syndromes in gestation, including gestational hypertension, preeclampsia, preeclampsia overlapping with chronic hypertension, eclampsia and hellp syndrome, emphasizing health education for pregnant women and health professionals.

Discussion

Blood pressure is a vital sign and thus needs continuous monitoring, especially during pregnancy. It is an indicator that allows the monitoring of health parameters that, when present or altered, predispose pathologies responsible for a large part of maternal and neonatal deaths and morbidity, such as pre-eclampsia and eclampsia [3, 5, 6].

The data entered are available for both personal and professional conferences, allowing for continuous monitoring of pregnant women's blood pressure values and early intercurrent screening, as well as better screening for high-risk pregnant women. To assist the health professional, the prototype of the hypertensive syndromes module, flowcharts and protocols on screening, diagnosis and management of SHGs are available.

The use of digital resources in preventive health presents itself as an opportunity to gather pregnant women, family and health professionals in a virtual and safe environment so that they can seek correct and relevant information and obtain a safe and continuous assistance among health professionals.

Conclusions

In this study we developed a prototype application that aims at hypertensive syndromes during pregnancy. The application will be incorporated into an obstetric telemonitoring system for prenatal follow-up. It is understood that the early identification of potential gestational complications through low cost and high impact interventions can reduce the number of morbidity and mortality among pregnant women worldwide.

Acknowledgements

Thanks to the Nucleus of Telehealth of the Federal University of Pernambuco for making this research feasible

References

[1] World Health Organization. *WHO recommendations for prevention and treatment of pre-eclampsia and eclampsia: implications and actions*, 2014.

[2] A.C.M. Oliveira, et al. Fatores maternos e resultados perinatais adversos em portadoras de pré-eclâmpsia em Maceió, Alagoas, *Arq Bras Cardiol* (2016), 106, 2, 113-20.

[3] J.M.G. Dias, et al. Mortalidade Materna, *Revista Médica de Minas Gerais*, (2015), 25, 2, 173–179.

[4] T.M. Calderon, et al. O uso da internet como ferramenta de apoio ao esclarecimento de dúvidas durante a gestação The use of the Internet as a support tool to clarify questions during pregnancy. *Journal of Health & Biological Sciences*, (2016) 4, 1, 18-22.

[5] Ministério da Saúde. *Protocolos da Atenção Básica : Saúde das Mulheres. 1a ed. Brasília - DF*: Ministério da Saúde. Secretaria de Atenção a Saúde. Departamento de Atenção Básica. Editora do Ministério da Saúde, 2016.

[6] World Health Organization. *Building a Future for Women and Children: The 2012 ReportWorld Health Organisation WHO.* World Health Organization, 2012.

Address for correspondence

Danielle Santos Alves

Federal University of Pernambuco. Nursing Department. Av. Prof. Moraes Rego, 1235 - University City, Recife - PE – Brazil – Zip Code: 50670-901 Phone: (+55 81) 981087638. dsa3@cin.ufpe.br

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190640

Terminology Gap in Continuous Care Between Acute and Long-Term Care Hospitals

Ryoma Seto[a], Toshitaka Inoue[b,c], Suemi Katayama[c]

[a] *Division of Healthcare Informatics, Faculty of Healthcare, Tokyo Healthcare University, Tokyo, Japan*
[b] *Department of Social Welfare, Nishi-Kyushu University, Saga, Japan*
[c] *Seishinkai Inoue Hospital, Fukuoka, Japan*

Abstract

This study aimed to elucidate the gap in terminology between acute and long-term care (LTC) hospitals. Fifty-seven hospital documents were analyzed using text mining. Each document contained a mean 194.2 terms. Acute care hospital documents often contain pharmacological information. LTC hospital documents often contain information related to patients' lives. Documents from both settings used local, non-standardized language. Our results suggest that expanding the national standard of nursing terminologies has potential for enhancing continuity of care.

Keywords:
Long-Term Care; Electronic Health Records; Standardized Nursing Terminology

Introduction

In a rapidly aging society, continuous care is becoming more and more important. Acute care hospitals are actively working to shorten the length of stay owing to health service reform initiatives. Acute care hospitals should provide patients with information using a "document for continuous care (DCC)" regarding the transition to long-term care (LTC) facilities (e.g. rehabilitation hospitals, nursing homes, and home care). Unfortunately, many LTC facility residents are readmitted to acute care hospitals because of changes in their general condition (e.g. loss of body water or inability to consume meals). LTC facilities therefore have to prepare DCCs for use by acute care hospitals.

The DCC content is not dictated by the Japanese government, but rather by local governments or public associations (local DCCs). However, local DCCs widely differ owing to the use of various formats, terminologies, or other items. Some DCCs are implemented as a component of the electronic health record (EHR).

In some areas, the DCC in the EHR is an effective means of sharing patient information with other settings [1-3]. Conversely, paper-based DCCs are more popular than electronic DCCs because most LTC facilities do not use EHRs. Most public sectors (e.g. local governments and/or public associations) prefer paper-based documents and forms compared with electronic documents, although paper-based documents are often generated using computer applications (e.g. Microsoft Office™).

Some public sectors provide a DCC template on their websites. These templates may prevent the expansion of the use of electronic DCCs as the national standard. This study aimed to elucidate the terminology gap in DCCs used in acute care and LTC hospitals.

Methods

1) Data collection

Twenty-five local areas in Japan that provided a DCC template on their websites were enrolled in the study. We therefore collected and examined 57 DCC forms from these areas.

2) Data analysis

First, the 57 DCCs were classified based on document user (sender/receiver). Next, each DCC was analyzed using a text-mining tool (*KH-Coder 2.00f*, Japan), and terms contained within each DCC were numbered. These terms were then categorized and analyzed to elucidate the terms that were most often used, relative to each setting (acute care or LTC hospital). Finally, frequently used terms were matched with those in nursing practice standardized terminology (NPST). NPST was approved to contain national standard terminologies in Japan by the Ministry of Health, Labor and Welfare in 2016. The NPST, as the national standard, covers the widest range of terminologies.

Results

1) Type of DCC forms

As for the sender of the DCCs, 26 of the 57 DCCs originated in medical settings (e.g. acute care hospitals), 22 originated from "care managers" (CM), and 21 in LTC settings. The total number includes common-use forms for both settings; thus, the total number is actually more than 57. A CM is a specialist who builds each patient's care plan under LTC insurance.

As for the receiver of the DCCs, 9 of 57 DCCs were CMs, 42 were acute care hospitals, and 17 were LTC settings (*Table 1*).

Table 1 – Sender and Receiver of DCCs (n=57)

User of DCCs	N	(%)
Sender		
Medical Settings (e.g. Acute hospital)	26	45.6
Long Term Care Settings	21	36.8
Care Manager	22	38.6
Others (e.g. Dental, Pharmacy)	4	7.0
Receiver		
Medical Settings (e.g. Acute hospital)	42	73.7
Long Term Care Settings	17	29.8
Care Manager	9	15.8
Others (e.g. Dental, Pharmacy, EMS)	5	8.8

2) Character of each DCCs

Each DCC form averaged 194.2 terms. The number of terms in DCCs from acute care hospitals or other medical settings, CM, and LTC settings was 187.8, 248.0, and 224.2, respectively. DCCs from acute care hospitals or other medical settings tended to contain pharmaceutical information (e.g. ability for drug self-management). DCCs from LTC settings and CMs tended to contain information related to patients' lives (e.g. having meals, toileting, etc.) (*Figure 1*).

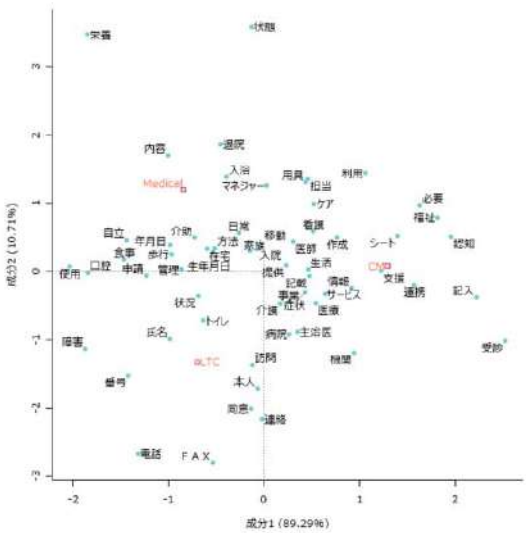


Figure 1. Character of each DCCs by correspondence analysis

3) Matching DCCs to NPST

The top 30 recurrent words found in each type of DCC were matched to the NPST. The match rate was 96.7% for DCCs from acute care hospitals or other medical settings (exception: "unnecessity"), 93.3% from CMs (expection: "smooth", "filling up"), and 86.7% from LTC settings (exception: "at a fine［nice, pretty, sad］pass" etc.).

Discussion

Some DCC forms have multiple functions, e.g. guiding the transition between (1) acute care hospitals to LTC settings, and (2) LTC settings to acute care hospitals. However, multi-function DCCs are not specialized, so the number of words was not enough. These forms tended to include free-text areas and were less structured.

DCCs from acute care hospitals contained some structured terms, but DCCs from CM and LTC settings tended to be unstructured. Unstructured DCCs are labor-intensive, and require senders to fill out the forms, possibility preventing future implementation of electronic DCCs [4].

DCCs from all settings (acute care hospitals, LTC facilities, and CMs) contain mostly local language and contain much standardized text. Even while working with paper-based DCC forms in the future, DCC terminology should consist of standardized terminology completely.

Conclusions

We found a terminology gap in the transition between acute care and LTC settings by analyzing DCC forms.

Our results suggest a need to use common terminology by expanding the NPST to cover terms that are frequently used in LTC settings. This is a realistic plan for improving the quality of continuous care in Japan.

Acknowledgements

This work is supported by Grant-in-Aid for Scientific Research (C) by Japan Society for the Promotion of Science (Grant Number 19K11180).

Reference

[1] Y. Zhang, P. Yu, and J. Shen. The benefits of introducing electronic health records in residential aged care facilities: a multiple case study, *Int J Med Inform*. **81** (2012), 690-704.

[2] H. Veseli, G. Kopanitsa, and H. Demski. Standardized EHR interoperability - preliminary results of a German pilot project using the archetype methodology. *Stud Health Technol Inform*. **180** (2012), 646-50.

[3] T.K. Colicchio, G. Del Fiol, and D.L. Scammon. Comprehensive methodology to monitor longitudinal change patterns during EHR implementations: a case study at a large health care delivery network, *J Biomed Inform*. **83** (2018), 40-53.

[4] R. Ramanayake. Structured printed referral letter (form letter); saves time and improves communication. *J Family Med Prim Care.* **2** (2013),145-8.

Address for correspondence

A/Prof. Ryoma Seto, Tokyo Healthcare University (3-11-3 Setagaya, Setagaya. Tokyo, 154-8568 Japan. e-mail: rseto-ind@umin.ac.jp)

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190641

Preconditions for Enabling Advanced Patient-Centered Decision Support on a National Knowledge Information Infrastructure

Line Silsand[a], Gro-Hilde Severinsen[a,1], Rune Pedersen[a,b], Gunnar Ellingsen[a,b]

[a] *Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway*
[b] *Telemedicine and E-health Research Group, The Arctic University of Norway, Tromsø, Norway*
[1] *Gro-Hilde has changed her surname from Ulriksen to Severinsen*

Abstract

In Western healthcare, an important goal is to provide clinical decision support "for the right healthcare personnel, in the right situation, at the right time". In this poster, we use a qualitative approach to outline the preconditions for enabling such advanced patient-centered decision support. This study indicates that establishing a national knowledge information infrastructure demands well-defined national standards, codes, and terminologies, as well as structured clinical data. An extensive governance structure is also required.

Keywords:

Knowledge bases, decision support systems, clinical

Introduction

In most developed countries, the demand for improved e-Health information and communication technology (ICT) solutions such as electronic patient record (EPR) systems has grown enormously [1; 2]. Well-designed EPR systems have the potential to support complex healthcare processes and subsequently raise the quality of treatment and improve patients' outcomes [3; 4]. In addition, it is important that healthcare personnel actively use evidence-based knowledge to ensure quality healthcare services. In the Norwegian context, the National eHealth Strategy (2017 – 2022) [5] emphasizes that EPR systems must be able to provide clinical decision support (CDS), where evidence-based knowledge is used to support patient treatment and care. Today, however, CDS capabilities are not widely used [1]. One reason for this is the lack of a central or national repository [1; 6] that enables access to evidence-based guidelines, procedures, and treatment protocols. Currently, the knowledge bases in Norway are scattered among different organizations. In addition, healthcare workers only have access to knowledge support through external publishing channels, which makes the process of finding and using the newest evidence-based knowledge complex and time-consuming [6].

Therefore, a national initiative has been started to explore how to establish a standardized "knowledge platform", the overall goal being to enable high quality knowledge support "for the right healthcare personnel, in the right situation, at the right time". In this poster, we explore the preconditions for establishing a national knowledge platform. We raise the following research question: How can a national knowledge information infrastructure be organized, and what are the preconditions to enable advanced patient-centered decision support?

We conceptualize the national platform for evidence-based knowledge as an information infrastructure (II) in order to analyze the complexity of organizing and using a large-scale knowledge base. The notion of II addresses how technology, like the "knowledge platform", is intertwined with users, work practices, and organizational structures. Therefore, standardization is an important prerequisite for communication within an II [7].

Methods

The study applied a qualitative approach based on literature searches and semi-structured interviews with EPR vendors, knowledge-base providers, and healthcare personnel and managers (see Table 1). The interviews followed a hermeneutic approach where we considered the informants' different viewpoints. All of the interviews were transcribed and systematized through thematic coding based on terms from the literature search [8], aiming to answer the research question and provide further recommendations for the national work with a "knowledge platform" [9]. When analyzing the data, II theory was used as a lens to make the findings relevant for other organizations beyond the Norwegian healthcare context [9].

Table 1 – Overview of the Data Collection

Organization	***Duration***	***Category***
EPIC	60 min	System provider
Cerner	90 min	System provider
DIPS AS	120 min	System provider
Inter Systems	60 min	System provider
Duodecim	60 min	Knowledge provider
Elsevier	60 min	Knowledge provider
UpToDate	45 min	Knowledge provider
Catalonia, Hospital	120 min	Clinical field
Finland (Pegasus)	60 min	Clinical field
HIMMS	45 min	

Results

This study presents three different categories of clinical knowledge support determined by the extent of the integration between different sources of evidence-based knowledge and the EPR system. Based on the empirical data, it is clear that the structure of and access to knowledge support varies extensively.

External passive knowledge support

In this most basic form of knowledge support and CDS, healthcare personnel are only able to access knowledge through external publishing channels outside of their EPR.

This makes the process of finding and using the newest evidence-based knowledge complex and time-consuming [6]. There exist both national knowledge sources, like the National Directorate of Health and the Norwegian Institute of Public Health, and international commercial knowledge providers that offer passive evidence-based knowledge support for healthcare. A representative from an international commercial knowledge provider stated, "We mainly provide external passive knowledge support and the knowledge is based on the best available evidence out there, without being tailored to specific countries, needs, or markets".

Almost every Norwegian healthcare organization has its own quality system which includes clinical procedures and guidelines. The knowledge content in these systems is developed and governed by the organizations themselves. Healthcare organizations abroad also work to develop and govern their own knowledge support systems. The CTO from Hospital A stated, "We have integrated some regional guidelines and procedures into the EPR system and work on a regional solution for real-time knowledge and decision support". This knowledge support provides such services as drug allergy warnings and advice on when to order blood tests.

Integrated passive knowledge support

In integrated passive support, CDS solutions are integrated with the EPR systems. However, the clinicians have to decide actively when to access the knowledge base. In Norway, there are platforms of medical treatment and care procedures that can be accessible directly from the EPR system or through a link to the procedure attached to a document or care plan. In addition, commercial knowledge providers deliver systems for integrated passive support. These CDS systems are not integrated directly with the EPR system. The clinicians have to decide which patient data from the EPR they want to combine with the CDS motor, whereupon the data is mapped and returned to the EPR system. This is particularly useful for patients with comorbidities because it is possible to combine different symptoms, medications and pathways to provide healthcare personnel with the best overall recommendations.

Active knowledge support

The most advanced form of evidence-based knowledge support is active knowledge support, where the knowledge systems are embedded in the EPRs. Health professionals automatically receive active notifications and recommendations, as well as access to patient-specific decision support in various clinical situations. Knowledge support may be triggered, for instance, when the clinician needs to choose between different treatment methods for a certain patient.

All four of the system providers interviewed are able to integrate their systems with any knowledge source or CDS. However, they argue that most healthcare organizations are not ready for active knowledge support. One system vendor stated, "The healthcare organizations must organize work processes and organizational structures like governance organizations in order to make use of effective knowledge and CDS". Neither the knowledge providers nor the system vendors offer governance related to regional or national knowledge support and CDS. In addition, knowledge and decision support are not automatically included in EPR systems. Some have partial support included in their EPR systems (e.g. for medication administration) but adjustments based on local conditions are always required. The degree of support also depends on how the healthcare organizations manage to organize and govern their knowledge base.

Conclusions

It is important to establish a national knowledge information infrastructure in order to enable advanced patient-centered CDS. Establishing a national knowledge information infrastructure demands an extensive governance structure spanning multiple organizational levels. There are three important preconditions necessary to achieve an active level of support. First, clinicians need to change their registration practice and document in real time in the EPR. Second, the EPR system must be based on structured clinical information, since CDS engines only extract structured data elements. Third, a national knowledge II requires a definition for what standards, codes and terminologies to use in order to integrate different knowledge sources into an II at a national level, as well as to ensure effective communication between the EPR and other healthcare ICT systems. This standardization encompasses both technical and organizational elements as a prerequisite for communication within the II. Finally, it is of great importance to involve the clinicians in the governance and standardization work, in order to ensure that the knowledge is updated and trustworthy to the clinicians.

References

[1] European Commission and Directorate-General for Health and Food Safety, Disruptive innovation – considerations for health and health care in Europe, Publications Office, Luxembourg, 2015.
[2] Ministry of Health and Care Services, Whitepaper no. 9 (2012-2013) "One citizen – one health record", 2012.
[3] M. Aanestad and T.B. Jensen, Building nation-wide information infrastructures in healthcare through modular implementation strategies, *The Journal of Strategic Information Systems* **20** (2011), 161-176.
[4] Ministry of Health and Care Services, Overall summary of ICT in Norwegian Healthcare, 2014.
[5] The Norwegian Directorate of eHealth, National action plan for e-Health 2017-2022 (2017).
[6] National ICT, Clinical documentation for overview and learning, Action 48, National ICT, 2014.
[7] O. Hanseth and K. Lyytinen, Design theory for dynamic complexity in information infrastructures: the case of building internet, *Journal of Information Technology* **25** (2010), 1-19.
[8] C. Robson, *Real World Research*, John Wiley & Sons Inc, 2011.
[9] H.K. Klein and M.D. Myers, A set of principles for conducting and evaluating interpretive field studies in information systems, *MIS Q.* **23** (1999), 67-93.

Address for correspondence

Gro-Hilde Ulriksen

Phone: 004790199457

E-mail: gro-hilde.ulriksen@ehealthresearch.no

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190642

CDSS for Documenting Blood Glycemia Critical Values at the POC

Camila Sofía Galván[a], Mariana Daus[a], Oscar Ignacio Jauregui[a], Matías Alejandro Álvarez[a], Carlos Otero[a], Daniel Luna[a]

[a] *Department of Health Informatics, Hospital Italiano de Buenos Aires*

Abstract

When a critical blood glucose value is detected, then it is necessary to inform, advice, communicate, act and document it in the patient´s medical record. We sought to describe the development of an alert system for glycemia values at the point of care integrated with an EHR, both in desktop and mobile devices.

Keywords:

Closed-loop communication, Point-of-care, Patient safety.

Introduction

A critical value is defined as an "indicator or result of a pathophysiological state, so far from normal that it can endanger the patient's life if it is not acted quickly, and for which it is necessary to adopt corrective measures" [1].

The emission and the reception of the critical value are equally important so the communication in the healthcare team is carried out effectively and safely [2]. The JCI and others organizations have made references to this communication, highlighting the need to notify critical results and readback of the professionals, and receiving the result; in this matter, standards and guidelines have been published. [3, 4].

When a blood glucose critical value is detected in this context, it is necessary to have a defined process to identify, advice, communicate, act and document it in the patient's clinical record [5]. EHR and CDSS represent an important tool in which patient safety oriented guidelines can be implemented throughout the organization. The Hospital Italiano de Buenos Aires (HIBA) intends to continue acquiring tools that facilitate the user experience, focusing on improving the quality of care and patient safety, adhering to current JCI recommendations, including the report of the critical value of blood glucose.

The objective of this article was to describe the design and development of a clinical decision support for documenting blood glycemia critical values at the point-of-care (POC).

Methods

The HIBA is a university hospital that covers the entire spectrum of healthcare. The institution holds a JCI certification since 2014, being re certified in 2018. Since 1998, it runs a homegrown Health Information System, that supports all the organization´s clinical and administrative operations. It has been certified by HIMSS as level 7 in the EMRAM model in 2017.

Glycemic critical value interface design and develop

For the prototyping, the usability area worked under the methodological framework “User-centered design” (UCD), adapted to the format and interface templates established in the design of EHR. The prototypes, both desktop and mobile were first designed in low fidelity reproducing the workflow and recording the critical values, and testing sessions were

performed with end users with qualitative analysis being done by members of the team. Then high fidelity prototypes were built, and again tested with users with same methodology (Figure 1).

Results

Process analysis and optimization

Some key findings were achieved in the pre-implementation process:

1. Nurses didn´t have a specific place in the EHR to register the glycemic value as critical.
2. The documentation about the communication of the critical value was registered asynchronously, most of the times after the 5 minutes stated in the policy.
3. Not all the information needed by the policy was present in the registries (e.g. physician communicated, read back, action taken).
4. Nurses had difficulties in finding the physician in charge to whom communicate the critical value.

Some modifications were suggested to address these difficulties and to comply with the organization's critical value policy, and the optimized process was plotted. The improvements consisted in having a real time documentation of the critical value communication; having a CDSS in which a number of rules were coded to interpret glycemic values and trigger the alert if a value was of range; create mandatory fields in the form to satisfy all required information by the policy; and offering the nurse information about the attending physician for the patient together with his contact number.

Also some modifications were done in the eMAR, highlighting the glycemic values recorded as critical in red for better information retrieval (Figure 1).

Interface design and development

The alert system should comply with four objectives:

5. Identify critical values of glycemia by glucometer.
6. Alert health professionals.
7. Communicate effectively and timely to the health team to provide a joint treatment.
8. Record the critical findings of blood glucose controlled at the POC.

A low fidelity mockup was built, and after user testing sessions a number of improvements were made when designing the high fidelity mockup, like adding the measured value to the pop up menu, showing specific fields like contacted physician and read back only if the option selected in the first field is affirmative, and changing the sides of the “save” and “back” buttons (Figure 2).

The final system had the following elements:

- **Input:** the trigger for the alert was the entry of the critical value in point of care glucose meter in the eMAR system.
- **Trigger:** rules were developed for the alert system to identify and alert health professionals of critical value when it is less than 40 mg/dL or greater than 500 mg/dL.
- **Output:** consisted in a pop up menu with identification of attending physician through Master Professional Index, a radio buttons for registering effective communication, and proper read back process.

Since medication administration and vital sign control are processes that are supported by our homegrown mobile eMAR application, android native interfaces were also designed and developed with the same methodology, objectives, and having the same elements as the desktop platform. Also, as both systems run in the same production environment, critical values recorded in one of them were seamlessly read in the other one.

Conclusions

This study presented the description of an iterative development with the objective of complying with the recommendations of the JCI that included, identify, alert, communicate and record the critical values of blood glucose by hemoglucotest, involving stakeholders in the validation of each of the tests. Establishing a protocol of action before critical values is a recommendable quality practice. Although there is a varied bibliography about the effective communication of these findings in areas such as clinical laboratory or diagnostic imaging, no literature was found that reports the development of blood glycemia critical values at the POC. This project only made the description of the glucose alert system in the electronic medical record, future lines of research could focus on the measurement of before and after implementation, and monitoring the tool.

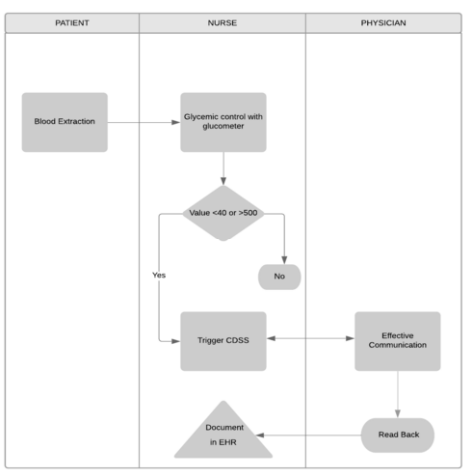


Figure 1 - Process analysis

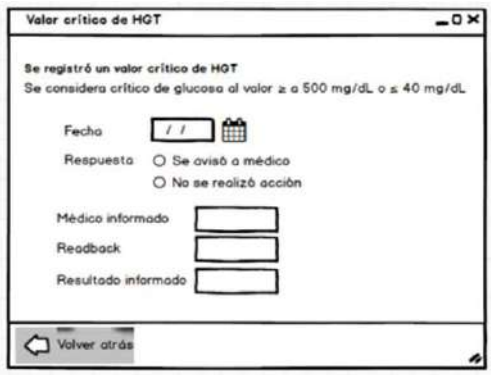


Figure 2 - Low fidelity mockup of the alert in the EHR

Figure 3 - High fidelity mockup of the alert in the EHR

References

[1] G.D. Lundberg, Critical (panic) value notification: an established laboratory practice policy (parameter), *JAMA* **263** (1990) 709.

[2] D. Hanna, P. Griswold, L.L. Leape, and D.W. Bates, Communicating Critical Test Results: Safe Practice Recommendations, *Jt Comm J Qual Patient Saf* **31** (2005) 68–80.

[3] F.F. Parl, M.F. O'Leary, A.B. Kaiser, J.M. Paulett, K. Statnikova, and E.K. Shultz, Implementation of a closed-loop reporting system for critical values and clinical communication in compliance with goals of the joint commission, *Clin Chem* **56** (2010) 417–423.

[4] J. Callen, A. Georgiou, J. Li, and J.I. Westbrook, The safety implications of missed test results for hospitalised patients: a systematic review, *BMJ Quality & Safety in Health Care* **20** (2011) 194.

[5] ScienceDirect, (n.d.). https://www.sciencedirect.com/science/article/pii/S0883944114003190 (accessed November 12, 2018).

Address for correspondence

Corresponding author: Camila Sofía Galván, Department of Health Informatics, Hospital Italiano de Buenos Aires; camila.galvan@hospitalitaliano.org.ar

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190643

Expectations in the Development of Computer Technology in Primary Care: A Multidisciplinary Delphi Study Among 23 French Experts

C. Sors[a], A. Bermes[a], J-B. Kern[a,b]

[a] *Department of General Practice, Grenoble Alpes University, Grenoble, France*
[b] *TIMC-IMAG Laboratory, CNRS UMR 5525, Grenoble Alpes University, Grenoble, France*

Abstract

The goal of this study was to determine a consensual proposition for the development of computer tools in primary care. A Delphi study using colored abaci was conducted among 23 French experts, some of whom were patients. The tools expected by the experts were: a customizable knowledge database integrated into an efficient clinical support system, a follow-up calendar designed as a collaborative patient-focused tool, and an information exchange data system.

Keywords:

Primary Health Care; Health Information Systems; Electronic Health Records

Introduction

Information and communication technologies are precious tools in dealing with present day health care challenges [1,2,3] but practical uses are often less efficient than expected. Their current use lies far below their potential and future possible applications.

In French primary care, general practitioners (GP) often use an electronic medical record but only in minor functionalities. A national e-health record is having difficulty getting off the ground. Patient portals are beginning to be deployed [4]. Health data is only partially collected using social insurance data, and just beginning to be opened with the "Système National des Données de Santé" (French health data organization) [5].

The purpose of this study is to determine a consensual proposition for expected computer tools to be used in primary care between different expert profiles. Three directions were explored: assisted systems for medical decision making, care coordination and communication, health data exploitation and the ethical issues involved.

Methods

A Delphi method associated with Regnier colored abaci [6] was conducted using an online written survey with the website *colorinsight*® (http://www.colorinsight.fr/).

Experts in medical and paramedical practical fields, care coordination, health project management, medical IT (information technology) systems, or patients' experts in health system use were recruited. Their expertise was assessed by their training and professional experience.

The first propositions were built from literature analysis supplemented by discussion with experts during recruitment. The second and third surveys were devised from analysis of the previous surveys content, to explore non-consensual points and find new ideas in view of obtaining a consensus. Finally, global answers were analysed by theme and summarized.

Surveys were given to experts with a glossary explaining technical words and the summarized personalized answers of previous survey when necessary. Experts were invited to express their opinion about each proposition by a colored vote, with an optional comment section, to enable them to justify their position. A comment was systematically requested in case of disagreement.

For each survey, the analysis was quantitative, with measurement of agreement rates, and qualitative, with the colored vote distribution analysis on the abaci and subsequent comments.

The primary objective was achieving a positive consensus, defined by an agreement rate of over 75%. The distribution of the other votes showed strength of consensus.

Expert consent was collected throughout the survey. French data protection authorities were notified (CNIL – "Commission Nationale de l'Informatique et des Libertés").

Results

Panel of experts and proceedings

Twenty-three experts, among thirty-seven originally contacted, agreed to participate. Refusals were related to lack of time for volunteer contribution. They were 8 primary health care professionals, 2 coordination professionals, 5 physicians involved in IT health projects, 3 medical IT experts and 5 patients promoted to patient experts.

Three surveys with 21, 20 and 18 propositions were given to experts between March and August 2018. Answers from respectively 23, 21 and 17 experts were collected.

Answer content

Main consensual **propositions** are presented in figure 1.

Clinical support system

The use of a knowledge database to gain access to organized information in order to support clinical decisions was consensual. It had to be integrated into professional workflow and to be personalized by using comments and personal documents.

The possibility for professionals to use a follow up calendar (including prescriptions, complementary investigations, secondary and tertiary health care feedbacks) for each patient was consensual. The calendar had to function collaboratively (between health professionals, social workers, other concerned parties and patients) and include major life events.

Data exchange

The need to consider interoperability in a global environment including health care facilities, the electronic health record, the pharmaceutical record and social facilities, was consensual. As

far as data standardization was concerned, experts suggested that standardized data was not accurate enough to be used as a means of supporting decisions and the only consensus was around consultation assessment.

Health data access by the patient and the possibility to exchange and share the data with other health professionals in a secured, structured way was a consensual point. Data exchange using a secured mailbox was consensual as was that he different data exchange methods be centralized around a single online access. Experts expressed the need for health professionals to manage electronic exchange in order to limit legal risk and preserve data treatment capacity.

Other communication systems seemed consensual: online prescription transfer to a designated pharmacist, an instant mailbox to communicate between health professionals and access to electronic exchange from the patient's electronic record in work IT software.

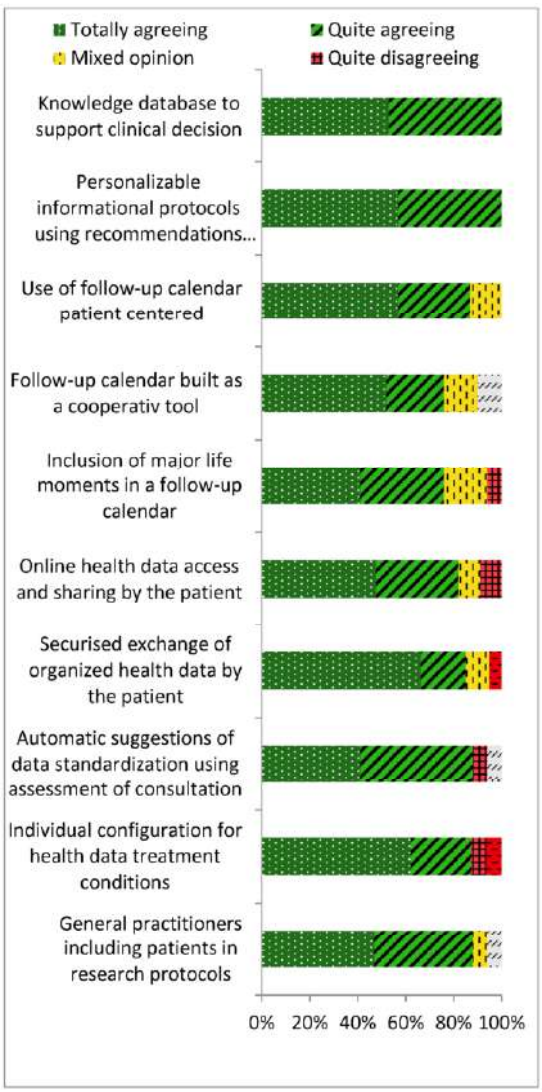


Figure 1 - Main results of Delphi rounds

Data treatment

Health data utilization and regulation was not agreed upon. Use of the data for research purposes and to improve health quality was not consensual. Some experts wanted a regulated control of data anonymization and a limitation of their utilization to pre-selected projects. Other experts were in favor of free access to the data to foster further research. Having a separate organization handle health data use was not consensual. Analysis of the data was described as difficult, because of disparity in electronic health record quality and in data collection.

The necessity for each patient to give his/her consent to the use of his/her health data was consensual.

That GP's should be able to offer the research protocol to a chosen patient was consensual.

Conclusions

This study shows common expectations for computer tools developing patient focused work. It highlights the need for a single and sharable medical record, thought around the patient, adapted to professional health needs and including training and communication features. Application of the consensual propositions of this study could have a major impact on primary health care organization. It should foster further research and question French national policies for e-health development.

Acknowledgements

The authors would like to thank Pr Moreau Gaudry MD PhD, Pr Letrilliart Laurent MD PhD and Dr Gaboreau Yoann MD-for their review and assessment of this study. The authors declare that they have no conflict of interest.

References

[1] C.S. Kruse, and A. Beane, Health Information Technology Continues to Show Positive Effect on Medical Outcomes: Systematic Review, J. *Med. Internet Res* **20** (2018).

[2] A.X. Garg, N.K.J. Adhikari, H. McDonald, M.P. Rosas Arllano, P.J. Devereaux, J. Beyene, J. Sam, and R.B. Haynes, Effects of computerized clinical decision support systems on practitioner performance and patient outcomes:a systematic review, *JAMA* **293** (2005) 1223–1238.

[3] J. Baek, and R.L. Seidman, Impact of Information Technology,Clinical Resource Constraints, and Patient-Centered Practice Characteristics on Quality of Care, *Health Serv. Res. Manag. Epidemiol* **2** (2015).

[4] J. Sauret, Economie des nouveaux systèmes d'information en santé, *Les tribunes de la santé* 2010/4(n°29): 59-68

[5] Mise en oeuvre du système national des données de santé et nouveau cadre d'accès aux données de santé - Ministère des Solidarités et de la Santé. Ministère des solidarités et de la Santé; 2016.

[6] P. Chapuy, Y. Louineau, and P. Mirenowicz, La méthode Delphi-Abaque. Un exemple d'application: la prospective du bruit, Futuribles (1990).

Address for correspondence

Dr Claire Sors MD : claire@sors.net. Dr Arthur Bermes MD arthur.bermes@gmail.com. Dr Jean-Baptiste Kern MD : jean-baptiste.kern@univ-grenoble-alpes.fr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190644

Electronic Image Documentation of Patient Reported Outcomes Using Mobile Technologies

Iñaki Soto-Rey[a], Tobias Hardt[a], Luca Hollenberg[a], Philipp Bruland[a], Sonja Ständer[b], Martin Dugas[a], Michael Storck[a]

[a] *Institute of Medical Informatics, University of Münster, Münster, Germany*
[b] *Center for Chronic Pruritus and Department of Dermatology, University Hospital Münster, Germany*

Abstract

Patient Reported Outcomes (PROs) provide essential clinical data for the diagnosis and treatment of patients. Mobile technologies enable rapid and structured collection of PROs with a high usability. MoPat is an electronic PRO system developed at the Münster University that enables patients to complete PROs in multiple languages. This research reports the further development of MoPat and the inclusion of features to document images electronically that will be evaluated in a multi-site clinical research.

Keywords:

Patient-reported Outcomes, Mobile Health, Electronic documentation.

Introduction

Patient-reported outcomes (PROs) are information provided by patients without any interpretation by a clinican or anyone else. Electronic systems and specially mobile devices provide a fantastic opportunity for the collection of these. In recent years, initiatives such as ResearchKit [1] have modified the way how PROs were being collected and also the kind of data that can be documented including new question types such as reaction time and tapping speed among others.

The Institute of Medical Informatics in Münster (Germany) has developed its own electronic patient-reported outcome system (ePRO): the Mobile Patient Survey (MoPat) [2]. MoPat is an interoperable and multi-lingual web-based ePRO that enables clinical data collection in clinical settings and at home.

The Translational Pruritus Research Group[1] aims to identify clinical and therapeutical relevant mediators for chronic itch, as well as related anatomical and functional neuronal changes. To achieve these aims, patient data regarding the regions where pruritus and pain appear need to be collected. A similar functionality is required by the McGill-Pain index PRO [3], where patients also need to mark painful regions on the image of a human body.

The objective of this research is to integrate an applicable way of reporting parts and individual marks in arbitrary images within the MoPat ePRO.

Methods

Before starting the implementation of the new features, it was necessary to perform a system analysis in order to define the software architecture of the new features and plan their integration into the existing code. At the same time, the development team identifyed available libraries that could be re-used to fulfill the requirements.

The major updates needed were the following: on the one hand, it was necessary to create and save the question data such as the images and the predefined body parts. On the other hand, it was crucial to display the new question types in a user-friendly way within a survey.

The image marker question uses an HTML5 <img>-tag that allows setting base64 strings as source of the image. This tag is embedded in a canvas element that facilitates the drawing of marks. The marks are stored in a list together with an additional pointer that indicates which marks should be displayed.

The body part question type was implemented using a class that contains an enum 'BodyPart', which defines the selectable regions. This class includes the multilingual message code, the path element, which defines the body region to be selected, and the enum ImageType that links the body part to the corresponding image. The path element is represented by an XML string that defines the region inside of a scalable vector graphics image.

The images in both question types are transmitted in a data transfer object to the client as base64 encoded strings. The new functionality was internally tested using automatic (JUnit) and semi-automatic tests.

Results

Image marker

The image marker question type enables MoPat administrators to upload arbitrary images to the MoPat server. When an image marker question is created, the file type, the dimensions and the resolution of the uploaded image are validated. If the size is unfitting or the resolution too low, warning messages are displayed. During a patient survey, image marker questions show the uploaded images together with a brushtool to select the color of the marks and the buttons to revert and redo changes (see Figure 1). Patients can then tap (or click) on the screen and draw "X" marks in black or white on it.

As MoPat enables answering questionnaires offline, the images had to be transfered at the beginning of the survey to the client without reloading it from the server, which explains why base64 encoding was chosen.

[1] www.prusearch.net

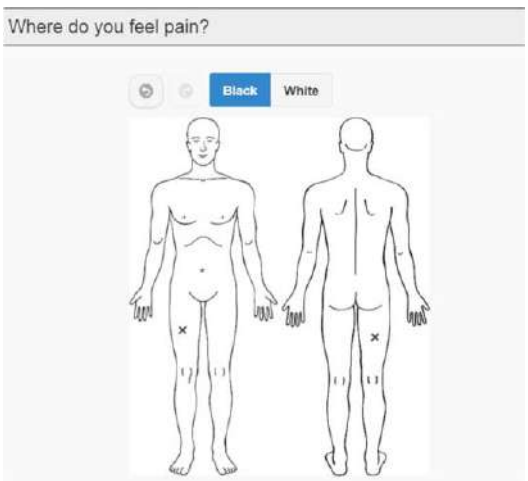


Figure 1– Snapshot of an image marker question with two marks on the right front and right back thigh

The marks included by the patients are stored as a percentage of the image width and height to ensure correct positioning if the survey is stopped in one device and restarted on another one with a different screen size or resolution.

The image marker question type was informally evaluated including 15 testers who completed a test script and filled out the System Usability Score questionnaire reaching an average score of 90.

Body Part

The body part question type enables MoPat administrators to include a predefined body part question and its selectable regions in a MoPat questionnaire. The administrator decides which body parts can be selected and these are displayed in light blue for the surveyed patient. When a body part is selected, the part's color is displayed in dark blue (see Figure 2).

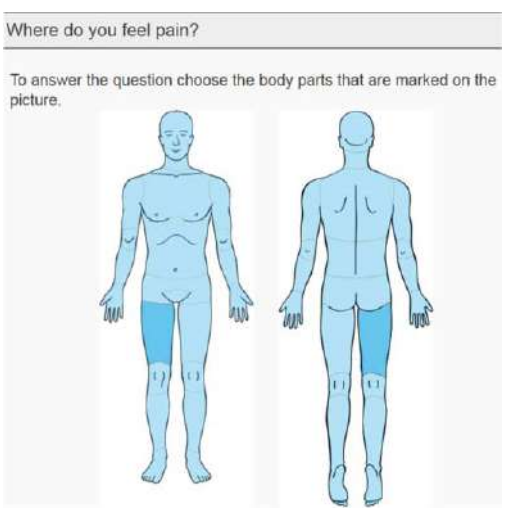


Figure 2– Snapshot of a body part question with two body parts selected (displayed in dark blue)

The implementation of body part question answering is similar to the multiple choice questions' one. The selectable regions are treated like multiple choice answers and can be exported as a nominal variable.

Discussion

The question types "Image Marker" and "Body Part" have been sucessfully implemented and tested within MoPat. This functionality will be evaluated in the context of a large multi site clinical study, the Translational Pruritus Research Group.

These two new question types entail some limitations: The image marker coordinates (where marks are included by patients within image marker questions) are not exported; however, they are being stored. This means that it would be straightforward to implement a coordinate export if required. The image marker question brushtool is limited to X marks and black and white colors. The colors were initially restricted in order to avoid possible categorisation of the marks, e.g. green marks are positive and red ones are negative. The body part question images and the parts that can be selected are hardcoded and can only be modifyed within MoPat's source code. The images are transmitted in base64 format, which represents a considerable amount of data. This could be an issue when mobile devices use network data.

The development of the image marker and the body part questions is an ongoing task. A new version of the image marker module should include a more complete brushtool enabling drawing of circles or elipses as well as a wider variety of colors. Likewise, the body part question type could be enhanced with the possibility to include images and define body parts within the administrators graphical user interface.

Conclusions

MoPat has been enhanced with two new question types: an image marker and a body part selector. These types will be tested and evaluated in the context of a multi site clinical study. The new question types will be further developed to improve their usability.

Acknowledgements

The authors declare that there is no conflict of interest regarding the publication of this article.

References

[1] Jardine J, Fisher J, Carrick B (2015) Apple's ResearchKit: smart data collection for the smartphone era? *J R Soc Med* 108, 294–296.

[2] Soto-Rey I, Dugas M, Storck M (2016) Implementation of an ODM and HL7 Compliant Electronic Patient-Reported Outcome System. *Stud Health Technol Inform* 228, 421–425.

[3] Melzack R (1975) The McGill Pain Questionnaire: major properties and scoring methods. *Pain* 1, 277–299.

Address for correspondence

Dr. rer. medic. Iñaki Soto Rey. Institute of Medical Informatics
Albert-Schweitzer-Campus 1, A11. 48149 Münster, Germany.
Email: Inaki.SotoRey@uni-muenster.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190645

Configuration of Input Forms in EHR Systems Using Spreadsheets, openEHR Archetypes and Templates

Erik Sundvall[a,b], Annika Terner[c], Helen Broberg[d], Carrick Gillespie[e]

[a]Department of Biomedical Engineering, Linköping University, Linköping, Sweden, [b]Region Östergötland, Linköping, Sweden
[c]Department of Public and Caring Sciences, Uppsala University, Uppsala, Sweden
[d] Department of Digitalization and IT, Region Skåne, Lund, Sweden
[e] Omniq AB, Stockholm, Sweden

Abstract

Differences in structure and semantics of data captured using screen forms in different Electronic Health Record (EHR) products and configurations is the root cause of many interoperability problems. We present a workaround enabling reuse of openEHR archetype and template semantics to configure forms in four surveyed, insufficiently standardized, EHR-products used in Sweden (Cerner Melior, Cerner Millenium, Cambio Cosmic and CGM TakeCare). Data from EHRs was then exported and queried using standardized query mechanisms.

Keywords:
electronic health records, semantics, medical record linkage

Introduction and Background

EHR systems provide both hardcoded non-configurable structures and other customer configurable parts. A significant part of EHR configuration is the creation of custom input forms, sometimes called templates, which consist of headings. However, the predefined headings are only to a small extent shared between the professional groups [1] within a healthcare provider or shared between providers [2]. This variation makes it hard or impossible to share structured information in an interoperable way. If the information is too different then it cannot be safely converted/mapped to structured exchange formats. Instead unstructured versions of exchange formats or telefax-transmission of printouts are often used, combined with time consuming reinterpretation and restructuring of the information by (already too busy) healthcare staff in order to fit into the receiving organization's IT-system and routines.

Standards targeted for EHR content standardization such as openEHR, HL7 CIMI, EN ISO 13606 combined with EN ISO 13940 can, if wisely used, contribute to support similar data capture at the source (data entry) making conversion processes unnecessary, simpler, or at least possible. In Sweden the implemented support for such standards in present and recently acquired EHR systems is limited. Customer-configurable forms do however exist.

The idea of using openEHR archetypes to configure input forms in a proprietary system by making limited modifications to the EHR System software has been described by Chen et.al. [3], and would likely have been useful if it had been implemented beyond research pilot. The work described here builds on similar ideas, but only uses (the more limited) mechanisms available in the studied current systems without vendor modifications of the software.

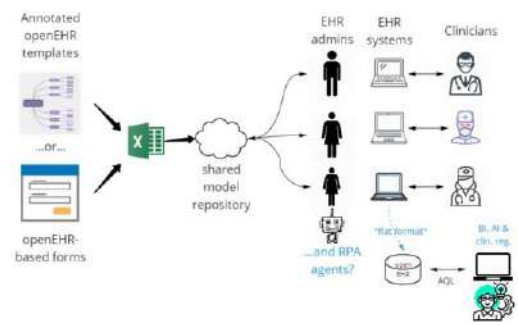


Figure 1 – Usage Context of Suggested Workaround

Several regions have signaled interest in using shared archetypes and openEHR-templates as building blocks when creating and maintaining configurable EHR input forms also in non-openEHR systems. That is the path explored here.

Methods

We asked what methods different healthcare regions employ when formalizing requirements and configuring EHR systems. Excel spreadsheets or documents with tables were received from six of the 21 Swedish healthcare regions. A new combined spreadsheet was compiled to fit all studied vendors' main datatypes for configurable input forms, see *Table 1*. A smoking-related spreadsheet example can be found in the "ehr-form-config" directory of https://github.com/modellbibliotek/standin/

A detailed description of openEHR-based modelling is out of scope, but a brief summary is that an openEHR template for a specific use-case can be constructed by joining and configuring/constraining and combining several (possibly repeated) archetypes (and terminology subsets/lists). Any node in such a template has a unique archetype based "path" and can also carry arbitrary "annotations" (key+value-pairs). We used annotations to mark which template nodes to expose in non-openEHR systems' forms and at what "indentation" level. The annotations could also allow automation of the (now manually executed) algorithm to convert from annotated template to the suggested spreadsheet format or ideally directly to control configuration APIs or interfaces. As part of the algorithm the paths of the used nodes are also written into corresponding spreadsheet rows (to enable later storage in the EHR system's or EHR-data export toolchain's data or metadata). The ID of the openEHR-template is also written into the spreadsheet.

Table 1 – A simplified view of the main datatypes found to be used for configuration in the four studied systems. In addition to the described datatypes the spreadsheet file, where applicable, contains descriptive help text intended to be easily available upon user request in form systems. Also columns relating the rows to openEHR templates and "paths" based on openEHR-archetypes were added to support openEHR based post-processing and potential querying of data. HL7 FHIR mapping guidance can also be manually added to the spreadsheet if available. A header level "depth" column is used for systems that support headings in hierarchical levels.

Datatype/ System	Free text	Date & Time	Numeric value	Calculation	Selection list	Coded term	Link
Cambio Cosmic	Yes	Yes, §F	Yes, §F, 1…2	Yes, §F	Yes, §F, +N, 2	Yes, 1…*	Yes
Cerner Melior	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cerner Millenium	Yes	Yes	Yes, +F	Yes	Yes, +F, 1S	Yes, 1...*	Yes
CGM TakeCare	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Abbreviation legend: +F = Free text field for comments can be added (configurable); §F = Free text field is always available for comment (not possible to turn off in configuration); +N = Numeric value can be added to the respective choices and can be used for calculation, e.g. for assessment instruments.; 2 = The number of choices in the selection list can be delimited to two choices by single selection, e.g. when using checkboxes for 'Yes' and 'No', respectively; 1...* = One to several code systems can be made available, e.g. both ICD-10 and ATC.; 1...2 = One or two numeric values can be specified in the same datatype. E.g. for blood pressure, two numerical values can be specified; 1S = If a special choice (in an otherwise multiple-choice type of) selection list is selected, then no more choices from the list can be made, e.g. if you choose 'No diseases' in a selection list that in addition to the choice 'No diseases' contains a selection of diseases.

Results

The combined spreadsheet format has seven different main datatypes, see Table 1. *Free text* implies narrative text without any additional structure. *Date & Time* (possibly partial) dates and times. *Numeric value* includes the value itself, unit, maximum value, minimum value and reference values. For *calculation*, a formula is given, combined with unit, maximum value, minimum value and reference values of the calculated value. For the *selection list*, the choices available to be selected are defined and whether single or multiple choice is allowed. *Coded term* carries codes that can be numeric or alphanumeric, and a meaning/display label of each code in a plain text. *Link* references other information, within or outside the EHR system

The spreadsheet method is suggested to be used as a way to "translate" annotated openEHR templates (or forms) based on archetypes to a form that resembles what current EHR system administrators are familiar with and using today. The configuration of the templates, in current EHR systems that lack an import mechanism, then has to be made manually based on the spreadsheets (as usually done today) or potentially assisted by RPA (Robotic Process Automation) tool scripts.

If/when openEHR modelling tools become more familiar in the regions and more automation is desired, then the spreadsheet step could be skipped and either annotated openEHR templates or form definitions could be used for configuration information. Templates were currently chosen instead of the form definition format already used by a couple of openEHR products, since there is not yet an official openEHR standard for that format. When standardized, it could potentially be used instead of (or as a complement to) annotated templates.

Figure 1 illustrates the workflow; templates of shared interest are collaboratively created and annotated (using normal openEHR toolchains) to fit use-cases and then algorithmically converted to the suggested spreadsheet format and uploaded to a shared repository (e.g. on GitHub) When the EHR administrators configure forms based on the spreadsheet they make sure to also carry over information regarding which openEHR template-ID the form is based on into the EHR. Information about which archetype-based path each spreadsheet row is associated with also needs to be stored in the EHR configuration or some other place that export mechanisms later can use to re-associate the EHR data with the path. Exact association mechanism may vary between systems, shorter Dewey-compressed [4] or hashed paths an also be used instead of full paths in EHR data.

The data in the (non-openEHR) EHR systems will not need to contain the full openEHR tree structures. As long as the openEHR-template-IDs and paths can be reconstructed during data export then e.g. an openEHR "flat" JSON format (https://github.com/ethercis/ethercis/blob/master/doc/flat%20json.md) can be used as import mechanism into an openEHR-compliant system and for example be queried using AQL [5].

The whole approach was tested by using an openEHR-template-derived spreadsheet to configure a CGM TakeCare EHR system. Export mappings to openEHR based "flat format" were then easily created and used to export data to an openEHR based open source EtherCIS EHR. The data (originating from the TakeCare system) could then be queried in the EtherCIS system using AQL (Archetype Query Language).

Conclusion

The described "spreadsheet" approach will of course not make existing non-standardized EHR content any easier to share in a structured way (no matter if it is user configured content or non-configurable content designed by the vendor). Organizations with an existing operational EHR installation will likely not suddenly switch all configurable EHR-entry forms from non-standardized to standardized, but can do it incrementally combined with other process and information change management and maintenance work.

References

[1] Terner, A., Lindstedt, H., & Sonnander, K. (2012). Predefined headings in a multiprofessional record system. J Am Med Inform Assoc 2012;19:1032-1038.

[2] Halilovic, A., & Terner, A. (2018). The Extent of Shared Headings in Electronic Health Records in Swedish County Councils and Regions. Journal of Medical Imaging and Health Informatics. Vol. 8, 1–5, 2018

[3] R Chen, GO Klein, E Sundvall, D Karlsson and H Åhlfeldt, Archetype-based conversion of EHR content models: pilot experience with a regional EHR system, *BMC medical informatics and decision making 9 (1), 33*

[4] E Sundvall, F Wei-Kleiner, SM Freire and P Lambrix, Querying archetype-based electronic health records using Hadoop and Dewey encoding of openEHR models, *Studies in health technology and informatics 235, 406-410*

[5] Archetype Query Language (AQL) specification https://www.openehr.org/releases/QUERY/latest/docs/AQL/AQL.html

Address for correspondence

Erik Sundvall, erik.sundvall@liu.se

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190646

Relationship Between Very Cold Outside Weather and Surgical Outcome: Integrating Shallow and Deep Artificial Neural Nets

Ahmad P. Tafti[a], Yue Dong[b], Elizabeth Habermann[a], Hongfang Liu[a], Vitaly Herasevich[b]

[a] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA,
[b] Department of Anesthesiology, Mayo Clinic, Rochester, MN, USA

Abstract

Patients' hospital length of stay (LOS) as a surgical outcome is important indicator of quality of care. We used EMR data to build artificial neural network models to better understand the impact of cold weather on outcome of first surgeries in a day in comparison to a matched cohort receiving surgical treatment in warm days. We found that LOS for first-in-a-day cardiac and orthopedic surgical cases are longer in very cold days.

Keywords:

Length of Stay, Machine Learning, Association Learning.

Introduction

High-quality surgical care is of paramount importance to the public health community as the volume of surgery has remarkably increased all across the world. Successive assessment of surgical outcomes (e.g., LOS, readmission rates, functional health status, patient satisfaction) improves the efficacy of hospitalization, keeping LOS to a minimum, and helps health care providers build effective programs to patient care [1; 5]. Baseline heath status, condition lead to surgical intervention and procedural factors that can impact surgical outcome; however, no knowledge exists to study exterior factors, such as weather conditions impacting the hospital LOS of surgical patients. The availability of large-scale multi-modal data along with artificial intelligence algorithms make it possible to estimate association of different factors in measuring surgical outcome. In the present study, we develop machine learning-based, data-driven models to analyze the impact of cold and freezing outside weather on LOS after surgery. Our hypothesis is that freezing outside weather can affect the first surgical case in a day as operating room (OR) staff could have cold hands and be rushing to work. We compare hospital LOS after similar surgeries in freezing and summer days. Our main contributions are three-fold: (1) Inspired by shallow and deep artificial neural networks, we integrate daily temperature data with surgical data, and begin to understand how does the cold weather affect LOS for a cohort exposed to cold weather in comparison to a matched, non-exposed to cold weather cohort; (2) We turn a descriptive analysis problem into a classification pipeline, and develop a deep-stacked, contractive auto-encoder to learn useful data representation and extract relevant features from the combined clinical and weather data; and (3) The present study is expected to open several research avenues to incorporate more data and more multi-modal data to inquiry of surgical outcome.

Methods

The present study was reviewed and approved by the Mayo Clinic institutional review board. The Mayo Clinic OR Datamart was used to identify first-in-a-day surgical cases between January 2005 and December 2017. All patients admitted during the study period to the OR in the Rochester campus of the Mayo Clinic were eligible for inclusion. The Rochester (MN 55902) temperature data was obtained using a publicly available data source, namely NOAA, the National Oceanic and Atmospheric Administration [2]. To ensure consistency, we cross-checked the temperature data using an alternative source, called Weather Underground [4]. Our proposed workflow and the software tiers we developed are shown in Figure 1. After data cleaning, we had 132,095 unique surgical cases to perform the analysis. 45 data fields across 4 different categories (e.g., patient data, surgical data, intervention data, weather data) were used. Although the temperature data were originally obtained in Fahrenheit in the format of decimal numbers, an additional field namely "Temperature_Ordinal" was added to the dataset, in which to also account for a categorical representation of temperature. We employed the facts demonstrated in [3; 7] to turn temperature data from continous to ordinal values, as follows: {Temperature $< 14\ ^{o}F$ ➔ '*Very Cold*'**;** $14\ ^{o}F \leq$ Temperature $< 50\ ^{o}F$ ➔ '*Cold*'**;** $50\ ^{o}F \leq$ Temperature $< 68\ ^{o}F$ ➔ '*Mild*'**;** $68\ ^{o}F \leq$ Temperature $< 85\ ^{o}F$ ➔ '*Normal*'**;** Temperature $\geq 85\ ^{o}F$ ➔ 'Warm'}. In regard to LOS, the minimum and maximum values among all surgical cases were 0 and 31 days, respectively. To turn LOS to a nominal value and have it as the target class, median and mean of daily basis LOS were separately calculated for each surgery procedure. We thus defined short-term, mid-term, and long-term LOS for every surgery procedure as: {LOS < LOS_median ➔ '*Short-Term*'**;** LOS_median ≤ LOS ≤ LOS_mean ➔ '*Mid-Term*'**;** LOS > LOS_mean ➔ '*Long-Term*'}. All surgical data records have been combined with temperature data (decimal and ordinal) and stored in a single data repository, called Aggregated Dataset (see Figure 1). Once we had all data records and variables (*X* in Figure 1) available, we fed the deep-stacked, contractive auto-encoder with the data (each record represented as a vector X), in order to better capture the local relationships and directions of variables dictated by the surgical data records. The stacked contractive auto-encoder [6] is an unsupervised learning strategy that automatically discovers useful data representation models out of the given data. As a mathematical function, it maps an input vector $X \in R^{dx}$ to a hidden representation $h(x) \in R^{dh}$, as: $h = f(x) = S_f(W_x + b_h)$, where S_f is a non-linear activation function, and it is parameterized through a $d_h \times d_x$ weight matrix W, and a bias vector $b_h \in R^{dh}$, and decoder function will map hidden

representation h to a reconstruction y as: $y = g(h) = S_g(W'_h + b_y)$. The contractive auto-encoder training tries to find parameters $\Theta = \{W, b_h, b_y\}$ that minimize the reconstruction error on a training set of instances D_n as follows [6]:

$$\delta(\Theta) = \sum_{x \in D_n} (L\left(x, g\left(f(x)\right)\right) + \lambda ||J_f(x)||_F^2) \quad (1)$$

Having LOS nominal variable as the target class [short-term, mid-term, long-term], we then utilized three different classifiers, namely Neural Network (NN), Decision Tree, and Random Forest (RF), to learn how relevant features could be associated to a class. We split all data records into different segments to first train the classifiers, and then validate and test how those classifiers are accurate. If the F-measure and ROC area we obtained using the classifiers plus those relevant variables that automatically extracted through stacked contractive auto-encoder were promising, then we took association rules with the help of Apriori algorithm to extract patterns and rules across the data. Otherwise, we were back to the stacked contractive auto-encoder and re-configured the internal parameters (e.g., penalty term for the reconstruction cost function, number of epochs) to iterate the process.

Results

We used 50% of the data records to train the contractive auto-encoder, 25% to validate, and 25% to test. For classification purposes, we used 5-fold cross validation to evaluate the models. We ran into learning curves and calculated the F-Measure and ROC area to see the performance of the supervised method (shallow nets, RF, and Decision Tree), and then iteratively tuned and refined the entire parameters (e.g., penalty term) of the feature extraction component, the deep-stacked, contractive auto-encoder. The best F-Measure and ROC area were achived by NN, with F-Measure of 0.831 and ROC area of 0.847. Table 1 presents some temperature-related rules that automatically extracted using the system. The current study showed some relationship between cold weather and LOS for cardiac and orthopedic surgeries. We are 95% confident that the cold and very cold day cardiac surgeries which classified as long-term LOS have fallen between 68.59% and 71.41%. We are 95% confident that the mean LOS difference between cold and very cold versus not cold and very cold day cardiac surgeries is between 2 and 2.2 days, and we are also 95% confident that the mean LOS difference between Long-Term LOS versus not Long-Term LOS for cardiac surgeries is between 6.5 and 6.7 days. For orthopedic surgeries, we are 95% confident that the cold and very cold day orthopedic surgeries which classified as long-term LOS have fallen between 71.37% and 74.63%. We are 95% confident that the mean LOS difference between cold and very cold versus not cold and very cold day orthopedic surgeries is between 1.7 and 2.3 days. We are also 95% confident that the mean LOS difference between Long-Term LOS versus not Long-Term LOS for orthopedic surgeries is between 6.5 and 7.0 days. Our study did not find any relationship in data between cold weather and some surgeries, including gynecology, plastic, general, and urology.

Discussion

In this data mining study, we found associations between LOS patients undergo surgeries, including cardiac and orthopedic as first case in a day and very cold outside weather. Some previous studies were done studying association between hot climate and perioperative outcome in elderly patients. But this study is first to investigate the impact of very cold weather. Besides the large number of surgical cases and 12-year observation period, our study has limitation as this is association study affected by number of biases.

Conclusions

We found that LOS for first-in-a-day cardiac and orthopedic surgical cases were statistically significantly longer in very cold days. In this study, we only matched type of surgeries. This was designed as an exploratory study that required more fundamental epidemiological approaches to eliminate biases and confounders. Traditional epidemiological research is required to support this conclusion behind simple association.

Table 1 – Temperature-Related Patterns and Rules, Automatically Extracted Using the Proposed System

Patterns/Rules	Support [Confidence]
Procedure = "Cardiac" AND {Temp_nominal ="Cold" OR Temp_nominal="Very Cold"} ➔ LOS_Nominal = Long-Term	68.59% - 71.41% [95%]
Procedure = "Orthopedic" AND {Temp_nominal ="Cold" OR Temp_nominal="Very Cold"} ➔ LOS_Nominal = Long-Term	71.37% - 74.63% [95%]

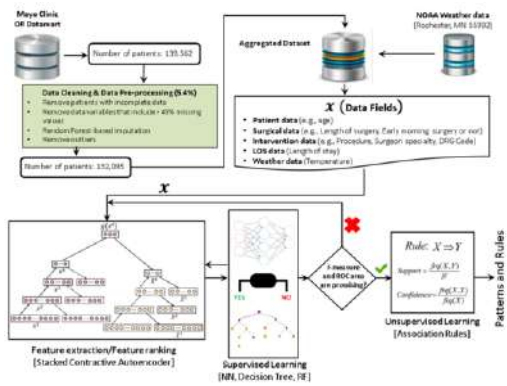


Figure 1 – The Proposed Workflow and Software Tiers

References

[1] NQF-endorsed measures for surgical procedures, 2017.
[2] National Oceanic and Atmospheric Administration, 2018. https://www.noaa.gov/.
[3] Think Metric, 2018. https://thinkmetric.org.uk/
[4] Weather Underground, 2018. https://www.wunderground.com/
[5] J.B. Dimick, J.D. Birkmeyer, and G.R. Upchurch, Measuring surgical quality: what's the role of provider volume?, *World Journal of Surgery* **29** (2005), 1217-1221.
[6] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio, Contractive auto-encoders: Explicit invariance during feature extraction, in: *Proceedings of the 28th International Conference on International Conference on Machine Learning*, Omnipress, 2011, pp. 833-840.
[7] B. Taylor, *Guide for the use of the International System of Units (SI): the metric system*, DIANE Publishing, 1995.

Address for correspondence

Ahmad P. Tafti, Vitaly Herasevich. tafti.ahmad@mayo.edu, vitaly@mayo.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190647

Integrating Heterogeneous Data Sources for Cross-Institutional Data Sharing: Requirements Elicitation and Management in SMITH

Kais Tahar[ac], Christoph Müller[b], Andreas Dürschmid[a] , Silke Haferkamp[b], Kutaiba Saleh[c], Patrick Jürs[d], Sebastian Stäubert[e], Jan Erik Gewehr[d], Sven Zenker[f], Danny Ammon[c], and Thomas Wendt[a]

[a] *University Hospital Leipzig, Data Integration Center, Leipzig, Germany*
[b] *University Hospital RTWH Aachen, Data Integration Center, Aachen, Germany*
[c] *University Hospital Jena, Data Integration Center, Jena, Germany*
[d] *University Medical Center Hamburg-Eppendorf, Research-IT, Hamburg, Germany*
[e] *Institute for Medical Informatics, Statistics and Epidemiology, Leipzig University, Germany*
[f] *Dpt. Of Anesthesiology & Intensive Care Medicine, University Hospital Bonn, Bonn, Germany*

Abstract

The digitization of health records and cross-institutional data sharing is a necessary precondition to improve clinical research and patient care. The SMITH project unites several university hospitals and medical faculties in order to provide medical informatics solutions for health data integration and cross-institutional communication. In this paper, we focus on requirements elicitation and management for extracting clinical data from heterogeneous subsystems and data integration based on eHealth standards such as HL7 FHIR and IHE profiles.

Keywords:
Data sharing , engineering

Introduction

The SMITH project [1], along with three other consortia, is part of the Medical Informatics Initiative (MII), funded by the German government. The aim of MII is to improve patient-oriented clinical research through optimized intra- and inter-institutional data exchange [2]. Therefore, a core dataset [3] and innovative medical informatics solutions for health data integration and cross-institutional data sharing are developed within the scope of MII. The SMITH consortium unites several university hospitals and medical faculties as well as industry partners in order to implement Data Integration Centers (DICs) at each hospital. These DICs are responsible for health data sharing and interoperability issues in the MII. Essentially, the implementation of such a trans-institutional health information system that enables the integration of heterogeneous data from various local hospital information systems (HISs) is a very complex and challenging task. In this paper, we focus on requirements elicitation and management for implementing this challenge.

Methods

We have adopted a generic approach that enables a collaborative requirement elicitation and a continuous refinement of the user requirements. Essentially, there were two conditions, among others, guiding the method selection for requirements documentation. 1) The team members belong to different institutions and have different levels of experience in requirements engineering. Consequently, they adopt different perspectives for requirement elicitation. 2) There was no specific tool for collaborative requirement engineering available and familiar to all team members. Hence, we decided to use a widespread word processor tool to create our technical requirements documents (TRDs). To support collaboration on the documents, we used SharePoint as a tracking system and repository. We additionally applied the method of Balzert [4]. This method helped us to structure specific requirements for heterogeneous data sources in the participating hospitals. Moreover, we modified this method by adding and linking local TRDs to a generic document that specify the core technical requirements (core TRD). Furthermore, conceptual models that describe the architectural requirements were included. They had been developed based on a DIC reference model [1] and the local TRDs, which take into account the local IT architecture and the resulting requirements. Finally, we adopted an ontology-based specification of the data requirements using Excel spreadsheets. To enable semantic interoperability, standard ontologies such as SNOMED CT and LOINC were used to code the data items required to implement SMITH use cases.

Results

We analyzed all functional requirements for data integration and validated them in cooperation with the participating hospitals. The resulting requirements of core business processes (BPs) were specified in the core TRD. Table 1 illustrates the core BPs. The code WP1-Core:BP50, for example, identifies UC data extraction as a generic BP requirement. The first three characters indicate the work package (here WP1). Table 2 illustrates an example that provides a generic specification for UC data extraction. Analogous to this example, we also defined requirements for all other core BPs. The implementation of this core process in DIC Leipzig differs from other locations. At the University Hospital Leipzig, i-Solutions LabCentre is installed as a laboratory information system whereas at the University hospital Jena Roche Swisslab is installed. For this reason, we also defined sub BP requirements specific to the local HISs. These

requirements were described in the separate local TRDs, which were linked to the core TRD. The TRD specific to DIC Leipzig contains, for example, five sub BP requirements that provide a detailed specification for the data extraction pipeline in DIC Leipzig. Each sub BP defines the relevant requirements for extracting UC data from different data sources such as LabCentre and i.s.h.med. I.s.h.med is a clinical documentation system provided by Cerner. To meet the requirement described in Table 2, a LabCentre interface has to be implemented. This interface extracts the UC data from LabCentre and transfers them to the data integration engine (DIE) using HL7 v2 [5]. The DIE is responsible for transforming the extracted data into interoperable standards such as HL7 FHIR [5] and CDA templates [5]. Finally, the Health Data Storage (HDS) stores the generated CDA documents in an IHE XDS [6] repository and loads the created FHIR resources into an FHIR server for discrete data.

Table 1– Core business processes

BP-Code	Core business process
WP1-Core:BP10	Extract calibration data for the UC tools and make it available for users
WP1-Core:BP20	Configure the DIE and activate DIC connectors to IHE-based components
WP1-Core:BP30	Configure the interface of DIE to clinical subsystems and establish connections to data sources
WP1-Core:BP40	Extract the MII core dataset
WP1-Core:BP50	Extract the UC data
WP2-Core:BP60	Transform extracted data into eHealth standards and integrate it in DIC HDS
WP9-Core:BP70	Supply UC data via standardized external interfaces such as IHE-profiles [6]

Table 2– An example of core business process requirements

/WP1-Core:BP50/ Core business process: extract the UC data	
Purpose	UC data can be extracted from clinical subsystem
Precondition	The DIE is ready for use and a connector to subsystem interface is established. The UC data are available in the subsystem to the extent necessary for the UC tools
Postcondition	UC data was successfully extracted from clinical subsystem and transmitted to DIE via a standard HL7 interface
Description	A connection to subsystem data source has to be established (see WP1-core:PB30) in order to extract the necessary data. The user can restrict the data quality and volume by certain parameters such as the number of patients and date. The user must also be able to sample the required data in a given time period (for example every 20s).

Once the data transformation and integration is finished, the next business process is data sharing. This core process requires IHE based data sharing services that connect all SMITH DIC information systems in order to provide access to anonymized clinical data stored in the local HDS.

Furthermore, we interviewed healthcare IT experts of the participating hospitals to get information about the primary data and to specify UC data requirements such as communication interfaces, source clinical subsystems, local IDs, local coding systems, and sampling rate. As a result, we created a spreadsheet file that combines all data items required to implement SMITH use cases. This file contains one spreadsheet for every partner hospital. To enable semantic interoperability, all data items specified in these spreadsheets were coded using standard ontologies such as SNOMED CT and LOINC. These spreadsheets will be used as a base for a metadata repository that represents all the data used and provided in SMITH.

Conclusion

We have adopted a generic requirement engineering method for supporting collaborative requirement elicitation and management in the SMITH consortium. This approach provided in our project a generic solution for aligning various requirements specific to different local hospitals. The resulting TRDs provide the basis for planning the system tests and future evaluations. Essentially, the resulting requirements are used to identify dependencies and to support the project management task.

In addition, the resulting spreadsheets provide a detailed specification of data items that should be extracted from local electronic medical records. Hence, our future work will include the development of corresponding FHIR profiles and the transformation of extracted UC data into FHIR resources.

Acknowledgment

This work is supported by German Federal Ministry of Education and Research (BMBF) under the project SMITH (FK: 01ZZ1609A, 01ZZ1609B, 01ZZ1609C, 01ZZ1803A, 01ZZ1803B, 01ZZ1803C, 01ZZ1803D, 01ZZ1803E, 01ZZ1803F, 01ZZ1803G, 01ZZ1803H, 01ZZ1803I, 01ZZ1803J, 01ZZ1803K, 01ZZ1803L, 01ZZ1803M, 01ZZ1803N).

References

[1] A. Winter et al., Smart Medical Information Technology for Healthcare (SMITH), *Methods Inf Med* 57*(S 01)*, 2018, e92-e105.

[2] SC. Semler, F. Wissing and R. Heyder, German Medical Informatics Initiative, *Methods Inf Med 57(S 01)*, 2018, e50-e56.

[3] MII core dataset, Available from: http://www.medizininformatik-initiative.de/sites/default/files/inline-files/MII_04_Kerndatensatz_1-0.pdf (accessed November 16, 2018)

[4] H. Balzert, *Lehrbuch der Software-Technik*, Spektrum Akademischer Verlag, Heidelberg, 2009.

[5] HL7 FHIR, HL7 v2, Available from: https://www.hl7.org/implement/standards (accessed October 16, 2018)

[6] IHE International, Available from: https://www.ihe.net/ (accessed October 16, 2018)

Address for correspondence

Corresponding Author, Kais Tahar, University Hospital Leipzig, Data Integration Center, Philipp-Rosenthal-Straße 27b, 04103, Leipzig, Germany; E-mail: tahark@medizin.uni-leipzig.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190648

Sweet Talking: Voice Technology and Virtual Assistants in Clinical Diabetes Management

Samuel Tan[a], Farhad Fatehi[b]

[a] *Centre for Online Health, Faculty of Medicine, The University of Queensland, Brisbane, Australia*
[b] *Australian E-Health Research Centre, CSIRO, Brisbane, Australia*

Abstract

Voice technology offers a range of novel and promising strategies for clinical diabetes management. Incorporation of voice-powered virtual assistants (such as Apple Siri and Microsoft Cortana) into diabetes care programs has the potential to improve patient awareness and adherence; facilitate comprehensive provider-patient integration and data collection; and expedite consultations, procedures, and meal preparations. This study will present a qualitative literature review on existing and speculative applications of voice technology in diabetes care.

Keywords:

Diabetes Mellitus, Mobile Applications, Speech Recognition Software

Introduction

Diabetes encompasses a number of related conditions chiefly characterized by dysregulations in glucose metabolism and resultant hyperglycemia. It is a globally recognized health issue with an estimated prevalence of over 400 million people worldwide [1]. Diabetes is primarily represented by two main subtypes: Type 1 diabetes mellitus (T1DM), an autoimmune condition; and Type 2 diabetes mellitus (T2DM), caused by deficient insulin secretion and/or insulin resistance [2]. Both forms of diabetes are lifelong and associated with increased mortality, largely due to cardiovascular and cancer-related causes [3].

Clinically, there are many issues that impair contemporary diabetes management. As diabetes is a chronic condition, patient education and adherence are closely intertwined with overall outcomes [4]. However, long-term adherence remains low, largely due to the complexity and consistency demanded by comprehensive management plans [5]. Surgery represents a prominent challenge in the diabetic population, both due to its fasting requirement and a greater rate of cardiovascular and infectious complications [6]. Additional issues in diabetes management include underdiagnosis, polypharmacy, drug interactions, wound care, and healthcare delivery [7].

Recent advances in digital and online technologies have presented a novel avenue to overcome these challenges [8], though reports on adoption and efficacy remain variable [9]. Smartphone applications have demonstrated efficacy at improving patient adherence and reducing blood glucose levels [10]. Similarly, artificial intelligence systems may allow patients and healthcare providers to better navigate the complexity of diabetes [11].

Voice technology, including virtual assistants such as Amazon Alexa, Apple Siri, and Google Home, offers a highly personalized and accessible approach to diabetes management. This study aims to summarize the current state of voice technology in diabetes management, as well as report on upcoming and potential applications.

Methods

A review of the literature was conducted to identify relevant peer-reviewed articles, conference abstracts, and technical reports. Databases searched included PubMed, Scopus, ACM Digital Library, and IEEE Digital Library. Google Scholar was used to identify relevant grey literature. The eligibility criteria for this study included English-language articles published from 01/01/2010 to 31/12/2018, as well as assessments of topical relevance, reliability, and data quality.

To find studies involving both diabetes and voice technology, a two-pronged search strategy was developed. Diabetes-related search terms included “diabetes”, “blood glucose”, and “blood sugar”. Voice technology search terms included “Amazon Alexa”, “Amazon Echo”, “Apple Siri”, “Microsoft Cortana”, “Google Assistant”, “Google Glass”, “Google Home”, “interactive voice”, “virtual assistant”, “voice assistant”, “voice intervention”, “voice recognition”, “voice interface”, and “voice response”. Overall, 28 peer-reviewed publications and 2 grey literature articles met the eligibility criteria for this study.

Results

Patient Adherence and Education

Interactive voice response (IVR) systems provide automated phone consultations through the use of voice recognition software [12]. Multiple studies have shown that IVR improves medication and lifestyle modification adherence in T2DM patients, ultimately resulting in HbA_{1C} and outcome reductions [12; 13]. By providing personalized advice and a range of answers to common medical questions, voice-driven technologies are also effective at educating and informing patients [14].

Insulin therapy is considered to be clinically underused in T2DM, largely due to difficulties in manually calculating one’s bolus dose [15]. In conjunction with a machine learning algorithm for nutrient determination [16], Foltynski et al. report on a smartphone application that can synthesize verbal descriptions of meals and accurately return the corresponding insulin bolus dose [17]. A similar voice-activated initiative was successfully trialled in patients with diabetic retinopathy [18]. Further development and distribution of these programs have the potential to greatly expedite insulin therapy for existing users, as well as promote its adoption in current non-users.

Healthcare Delivery

IVR and virtual assistants allow patients to easily book health appointments, as well as rapidly notify care providers about hypoglycemic and cardiovascular events [14]. Voice recognition software has successfully been used to transcribe endocrine consultations, improving healthcare efficiency and continuity [19]. Taken together, these systems allow for

doctors' advice and recommendations to be repeated at home, improving patient recall and adherence [14].

Surgeries in diabetic patients are often complex and demand significant expertise [6]. Armstrong et al. document the use of proprietary Google Glass software during diabetic limb salvage surgery to facilitate clinician communication and consultation, reducing errors and improving patient outcomes [20].

Surveillance and Diagnosis

Notably, voice recognition software appears capable of identifying early diabetic [21] and acutely hyperglycemic [22] patients from alterations in their vocal characteristics. Basatneh et al. additionally propose integrating virtual assistants with wearable sensors, such as foot mats and 'smart' wound dressings. This would facilitate wound self-care, early notification of risky behavior, and timely alerts to health providers [14].

Conclusions

Voice technology is swiftly gaining traction in contemporary diabetes management, with studied applications in intervention adherence, medication guidance, healthcare access, intra-operative consultation, and surveillance. Future expansions include the potential for further integration with existing diabetes technologies and machine learning systems, improving symptom detection and reducing intervention complexity. While voice technology currently holds a relatively niche role in diabetes management, the rapidly increasing prevalence of virtual assistants and speech recognition software heralds far greater adoption in the years to come.

Acknowledgements

No funding has been acquired for this study.

References

[1] K. Ogurtsova, J.D. da Rocha Fernandes, Y. Huang, U. Linnenkamp, L. Guariguata, N.H. Cho, D. Cavan, J.E. Shaw, and L.E. Makaroff, IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040, *Diabetes Res Clin Pract* **128** (2017), 40-50.

[2] T. Tuomi, Type 1 and Type 2 Diabetes, *Diabetes* **54** (2005), S40.

[3] O.H.Y. Yu and S. Suissa, Identifying Causes for Excess Mortality in Patients With Diabetes: Closer but Not There Yet, *Diabetes Care* **39** (2016), 1851.

[4] S.E. Curtis, K.S. Boye, M.J. Lage, and L.E. Garcia-Perez, Medication adherence and improved outcomes among patients with type 2 diabetes, *Am J Manag Care* **23** (2017), e208-e214.

[5] W.H. Polonsky and R.R. Henry, Poor medication adherence in type 2 diabetes: recognizing the scope of the problem and its key contributors, *Patient preference and adherence* **10** (2016), 1299-1307.

[6] S. Dagogo-Jack and K.G.M.M. Alberti, Management of Diabetes Mellitus in Surgical Patients, *Diabetes Spectrum* **15** (2002), 44.

[7] L. Blonde, Current challenges in diabetes management, *Clin Cornerstone* **7 Suppl 3** (2005), S6-17.

[8] F. Fatehi, A. Menon, and D. Bird, Diabetes Care in the Digital Era: a Synoptic Overview, *Current Diabetes Reports* **18** (2018), 38.

[9] B. Holtz and C. Lauckner, Diabetes management via mobile phones: a systematic review, *Telemed J E Health* **18** (2012), 175-184.

[10] P.Y. Benhamou, V. Melki, R. Boizel, F. Perreal, J.L. Quesada, S. Bessieres-Lacombe, J.L. Bosson, S. Halimi, and H. Hanaire, One-year efficacy and safety of Web-based follow-up using cellular phone in type 1 diabetic patients under insulin pump therapy: the PumpNet study, *Diabetes Metab* **33** (2007), 220-226.

[11] I. Contreras and J. Vehi, Artificial Intelligence for Diabetes Management and Decision Support: Literature Review, *J Med Internet Res* **20** (2018), e10775.

[12] E.D. Williams, D. Bird, A.W. Forbes, A. Russell, S. Ash, R. Friedman, P.A. Scuffham, and B. Oldenburg, Randomised controlled trial of an automated, interactive telephone intervention (TLC Diabetes) to improve type 2 diabetes management: baseline findings and six-month outcomes, *BMC Public Health* **12** (2012), 602.

[13] J.D. Piette, N. Marinec, E.C. Gallegos-Cabriales, J.M. Gutierrez-Valverde, J. Rodriguez-Saldaña, M. Mendoz-Alevares, and M.J. Silveira, Spanish-speaking patients' engagement in interactive voice response (IVR) support calls for chronic disease self-management: data from three countries, *Journal of telemedicine and telecare* **19** (2013), 89-94.

[14] R. Basatneh, B. Najafi, and D.G. Armstrong, Health Sensors, Smart Home Devices, and the Internet of Medical Things: An Opportunity for Dramatic Improvement in Care for the Lower Extremity Complications of Diabetes, *Journal of diabetes science and technology* **12** (2018), 577-586.

[15] M.B. Davidson, Insulin Therapy: A Personal Approach, *Clinical diabetes: a publication of the American Diabetes Association* **33** (2015), 123-135.

[16] P. Ladyzynski, J. Krzymien, P. Foltynski, M. Rachuta, and B. Bonalska, Accuracy of Automatic Carbohydrate, Protein, Fat and Calorie Counting Based on Voice Descriptions of Meals in People with Type 1 Diabetes, *Nutrients* **10** (2018), 518.

[17] P. Foltynski, P. Ladyzynski, E. Pankowska, and K. Mazurczak, Efficacy of automatic bolus calculator with automatic speech recognition in patients with type 1 diabetes: A randomized cross-over trial, *J Diabetes* **10** (2018), 600-608.

[18] M.H. Radfar, M. Hamilton, and A. Ming, A voice activated device for insulin dosage calculations for visually impaired diabetes, in: *2011 24th Canadian Conference on Electrical and Computer Engineering(CCECE)*, 2011, pp. 000904-000907.

[19] R. Peringattuthodiyil, T. Holdsworth, and P.C. Johnston, A PILOT OF THE USE OF VOICE RECOGNITION SOFTWARE IN AN ENDOCRINE OUTPATIENT CLINIC, *The Ulster medical journal* **86** (2017), 125-125.

[20] D.G. Armstrong, T.M. Rankin, N.A. Giovinco, J.L. Mills, and Y. Matsuoka, A heads-up display for diabetic limb salvage surgery: a view through the google looking glass, *Journal of diabetes science and technology* **8** (2014), 951-956.

[21] D. Chitkara and R.K. Sharma, Voice based detection of type 2 diabetes mellitus, in: *2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)*, 2016, pp. 83-87.

[22] C. Tschöpe, F. Duckhorn, M. Wolff, and G. Saeltzer, Estimating Blood Sugar from Voice Samples: A Preliminary Study, in: *2015 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2015, pp. 804-805.

Address for Correspondence

Samuel Tan. E-mail Address: samuel.tan@uq.net.au

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190649

The Evaluation of the Medical Information Exchange on 24-Hour Operation at North Kyushu Area in Japan

Naota Taura[a], Takehiro Matsumoto[a], Masayuki Honda[a]

[a] *Department of Medical Informatics, Nagasaki University Graduate School of Biomedical Sciences, Japan*

Abstract

We report on triage before transportation by sharing patient information between hospitals, making it possible to use the Ajisai-net for 24 hours. Fifty-six times hospital collaboration was conducted during the period from start of hospital cooperation operation until October 2018. Transportation cost reduction of 3,935,000 yen (34,620 dollars) was estimated. It is expected to contribute to improvement of efficiency of emergency medical care in North Kyushu.

Keywords:
Medical information exchange, triage before transportation.

Introduction

Health Information Exchange (HIE) has been gradually spreading in Japan [1]. The Japan Ministry of Health, Labour and Welfare reported that the number of these networks were spreading to 26 prefectures (55%) in Japan. Ajisai (hydrangea)-net, one of the most popular health information exchange (HIE), has been used since 2004 [1]. The electric medical records of 35 large hospitals of North Kyushu area (Fukuoka, Saga and Nagasaki prefecture) are shared among hospitals, clinics, pharmacies and other medical institutions. However, since regional collaborating room of each information provision hospital which implements access right setting cannot respond outside weekends, public holidays and weekday hours, it is difficult to operate for 24 hours, and lack of function has been pointed out in emergency medical care. We report on triage before transportation by sharing patient information between hospitals, making it possible to use Ajisai (hydrangea)-net 24 hours a day, 7 days a week.

Methods

The electric medical records of 84,332 patients were shared with their consent in Ajisai (hydrangea)-net from 2009 to 2018. This network system when utilized at the right setting is required to carry out strict protection of personal information of each patient.

The regional collaboration room of the information provision hospital's inability to set access right outside weekends, public holidays, weekday hours and lack of function in overtime medical examination has been pointed out. In addition, despite acute phase hospital collaboration with secondary and tertiary hospitals, spoil cases have been found to be appropriate for patients who really need transport outside of the indication. For these problems, it was decided to outsource weekends, public holidays and outside of weekday hours access right setting. To obtain written consent from the patient or a representative at the hospital, fax is sent to outsourced facility and consent form is confirmed by outsourcer. It performs setting of access right by outsourcer and makes completion report to requesting facility. Outsourcing facility further sends work report and consent form to regional collaboration office of each information provision hospital. The work of outsourcer is then terminated. Fax is sent to regional collaboration room of each hospital from the outsourced facility and work of outsourcer terminated. In addition, associated outsourcing costs were 1.2 million yen (10,909 dollars) per year. We examined triage before transportation using Ajisai (hydrangea)-net by cases where hospital-to-hospital cooperation was utilized.

Results

Fifty-six times hospital collaboration was conducted during the period from start of hospital cooperation operation until October 2018. Details were 39 cases (70%) before hospital transfer, 9 cases (16%) consultation, and 8 cases (14%) unknown (see Table 1).

Table 1 – The reason for between hospital cooperation

	Number	Proportion (%)
All	56	100
Before hospital transfer	39	70
Consultation	9	16
Unknown	8	14

Further investigation revealed that of the 39 cases conveyed before triage, 27 cases were transported [16 cases by helicopter (41%) and 11 by ambulance transportation (28%)], and 12 cases (31%) were not transported (see Table 2). We examined cases that were not transported further, and we found 7 cases scheduled to be conveyed by helicopter and 5 patients scheduled to be transported by ambulance.

Table 2 – Hospitalty transfers by between hospital cooperation

	Number	Proportion (%)
All	39	100
Transported:	27	
Helicopter	16	41
Ambulance	11	28
Not transported	12	31

According to previous reports, we estimated 530,000 yen (4,660 dollars) for helicopter transportation per case and 45,000

yen (395 dollars) for ambulance transportation in Japan [2,3]. Estimated total transportation cost reduction was 3,935,000 yen (34,620 dollars), helicopter transport cost reduction of 3,710,000 yen (32,640 dollars), and ambulance transport cost reduction of 225,000 yen (1,980 dollars).

Discussion

The medical network system started from paper media and has shifted from the fax using telephone communication network to the internet system. With these advances, it has become possible to confirm patient information in real time, which has greatly contributed to telemedicine and medical treatment support [4,5]. However, protection of patient's personal information, response at time of remote medical care, response to overtime, and so on still often must be solved.

We reported that to construct a method for registering out of time to the online system. By using outsourcing to solve the problem posed by holidays and weekday outside hours, it is possible to share information at emergency medical scene by enabling the use of hydrangea 24 hours a day, 7 days a week.

In Japan, several hospitals in the region cooperate with each other for emergency medical care. It is important to share imaging findings and laboratory data when linking between hospitals. In our study, by establishing a 24-hour hospital-to-hospital linkage, these results could be shared to allow treatment decisions to be made prior to transport, thus reducing emergency ambulance or helicopter transport by 31%. As a result, cost savings estimated at 3,935,000 yen (34,620 dollars) was reported. The role sharing of hospitals by triage before transportation, and reduction of medical cost is expected to contribute to improvement of efficiency of emergency medical care in North Kyushu.

Conclusions

It was possible to review the trial and treatment method before transporting by confirming the medical information in the receiving hospital before medical examination.

References

[1] K. Hara, The present condition and future view of information network for regional medicine, *The Medical Frontline* **58** (2003), 103-116.

[2] T. Matsumoto, M. Honda, The evaluation of the health information exchange with the number of usage and the introduction of outpatient to the hospitals at Nagasaki Japan, *Stud Health Technol Inform* **245** (2017), 1373.

[3] M. Kuriki, Simulation using the probabilities of death by ambulance transport in trauma patient, *Journal of the Japan Society for Healthcare Administration*, **46**, 241-250.

[4] Yamashita H, The problem and the strategy of increased emergency transport in the geriatric emergency medicine: especially about charged ambulance, *Journal of Japanese Society for Emergency Medicine* **19** (2016), 1-6.

[5] K. Hara, Medical information – about medical cooperation by 4 prefecture electronic medical record network, *Laboratory and Clinical Practice* **20** (2002), 28-33.

Address for Correspondence

Naota Taura, Department of Medical Informatics,
Graduate School of Biomedical Sciences Nagasaki University,
E-mail: ntaura-g@nagasaki-u.ac.jp

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190650

Improving Postural Balance in Dentoalveolar Malocclusion Patients Using a Vibrotactile Posture Trainer Device

Bhornsawan Thanathornwong[a], Siriwan Suebnukarn[b]

[a] *Faculty of Dentistry, Srinakarinwirot University, Bangkok, Thailand,*
[b] *Faculty of Dentistry, Thammasat University, Pathumthani, Thailand*

Abstract

The biomechanical relationship between the body and teeth suggested that posture imbalance can lead to dentoalveolar malocclusion. A randomized crossover trial was conducted to compare body posture measure using tilt angles of the neck and center of pressure of the malocclusion patients that received vibrotactile biofeedback from the posture trainer device with those who received no feedback. The results showed that the system is associated with quantitative improvements of the body posture in malocclusion patients.

Keywords:

Malocclusion; Postural Balance; Cross-Over Studies.

Introduction

Occlusion is a term used to refer to the alignment of teeth. Deviations from an ideal occlusion are called malocclusion. Orthodontics is a branch of dentistry that treats malocclusion. Some patients can maintain satisfactory conditions long after the treatment. However, in some patients, the symptoms reappear. Relapses can be interpreted as a return to an earlier posture expressing the patient's initial postural pattern [1]. A number of studies have been conducted on body structures affecting occlusion, focusing on genetically abnormal and unvaccinated body structures. The results of these studies indicate that, in general, individuals with unusual body structures are also associated with abnormal facial structure and occlusion [2, 3] due to the fact that both structures are connected by muscles. To be in a balanced position both statically and dynamically, it is necessary to balance the neck as well as the facial muscles and the occlusion plane [3].

Studies on the relationship between biomechanical and neurological aspects of the oral cavity and jaw have continued to multiply in the literature [4-6]. The study of body balance in patients with malocclusion using the equilibrium platform showed that it was instability in patients with severe class II or III malocclusion [5]. Body inclination has been found in cases of severe malocclusion, such as prominent protrusion or short jaw [6].

The use of biofeedback has been offered in the past as an instrument for training that enables an individual to learn how to change physiological activity or behavior for the purposes of improving performance. In therapeutic applications, biofeedback training of balance and posture has shown to be effective for posture control in adolescent scoliosis [7], and has also decreased the fall rate in elderly patients with peripheral neuropathy [8]. In patients with bilateral vestibular loss [9], biofeedback training was also found useful in enhancing postural stability even under challenging standing conditions, beyond the effect of practice alone.

This study evaluated the effectiveness of a vibrotactile posture trainer device that comprised of a wearable device containing an accelerometer sensor for measuring the tilt angle of the body (input) and provided real-time vibrotactile biofeedback (output) when the body is unbalanced measured using center of pressure (COP), the point where the total sum of a pressure field acts on a body, causing a force to act through that point.

Methods

The wearable device developed by our group consisted of an ADXL345 3-axis accelerometer sensors with high-resolution (13-bit) measurement at up to ±16g and a 12.5-400Hz bandwidth response. An accelerometer sensor was used for tracking body tilt angles. The accelerometer sensor presented two reading outputs, one for the X_{out} and another for Y_{out}, and a power supply voltage input of 2.0-3.6V. The expected values for X_{out} and Y_{out} were in the digital IO voltage range of 1.8-2.5V. The sensor consisted of a structure with a capacitive sensing cell (g-cell) and signal conditioning to detect small displacements. The signals from the accelerometer sensors were amplified and converted into digital signals through a data acquisition card (13-bit resolution) connected to a computer. The accelerometer mounted on a circuit card was used as an inclinometer to calculate body tilt angles. The vibrotactile biofeedback was provided to the user via a wearable device when the tilt angles were higher than 15 degree (Figure 1).

Figure 1– The wearable device containing an accelerometer sensor and vibrotactile biofeedback module was attached to the participant to measure the tilt angles of the body.

A randomized crossover trial was conducted to compare COP of the participants that received vibrotactile biofeedback from the system with the control group who received no feedback. The COP (Newton meter) was measured using Bertec forceplate (BERTEC, USA). We recruited eight Class II malocclusion (Maxillary protrusion) and eight Class III malocclusion (Mandibular protrusion) patients aged between 20-30. The study was approved by the institutional Ethical Review Board. A written consent form was provided by all participants.

Participants in the experiment group received vibrotactile biofeedback from the system after standing for 20 second, while those who were in the control group received no feedback. The primary outcome measures were the mean values of COP. The effects of a vibrotactile posture trainer (control vs. feedback group) were evaluated using the Paired t- test and were assumed to be significant at $p < 0.05$ (two-side). All analyses were conducted with the statistical package for the Social Science version 21.0 (SPSS, Chicago, IL).

Results

Participants who have Class II and Class III malocclusion that received feedback from the system had lower center of pressure which were statistically different ($p < 0.05$) from those who received no feedback from the system (Table 1).

Table 1– COP of feedback and no feedback group

Mal Occlusion	*n*	No Feedback Mean COP±SD	Feedback Mean COP±SD	*p*
Class II	8	21.14±6.92	17.08±14.38	0.007
Class III	8	20.76±11.09	14.74±3.04	0.005

Discussion

The findings indicated changes or modifications of the parts that affect any other body parts connected with muscles ensure a reaction by the nervous system as well, which influences the center of pressure and gait stability as exhibited through different body postures [10, 11]. A study by Lippold et al. [12] found a relationship between class II molar relationship occlusion and the balance platform. The children in the group had a poorer balance platform than other children. It was therefore recommended to observe occlusion during childhood so as to prevent future disorders of the body structure.

Several studies have used vibration-type feedback data and found it to be highly effective. For example, a study by Gopalai and Senanayake [13] conducted an experiment on a group of 12 people, comprising 6 males and 6 females. Vibration-type of feedback data was used by attaching a sensor on the back of participants while they were standing. The results indicated that the group with feedback data had a better balance platform with less movement. This was reiterated by a study carried out by Alahakone and Senanayake [14], who found that feedback data with vibration could help improve the walking patterns of patients who were unable to walk in a straight line because of tandem gait, and was better than letting the patients practice alone.

Conclusions

The results presented here demonstrate that the vibrotactile posture training system is associated with quantitative improvements of postural balance. This may be viewed as a promising first step to implement the system to minimize unbalanced posture that cause relapses in orthodontic treatment.

Acknowledgements

This study was supported by the Strategic Wisdom and Research Institute, Srinakharinwirot University and the Faculty of Sports Science, Chulalongkorn University.

References

[1] M. Levinkind, Consideration of whole-body posture in relation to dental development and treatment of malocclusion in children, *BDA Journal Supplement* (2008), 1-6.
[2] T. Shimazaki, M. Motoyoshi, K. Hosoi, and S. Namura, The effect of occlusal alteration and masticatory imbalance on the cervical spine, *Eur J Orthod* 25 (2003), 457-463.
[3] A. Michelotti, G. Buonocore, P. Manzo, G. Pellegrino, and M. Farella, Dental occlusion and posture: an overview, *Prog Orthod* 12 (2011), 53-58.
[4] S. Wakano, T. Takeda, K. Nakajima, K. Kurokawa, and K. Ishigami, Effect of experimental horizontal mandibular deviation on dynamic balance, *J Prosthodont Res* 55 (2011), 228-233.
[5] P. Arumugam, S. Padmanabhan, and A.B. Chitharanjan, The relationship of postural body stability and severity of malocclusion, *APOS Trends Orthod* 6 (2016), 205-210.
[6] A.M. Alwarawreh, S.A. Sarayreh, H.F. Rabadi, E.A. Albdour, and A.M. Marzouq, Effect of body posture on malocclusion, *Pakistan Oral & Dental Journal* 34 (2014), 635-639.
[7] M.S. Wong, A.F. Mak, K.D. Luk, J.H. Evans, and B. Brown, Effectiveness of audio-biofeedback in postural training for adolescent idiopathic scoliosis patients, *Prosthet Orthot Int* 25 (2001), 60-70.
[8] G. Wu, Real-time feedback of body centre of gravity for postural training of elderly patients with peripheral neuropathy, *IEEE Trans Rehabil Eng* 5 (1997), 399-402.
[9] M. Dozza, L. Chiari, and F.B. Horak, Audio-biofeedback improves balance in patients with bilateral vestibular loss, *Arch Phys Med Rehabil* 86 (2005), 1401-1403.
[10] C. Sforza, G.M. Tartaglia, U. Solimene, V. Morgun, R.R. Kaspranskiy, and V.F. Ferrario, Occlusion, sternocleidomastoid muscle activity and body sway: a pilot study in male astronauts, *Cranio* 24 (2006), 43-49.
[11] P. Bracco, A. Deregibus, and R. Piscetta, Effects of different jaw relations on postural stability in human subjects, *Neurosci Lett* 56 (2004), 228-230.
[12] C. Lippold, L. van den Bos, A. Hohoff, G. Danesh, and U. Ehmer, Interdisciplinary study of orthopedic and orthodontic findings in preschool infants, *J Orofac Orthop* 64 (2003), 330-340.
[13] Gopalai AA, Senanayake SMNA. (2010). A wearable real-time intelligent posture corrective system using vibrotactile feedback. *Mechatronics*. 16: 827-834.
[14] A.U. Alahakone, S.M.N.A. Senanayake, A real-time system with assistive feedback for postural control in rehabilitation, *Mechatronics* 15 (2010), 226-233.

Address for correspondence
Bhornsawan Thanathornwong, ppeetakul@hotmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190651

A Descriptive Review: Six Pediatric Personal Health Records

Cori Thompson[a]

[a] *Global ICT Consulting, Victoria, British Columbia, Canada*

Abstract

This descriptive review narratively synthesized themes and concepts from academic studies on pediatric personal health records (PHRs) from Medline EBSCO and CINAHL published January 1, 2007 to April 7, 2017. Reported features were summarized into a frequency table. Six studies in four countries reported PHRs for children with special health care needs, well-children, and adolescents. All studies advocated further development of such tools to improve healthcare. Such development might include children in the design phase.

Keywords:
Child, Personal health record, Review.

Introduction

This poster presents main findings from a descriptive review of academic literature on personal health records (PHRs) for children and adolescents (children) as a first step towards summarizing the evidence. PHRs have the potential to improve patient health outcomes by supporting patient interaction with and management of their own health care. PHRs may be defined as "a record controlled by the individual and may include health information from a variety of sources, including multiple health care providers and the patients themselves. The PHR is separate from, and does not replace the legal record of any health care provider… and may be stand alone or connected" (Office of the National Coordinator [ONC])[1].

Methods

A descriptive review methodology was chosen to identify any "trends or patterns" from published academic empirical studies, to "collect, codify and analyze numeric data" for frequency in pediatric PHRs, and to review a "representative sample" rather than being comprehensive [2]. For further rigour, the review relied on additional items from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (i.e., rationale, eligibility criteria, information sources, search, study selection, data collection process, data items, synthesis of results, study characteristics, limitations) [3]. The topics were the reported key information types and functionalities - or features - in pediatric PHRs. From the included studies, reported features were abstracted into a frequency table and the themes and concepts were narratively synthesized. The concept of pediatric PHR using terms "child" AND "personal" AND "health record" was searched in Medline EBSCO and in the Cumulative Index to Nursing and Allied Health Literature (CINAHL) from January 1, 2007 to April 7, 2017. Excel tables were used to screen the returns, identify duplicates, and abstract and summarize findings from included studies. Inclusion criteria were (1) on patient-controlled integrated or non-integrated PHRs as defined by ONC, (2) designed for or tailored to children, (3) reported on key information types and functionalities, (4) primary studies or used secondary data, and (5) in English. Exclusion criteria included (1) PHRs designed and/or used for adults, (2) records not controlled by patients such as school health records, (3) pediatric medical records to avoid skewing the frequency counts, (4) commentaries, grey literature, and policy statements, and (5) although a formal qualitative assessment was not conducted to avoid excluding any of the few studies, an informal assessment excluded studies that reported too few features to avoid skewing frequency counts.

Results

Of 49 results, 3 duplicates were excluded, and 36 articles did not meet the eligibility criteria. Ten articles were reviewed in full. Subsequently, four were excluded: three reported too few key information types and one was not about a PHR. Table 1 compares key study characteristics of the six included studies from four countries. The studies were found to use 15 different terms for pediatric PHR. The frequency analysis listed forty-eight features: all six studies reported on reminders or notifications, privacy (confidentiality), and security; five reported on patient demographics, immunizations, problem list/diagnosis, medications, and secure messaging; and four reported on lab results, calendaring/appointments, and ability to grant limited access. Evidence from the six studies was synthesized into four main themes and concepts for pediatric PHR design and usage. First, the main value of a PHR was the potential for communication and care coordination with the care team. Many children have multiple care providers in different locations and specialities [4-7, 9]. Parents used the PHR as a tool to support care coordination. Second, intrinsic motivation to use the PHR is needed for empowerment. Empowerment includes a balance of taking and relinquishing control (e.g., a patient may choose not to view a test result prior to the visit, a patient may review the record for any missing information) [7]. Third, the adoption, usage, and usability of the records depended on the parents' coping ability with the child's condition [5-7]. The PHR could also be a vehicle for knowledge transfer about the child's condition between provider-parent and specialist-primary care provider, as well as a means for parent-parent support and identifying parents as experts in their child's condition [4-7, 9]. Pediatric PHRs also need to support parental responsibility for their child's health decisions, right to reliable, complete health information, parental control and accountability of the record, and the need for understandable information [6]. While there was mixed evidence for PHR adoption by parents [7], most studies provided evidence of PHR adoption. Further research is needed as PHR adoption might be

by "parents more engaged in their child's care" [9]. Last, privacy (confidentiality) and security were key especially for adolescents' PHRs where both legislation and meaninful use criteria needed to be met. Organizational policies may help ensure these criteria are met in the PHR design and use [8].

Table 1 – Characteristics of the six reviewed studies

First author (Year) ***Country*** **[citaton]**	**Study type** ***Data source***	**Record name (Stage)** ***Integrated or Stand alone***	**Population/ Condition (Record control)**
De Graaf (2014) *NL* [4]	Evaluative *Primary data*	huidhuis.nl PHR with treatment plan (Implemented) *Integrated*	Congenital conditions (Patient owned, provider accessible)
Popkin (2009) *CA* [5]	Evaluative *Primary data*	eFOSTr PHR (Prototype) *Stand alone*	Organ transplants (Partially by parents)
Rocha (2007) *USA* [6]	Descriptive *Primary data*	PedMHR – medical home record (Design) *Stand alone*	Special health care needs (Partially by parents)
Schneider (2016) *UK* [7]	Descriptive *Primary data*	Patient Knows Best (PCEHR) (Implemented) *Stand alone*	Serious chronic conditions (Partially by parents)
Thompson (2016) *USA* [8]	Evaluative *Secondary data*	MyUFHealth Portal (Implemented) *Integrated*	Adolescents (Adolescent controlled, proxy granted to parents)
Tom (2012) *USA* [9]	Evaluative *Secondary data*	Kaiser Permanente child PHR (Implemented) *Integrated*	Well-child (visits, immunization) (Parent proxy via own PHR)

While four studies described participatory design with parents, providers, and/or organizational representatives [5-8], none reported involving children or adolescents in the design phase. This maybe an area for future research [8]. Future recommendations also include providing training and real-time technical support within a supportive health care organization environment, and tailoring information to the parents' ability to cope with their child's condition [6-7]. There may be differences in PHR use for well-children compared to PHR use for children with complex health care needs [5-6, 9]. Further research is needed.

Conclusions

This review attempted to synthesize academic literature on pediatric PHRs. Limitations included: one researcher designed and conducted the review, the studies' findings might not be generalizable to other jurisdictions, it is possible the studies did not report all features, and, last, only one expression of the key concept and two databases were searched. https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=75346 has been registered to address the limitations. All six studies advocated future research of pediatric PHRs.

References

[1] Office of the National Coordinator [ONC], Personal health records: What health care providers need to know. Accessed March 29, 2019: https://www.healthit.gov/sites/default/files/about-phrs-for-providers-011311.pdf

[2] G. Paré, M.C. Trudel, M. Jaana, and S. Kitsiou, Synthesizing information systems knowledge: A typology of literature reviews *Information & Management* **52** (2) (2015), 183-199.

[3] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, and the PRISMA Group, Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement, *Ann Intern Med* **151** (4) (2009), 264-269.

[4] M. de Graaf, J.E.E. Totté, H. van Os-Medendorp, W. van Renselaar, C.C. Breugem, and S.G.M.A. Pasmans, Treatment of infantile hemangioma in regional hospitals with eHealth support: Evaluation of feasibility and acceptance by parents and doctors, *JMIR Res Protoc* **3** (4) (2014), e 52.

[5] J. Popkin, A. Kushniruk, E. Borycki, D. Guarin, L. Mozley, N. Kilarski, L. Robson, and W. Creed, The eFOSTr Project: Design, implementation and evaluation of a web-based personal health record to support health professionals and families of children undergoing transplants, *Stud Health Technol Inform* **143** (2009), 358-363.

[6] R.A. Rocha, A.N. Romeo, and C. Norlin, Core features of a parent-controlled pediatric medical home record, ISO Press 129 (Pt 2) (2007), 997-1001.

[7] H. Schneider, S. Hill, and A. Blandford, Patients Know Best: Qualitative study on how families use patient-controlled personal health records, *J. Med. Internet Res* **18** (2) (2016), e43.

[8] L.A. Thompson, T. Martinko, P. Budd, R. Mercado, and A.M. Schentrup, Meaningful use of a confidential adolescent patient portal, *J Adolesc Health* **58** (2) (2016), 134-140.

[9] J.O. Tom, C. Chen, and Y.Y. Zhou, Personal health record use and association with immunizations and well-child care visits recommendations, *J. Pediatr* **164** (1) (2014), 112-117.

Address for correspondence

cori.thompson@ictconsultingltd.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190652

Is Teledermoscopy Improving General Practitioner Skin Cancer Care?

Femke van Sinderen[ab], Esmée Tensen[ab], Job P. van der Heijden[b], Leonard Witkamp[ab], Monique W.M. Jaspers[a], Linda W.P. Peute[a]

[a] *Amsterdam UMC, University of Amsterdam, Department of Medical Informatics, Amsterdam Public Health Research Institute, Meibergdreef 9, Amsterdam, the Netherlands*
[b] *Ksyos Health Management Research, Prof. J.H. Bavincklaan 2-4, Amstelveen, the Netherlands*

Abstract

Skin cancer incidences have tripled in the Netherlands for the last twenty years and are expected to increase even more in the coming years. Teledermoscopy (TDsc) is implemented in Dutch practice to support and enhance early skin cancer detection by general practitioners (GPs) through remote consultation with dermatologists. This study assesses the effect of TDsc consultation on the quality and efficiency of skin cancer care in the primary setting by analyzing 10,184 TDsc consultations.

Keywords:

Telemedicine, Dermatology, Skin Cancer

Introduction

Incidences of the most common skin cancer types have tripled in the Netherlands for the last twenty years [1]. Skin cancer accounts for 15% of all newly diagnosed cancers. Specifically, melanoma incidences are the highest in the Netherlands compared to other European countries, thereby likely increasing the burden on the general practice [1; 2].

Teledermoscopy (TDsc) has been suggested as a method to enhance the quality of care by supporting patient treatment and/or referral decisions. TDsc is defined as the provision of a consultation with a remote dermatologist based on digitally available dermatoscopic images [3]. As a result, TDsc could support GPs in primary dermatology healthcare in early diagnosis and referral of patients with skin problems, thereby augmenting their expertise in melanoma detection. Also, it can improve the efficiency of primary dermatologic care by preventing unnecessary referrals to secondary care, as well as accelerate the time to diagnosis.

Store-and-forward TDsc has been reimbursed and integrated into the Dutch healthcare system as a regular health service since 2009 by Ksyos Telemedical Center. Earlier studies in the Netherlands mainly focused on regular teledermatology [4; 5]. The aim of this study was to assess how TDsc affects the quality and efficiency of Dutch general practitioner dermatology care.

Methods

TDsc consultations data performed between February 2009 and March 2017 from routine clinical practice sent between GPs and dermatologists in the Netherlands using Ksyos Teledermoscopy services (Ksyos, Amstelveen, the Netherlands) was analyzed descriptively. The TDsc flow can be seen in Figure 1.

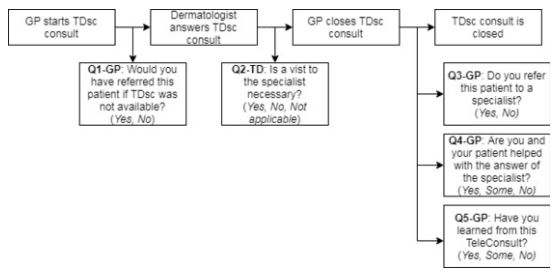


Figure 1. Overview of the five evaluation questions in the TDsc service. Q = Question, TD = teledermatologist.

Patients gave oral informed consent for a TDsc consultation. A TDsc consultation included among others, a maximum of four (dermatoscopic) pictures, an anamnesis, differential diagnosis by ICD-10 classification, and medical history. Questions to the dermatologist were asked in a free text field. As a result, the teledermatologist provided the diagnosis by providing an ICD-10 code and advice on diagnostic management by free text.

Five mandatory evaluation questions implemented in the Ksyos teledermoscopy service were asked to the GP and dermatologist (Figure 1). Answers to these evaluation questions were translated to indicators to assess the quality and efficiency of TDsc. Several of these indicators have also been used in a prospective study on the quality of general teledermatology [5]. In this study, dermatologists' response time was seen as an efficiency of care indicator related to the efficiency in the 'time to diagnose' process. The following indicators were subject in this study:

Quality of Care

1. Extra dermatologic advice requests: when Q1 = No.
2. Additional required referrals: when Q1 = No AND Q3 = Yes.
3. GP valuation of TDsc: when Q4 = Yes, Some. And when Q5 = Yes, Some.

Efficiency of Care

1. Prevented physical referrals: when Q1 = Yes AND Q3 = No.
2. Overall percentage of prevented physical referrals: 1 – (numerator is Q3 = Yes and the denominator is Q1 = Yes) * 100%.
3. Median dermatologists' response time: this was measured based on working days (8.30 am to 5.30 pm) from a 5-day working week.

Results

A total of 10,184 TDsc consultations performed between 26th February 2009 and 27th March 2017 was analyzed. Those were sent by 730 GPs from 555 practices and answered by 95 dermatologists. TDsc consultations with missing responses were excluded from the analysis (N = 2,345). Table 1 represents the interpretation of those outcomes concerning the indicators described below.

Table 1. The TDsc quality and efficiency indicator outcomes.

TDsc quality of care indicators	
Of the TDsc consultations:	
1.	**30.3%** were performed by GP due to the availability of TDsc, to gain extra dermatological advice.
2.	**17.3%** led to physical referrals of patients otherwise not referred by the GP, these included pre-diagnosed skin cancer cases.
3.	**97.4%** was reported as helpful by the GP and **95.3%** were considered educational/instructive.
TDsc efficiency of care indicators	
Of the TDsc consultations:	
1.	**69.6%** prevented a physical referral, decreasing the number of patients needed to be seen face-to-face by the dermatologists.
2.	Overall **62.1%** physical referrals were prevented.
3.	**2.23** hours was the median dermatologist response time increasing efficiency of the diagnostic process (mean; 6.58h, standard deviation; 26.8h).

Quality of Care

Regarding the first indicator, 7,839 of 10,184 of the TDsc consultations were included and had responses on both Q1 and Q3. Of the included consultations, 2,379 (30.3%) were requested by the GP to gain supplementary dermatological advice (Table 1).

Concerning the second indicator, 411 patients (17.3%) who would otherwise not have been referred were physically referred after a TDsc consultation to the dermatologist (Q1 = No and Q3 = Yes). Teledermatologists provided an ICD-10 code for 149 of those consultations, including skin cancer diagnosis. In 7,637 TDsc consultations (97.4%) GPs reported that they and their patients were helped by the dermatologist's response (Q4 = Yes, Some). Assessment of the reported learning effect by GPs showed that 7,474 consultations (95.3%) were considered instructive (Q5 = Yes, Some).

Efficiency of Care

Of the included consultations 5,460 (69.7%) were intended to prevent a physical referral (Q1 = Yes). Of these, 3,800 (69.6%) were actually prevented (Q1 = Yes and Q3 = No). Without TDsc 5,460 patients would have been referred to the dermatologists. After TDsc 2,071 patients were referred, resulting in 62.1% of prevented physical referrals overall.

Timestamps have been stored since July 2011. Hundred-eighteen consultations were excluded due to missing timestamps. The median response time of the dermatologist was 2.23h (Figure 2).

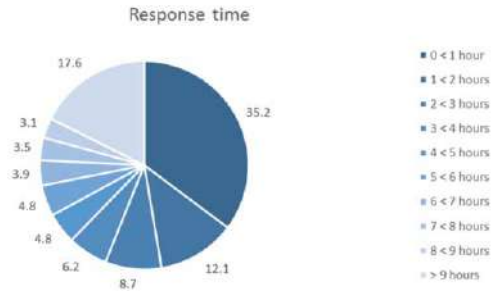


Figure 2. The distribution (%) of response times of TDsc consultations (N=10,066)

Conclusions

The results from this study give a strong indication that TDsc is improving Dutch dermatologic general practitioner health care regarding its quality and efficiency. TDsc availability is valued by GPs in gaining additional and helpful dermatologic advice and prevented a high percentage of physical referrals. Moreover, TDsc resulted in the detection of potential skin cancer cases of patients who would not have been referred if TDsc had not been available. TDsc also improves the efficiency of GP dermatology healthcare by obtaining dermatological advice within one working day.

Conflicts of Interests

E.T., F.v.S, and J.P.v.d.H. are employed (part-time) by Ksyos, and L.W. is the director of Ksyos. The remaining authors state no conflicts of interest.

References

[1] Rijksinstituut voor Volksgezondheid en Milieu, Huidkanker, Regionaal & Internationaal, Internationaal, https://www.volksgezondheidenzorg.info/onderwerp/huidkanker/regionaal-internationaal/internationaal#node-incidentie-huidkanker-eu-landen (accessed 2017).

[2] C.J. Koelink, B.J. Kollen, F. Groenhof, K. van der Meer, W.K. van der Heide, Skin lesions suspected of malignancy: an increasing burden on general practice, *BMC Fam Pract* **15** (2014).

[3] J. Dahlen Gyllencreutz, E. Johansson Backman, K. Terstappen, J. Paoli. Teledermoscopy images acquired in primary health care and hospital settings - a comparative study of image quality, *J Eur Acad Dermatol Venereol* **32** (2018), 1038-1043.

[4] E. Tensen, J.P. van der Heijden, M.W.M. Jaspers, L. Witkamp, Teledermatologie, *Bijblijven* **33** (2017), 101-113.

[5] J. van der Heijden, N. de Keizer, J. Bos, P. Spuls, L. Witkamp, Teledermatology applied following patient selection by general practitioners in daily practice improves efficiency and quality of care at lower cost, *BJD* **165** (2011), 1058-1065.

Address for Correspondence

Femke van Si9eren, f.vansinderen@amc.uva.nl, +31 20 5666204

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190653

Standards, Processes and Instruments for Assessing Usability of Health Mobile Apps: A Systematic Literature Review

Francisco Vera[a], René Noël[a], Carla Taramasco[a]

[a] *Escuela de Ingeniería Civil Informática, Universidad de Valparaíso, Valparaíso, Chile*

Abstract

mHealth applications have had sustained growth in recent years. However, there is scant and disorganized scientific evidence of methods for evaluating their usability and adoption. We conducted a systematic literature review to describe standards, processes, methods, and tools for evaluating mobile health software. We analyzed ten studies: two identify standards for evaluation, and seven specify tools or instruments for supporting assessment. PSSUQ, SUS, and MARS were the most referenced instruments for usability and adoption evaluation.

Keywords:
Mobile Applications, Information Systems, Review Literature as Topic

Introduction

The World Health Organization defines mHealth as "the practice of medicine and public health supported by mobile devices such as mobile phones, patient monitoring devices, personal digital assistants and other wireless devices" [1]. Mobile applications have proliferated in recent years, addressing health education, patient monitoring, or as part of health information systems. It is estimated that there are more than 325,000 mHealth applications for the most popular mobile platforms (iOS and Android) [2]. One of the challenges associated with the large number of applications is quality. In health, usability has been considered quality factors related to patients' safety and care quality [3-5]. There are specific instruments for the evaluation of usability in mobile applications [6] and recommendations of global organizations, such as HIMMS of processes and instruments [7]. Secondary studies have also been conducted looking for evidence of usability processes for mobile applications, identifying that the majority of usability evaluations are based on questionnaires [8]. Although these initiatives are varied, there is scant and disorganized scientific evidence of the standards, processes, and tools used in real life for the evaluation of quality from this perspective. We conducted a Systematic Literature Review (SLR) regarding the use of quality assessment standards, processes, methods, and tools with respect to the usability and adoption, from the point of view of the researcher in the context of trending scientific publications on health and information technologies.

Methods

Based the purpose of this review, we defined three research questions:

RQ1. What standards and norms are used for the certification of the attributes of quality in usability, for the development of mobile applications for health?

RQ2. What processes or methodologies have been used in the development of mobile applications for health in which they include elements of usability?

RQ3. What tools to assess the usability of mobile health applications have been used by software developers?

The search string was created according to the "Population, Intervention, Comparison and Output" (PICO) structure [9,10]. We defined population as mobile health, intervention as use and usability of health information systems, and ommitted comparison and outcomes due to lack of prior studies in the same context. The final search string was: *((Mobile OR smartphone) AND (app OR application) AND ("mHealth" OR "health software")) AND ("Adoption" OR "Usage" OR "Usability" OR"Assessment" OR "Evaluation" OR "Study" OR "Standard" OR "Metric")*. We selected two main scientific databases as sources: SCOPUS and Web of Science (WoS). The search string presented in the previous section was modified according to the syntax of the search engines of each database. Studies with at least one of the below exclusion criterion were not selected:

- Assessment target was not related to mobile apps,
- Domain of mobile apps was not related to health,
- Assessed quality attributes were not related to usability nor quality in use,
- Documents type is different from article, non-English, or published before 2000 or after 2017, or
- Article does not present experiences, processes, methods, tools, instruments, surveys, or literature reviews related to assessment of usability or quality.

Results

Figure 1 presents the results of each stage. Ten studies were included in the final analysis and classified according to the proposed protocol. Regarding RQ1, only two studies [11,12] mentioned standards for the evaluation of usability of health information systems, including ISO 9241-11: 2018 [13], ISO 9241-12:1998 [14], and ISO / IEC 9126-1) [16]. Regarding RQ2, 50% of the articles specified a process or methodology, while just 10% specified an ad-hoc approach. Regarding RQ3, 70% informed the use of tools and instruments; PSSUQ[18], mHIMMS[7] and MARS[4] were the most referenced.

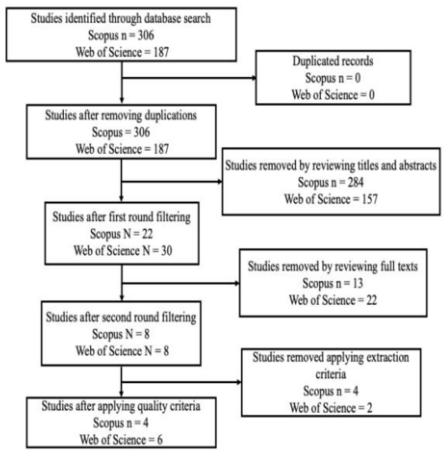


Figure 1 – Flow Diagram for the Study Selection Process.

Conclusions

We performed a SLR regarding usability assessment in mHealth mobile apps. From 493 initial articles, 10 were included and analysed. Main findings are consistent with previous similar reviews: usability assessment is mainly based on polls or questionnaires (only two studied report the use of standards), although specific mobile platform guidelines [19] and specific evaluations frameworks [11,17] were also found. All of the studies mentioned usability as a quality characteristic, 30% specified other attributes such as security and efficiency, and 30% reported a usability sub-characteristic (memorability). Given that selected studies are recent (3 were published in 2014, 4 in 2016, and 3 in 2017), it is likely that the evaluation of usability of mobile applications is in an incipient period.

Acknowledgements

This work was supported by CORFO - CENS 16CTTS-66390 through the National Center on Health Information Systems and by Centro Interdisciplinario de Estudios en Salud (CIESAL) de la Universidad de Valparaiso. This work was also supported by the National Commission for Scientific and Technological Research (CONICYT) through the program STIC-AMSUD17STIC- 03: “MONITORing for ehealth” and FONDEF ID16I10449 “Sistema inteligente para la gestión y análisis de la dotación de camas en la red asistencial del sector público.”

References

[1] M. Kay, J. Santos, and M. Takane, mHealth: New horizons for health through mobile technologies. *World Health Organization* **64** (2011), 66-71.

[2] Research2Guidance. (2017). mHealth App Economics 2017: Current Status and Future Trends in Mobile Health.

[3] J. Brooke, SUS: A quick and dirty usability scale. *Usability Evaluation in Industry* **189** (1996), 4-7.

[4] S.R. Stoyanov, L. Hides, D.J. Kavanagh, O. Zelenko, D. Tjondronegoro, and M. Mani, Mobile app rating scale: A new tool for assessing the quality of health mobile apps. *JMIR mHealth and uHealth* **3** (2015).

[5] B. Middleton, M. Bloomrosen, M.A. Dente, B. Hashmat, R. Koppel, J.M. Overhage, ... and J. Zhang, J. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: Recommendations from AMIA. *Journal of the American Medical Informatics Association* **20** (2013), 2-8.

[6] R. Schnall, H. Cho, and J. Liu, Health Information Technology Usability Evaluation Scale for Usability Assessment of Mobile Health Technology: Validation Study. *JMIR mHealth and uHealth* **6** (2018).

[7] P. Arellano, J. Bochinski, B. Elias, S. Houser, and H. Head, Selecting a Mobile App: Evaluating the Usability of Medical Applications mHIMSS App Usability Work Group. *mHIMSS App Usability Work Group* **1**(2012), 1-28.

[8] B.C. Zapata, J.L. Fernández-Alemán, A. Idri,and A. Toval, Empirical studies on usability of mHealth apps: A systematic literature review. *Journal of Medical Systems* **39** (2015).

[9] B. Kitchenham, Procedures for Performing Systematic Reviews, Keele University, Keele, Staffs, UK, 2004.

[10] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, Systematic Mapping Studies in Software Engineering. *EASE* **8** (2008), 68-77.

[11] T. Aljaber and N.A. Gordon, Evaluation of mobile health education applications for health professionals and patients. In *8th International Conference on e-Health* (2016), 107-114.

[12] R. Gunter, S. Fernandes-Taylor, A. Mahnke, L. Awoyinka, C. Schroeder, J. Wiseman, ... and K.C. Kent, Evaluating patient usability of an image-based mobile health platform for postoperative wound monitoring. *JMIR mHealth and uHealth* **4** (2016).

[13] ISO 9241-11:2018 - Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts, *International Organization for Standardization*. https://www.iso.org/standard/63500.html (accessed August 1, 2018).

[14] ISO 9241-12:1998 - Ergonomic requirements for office work with visual display terminals (VDTS) – Part 12: Presentation of information. https://www.iso.org/ standard/16884.html (accessed August 1, 2018).

[15] ISO 9241-125:2017 - Ergonomics of human-system interaction – Part 125: Guidance on visual presentation of information. https://www.iso.org/standard/ 64839.html (accessed August 1, 2018).[16] ISO/IEC 9126-1:2001 - Software engineering – Product quality – Part 1: Quality model. https://www.iso.org/standard/22749.html (accessed on August 1, 2018).

[17] R. Schnall, M. Rojas, S. Bakken, W, Brown III, A. Carballo-Dieguez, M. Carry, and J. Travers, A user-centered model for designing consumer mobile health (mHealth) applications (apps). *Journal of Biomedical Informatics* 60 (2016), 243-251.

[18] J.R. Lewis, Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **36** (1992), 1259-1260.

[19] B. Martínez-Pérez, I. de la Torre-Díez, and M. López-Coronado, Experiences and results of applying tools for assessing the quality of a mHealth app named Heartkeeper. *Journal of Medical Systems* **39** (2015).

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190654

Oncotherapy: A Decision Support System to Validate Oncological Treatments

María Laura Vera Righi[a], Pablo Martinez[b], Alcides Silva[b], Cati Umpierrez[c], Robinson Rodriguez[d]

[a]Department of Medical Informatics, National Cancer Institute, Uruguay
[b]ADN Engineers, Uruguay
[c] National Bank of Cytostatic, National Cancer Institute, Uruguay
[d]Medical Oncology, National Cancer Institute, Uruguay

Abstract

Making therapy decisions in oncology is a challenging task in the medical precision era. Oncotherapy is a decision support system that provides oncologists with suitable therapies for the patient within the national guidelines. The system is capable, on the one hand, to help the oncologist to maximize fitting therapy to the patient and, on the other hand, to provide control tools for the country´s sanitary authorities.

Keywords:

Clinical decision support system, Decision making, Medical oncology

Introduction

A clinical decision support system is a technological tool to help the physician to make decisions. Nowadays, choosing the best available therapy for a patient is a difficult task. We wanted to make an information system to help physicians to select the best option for each particular case, taking information from the accurate knowledge of the expert in the field enriched by the opinion of the oncologic community. It is a valuable assistance to reduce the occurrence of errors [1] in the physician's decision making process. This type of error is not the most frequent, but surely one of the most serious, because in most cases it does not allow the possibility of correction before any damage occurs. Oncotherapy also guarantees patients receive therapies in equal terms in the public health system. Besides, it allows the authorities to model the available therapies in the National Health System.

The objective of this study is to show a decision-making process in the prescription of oncological protocols assisted by a clinical decision support system called Oncotherapy, version 1.9.6.2.

Methods

Until June 2016, the National Cancer Institute collected, processed and sent by hand more than 2500 requests of series of therapy protocols every month. We have created an oncologic therapy engine with the guidelines of oncology therapies, based on the knowledge provided by the Guidelines Committee, guided by state-of-the-art oncology therapies. This engine will provide the available therapies according to the following three types of data: clinical diagnosis, clinical stage, and therapeutic criteria. The requests go through a designed workflow and may be automatically approved by the system if they match the criteria established by the engine. If the data do not match, the request is submitted to the consideration of a Therapy Committee and may be either approved or rejected. If it is rejected, the petitioner could resend this petition to the Coordinator of the Therapy Committee, who could approve or reject it definitely (see Figure 1). The main reason for rejection is to have another similar therapy in progress. Patients could receive more than one treatment at the same time or at different times, requested by either the same doctor or a different one. The procedure is allowed and controlled by the system. Each request is processed, by computing drugs, presentations, doses by patient which are packaged and sent to the medical centers where the patients receive oncological medical treatment. The request comprises a complete treatment and not one treatment cycle as it used to be done before. Every treatment is modeled within a period of time and it is required to confirm the request at certain intervals. The doctor receives an alert signal to proceed. In this way we are certain to send therapies that are still needed by the patient. The request could be ordinary or express depending on the timing that the treatment is needed according to the clinical situation. To date, we have processed 2260 urgent requests. The treatment could be suspended if it has already begun or cancelled if it has not been sent yet. In case the treating oncologist should be unable to find the intended therapeutic protocol by using the engine, a request to the Guidelines Committee could be sent, which may then enter new knowledge into the engine to enable the new option. In this way, the system allows a dynamic interaction between the Guidelines Committee and the country's oncologic community, providing the engine with feedback. One messenger service is activated inside the system to facilitate the communication among the oncologic community members, as well as with the Guidelines Committee. The Administrator profile is allowed to model the engine, feeding it with permits and constraints depending on the existing therapies in the National Health System. The solution is presented in oncological language and is assisted by colors and images to help the physicians to reach the desired function promptly.

The sanitary authorities could control the procedure and indirectly they also control the expenses in medical oncology.

Oncotherapy is made in a distributed environment, using web-based technology, supported with Java development and uses the PostgreSql® database engine. The application is supported and administered on a TomCat ® server.

Results

The decision-making about oncology therapy is a very challenging aim to achieve. Physicians participate in Tumor Committees, but these are not always possible in places far from the oncology centers. This type of clinical decision support system brings the knowledge universally in the same conditions wherever the patient is.

From July, 2016 to August 2018, Oncotherapy processed 14905 requests for 9300 patient made by 78 medical oncologists

distributed throughout the country. The workflow of the requests is shown in Table 1. 80.5% of requests follow the automatic approval without human intervention. Therefore, the administrative manual work has been eliminated and the number of members in the Therapies Committee has been reduced. More than 2000 therapies are currently in progress in the whole country and the medical request error is nonexistent (see Table 2). The diagnostic area of therapeutic intervention was distributed as Table 3 shows. The Guidelines Committee has a continuous intervention in the update of the engine model. The administrator sets the model depending on the Guidelines Committee resolutions and the sanitary authorities criteria. Indirectly, the cost of therapies was reduced by almost 20%, in the same period of time the number of patients increased due to the possibility to request therapies across the whole country. The medical informatics solution will contribute to creating a robust procedure to request oncology therapies according to the national guidelines in the National Health System.

Table 1 - Summary of Requests

Status	Request
Request	14905
Automatically Approved	12020
Requesting	145
Under Evaluation	10
Approved by the Therapies Committee	2475
Rejected	130
Temporary Rejection	15
Request after temporary rejection	12
Rejoinder awaiting answer	19
Approved by the Coordinator of the Therapies Committee	60
Rejected by the Coordinator of the Therapies Committee	40
Guidelines Committee Request approved	100

Table 2 – Summary of Therapies

Status Therapies	Sum
Under therapy	2105
Awaiting confirmation	525
Finished	7060
Suspended	120
Cancelled	705
Confirmed	3670

Table 3 – Summary of Diagnostic Area

Area	Sum
Gynecological cancer	7312
Urological cancer	3923
Digestive cancer	1585
Lung cancer	350
Head and Neck cancer	239
Skin cancer	66
Brain cancer	49
Soft Tissue cancer	41
Primitive unknown	10

Discussion

It is not an easy goal to create an IT (Information Technology) tool that helps physicians to do part of their technical job. Finding the suitable windows where the information system could be part of the routine without interrupting the clinical process is a very challenging work for medical informatics. It requires deep knowledge of the business and the technology to make it part of the clinical process. In order to reach this objective, it is very important to understand that technology is not the most important thing. It is better to focus on the involvement with the clinical situation with the doctor.

Until now we were working in terms of different dimensions such as patients, treatment, drugs, protocols of therapies and guidelines. However, our work has not finished yet. We are working on user profiles to use the physician´s dimension as a new filter to expand the use of the software to other cancer specialities and to other medical specialities, because the engine is useful in some other diseases. This will be included in the new version before the end of 2019.

The next step is to create the interfaces of the software to talk with other IT solutions in the national information system.

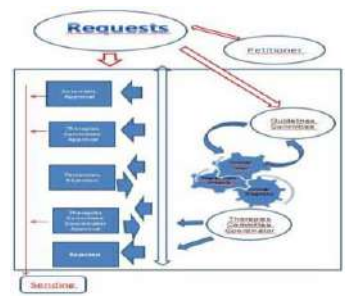


Figure 1 – Workflow of Request

Conclusions

We have created a clinical support system, where the oncological therapies engine is the most important input to the systematization of the decision-making process. This has been working successfully for 2 years in the control area of the National Cancer Bank of antitumor therapies. At the same time it is beneficial for the oncologic community to eliminate prescription errors. We have created a tool for the authorities to control the cost of the therapies, controlling the model and the spread, avoiding the waste of drugs.

Acknowledgements

To the authorities of the National Cancer Institute, specially to the director, Dr. Alvaro Luongo.
To Roche Industry Laboratories which contributed part of the cost of the development of the system.
To the Oncologic Community for contributing to the best usefulness of the engine.

References

[1] Bates DW, Gawande AA. Patient safety: Improving safety with information technology. *New Engl J Med*. **348** (2003), 2526-2534.

Address for correspondence

Laura Vera. E_mail: lverarighi@gmail.com.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190655

Age and Nationality: Two Variables That Challenge the Univocal Identification of People in the Health System of the Autonomous City of Buenos Aires

María Victoria Risoli, Mora Ruffo, Manuel Rodríguez, Belén Islas, Micaela Zapata, María Victoria Giussi, Analía Baum

Health Information Systems Office, Ministry of Health, Buenos Aires City, Argentina

Abstract

The univocal identification of people is fundamental for the longitudinal monitoring of health status. The impossibility of all people to credit their identity via their identity document generates temporary records that make it difficult to record the continuity of care. This paper analyzes the Patient Index of the Ministry of Health of the Buenos Aires City and describes how the age and nationality variables influence the creation of temporary records.

Keywords:

Health Information System, Health Plan Implementation, Data Collection.

Introduction

The univocal identification of a patient in the health information systems is key to guarantee the continuity of care [1-4]. The situation of the Autonomous City of Buenos Aires is not replicable to other places in the world where accreditation is a common and indisputable fact. In our context, the accreditation of identity is not a mandatory requirement since it is imperative to comply with current health laws. Law 153 “Basic Health Law of the City of Buenos Aires” implies the obligation to provide care to all citizens, residents or persons in transit within the City of Buenos Aires (CABA) [5]. There are many reasons that make it difficult to prove the identity of the entire population attending the facilities of the public system: without documentation [6], no identification at the time of contact or not willing to present it to State institutions.

In any case, it is known that the problem is not simply finding a unique patient identifier and its attributes, but the creation and maintenance of a Master Patient Index (MPI) with an associated patient identification service that supports and maintains [1-3]. In Buenos Aires City, the procedure through which the opening of an Electronic Health Record (EHR) is generated is called “registration” and is based on specific standards for the univocal identification of people [6]. These standards imply the search of candidates in the MPI before entering a new patient to avoid duplicates [1, 2, 6]. When the patient cannot prove identity, an EHR called “Temporary” is created. This condition should be exceptional in order to proceed to the patient's care and then be rectified by an admission process that identifies the patient and captures a document that enables the EHR to be “Permanent” [1]. The aim of this poster is to investigate the variables that are linked to the opening of EHR in a Temporary state and the characteristics of its use, particularly in cases where it does not act as a temporary situation and it is extended over time crystallizing the risks of the longitudinally of a non-valid record.

Methods

The Buenos Aires city healthcare network is conformed by Ministry of Health, 35 Hospitals, 74 Primary Care Centers (CESAC, CMB), 1 Ambulatory Reference Medical Facility (CEMAR), 2 children's dental centers, 2 health mental centers, It is structured into 12 geographical areas in order to organize health care delivery. Since June 2016, an EHR has been gradually implemented in the outpatient setting [7].

A mixed, quantitative and qualitative methodology was designed. In the first phase, data were requested and collected to confirm or rule out variables that could influence the creation of temporary EHRs provided by the Information and Statistics Management sector of the Ministry of Health.

We analyzed the nationality, the place of residence (divided into the Autonomous City of Buenos Aires, the Province of Buenos Aires and others), the age range, plus the place where it was registered (Primary Care Center or hospital). In addition, the amount of EHRs in the temporary state and the percentage of temporal registers rectified were analyzed.

Results

Among temporary records, the variables nationality and age range present differences in relation to the total number of registered patients, while patient's address and the type of facility did not show significant differences in terms of the percentage of temporary registrants.

Nationality

In August 2018, the proportion of temporary records for people with a foreign document was 19.49% compared to 9.07% of records with an Argentine document.

Age

In the last six months, an increase in number of the temporary records has been observed, especially in children under one year of age. In the last six months there has been an increase in Temporal EHR (Figure 2): in August 2018 the percentages of registered as temporary in children under 1 year, children between 1 and 6 years, children in 6 and 18 years and over 18 years old, were 18%, 7.14%, 9.59% and 8.72%, respectively.

This coincides with the implementation of SIS in acute general hospitals that have maternities.

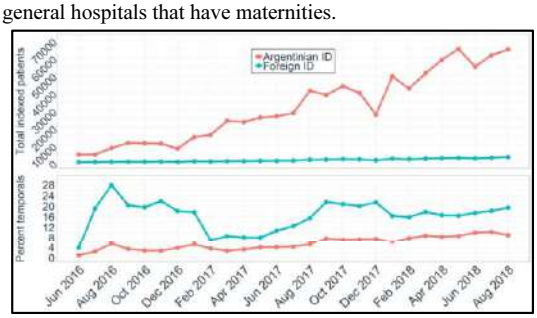


Figure 1: Total number of patients registered as temporal and percentage of the total number of registered patients per month and nationality

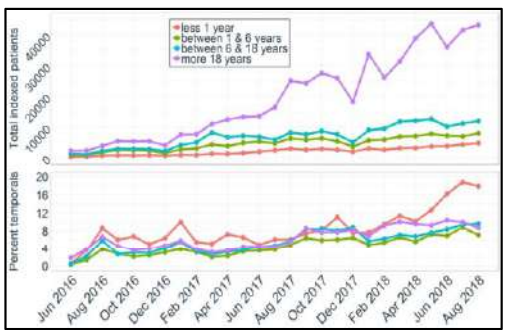


Figure 2: Total number of registered patients and percentage of temporal records per month and age range.

Of the total number of patients registered as temporary (N = 65734 = 6.6%), 10.3% was rectified (N = 6751). 36.4% had at least one clinical note in their Temporary EHR (Table 1).

Table 1 - Amount of temporary and rectified records without clinical notes, with one note and two or more notes

	Temporary		Rectified	
Consultations	**n**	**%**		**%**
0	41,809	63.60	3485	8.34
1	11,169	16.99	1131	10.13
>2	12,767	19.41	2135	16.74
	65,734		6,751	

Conclusions

Regarding age, children under one year usually do not have a document because the parents have not yet completed the process or sometimes because they do not bring it with them to the pediatric outpatient clinic [7]. In order to minimize this fact, the administrative staff were asked to register the child with a permanent status using the parent's document. Regarding the temporal-nationality relationship, it may be due to the fact that in our country, immigrants are in a very vulnerable social situation, but the true cause should be deepened. We were not able to explain the fact that some temporary records are not associated to a visit, but we believe that the creation of a temporary records could be shortcut used by the administrative staff to expedite attention times, since this process is faster than creating a permanent patient record [1-6].

Limitations and future directions

We need to perform qualitative explorations that allow us to determine and better understand why the temporary records are created. Understanding the causes will allow us to develop communication strategies and training or modification in the processes so that these obstacles are not an impediment that lasts over time. Further qualitative explorations are needed to better understand this phenomenon.

References

[1] L. Garfi, P. Navajas, A. Gómez, D. Luna, and F. González Bernardo de Quirós, Implementación de un sistema centralizado para la identificación de pacientes en un hospital de alta complejidad, 2002. https://www.hospitalitaliano.org.ar/multimedia/archivos/servicios_attachs/1125.pdf (accessed November 8, 2018).

[2] M. Maira, P. Navarro, L. Biaggini, G. Morales, et al., Rediseño del Sistema de Información de una Red Integrada de Salud de cobertura nacional en Chile, Proyecto SIRIS, 2008. https://www.hospitalitaliano.org.ar/multimedia/archivos/servicios_attachs/4821.pdf (accessed November 8, 2018).

[3] World Health Organization, Identificador Único de Pacientes, World Health Organization, 2018. http://www.paho.org/relacsis/index.php/noticias/item/804-identificador-unico-de-pacientes[3] (accessed November 8, 2018).

[4] A.W. Group, Managing the Integrity of Patient Identity in Health Information Exchange (2014 update), *J. AHIMA*. **85** (2014) expanded web version. http://library.ahima.org/doc?oid=300436#.W-SKqpNKiUl (accessed November 8, 2018).

[5] Gobierno de la Ciudad de Buenos Aires, Ley 153 - Ley básica de salud de la ciudad de Buenos, Ciudad de Buenos Aires, 1977. http://www.buenosaires.gob.ar/sites/gcaba/files/ley_basica_de_salud__153.pdf (accessed November 8, 2018).

[6] C. Burgo González, M.V. Giussi Bordoni, A. Stieben, and A.J. Baum, Estrategias en la implementación de estándares para la identificación de pacientes sobre un software legacy en el Sistema de Salud de la Ciudad de Buenos Aires, (2018) 21–27. http://47jaiio.sadio.org.ar/sites/default/files/cais-4.pdf.

[7] M.V. Giussi, A. Baum, F. Plazzotta, P. Muguerza, and F. González Bernaldo De Quirós, Change management strategies: Transforming a difficult implementation into a successful one, *Stud. Health Technol. Inform.* **245** (2017) 813–817. doi:10.3233/978-1-61499-830-3-813.

Address for correspondence:

María Victoria Risoli

Ministry of Health of Buenos Aires city

victoriarisoli@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190656

Design, Implementation and Adoption of an Electronic Dental Record Within an Electronic Health Record in the Public Healthcare System of Buenos Aires City

Stok Capella JF, Zubillaga MJ, Nero FG, Muguerza P, Lanuza J, Alassia L, Rodriguez Tablado M, Giussi MV, Baum A

Health Information Systems Office, Ministry of Health, Buenos Aires City, Argentina

Abstract

This paper describes the development and implementation of a Dental Record within an Electronic Health Record in the public Primary Health Care network of Buenos Aires City. In the five months of implementation, the adoption was progressively increasing but with a great deal of variability among the health facilities. Pediatric population was the most common assisted and the most frequent oral health problem was caries.

Keywords:

Dental Health Services, Electronic Health Record, Dental Record

Introduction

Oral health has been an issue of great concern for many years. Reduction of risks to oral diseases is only possible if services are oriented towards primary health care and prevention [1-2].

Control of oral disease depends on availability and accessibility of oral health systems. In developing countries, oral health services are mostly offered from regional or central hospitals of urban centers and little, if any, priority is given to preventive or restorative dental care. Even more, many countries in Latin-America have a shortage of oral health personnel, and by and large, the capacity of the systems is limited to pain relief or emergency care [1].

International statistics indicate that the main problem of oral health in underdeveloped countries is dental caries, 60% and 90% of young students and 92% of the adult population [3].

There are no known studies that provide data on the oral health status of the adult Argentinian population [4].

A dental electronic health record has become a priority in many countries. However, experience shows that any dental electronic health record system cannot be subordinate to, or a subset of, a medical record.

Optimal, oral health care needs integrated dental, medical, and behavioral health information into an electronic health record and be readily available to health providers [2-4]. The Buenos Aires City public healthcare network has a unique opportunity for the integrated Electronic Dental Record because oral health professionals work at a primary care level where an Electronic Health record has been implemented for the last two years [5].

Methods

Setting

The Buenos Aires city healthcare network is conformed by a Ministry of Health, 35 Hospitals, 74 Primary Care Centers (CESAC, CMB), 1 Ambulatory Reference Medical Facility (CEMAR), 2 children´s dental centers, 2 health mental centers, and its structured into 12 geographical areas in order to organize health care delivery. Since June 2016, an Electronic Health Record (EHR) has being gradually implemented in outpatient. [6] Until October 2018, 60 healthcare facilities have been using EHR and more than 2.5 million of clinical notes had been registered. This EHR incorporated dental record capabilities in June 2018. The dental network [7] delivers care in 44 primary care facilities, 26 general hospitals and 3 specialized dental care hospitals which provide emergency services 24 hours a day. There are approximately 250 dental care professionals supporting health care in this network.

Software Development Process

There was an interdisciplinary team focused on designing a tool based on the professionals´ needs. Then, three iterative cycles of design, development and testing of the electronic dental record took place. The Electronic Dental Record (EDR) was developed within the Electronic Primary Care Health Record System.

In order to know the users' needs, interviews of dentists and observations of dental practices were carried out in primary care facilities. Then, a tailor-made implementation plan was designed taking into account the dental care processes within the different facilities of the network.

Results

Between January and June-July of 2018, three iterative cycles of design, development and testing of the electronic dental record were completed. The design and development of the tool was divided into three phases. The first phase required two full iterations. In the first iteration, the goal was to complete the minimal required functionalities needed by dentists to replace a paper based, dental record, allowing the registration of the current and historical status of the denture as well as the

planned treatment. As a result of this first iteration, an alpha version was released and evaluated by an interdisciplinary team and a backlog of minimal required corrections and improvements were defined. A second iteration was completed in order to incorporate these changes, resulting in a production-ready version by June 2018.

In the second phase, which required one full iteration, non-blocking but desired features were added to the EDR: the possibility to register supernumerary teeth and the live calculation of epidemiological indicators for dental cavities.

The third phase, which is under development, is aimed to the specific registry of oral mucosal health.

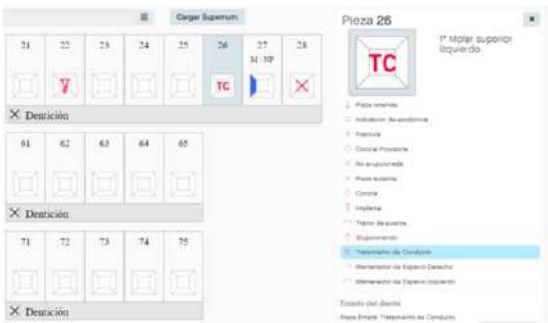


Figure 1. Dental Record Interface

Interdisciplinary team and implementation strategy

Participation in the interdisciplinary team included 2 computer scientists and 3 specialists in health informatics with different disciplines (sociology, dentistry and medicine).

In order to understand the needs of the oral health registry and gather information of importance for the design and development of the electronic dental registry, 30 interviews with dentists were conducted.

Chiefs of the primary care centers and dental network coordinators also participated in the interviews to find out what information was critical for public policy decision-making.

Fifteen implementers were trained to give instructions about the use of the dental record to 100 primary care dentists from June to October 2018. The trainings were carried out in a personalized way with each professional and a passive support time was destined to evacuate doubts from the use of the tool.

Electronic Dental Record Adoption

Between June and October of 2018, the EDR was implemented in 30 Primary Care Facilities (CESAC).

Conclusions

Oral disease is a health care problem and not solely a dental one. This is why the dental record cannot be isolated from the health record. Optimal, oral health care needs integrated dental, medical, and behavioral health information into an electronic health record and be readily available to health providers. We did that integration but future investigation is needed to understand causes associated with oral health and design policies that improve the oral health care of the population in this sense.

References

[1] Strategies and approaches in oral disease prevention and health promotion, WHO. (2010). http://www.who.int/oral_health/strategies/cont/en/ (accessed November 20, 2018).

[2] HHS, and CDC, Oral Health Strategic Framework, 2014-2017, 2014. https://www.hrsa.gov/sites/default/files/oralhealth/oralhealthframework.pdf (accessed November 20, 2018).

[3] Argentina Investiga, Un estudio epidemiológico sobre salud bucal inédito en el país, (2014). http://argentinainvestiga.edu.ar/noticia.php?titulo=un_estudio_epidemiologico_sobre_salud_bucal_inedito_en_el_pais&id=2081 (accessed November 20, 2018).

[4] E. Kalenderian, J.D. Halamka, and H. Spallek, An EHR with Teeth., Appl. Clin. Inform. **7** (2016) 425–9. doi:10.4338/ACI-2015-09-LE-0124.

[5] M. Bordoni Giussi, F. Plazzotta, A. Baum, C. Ilc, and F. González Bernaldo de Quirós, Elaboración e Implementación de una agenda digital en atención primaria en la ciudad autónoma de buenos aires, Goiânia - Brasil, 2016. www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis (accessed November 8, 2018).

[6] M.V. Giussi, A. Baum, F. Plazzotta, P. Muguerza, and F. González Bernaldo De Quirós, Change management strategies: Transforming a difficult implementation into a successful one, Stud. Health Technol. Inform. **245** (2017) 813–817. doi:10.3233/978-1-61499-830-3-813.

[7] Odontología | Buenos Aires Ciudad - Gobierno de la Ciudad Autónoma de Buenos Aires, (n.d.). http://www.buenosaires.gob.ar/salud/odontologia (accessed November 20, 2018).

Address for correspondence

Juan Francisco Stok Capella
Ministry of Health, Buenos Aires city
jstokcapella@buenosaires.gob.a

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190657

Development of a Virtual Reality System for Early Mobilization of Critically Ill Patients

Jiani Wang[a], Chunyan Zhang[b], Yanrui Jia[b], Chenxi Shi[b], Thomas Choi[c], Qian Xiao[a]

[a] *School of Nursing, Capital Medical University, Beijing, China*
[b] *Beijing Chaoyang Hospital, Capital Medical University, China*
[c] *Centre for Smart Health, School of Nursing, The Hong Kong Polytechnic University, HKSAR, China*

Abstract

More and more researchers have recommended critically ill patients to start mobilization as early as possible. However, the clinical utilization rate of early mobilization remains low for patients in the intensive care units (ICU) because of various factors. In order to promote the rehabilitation of critically ill patients, a multidisciplinary research team, including academic researchers, ICU head nurses, respiratory therapists, and a software engineer, has developed a virtual reality system for early mobilization in ICU. This system has four main features-the diverse forms of mobilization based on muscle strength, the integration of exercise and cognitive training, the visualization of the mobilization process and the record of the trajectory during mobilization exercises. This paper presents and discusses the development process of this system.

Keywords:

Critical illness, early mobilization, virtual reality

Introduction

With the advancement of medical technology, researchers concern not only the increase in survival rate with technology, but also the quality of rehabilitation of critically ill patients[1]. Traditionally, rehabilitation begins after patients are discharged from ICU, which could be too late for the recovery of physical and cognitive functions, causing a majority of ICU patients to lose their ability of independent living and the quality of life[2]. In recent years, researchers have suggested that early mobilization for critically ill patients during treatment in ICU can circumvent many complications, reduce the hospitalization length and improve their quality of life[3]. However, the current implementation of early mobilization in ICU is not well administered. The main reasons are due to the shortage of medical staff, equipment and other resources, and more importantly, patient's unwillingness to exercise. Therefore, it is necessary to explore methods to improve the motivation and compliance of critically ill patients towards early mobilization. Recently, many researchers have used information technology to assist patients in the rehabilitation process, where virtual reality technology, as novel information technology, has the potential to meet the needs of promoting early mobilization concerned with this research. Hence, our team has developed a virtual reality system to assist ICU patients and record their progress in early mobilization. The system is presented and discussed in this paper.

Methods

A multidisciplinary research team including academic researchers, ICU nurses, respiratory therapists, and a software engineer was set up. The team had been involved in designing early mobilization strategies for ICU. ICU nurses and respiratory therapists have rich experience in critical care and early mobilization.

Stage 1: Alpha Version of the System

After reviewing literatures on ICU early mobilization and making reference to clinical experience, our research team came to the conclusion that early mobilization of ICU patients would be administered based on the muscle strength of their upper and lower extremities as measured by the MRC Muscle Scale; and that mobilization would be performed step by step. Then, the functions of the system were determined, which included diverse forms of mobilization based on muscle strength, integrating exercise and cognitive training, the visualization of the mobilization process and the record of the trajectories during mobilization exercises.

Stage 2: Beta Version of the System

The alpha version of the system was refined based on the result of stage 1. It was re-programmed to increase the ease of use by providing a user-friendly interface. The fidelity of the system was also improved by modifying the algorithms and parameters of the computer program so as to better capture the motion of the lower limbs.

Results

The virtual reality early mobilization system for critically ill patients has been developed after several rounds of discussion and modification. Before using this system, the medical staff is required to assess the patient's basic vital sign and muscle strength. In the first page of the system, medical staff should fill in patients' name, age, gender and muscle strength (Figure 1). Based on the inputs, they are provided with suggestions on the form of early mobilization, the dosage, and intensity that are suitable for the patients. The four basic functions of the system are described as follows.

Figure 1– First page of the system

Function 1: Diverse forms of mobilization based on MRC

This function is to identify the best form of early mobilization for patients according to their muscle strength, which is depicted in Figure 2. For example, patients with an MRC Score

of 2 for their upper extremity muscle strength need to complete plane movement, fisting activity, and wrist flexion.

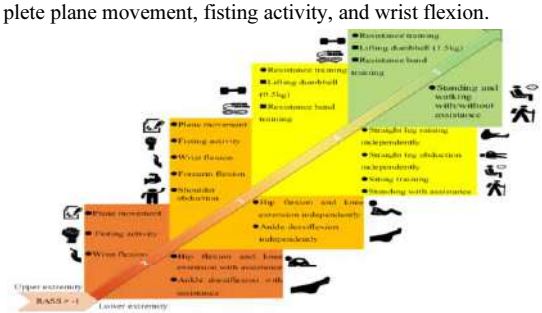


Figure 2–Different mobilization strategies

Function 2: Integration of exercise and cognitive training

With this system, every patient is suggested to receive various physical and cognitive training. For instance, patients with an MRC Score of 2 for upper extremity muscle strength need to complete plane movement integrated with cognitive training, in which patients are guided to write or draw using a visual sensing device. When the MRC Score for the upper limb is greater than 3, a variety of exercise games are recommended for the patients to choose. The games require patients to make a decision and stimulate their thinking while motivating them to exercise more and making it an enjoyable process as a whole.

Function 3: Visualization of the mobilization process

After adjusting the devices and computer, data about the movement of the critically ill patients can be transmitted to a laptop, where all the processes of the mobilization can be visualized. In some games, the virtual person appearing on the screen could follow the actions of the patient. Standard exercise indicators are displayed on the screen to assist patients to judge by themselves whether their movements are correctly performed. Furthermore, in the games of playing piano, jumping frog and others, patients can gradually increase the intensity of exercise and target at the higher game score, which is helpful to increase the interest and motivation of early mobilization. By adding algorithms into the program, the ability of patients in meeting the criteria of mobilization can be judged and the frequency of exercise can be recorded.

Function 4: Record of trajectory during exercises

The trajectory during mobilization exercises can be recorded by the system by capturing the patient's movement with a camera. In the later stage, the trajectories of daily mobilization exercises can be compared, which can fully reflect the rehabilitation trend of the patient.

Discussion

This is the first study to develop a virtual reality system on early mobilization for critically ill patients. It is well known that there are many barriers in implementing early mobilization for patients in ICU, such as patient-associated factors, structural issues, process factors, and cultural factors. Among these factors, patient's willingness is the most important factor that is difficult to deal with. According to clinical investigation, patients with critical illness often show frailty, indifference, and melancholy, which are partly due to the physical diseases and the psychological status of the patients. Medical staff should recognize the true state of patients, and then remove the discomfort from patients, and actively encourage them to exercise as early as possible. However, ICU medical staff already have a heavy workload. It is difficult to carry out early mobilization for the patients. The development of this system can help improve compliance of critically ill patients on early mobilization. The usual care in ICU could make patients feel boring, making it difficult to adhere to and achieve the effect of rehabilitation. The proposed system utilizes virtual reality technology to enable patients to gain interesting experience with various forms of exercise, making them more willing to participate actively in rehabilitation therapy. This system can also reduce the workload and pressure on the medical staff.

One of the innovations of the system is the integration of exercise and cognitive training. Although some studies tried to combine early mobilization and cognitive training, the two interventions were separate and patients have to receive them independently. In addition, exercise games are goal-oriented, giving patients a good sense of achievement when they achieve their goals. More importantly, the usage of the system does not require patients to wear or carry any objects. Patients can use their fingers as a pen to interact with computer programs, thus making them feel more natural when doing the exercises.

Another innovation of the system is to record the trajectory of exercises for critically ill patients. The dynamic changes of patients' movement are often difficult to capture and describe clearly. This system allows for the visualization of patients' movement during the whole process, which can also be recorded for analysis. With this function, the patient can understand their progress and the effect of rehabilitation more intuitively, which is beneficial to enhance their belief and confidence in early mobilization. For medical staff, the system provides more detailed data and record about the progress of recovery.

A limitation of this study is that it targets only at patients who are awake and with moderate muscle strength, For patients in coma or with lower MCR Score, medical staff is still required to perform passive or assistive exercise intervention for the patients. We will conduct research on this aspect in the pursuit of research breakthroughs.

Conclusions

This study described the development process of a virtual reality system by a multidisciplinary research team for early mobilization in ICU. The system is developed with novel design concepts and practical features. It can be applied to promote early mobilization of critically ill patients. This system motivates patients to exercise with engaging designs and records the trajectory of mobilization and rehabilitation. A further improved version of the system will be developed after addressing the comments from clinical experts and patients. The system has the potential to enhance the rehabilitation and the quality of life for ICU patients. It is expected to play an important role in clinical applications.

References

[1] A. Messer, L. Comer, and S. Forst, Implementation of a Progressive Mobilization Program in a Medical-Surgical Intensive Care Unit, *Crit Care Nurse* **35** (2015), 28-42.

[2] D. McWilliams, C. Jones, G. Atkins, J. Hodson, T. Whitehouse, T. Veenith, E. Reeves, L. Cooper, and C. Snelson, Earlier and enhanced rehabilitation of mechanically ventilated patients in critical care: A feasibility randomised controlled trial, *J Crit Care* **44** (2018), 407-412.

[3] M. Paton, R. Lane, and C.L. Hodgson, Early Mobilization in the Intensive Care Unit to Improve Long-Term Recovery, *Crit Care Clin* **34** (2018), 557-571.

Address for correspondence

Qian Xiao, Email: julia.xiao@163.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190658

Twenty Plus Years of Distance Learning: Lessons Learned

Elizabeth Weiner[a], Ryan McNew[a], Jeffry Gordon[a], Patricia Trangenstein[a], Keith Wood[a]

[a] *Frist Nursing Informatics Center, Vanderbilt University School of Nursing, Nashville, TN, USA*

Abstract

Private institutions of higher education in the United States were hesitant to institute programs of distance learning for fear that they could not maintain the quality of the education they had delivered in face-to-face programs. Vanderbilt University allowed their School of Nursing to embark on such an endeavor in 1996, as long as quality measures were incorporated. The result has been a comprehensive resource support team using Quality Matters and an increase in program rankings.

Keywords:
Distance Education, Online Learning, Online Education

Introduction

In the 1990s, the Masters of Science in Nursing (MSN) program at Vanderbilt University School of Nursing (VUSN) was a relatively small, high quality, boutique program solely aimed at students within commuting distance of the campus. Growing the program required steps to accommodate a different type of student—the distance learner—who in turn brought about a completely new set of challenges.

The first specialty to implement a modified distance option was the Psychiatric Mental Health Nurse Practitioner program in the fall of 1996. They started by video streaming classes while simultaneously recording sessions for later distribution. Bandwidth in many of the rural areas prohibited quality viewing, so the IT support staff of five resorted to pressing CDs and mailing them out within 24 hours of the class.

Dean Colleen Conway-Welch determined that leadership was needed to guide the school through the various barriers to deliver quality online education. She had the vision to hire a Senior Associate Dean for Informatics. That individual assumed the responsibility of guiding the expansion of the program using "at a distance" tools and then identified and hired the additional staff needed for support. Support and quality became the cornerstones of all subsequent planning, implementation, and evaluation.

While the initial intent was to expand enrollments because the regional area had been saturated with nurse practitioner graduates, it soon became clear that other programs were seeing rising student satisfaction studies with online learning as faculty became more skilled with the online learning environment [1]. Furthermore, the quality of online courses was becoming the same as, or better than, traditional face-to-face courses [2].

This poster will explore the lessons learned during this 23-year journey. The design of the poster will be completed using innovative interactive software that will allow users to use smartphone technologies to further explore various subjects using augmented reality to visualize additional multimedia materials.

Methods

Staffing

When the Senior Associate Dean for Informatics was hired, she shared her vision for a comprehensive technology and informatics support team. As a condition of her employment, she received institutional commitment that a team approach would be employed to provide the best support model, prepared in a range of skill sets from networking to instructional design.

Technology Tool Set

Each year the informatics support team has chosen the hardware and software solutions that best meet the needs of the faculty. Decisions are made on which tools will be supported so that duplicate videoconferencing solutions, for example, do not exist. All software purchases have been centralized under the informatics area, including grant purchases. This decision has resulted in more effective support and more efficient purchasing.

Student and Faculty Orientation

Orientation activities are crucial to every academic year, and are presented both online and during the face-to-face sessions at the beginning of the academic year. Configuration sessions are critical so that all devices are equipped with the appropriate software tools and can attach to the university network. Once students are at home, they need to be able to make the appropriate connections so that they can successfully use their technology tools for learning. Students are instructed to call or email for any support or connection issues once they are back in their home environment.

Quality Standards

For the last ten years, VUSN has subscribed to the Quality Matters (QM) framework. Skiba describes this framework as being built upon eight standards derived from current research [3]. The framework is based on the assumption that quality online education is represented by institutional commitment to quality. Two instructional designers trained in the QM framework work with faculty to help them meet the criteria incorporated in the QM rubric [4]. The rubric is used to examine eight general standards for course design. The rubric also incorporates the concept of alignment, which refers to several essential course components working together to achieve desired outcomes.

Results

Staffing

Five technology support people were at VUSN when the Senior Associate Dean for Informatics was hired in 2000. Implementing her vision for comprehensive support, the team has now grown to 28. Positions include classroom support, computer lab coordinator, network manager, web developer, graphic artists, general IT support, instructional designers, program coordinator, videographers, materials coordinator, videographers, media services, programmers, simulation staff, and informatics faculty. Enrollments have more than tripled, reaching as high as 1000 students in the MSN, DNP, and PhD programs.

Technology Tool Set

In order to effectively manage on-site delivery of the technology, the network infrastructure was strengthened with all classrooms and offices having wired and wireless connections. Electronic classrooms have been updated to include the Crestron control system [5]. Hardware guidelines are agreed upon annually and communicated to students along with the software.

Software solutions include:

- digital video streaming via Mediasite,
- web conferencing via BlueJeans,
- clinical log via Medatrax,
- secure large file exchange via Box, s
- ecure testing via Remote Proctor,
- video case studies via ReelDx ,
- simulation software via B-Line Medical,
- instructional design software via Lectora,
- plagiarism checker via Turn-It-In,
- digital signatures via Adobe Acrobat,
- remote support via Team Viewer,
- course management via Brightspace, and
- Microsoft Office Productivity suite.

Student and Faculty Orientation

VUSN teaches courses year-round so that the August orientation starts the beginning of all classes. During this time, each program has week-long orientation activities. Reviewing technology standards is important to all programs, along with the configuration of all their devices to connect back to VUSN from anywhere in the world. New faculty are oriented online via technology support and solutions, and returning faculty are updated at the annual Fall Faculty meeting. Further support is provided online including a Tech Tools section of the VUSN website along with a Knowledge Base.

Quality Standards

The Institute of Medicine defines a learning healthcare system as a system in which "science, informatics, incentives, and culture are aligned for continuous improvement and innovation, with best practices seamlessly embedded in the delivery process and new knowledge captured as an integral by-product of the delivery experience [6]." The Learning Health System Cycle is further described by Friedman as being cyclical in nature with the steps of "assemble, analyze, feedback, and change [7]." This same process has been used by VUSN through course evaluations at the end of each course, curriculum evaluations through course portfolio reviews across all programs, and surveys regarding distance learning support completed by faculty and staff every two years.

All MSN and DNP courses have introduced an interactive course template that guides them through the appropriate planning stages that maps the objectives with outcomes. Current efforts are on increasing course interactivity, and the QM rubric is being used internally to help all faculty improve the quality of their courses. Feedback from students is received via standardized course evaluations along with a survey sent out to students every other year about the technology support. All evaluations have been extremely positive (4.95 on a 5 point scale) with minor suggestions for improvement.

Program rankings have all risen during the 23 years of online learning. Rankings in 2019 for VUSN were eight for the MSN program, and five for the DNP program, marking the first time that both programs have been in the top 10 [8]. In addition, throughout the 23-year period all VUSN programs have maintained accreditation by national nursing bodies, demonstrating that by offering education in an online fashion quality was at least equal to, if not higher than, face-to-face delivery of programs.

Conclusions: Lessons Learned

The lessons learned from over twenty years of experience can be summarized by the following important considerations. Quality (from an organizational commitment and framework), comprehensive support team, leadership at the Senior Associate Dean level, a strong network infrastructure, standardized hardware and software, student and faculty orientation activities followed by web tools, and feedback from users including both faculty and students followed by revision of activities. Late lessons included the need for a remote proctor tool with browser lockdown, the addition of a searchable knowledgebase, and the need to be active participants at the university level for campus-wide selection of online tools. Maintaining faculty engagement continues to be a challenge due to a number of off-site classes rather than being taught from faculty offices.

References

[1] C.J. Thompson, Disruptive Innovation in Graduate Nursing Education: Tipping the Scales for Online Education., *Clinical Nurse Specialist.* **30** (2016), 304–307.
[2] K. McCutcheon, M. Lohan, M. Traynor, and D. Martin, A systematic review evaluating the impact of online or blended learning vs. face-to-face learning of clinical skills in undergraduate nurse education, **71** (2015), 255-270.
[3] D.J. Skiba, Quality Standards for Online Learning., *Nursing Education Perspectives.* **38** (2017), 364–365.
[4] Quality Matters, https://www.qualitymatters.org/, 2018. Accessed March 25, 2019.
[5] Crestron, https://www.crestron.com, 2018. Accessed March 25, 2019.
[6] Institute of Medicine, *The Learning Healthcare System: Workshop Summary*, The National Academies Press, Washington, DC, 2007.
[7] C. Friedman, J. Rubin, J. Brown, M. Buntin, M. Corn, L. Etheredge, C. Gunter, M. Musen, R. Platt, W. Stead, K. Sullivan, and D. Van Houweling, Toward a science of learning systems: a research agenda for the high-functioning Learning Health System, *JAMIA* **22** (2015), 43-50.
[8] US News and World Reports Rankings, https://www.usnews.com/best-graduate-schools/top-nursing-schools, 2019. Accessed March 22, 2019.

Address for correspondence

Elizabeth (Betsy) Weiner, Vanderbilt University School of Nursing, 461 21st Ave. S., Nashville, TN 37240-1104, USA; E-mail: Betsy.Weiner@Vanderbilt.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190659

A Precision Post-Operative Wellness Monitoring Solution

Nilmini Wickramasinghe[a,b], Vijay Gehlot[c], Elliot Sloane[c], Phil Smart[b,d] and Jonathan Schaffer[e]

[a] *Swinburne University, Australia*
[b] *Epworth HealthCare, Australia*
[c] *Villanova University, USA*
[d] *The Austin, Australia*
[e] *The Cleveland Clinic, USA*

Abstract

Multiple orthogonal challenges around escalating costs and providing quality care plague healthcare delivery, especially in OECD countries. This research in progress paper addresses the post-operative discharge phase of the patient journey and proffers a technology-enabled model that both supports a quality care experience post-discharge but also prudent management to minimise costly unplanned readmissions and thereby subscribe to a value-based care paradigm. The chosen context is stoma patients but the solution can be easily generalised to other contexts. Next steps include the conduct of clinical trials to establish proof of concept, validity and usability.

Keywords:

Patient Readmission, Patient-Centered Care, Patient Monitoring

Introduction

Given the challenges facing private healthcare today, there is an increasing pressure on private healthcare organisations to provide high value, high-quality patient-centred care across the acute-care continuum (Australian Commission on Safety and Quality in Healthcare 2010 [1]). While the recognition for the need of care delivery to be patient-centric is growing, the appropriateness of the approaches adopted to achieving this endeavour remains questionable [2].

An integral enabler is without question Information Technology (IT) solutions (see for example [3]). The limitation with many current systems is their limited coverage across the acute-care continuum [4], where both pre-admission and post-discharge phases are not seamlessly connected to the hospitalisation phase in the patient journey.

In Australia, as in other OECD countries, most notably the US, unplanned readmissions are now becoming more carefully scrutinised. In most instances in Australia, unplanned readmissions are considered to be readmissions for issues relating to the primary diagnosis within 28 days of the initial treatment for that primary diagnosis. We believe that it may be possible to reduce the number of unplanned readmission by developing a precision post discharge wellness monitoring solution and the following serve to outline this solution.

We select stoma patients as a pilot study for this solution because we note that based on hospital data gathered from a large not-for-profit tertiary institute in Melbourne, Australia a common and avoidable unplanned readmission relating to stoma patients is around lack of hydration. This is particularly problematic during the hot dry Australian summer months.

This study proposes a generic, open-source-based starting point for a customizable modeling, simulation and testing framework for mobile Patient Care Devices (PCDs) that can support relatively complex coordination of care for post-surgical patients. This system uses an architecture that can be modified to suit each patient's precise needs or widely differing clinical protocols. These tools allow simulation of expected ranges of safe and reliable performance, and can also be used to explore or simulate likely failure modes and potential safety or health risks due to communication system or staffing overload, errors, or other complications

Clinical Use Case

In order to provide proactive patient care post-discharge following stoma surgery, several remote medical devices can be employed. For example, elevated patient temperature or pulse can indicate an emerging infection at home which might easily and inexpensively be treated by early intervention with an appropriate antibiotic. High blood pressure and pulse might be an indicator of patient pain or discomfort, which might be initially treated with basic over-the-counter anti-inflammatory medications. Low blood pressure and elevated pulse might be predictors of dehydration, which might readily be treated by drinking more fluids and electrolytes. The stoma bag may also be monitored with simple sensors that keep track of filling and emptying rates, providing an indication of inadequate food and liquid intake.

In all cases, the monitored patient wellness parameters can be routed to a central patient homecare coordination team. The care coordination team can invoke appropriate rules and treatment actions. For example, a care coordinator could call the patient/family, provide remediation guidance, education, and/or prescriptions. If needed, a visiting nurse or physician could be dispatched to provide in-home care.

In addition, the monitors and the data monitoring systems can have an alarm and/or alert level triggers pre-set or remotely adjusted. Thus, if a low-grade fever appears to be emerging, the patient, family, or care coordination team could increase the temperature alarm by one degree, to notify them if/when the fever becomes more severe. Similarly, a high and low blood pressure alarm could be pre-established based on the patient's

discharge condition, in order to alert caregivers of unusual emerging risks.

Once a patient's physiological data is available to the care coordination system and team, clinical decision support algorithms and systems can be used to enhance, accelerate, and escalate emerging patient risks to staff who are appropriately trained to support high-quality, safe home care for discharged patients. Intervention at the home will usually be far less expensive than re-admission to the hospital, and care plan changes can often occur quickly. The care coordination team will be able to arrange emergency care or transport if home-care turns out to be inadequate for a specific patient situation.

Because many physiologic monitoring devices are now becoming rather inexpensive and ubiquitous, and they rely on consumer-grade internet or cellular communication channels, the incremental cost of deploying such systems is falling rapidly. In addition, many of these devices can be cleaned and re-used for subsequent patients.

In the following sections, we illustrate the system design and simulation of a flexible home-care monitoring and care coordination system. We identify representative monitoring devices, but the model is extensible. Additional monitoring devices and data can easily be added. e.g., in some cases, patient weight may be a valuable indicator of dehydration, or of congestive heart failure complications. Adding a patient scale and decision support rules is very easy with this system.

In addition, this system could easily be extended to include patient co-morbidities. For example, a severe industrial or automobile trauma patient could conceivably be sent home with both a stoma bag and a hip replacement. If that were the situation, additional physiologic channels may be added, such as gait and PT/exercise/mobility tracking and analysis.

Results and Discussions

The previous section has served to proffer an appropriate solution leveraging the capabilities of various technologies. The aim is to monitor patients as unobtrusively as possible so that alerts can be triggered to inform the designated healthcare professional if a trigger incident has occurred that should be addressed. In this way, the patient is able to navigate the post-discharge phase of their treatment effectively and efficiently and with the highest level of a positive patient and caregiver experience. In addition, by having alerts triggered no sooner a trigger situation arises, it is also possible to act as quickly as possible, thereby averting a more complicated, dangerous, and/or expensive problem further down the track. In this way, we believe we are addressing two critical objectives of healthcare delivery simultaneously; namely providing a high-quality patient experience as well as providing a high-value solution as we are trying to mitigate the need for unplanned readmissions by catching trigger situations as early as possible and then addressing them.

The next step is to run a two-arm non-blinded clinical trial to test the full benefits of the proposed solution. The control arm will continue to have patients exposed to standard practices around discharge and post-discharge follow up and monitoring while the intervention arm will focus on utilising the developed technology solution in addition to standard care practices around discharge and post-discharge. In particular, we plan to monitor levels of hydration, using triggers of blood pressure, pulse and weight to trigger levels going below an appropriate threshold. In addition, patients will receive via their mobile phones, education, reminders and other important information about maintaining appropriate levels of hydration. We focus on hydration as this has been identified as the singular most frequent reason for unplanned readmission with stoma patients at the chosen healthcare facility. We have secured ethics committee approval, and plan to run this trial as soon as final clinical post-discharge protocols are complete, validated, and approved and patient recruitment is complete.

Conclusions

This research in progress has served to proffer a solution to address the post-discharge phase of stoma patients. The proffered technology-enabled solution support both a high-quality patient experience as well as supporting a value-based care paradigm. We contend that such solutions are not just useful in the context as we have presented; i.e., stoma patients, but can be applied more generally so that the post-discharge phase of the patient journey is also monitored and managed, enabling both a high-quality patient experience as well as prudent management of likely trigger situations that may if not addressed lead to unplanned readmissions. Given the challenges faced by all OECD countries with respect to the exponential costs to provide quality care and the increasing pressures on healthcare organisations to deliver high quality and high-value care, we believe such technology solutions are strategic necessities and must be carefully considered. Our future work will focus on conducting clinical trials to establish the validity, usability and proof of concept of the proffered solution.

Acknowledgements

We acknowledge the healthcare organizations who have assisted us in the work to date and will continue to assist us as we move forward with planned clinical trials.

References

[1] Australian Commission on Safety and Quality in Healthcare, *Patient-Centered Care: Improving Quality and Safety by Focusing Care on Patients and Consumers Discussion Paper*, 2010.

[2] A. Kitson, A. Marshall, K. Bassett and K. Zeitz, What are the core elements of patient-centred care? A narrative review and synthesis of the literature from health policy, medicine and nursing, *Journal of advanced nursing* **69** (2013), 4-15.

[3] J.H. Hibbard and J. Greene, What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs, *Health affairs* **32** (2013), 207-214.

[4] M.F. Collen and M. J. Ball (Ed.), *A history of medical informatics in the United State*s, Springer, New York, 2015.

Address for Correspondence

Corresponding author: Nilmini Wickramasinghe
nilmini.work@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190660

Continuous Video Recording of Electronic Health Record User Sessions to Support Usability and Safety

Adam Wright[a,b,c], Skye Aaron[a], Gianna Zuccotti[a,b,c]

[a] *Brigham & Women's Hospital, Boston, MA,*
[b] *Harvard Medical School, Boston, MA,*
[c] *Partners HealthCare, Boston, MA*

Abstract

Electronic health records (EHRs) have been shown to improve safety and quality. However, usability and safety issues with EHRs have been reported. The current state of the art in usability testing is to have clinicians conduct simulated activities in a usability lab. In this poster, we describe our experience with continuous recording of real-world EHR use to improve safety and usability.

Keywords:
Electronic Health Records, Patient Safety

Introduction

There is ample evidence that electronic health records (EHRs), when implemented and used effectively, are associated with increased quality of care and safety [1]. However, significant usability and safety issues [3] with EHRs have been reported, including instances of patient harm [4]. Usability has been found to play a major role in the exacerbation and prevention of health information technology (HIT) hazards, and has been identified as a critical component of safe HIT use [3]. Poor usability has been shown to put patients at risk by impeding HIT adoption and workflow, decreasing clinician satisfaction and productivity, and directly causing clinical errors [2].

The current state of the art for usability evaluation of clinical information systems is to have clinicians conduct simulated activities in a usability lab. Less commonly, a usability researcher observes the clinician during routine care. These approaches are useful, but limit the amount of data that can be collected. This is particularly challenging for assessing the usability of infrequently used workflows (such as organ transplant or cardiac arrest documentation) that are unlikely to be observed naturalistically.

Citrix provides application virtualization services and is used with many EHRs, also offers session recording software. This feature allows for efficient capture and archiving of full-screen videos of applications running through Citrix. We describe a pilot study to assess the usefulness of this functionality for usability and safety research.

Methods

We installed and configured session recording on our Citrix environment, and enabled recording for a group of pilot users. Videos are captured whenever users use our Epic EHR. Only Epic activity is recorded; local application activity such as email or web browsing is not captured. A system diagram is shown in Figure 1.

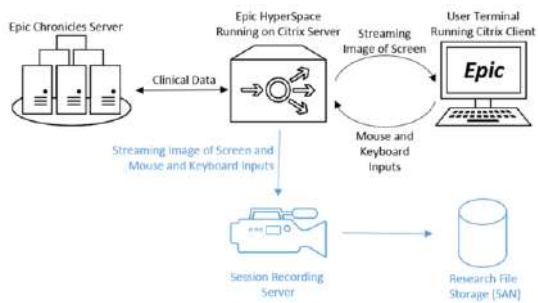


Figure 1– System Diagram of Screen Capture

Results

In early pilot testing, we found the videos to be useful, and there was no measurable impact on system performance or reliability. Videos consumed 14 megabytes per hour, which exceeds Citrix's published estimate of 4 megabytes per hour but is still reasonable. Extrapolating across Partners HealthCare system, there are approximately 22 million Epic sessions per year, with a typical session lasting about 10 minutes (with considerable variability). This would yield 3,704,458 hours of video per year, hence a year's worth of videos would require approximately 51.8 terabytes of storage. We currently pay 12.96 cents/gb/yr for storage, yielding a cost of $6,713 to store videos internally. If we used an archival storage system, costs would fall to a third of this amount.

Figure 2 shows our current approach for identifying videos to review, indexing, and analyzing them. Our approach adopts a series of general and EHR-specific usability tools, standards, and processes. In pilot testing, videos were manual retried, and could be tedious; we are in the process of developing indexing tools to make video retrieval more straightforward.

Discussion and Conclusion

In early pilot testing, we found Citrix session recording feasible and useful. Special challenges included system upgrades and dual monitors. We plan to extend our pilot to additional users and collect more data. We are also planning to

improve indexing capabilities to make it easier to retrieve videos that correspond to specific events in the EHR transaction log. When complete, we will be able to launch directly to the video of a workflow under investigation. For example, the video of a user adding a particular allergy or using a particular function. We anticipate this functionality will have broad-based application for quality and safety, as well as for usability and system redesign.

Acknowledgements

The authors appreciate the assistance of the Partners HealthCare Citrix team in implementing the tool, as well as valuable feedback from usability personnel at Partners, who have provided feedback on the tools, processes and videos.

References

[1] M.B. Buntin, M.F. Burke, M.C. Hoaglin, and D. Blumenthal, The benefits of health information technology: a review of the recent literature shows predominantly positive results, *Health Affairs* **30** (2011), 464-471.

[2] R. Koppel, J.P. Metlay, A. Cohen, B. Abaluck, A.R. Localio, S.E. Kimmel, and B.L. Strom, Role of computerized physician order entry systems in facilitating medication errors, *JAMA* **293** (2005), 1197-1203.

[3] B. Middleton, M. Bloomrosen, M.A. Dente, B. Hashmat, R. Koppel, J.M. Overhage, T.H. Payne, S.T. Rosenbloom, C. Weaver, and J. Zhang, Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA, *Journal of the American Medical Informatics Association* **20** (2013), e2-e8.

[4] E. Sparnon, Spotlight on Electronic Health Record Errors: Errors Related to the Use of Default Values, *Pennsylvania Patient Safety Authority* **10** (2013), 92-95.

Address for correspondence

Adam Wright, PhD
Brigham and Women's Hospital
75 Francis St.
Boston, MA 02115, USA
awright@bwh.harvard.edu

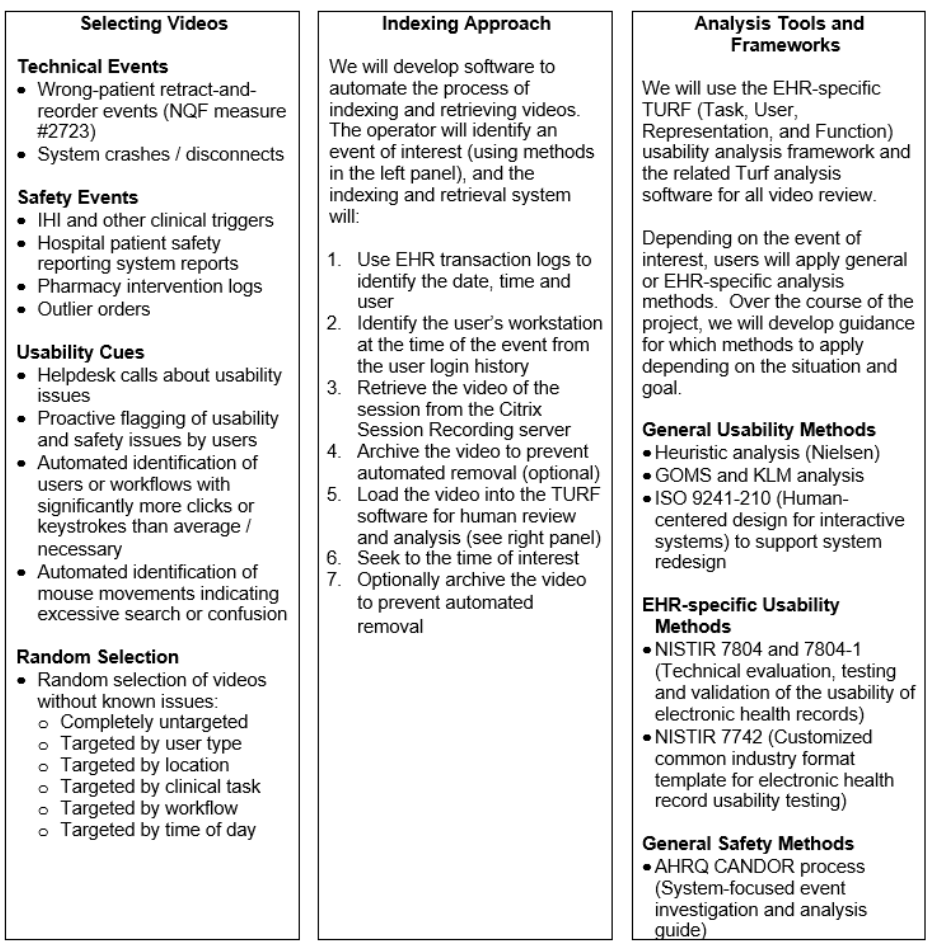


Figure 2-Overview of video analysis process and tools

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190661

Intelligent Conversational Agents in Patient Self-Management: A Systematic Survey Using Multi Data Sources

Zhaopeng Xing[a], Fei Yu[a,b], Yousef A. Mustafa Qanir[c], Ting Guan[d], Jennifer Walker[b], Lixin Song[c,e]

[a] *Carolina Health Informatics Program, University of North Carolina, Chapel Hill, North Carolina, United States,*
[b] *Health Sciences Library, University of North Carolina, Chapel Hill, North Carolina, United States;*
[c] *School of Nursing, University of North Carolina, Chapel Hill, North Carolina, United States;*
[d] *School of Social Work, University of North Carolina, Chapel Hill, North Carolina, United States;*
[e] *Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina, United States*

Abstract

Intelligent conversation agents (ICAs) have been used for patient self-management and support in recent years. This study systematically reviewed ICA resarch and innovation in academic and industrial institutions using bibliometric and patent analysis. We reported the types of diseases and patients, and the ICAs delivery approaches for patient-self-management. We identified the gaps in the productivity and focused areas.

Keywords:
self-management, patients, conversational agent

Introduction

The Intelligent conversational agent (ICA) (e.g., Amazon Alexa and Apple Siri) is an application that enables natural language communication with users [1]. The human-like communication nature of ICAs is making themselves very popular. Clinical analysts foresee that ICAs can potentially improve patient care in terms of symptom management, caregiver engagement, and patient support [2]. A recent systematic review on conversational agents in healthcare has shown that ICAs has addressed health issues related to psychotherapy support, education, self-monitoring, and data collection for consumers (e.g., patients), caregiver, or healthcare professionals [3]. However, no study to-date has explicitly focused on how ICAs support patient self-management in a non-clinical setting.

Patents are associated with research and mirror technology development and investment trends in a particular domain [4]. Analyzing patent records, thus, will help understand the state of the art of technology development and the landscape of the innovations. This study aimed to use data from both publications and patents to systematically investigate ICA research and innovation for patient self-management in home settings.

Methods

Data Collection

The research publications were retrieved from five citation databases (i.e., PubMed, Scopus, CINAHL, EMBASE, ACM D-Library). The granted patent and published applications were retrieved from the Derwent Innovation Index (DII) which has a comprehensive coverage of 40 worldwide patent-issuing authorities. The searches in citation databases and DII were executed with the same search strategies in August 2018. The search term consists of three parts, the ICA related terms, self-management related terms, and patient related terms. The ICA related terms included about 100 synonyms and term variations (e.g., voice-activated interface, conversational assistant, chatbots). The self-management related terms included about 20 synonyms and term variations (e.g., self-management, self-care, self-efficacy, supportive care). The patient related terms included 13 synonyms and term variations (e.g., patients, caregivers, sufferer, and survivor). In addition, this study collected publications identified through citations of retrieved research papers and from grey literatures.

Screening Critiria

The title and abstract fields of the retrieved publications and patent were screened by three of researchers (i.e., ZX, FY, & YQ) via Covidence [5]. Disagreement was resolved among researchers in group discussion. A publication or patent was included if it 1) adopted natural language interaction either in written, graphical or voice-activated communication modality, 2) was designed for patients, caregiver or customers. A publication or patent was excluded if it 1) discussed only partial techniques of ICAs such as automatic speech recognition, natural language understanding, dialogue management, response generation, or text-to-speech synthesis; 2) had limited and predefined interaction model (e.g., Voice-Activated-Dialing which only allows number as input); 3) had missing titles or abstract.

Data Abstraction and Analysis

For publications, we extracted the following data fields: authors, publication year, affiliation, title, and abstract. For patent, we extracted the inventor, priority date, assignee, title, and abstract. All included patent applications were filed before 2018 because patent applications are generally published 18 months after the earliest priority date.

In this study, VOSviewer [6] was employed to extract and visualize the key terms from both the title and abstract fields for included publication and patent records. We set the threshold of minimum key term occurrence as 3 for publication records and as 1 for patent records in mapping graph, for appropriate visual effects. In addition, the synonyms and variations of an extracted key term were merged and controlled for visualization using a thesaurus (e.g., "child" and "kid" merge to "child").

Results & Discussion

Data Description

We retrieved a total of 1835 publications from PubMed (N=751), EMBASE (N=137), CINAHL (N=29), Scopus

(N=687), ACM Digital Library (N=231), 535 patent records from DII, and 12 additional records from the citations of the patents. We included 91 publications and 21 patents in the analysis after title-abstract screening and removal of duplicate records. The publications were contributed by 128 institutions from 28 countries, whereas the patents were from 22 patent assignees (i.e., 16 companies and 6 individuals) from 5 countries. The United States is the leading country in the numbers of publications and granted patents. Particularly, the Northeast University and Boston University have had the most research productivity. Regarding the patents, Koninklijke Philips N.V. from Netherlands had three patents and each of the rest assignees have had less than three.

Topic Mapping

The key terms extracted from the publications and patents were visualized and compared based on types of diseases and patients and delivery approaches to patent self-management.

The Types of Disease and Patient

Compared with patents, the publications reported more types of diseases and patients (Fig 1). The publications reported that ICAs were developed to address chronic conditions (e.g., "diabetes", "cancer", "stroke", "hypertension"), mental disorders (e.g., "autism", "dementia", "post-traumatic stress disorder"), and emotional issues (e.g., "depression", "anxiety", "stress", "mindfulness"). The targeted users included "child", "elderly", "young", "woman", "mother", and "pregnant (women)". As displayed in Fig 2, the patents have focused on health conditions such as diabetes, communication disorders, and skin issues (e.g., "lesion"). Patterns have not specified the types of targeted patients but used a general term "user" instead.

The Delivery Approach of Self-Managment Support

The topics related to health interventions in the publications included "psychotherapy", "neurorehabilitation", "virtual interviewer" whereas the patents tended to be uniquely designed with customized support (e.g., "interaction plan", "first conversation", "user profile", "patient relationship model") based on user data ("user data", "medical record") and dialogue context (e.g., "dialogue context", "context pattern"). This customized support has not been observed in publications.

Nevertheless, the publications and patents under review demonstrated common interests such as patient's decision-making support (e.g., "patient coaching" and "health consultation") and user/patient behaviors (e.g., "behavior change", "behavioral intervention" in the publication and "dialogue behavior" in the patents). They also discussed multi-modal interaction modality in ICAs. Specifically, publications described "multimodal interaction", "embodied agent" and the patents discussed "graphical interface", "voice interface", "animation".

Conclusions

To our knowledge, this is the first study that has systematically examined ICA research and innovations for patient self-management using data extracted from both publications and patents. We have found that the ICA research and innovation in patient self-management is in its infancy. The total numbers of publications and patents are small. The ICAs to-date have mostly focused on patient coaching, consultations for decisional support and behavioral changes. In addition, the academia and industry have focused on different types of diseases and patients, and on various delivery approaches to self-management support by ICAs. However, they also share some common topic interests, such as multi-modal interaction ICAs. The diversity may lead to a broad spectrum of ICA applications in health care whereas the commonality may indicate the topical trends in the near future.

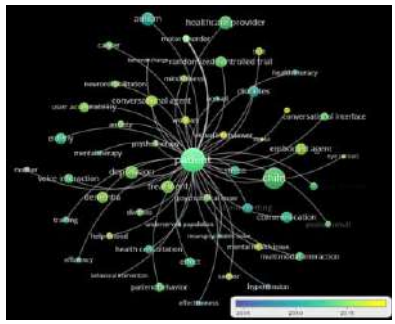


Figure 1 – Publicaiton Keywords Mapping

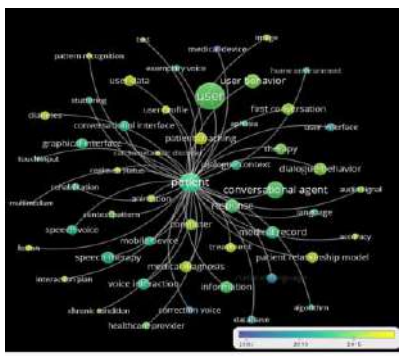


Figure 2 – Patent Keywords Mapping

References

[1] M. McTear, Z. Callejas, D. Griol, The dawn of the conversational interface, in: The Conversational Interface, Springer International Publishing, Cham, 2016: pp. 11–24. doi:10.1007/978-3-319-32967-3_2.

[2] M. Coyne, L. Matchstick, C. Franzese, The Promise of Voice: Connecting Drug Delivery Through Voice-Activated Technology, (2017).

[3] L. Laranjo, A.G. Dunn, H.L. Tong, A.B. Kocaballi, J. Chen, R. Bashir, D. Surian, B. Gallego, F. Magrabi, A.Y.S. Lau, E. Coiera, Conversational agents in healthcare: a systematic review., *J. Am. Med. Inform. Assoc.* **25** (2018) 1248–1258. doi:10.1093/jamia/ocy072.

[4] B. Fabry, H. Ernst, J. Langholz, M. Köster, Patent portfolio analysis as a useful tool for identifying R&D and business opportunities—an empirical application in the nutrition and health industry, *World Patent Information*. **28** (2006) 215–225. doi:10.1016/j.wpi.2005.10.004.

[5] J. Babineau, Product Review: Covidence (Systematic Review Software), *J. Can. Health Libr. Assoc.* **35** (2014) 68–71. doi:10.5596/c14-016.

[6] N.J. van Eck, L. Waltman, VOS: A new method for visualizing similarities between objects, in: R. Decker, H.-J. Lenz (Eds.), Advances in Data Analysis, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007: pp. 299–306. doi:10.1007/978-3-540-70981-7_34.

Address for correspondence:

Zhaopeng Xing, zhaopeng@live.unc.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190662

Natural Language Processing Based Approach for Identification of Problems in Medical Image Management Using PACS

Ayako Yagahara[a,b], Takumi Tanikawa[a,b], Akihisa Fukuda[b,c], Daisuke Ando[d], Tastuya Suzuki[e], Kohei Harada[f], Shuichi Karata[g], Masahito Uesugi[h]

[a] Faculty of Health Sciences, Hokkaido University of Science, Sapporo, Hokkaido, Japan,
[b] Faculty of Health Sciences, Hokkaido University, Sapporo, Hokkaido, Japan,
[c] Department of Radiology, Hokkaido Medical Center, Sapporo, Hokkaido, Japan,
[d] Department of Radiological Technology, Hokkaido University Hospital, Sapporo, Hokkaido, Japan,
[e]Department of Radiological Technology, Asahikawa Medical University Hospital, Asahikawa, Hokkaido, Japan,
[f] Department of Radiological Technology, Sapporo Medical University Hospital, Sapporo, Hokkaido, Japan,
[g] J-MAC SYSTEM, INC., Sapporo, Hokkaido, Japan,
[h] Department of Medical Management and Informatics, Hokkaido Information University, Ebetsu, Hokkaido, Japan

Abstract

We investigated problems concerning medical imaging management using PACS in medical settings through text analysis. We conducted a questionnaire survey in Hokkaido, Japan, where PACS related problems were described by radiological technologists. After analyzing the descriptions in response to the questionnaire, we detected four main problems in PACS using a word co-occurrence network; i) image management, ii) image fetching error, iii) DICOM violation and iv) name notation.

Keywords:

Management Information Systems, Educational Measurement, Computers

Introduction

DICOM (Digital Imaging and Communications in Medicine) standard, which is the standard in the medical information field, provides a seamless connection, clinical image, and data sharing when introducing new devices to medical institutions. However, the details of the functions of PACS (Picture Archiving and Communication Systems) depend on the specifications of each medical equipment manufacturer. In Japan, especially in local cities, medical professionals, such as radiological technologists, who are not specialized in medical information, often struggle to manage PACS during their professional work. Occasionally, serious problems with the PACS system occur when medical staff are unable to confirm the interface with image scanners or hospital information systems. When such an incident occurs, it disrupts the delivery of routine examinations. Several studies about the introduction factors of PACS have been analyzed [1]. However, there are few studies investigating problems concerning medical imaging management using PACS in medical settings.

The purpose of our study was to identify problems of image transfer and storage related to PACS specifications to avoid any serious trouble.

Methods

261 hospitals have introduced PACS in Hokkaido prefecture, which is the northern part of Japan. We sent questionnaires by mail to presidents and leaders of radiologic technologists in all of the hospitals, and they input the answers on the specialized web site we developed (from February to March in 2018). In the questionnaire, there were some questions about the hospital (Table 1) and we made open-ended statements (seeking arbitrary endings) about troubling cases related to the operation of PACS and critical points of operation. Table 1 shows examples of items in the questionnaire.

Table 1 Items of the questionnaire in hospital attribution

Items
The number of beds
Having or not medical information system correspondents in radiology department
Who are PACS correspondents in the hospiral
Introduction time of PACS
Capacity of PACS
Problem cases and issue related to operation of PACS

After collecting questionnaires, we added up numbers and analyzed the open-ended statements by natural language processing (NLP). In NLP, we conducted morphological analysis using ChaSen [2] at first. Since ChaSen is a dictionary-based morphological analyzer, words that are not included in the dictionary are output as unknown words. In this study, we used ipadic-2.4.4 as Japanese dictionary for morphorogical analysis. Then, the number of words were counted. Next, we developed a word co-occurrence network using the co-occurrence of the words that appeared more than twice, to clarify the problems related to image transfer and storage in PACS. Jaccard index was used as a co-occurrence index.

$$\text{Jaccard index} = \frac{a}{F_1 + F_2 - a}$$

Where, F_1 represents the total number of sentences where word "A" appears, F_2 represents the total number of sentences where word "B" appears, a represents the total number of sentences where A and B both appear.

We used a free software KH coder 2 [3] for a series of procedures from morphological analysis to developing the co-occurrence network. The questionnaire survey was conducted with the approval of the Hokkaido University of Science Ethics Committee.

Results

There were 94 responses (collection rate: 36%). The number of completed opened ended responses was 39.

About half of the hospitals had fewer than 100 beds. The number of radiological technologists in each hospital was 1~58 and 20% of the hospitals had fewer than 10 technologists. 84% of the hospitals had radiological technologists who manage PACS.

In the results of morphological analysis, we extracted 441 words. Regarding the high frequency of words, image (34), PACS (12), fetching (9), DICOM, (6) and data (6) were recorded (Table2).

Table 2 Example of word appearance in each part of speech

noun	n.	verb	n.	Unknown word	n.
画像 (image)	34	行う (conduct)	7	PACS	12
取り込み (fetching)	9	取り込める (can fetch)	4	DICOM	6
データ (data)	7	取り込む (fetch)	3	CD	3
患者 (patient)	6	異なる (differ)	2	ID	3
情報 (information)	6	含む (include)	2	Tag	3
モダリティ (modality)	5	考える (consider)	2	CS	3
管理 (manegement)	5	困る (difficulty)	2	PC	2
医師 (doctor)	4	思う (think)	2	RIS	2
サーバー (server)	3	取り交わす (exchange)	2	パブリッシャー(Publisher)	2

We represented the word co-occurrence network using 105 words to visualize co-occurrence relations among words. Common problems reported were regarding "image management," "image fetching error," "DICOM violation," and "name notation".

Table 3 items of the questionnaire

Category	Example
image management	Best storage period Defect of storage, education about virus Fetching video data
image fetching error	Can not import patient's data from CD made by other hospitals
DICOM violation	Patient data from other hospitals are DICOM violation
name notation	Handling of names of international people in kanji and kana characters which are Japanese characters

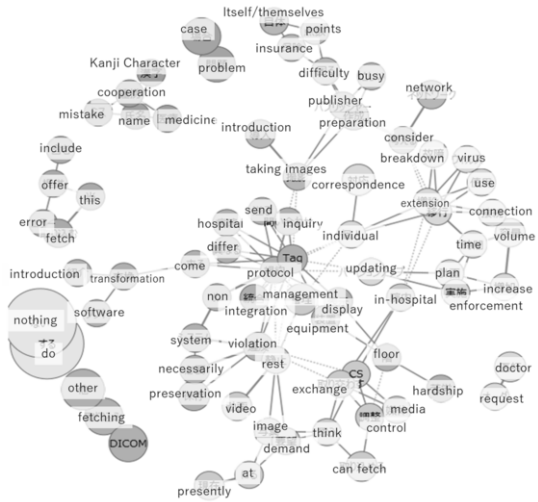


Figure 1 word co-occurence network

Conclusions

We clarified that radiological technologists engaged in PACS management in most of the hospitals, and there were four categories of problems when applying the guidelines in the medical setting; image management, image fetching error, DICOM violation and name notation. This study will contribute to developing user-friendly systems and consider making and revising guidelines or standardization.

Acknowledgements

This study was supported by a Japanese Society of Radiological Technology, Hokkaido Radiological Technology Research Grant. The authors would like to express their deep appreciation to the members of the Japanese Society for Radiological Technology, Hokkaido.

References

[1] Tzeng WS, Kuo KM, Lin HW, Chen TY. A socio-technical assessment of the success of picture archiving and communication systems: The radiology technologist's perspective. BMC Med Inform Decis Mak.**13** (2013), 109. doi:10.1186/1472-6947-13-109.

[2] Nara Institute of Science and Technology. ChaSen. [cited 2018 Nov 12]; Available from: http://chasen-legacy.osdn.jp/.

[3] Koichi Higuchi. [cited 2018 Nov 12]; Available from: http://khcoder.net/en/ .

Address for correspondence

Corresponding author: Ayako Yagahara

Mailing address: 7-Jo 15-4-1 Maeda, Teine, Sapporo, Hokkaido, Japan 006-8585

Email: yagahara-a@hus.ac.jp

Phone: +81-11-676-8504

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190663

Development of In-Hospital Infection Management Using IoT

Yoshinori Yamashita[a], Hiromici Iwasaki[b],

Yoko Muroi[b], Masao Hida[c], Hiroko Shigemi[b,d]

[a] Department of Medical Informatics, University of Fukui, Fukui, Japan,
[b]Center for Patient Safety and Infection Control, University of Fukui, Fukui, Japan,
[c] Department of Clinical Laboratory, University of Fukui, Fukui, Japan,
[d] Carecom co.,ltd, Tokyo,Japan

Abstract

As one of the countermeasures against infection at medical institutions, thorough hand hygiene is extremely important. In Japan, these controls are not sufficient. In order to do this management, it is necessary to track the hand washing situation. *Therefore, we decided to monitor the condition of hand washing by utilizing IoT.* Since there are environments where we can use IoT in our hospital, we decided to follow up using these environments. *As a result, it is possible to collect data continuously for 24 hours, 365 days, and evaluate infection risk based on data. In addition, our hospital can also obtain location information on smartphone, so we can also track work. We are considering support for medical staff by utilizing smart devices.*

Keywords:
Hand Hygiene, Infection Control, Medical Staff

Introduction

The World Health Organization (WHO) created hand hygiene guidelines in 2009. It saught compliance with hand hygiene at five timing. However, in the announcement of the Japan Society for Environmental Infections in 2017, our hand hygiene compliance rate is 38% on average, which is lower than much of the world, and the occurrence rate of resistant bacteria is very high compared to the Northern Europe such as Sweden. For this reason, we consider it necessary to reduce the infection rate.

Nurses who have many contact opportunities with patients are actively engaged in nosocomial infection prevention. They always carry disinfectants and maintain cleanliness of fingers, are cooperative in countermeasures against nosocomial infections, and the compliance rate of hand hygiene is higher than other medical treatments It shows a high price compared with the person. However, medical personnel other than doctors and nurses do not adequately comply with the WHO hand hygiene guidelines and are low consciousness.

Since Internet of Things (IoT) devices using Bluetooth and others can be used cheaply, the use of radio wave networks has begun to spread to solve medical problem problems. In our hospital, we can also use it for patient observation, and the positioning of medical staff is possible in the whole hospital as well. Using this kind of communication environment, we attempt to solve the problem of nosocomial infection control using hand hygiene disinfectant with IoT.

By visualizing the present state of hand hygiene using IoT and supplementing the management of infection control by utilizing the collected data, it is believed to lead to the creation of a safer medical environment. In addition, by evaluating hand hygiene for each medical doctor and giving feedback to individuals of medical personnel in real time, if there is a risk, educational effect can be obtained, and expectation can be expected for habitualization of behavior change of handicap and hand hygiene. Furthermore, we decided to aim for a method to use on a daily basis with smartphones and others.

Methods

In our hospital, we are strengthening the use of wireless communication as part of promoting medical ICT. Therefore, communication such as Bluetooth and ZigBee can be used. We have created and installed modules so that these communications can be used in hospitals. The configuration of the IoT gateway module is for converting from communication such as Bluetooth, Zigbee etc. to a Wifi connection to Wi-Fi. (Figure 1)

We developed this IoT gataway to acquire data and location information including alarms from medical devices and various sensors. We have already installed 1,500 IoT gataways in the hospital. In addition to this device, the staff distributes the smartphone as an extension, so this smartphone is used for location detection and personal identification.

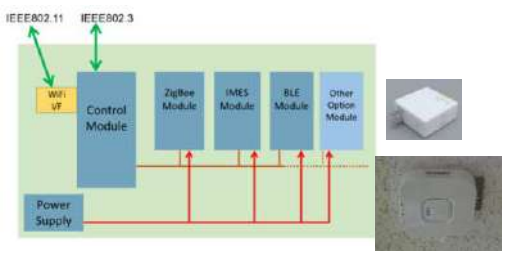


Figure 1 – Structure of IoT Hateway

We use a combination of smartphones and IoTgateway to detect medical staff locations. As a positioning method, Bluetooth low energy (BLE) beacon is performed by position calculation from multiple RSSI radio wave intensity measurements. In order to detect hand hygiene, it is necessary to detect the operating state of the disinfectant. For this reason, we created a mechanism to output a signal when the disinfectant was operated (Figure 2).

Figure 2 – Disinfection Pump Operation Detection Sensor

By combining the staff's flow line and disinfection pump operation, it is possible to detect the hand disinfection situation. In this way, it is possible to detect the hand disinfection situation at the time of entering the area or approaching the patient. By collecting these data continuously, it is possible to measure the required hand hygiene performance.

Results

In this system, positions and flow lines of staff are acquired with accuracy of approximately one meter (Figure 3). For this reason, we are making it possible to receive multiple BLE beacons at any location.

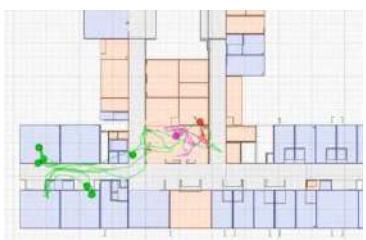


Figure 3 – Acquisition of Position and Flow Line of Staff

The fixed disinfection pump maps the coordinates of the installation position, and the carried disinfection pump is made to correspond to the person. Timing is determined by matching the operation signal of the disinfection pump with the position information. As a result, continuous measurement is possible , and it is possible to present collected hand hygiene data.

One result is the measurement result of the ICU. The status of hand hygiene compliance when entering the ICU and entering the patient area was (94.5%, 21.3%) by nurses (37 people), (64.3%, 12.5%) by anesthesiologists (5 people), (28.0%, 13.5%) by cardiac surgeons(6 person), (48.0%, 15.7%) by circulatory physicians (1 person), (100%, 18.2%) by physiotherapists (1 person), (100%, 100%) by nurse assistants (6 persons).

This is an ICU situation, but similar measurements are possible in hospital wards. By feedback the results of hand hygiene measurement, the hand hygiene action of the medical staff has improved. However, we believe that infection control is an inadequate condition.

Discussion

With this method, the compliance rate of hand hygiene can be measured steadily. However, simply presenting this compliance rate is insufficient for continuous improvement. Although there are advantages in terms of visualization of compliance status, there are many cases where it is limited to presenting data. In order to make this improvement, we need to develop a method that reflects our actual behavior. we think that real-time assistance should be provided in important situations for infection control.

Conclusions

By using this method, it is possible to measure the behavior of medial staff in real time, so it is necessary to consider a method to utilize this result. In our hospital, one staff member uses smartphones, so we believe that hand hygiene can be improved by adding a mechanism to notify warnings when hand hygiene is needed. In the future, we will consider implementation with smartphone notification function. We plan to develop from the location of the medical staff's smartphone to a system that can provide appropriate assistance notifications. This method can be used in the same way as patient monitoring in the future, so it can be used together with these devices. We are also considering combining it with the sensor used in the nurse call system. In such a combination, we believe that this method is widely available.

Acknowledgements

This research is conducted using SCOPE which is the research expenses of the Ministry of Internal Affairs and Communications.

Address for correspondence

Yoshinori Yamashita
yyama@u-fukui.ac.jp,+81-776-8140.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190664

A Survey on Health Care and Health Concerning Workers for Considering Appropriate Personal Health Record Service

Mayumi Yoshida[a], Ryuichi Yamamoto[a]

[a] *Medical Information System Development Center, Tokyo, Japan*

Abstract

To prevent worsening of diabetes mellitus, we conducted a developmental research of personal health record (PHR), whereby affected individuals, medical staff and an insurer cooperate and manage effective treatment. Medical expenses can be suppressed by preventing the onset of lifestyle-related diseases. Companies can benefit from improved employee productivity by promoting healthy and efficient working styles. We conducted a health and medical consciousness survey to examine PHR models that are attractive for both employers and employees.

Keywords:
Personal Health Record, Survey

Introduction

In Japan, there is a universal insurance system that allows most people to have one health insurance. [1]. Company employees engage in corporate health insurance and mutual aid insurance coverage that are affiliated by employers. In addition, retirees and individual business owners subscribe to the National Health Insurance. Till date, the social security of Japan has been substantial, with the citizens receiving equally high-quality treatment, regardless of which medical institution they attended. However, with an increase in lifestyle diseases and acceleration of the ageing population, medical expenses are increasing and insurers are also being dealt with.

To prevent worsening of diabetes mellitus (DM), we conducted a developmental research of personal health record (PHR), whereby the affected persons, medical staff and insurer cooperate and manage effective treatment by themselves [2,3]]. In the case of mild diabetes, there is not much burden on patients and insurers. Patients can visit medical institutions once a month and receive appropriate treatment after examination and through medication at hospitals. However, when diabetes becomes severe, it results in renal failure, blindness, etc. For example, it is inevitable for patients to worsen their quality of life owing to the requirement of artificial dialysis. The burden of medical expenses is increased on both the patients and insurers. Alternatively, in Japan, companies tend to be concerned about their employee's health management.

Companies are also believed to contribute towards the improved productivity of their employees by promoting healthy and efficient working. To investigate the development of an attractive PHR model for both employers and employees, we conducted a preliminary survey on health and medical attitudes of 10,000 people residing in the country.

Methods

The survey included questions regarding the current health status, status of medical examination attendance, meals and sleep pattern, lifestyle habits, job types, working hours, preferred leisure activities and consciousness of own health. This web questionnaire survey was conducted in March 2018 after acquiring the consent of 10,000 workers aged >20 years in Japan. We allocated participants to different groups based on their age and sex. For each sex in their 20's, 30's, 40's and 50's, as a unit of 1,000 at the age of 10, >60's were taken as a group of 1,000 each, and total 10,000 people were targeted.

Results

Respondents included 41.7% children and 28.5% single people, with 32.7% involved in office work, 23.3% in professional and technical work and 12.9% in the service industry. Approximately 72.4% of the participants worked for 5 days a week and 13.5% for 6 days a week. Of 28.5% individuals currently undergoing treatment, 47.3% had confidence in their own health, 22.9% smoked regularly, 57.5% consumed alcohol and 46.8% preferred to sleep to rest. The main questionnaire results are listed below. When N is not displayed, N = 10000.

1. Regarding discomfort in the most concerned body part, the following was found: stiff neck in 36.1%; headache in 27.8%; lower back pain in 25.7%; and obesity in 10.8% patients.
2. Diseases under treatment included: hypertension 25.4%; dyslipidaemia 12.5%; lower back pain 11.1%; DM 10.7% patients (N = 2851).'.
3. For illness prevention, 70.4% received medical check-ups conducted by companies and municipalities.
4. The reasons for not receiving medical check-ups included the following (N = 2,959): 'Proceedings are troublesome' for 26.5%, 'Because I do not want to spend on such things' for 16.2% and 'I was unaware of health check-ups as such' for 12.0%.
5. Regarding whether paper inspection results or prescriptions were preserved, 51.4% said 'We have not kept it at all or I do not know', 13.8% said 'all are kept', 10.5% said 'I have stored test results on chronic disease', 24.4% said 'I keep results of important test results and prescription'.
6. Approximately 59.8% of persons retain the results of past health check-ups for the following reasons: 'Because it is custom' by 38.7%, 'I want to use it when

I become sick in the future' by 32.8% and 'to review my own lifestyle' by 24.6% (N = 5989).

7. Approximately 60.4% persons regularly measure their weight (daily); 21% (N = 6041) preserved records of their weight.
8. Of the 27.9% participants that used pedometers and activity meters, 38% (N = 2787) record their results.
9. Regarding physical exercising, 35.8% said 'I always liked exercising', 23.5% said 'I do not dislike exercise, but I enjoy it' and 40.7% said 'I do not prefer to exercise'.
10. Regarding regular physical exercise, 21.9% responded with 'I do it' and 78.1% with 'No, I do not'.
11. Regarding eating habits, 'meals considering physical condition and nutrition balance are important to them' was responded by 59.3%, 'Meals that are balanced and nutritious are not fun' by 21.1% and 'I am not interested in meals' by 19.6%.
12. About using a healthcare application in smartphones (such as applications for blood pressure and diet management), 1.6% said 'I use paid apps', 13.3% said 'I use free apps' and 85.1% said 'I do not use apps'.
13. On things necessary for maintaining health, 26.4% voted for 'strong will', 25.1% for 'target with deadline', 14.5% for 'everyday encouragement' and 'Healthcare IoT (Internet of Things) device that automatically measures weight and step count' by 14.2%, and 'Comrade' by 13.8%.
14. As for PHR using a smartphone that holds all data on health and medical care, such as daily weight and blood pressure as a result of examination result and prescription, periodical health check-up and can be used when necessary; 65.6% responded as 'I want to use it for free', 4.5% as 'I want to use it for a fee' and 29.9% as 'I am not interested or do not want to use it' (multiple answers).

Discussion

Based on the survey results, most participants worked for 5 or 6 days in a week and 70% of the participants were healthy. The participants who were suffering from illnesses tended to have more number of lifestyle diseases and pre-diabetes mellitus. Stiff shoulder, headache and back pain were believed to be caused by labour and lifestyle habits.

More than 50% of participants had drinking habits and 20% were smokers. In addition, half of the participants did not get enough sleep and half were not confident about their health.

Approximately 70% of the participants underwent periodic health check-up for prevention, with more than half of them preserving health check-up results. In addition, 60% of the participants weighed themselves regularly and 20% of them recorded their weight.

The results showed that diet for nutritional balance was considered to be important.

Despite 60% of the participants believing that physical exercise was important, only 30% exercised daily. Regarding the necessary factor in maintaining health, 'strong will' and 'goals that decided the deadline' were rated higher than 'the need for convenient healthcare IoT devices'.

Approximately 25% of the participants who used health applications, 65.5% wanted to use PHR service by smartphone for free, whereas 4.5% were willing to use it for a fee. Companies can thus consider providing health apps and PHR to employees as an effective step towards health management.

Conclusions

Based on the results of our questionnaire survey, workers suffering from illness had lifestyle diseases and pre-diabetes mellitus. Approximately 70% of the workers had mild disabilities and several of them had bodily malfunctions owing to stiff shoulder, headache and back pain, which are believed to be caused by labour and lifestyle habits.

Based on the results of the questionnaire survey, it is expected that the development of PHR using an appropriate Healthcare IoT device can help prevent or delay the onset of lifestyle diseases. For employers, it is expected that employees who work in a healthy and efficient manner show improvement in productivity and face reduced burden of medical expenses and bear a part of the introduction cost of PHR.

Acknowledgements

This study is supported by AMED: Japan Agency for Medical Research and Development, Research assignment ID 17943075.

References

[1] N. Ikegami, M.D., and J.C. Campbell, Ph.D., Medical Care in Japan, The NEW ENGLAND JOUNAL of MEDICINE, URL:https://www.nejm.org/doi/full/10.1056/nejm199511093331922, accessed 2019/03/27

[2] M. Yoshida, and R. Yamamoto, Study on secondary usage of information accumulated in PHR based on consciousness survey divided among physician, general public, corporate research developer, 1016-1019.

[3] N. Nakashima, M. Yoshida, and R. Yamamoto, Personal health record system based on standardized item sets generates personal advice in accordance with clinical guidelines, ATTD2018, 2018/2/14-17, URL: http://simul-europe.com/2018/attd/Files/(nnaoki@info.med.kyushu-u.ac.jp)ATTD2018NN.pdf. accessed 2019/03/27

Address for correspondence

Corresponding Author; Mayumi Yoshida.
Medical Information System Development Center, 1-1 Kagurazaka, Shinjuku, Tokyo, Japan; E-mail: myoshida@medis.or.jp, phone:+81 3 3267 1923

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190665

Pilot Testing of an ICT-Based Care Management Support System to Deliver Integrated Community Care

Nayoung Lee[a], Minjung Kwak [b], Miri Jeong [b], Eun Jeong Choi [b], Eunjung Lim[a], Il Bum Kwon[c], Wonpyo Lee[d], Hanwool Ku[e], Dongil Kim[e], Haesung Nam[f], Junsik Na[c], Myonghwa Park [b]

[a] Research Institute of Nursing Science Chungnam, National University, Daejeon, Republic of Korea,
[b] College of Nursing, Chungnam, National University, Daejeon, Republic of Korea,
[c] Mindle Health & Care Social Coop, Daejeon, Republic of Korea,
[d] Social Economy Institute, Daejeon, Republic of Korea,
[e] Department of Computer Science and Engineering, Chungnam National University, Daejeon, Republic of Korea,
[f] Department of Preventive Medicine, School of Medicine, Chungnam National University, Daejeon, Republic of Korea

Abstract

The purpose of this study was to develop an ICT-based care-management support system and to assess its validity and applicability through pilot testing. The system links users' health and social needs to community resources. Users and care managers participated in each step of care management through an interactive decision support system. The system contributes to facilitate person-centered community care.

Keywords:
Case management, community health Services.

Introduction

Care systems in Korea face significant challenges due to an aging population and the increasing complexity of the health care system. Needs continue to grow for services and care from an integrated and coordinated system that is easy to navigate, efficient, and communicates well.

According to KISA (Korea Internet Security Agency), all households in Korea are able to access the Internet through mobile devices. To respond to these changes, we developed an Information Communication Technology (ICT)-based care-management support system for care managers and citizens in the community.

The purpose of this system is to support efficient care management and to provide tailored service information, based on the individual health and social needs of citizens. The main study objectives were to develop a care-management support system based on ICT by modifying the case-management to be more user-centered and to test the suitability and validity of the developed system.

Methods

The study developed a prototype of an ICT-based care-management support system. The system involved the following modules: 1) Engagement and relationship building with users; 2) Basic information collection and assessment; 3) Patient decision to participate in care management; 4) Comprehensive assessment, including physical condition and health, activity and engagement evaluation, and environment evaluation; 5) Comprehensive profile of users' needs based on standardized ICF classifications on functioning, disability, and health; and 6) An individualized service plan.

This ICT-based care-management support system will help service users and care managers collaborate in each stage. In addition, the user-friendly diagrams and infographics guided each decision step. A pilot test of this system was performed with 30 older citizens.

Results

The majority of users were positive about the usability of the ICT-based care-management support system. Users responded that they felt actively involved in the care-management process by sharing health and social care information on the web. The ICT solution enabled efficient collection of data through standardization and individualization of care management and provided user-friendly individualized recommendations for users.

Discussion

The ICT-based care management support system was developed for the Korean community case management users and reflected feedback from the case management users and community residents to assess its usefulness and feasibility. As the results show, users actively engaged in care management through a user-centered care-management system. However, various incentives are needed to keep ICT-based case managers involved. Further training is needed to make it easier for users to use the ICT system.

Conclusions

The overall prospect was that the ICT-based case management system would enable service recipients to become involved in the decision-making process of care management. Therefore, the care-management system using ICT can serve as the basis for person-centered community care.

Acknowledgements

This work was supported by the Community Business Development Program through the Korea Institute for Advanced Technology (KIAT) funded by the Ministry of Trade, Industry, and Energy (Project No. P0004335).

References

[1] J.Y. Joo and D.L. Huber, Case management effectiveness on health care utilization outcomes: A systematic review of reviews, Western journal of nursing research (2018), 0193945918762135.
[2] G.T. McMahon, H.E. Gomes, S.H. Hohne, T.M.J. Hu, B.A. Levine and P.R. Conlin, Web-based care management in patients with poorly controlled diabetes. Diabetes care (2005), 28(7), 1624-1629.
[3] S. Willard, G. Cremers, Y.P. Man, E. van Rossum, M. Spreeuwenberg and L. de Witte, Development and testing of an online community care platform for frail older adults in the Netherlands: a user-centered design. BMC geriatrics (2018), 18(1), 87.

Address for correspondence

Minjung Kwak
College of Nursing, Chungnam National University
266, Munhwa-ro, Jung-gu, College of Nursing, Chungnam National University, Daejeon
35015, Republic of Korea
Telephone number: +82-42-580-8407
Fax number; +82-42-580-8309
E-mail address: qufquf1029@naver.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190666

Implementation of a Novel User Interface for Review of Clinical Microbiology Results

Gianna Zuccotti[a,b,c], Skye Aaron[a], Adam Wright[a,b,c]

[a] *Brigham & Women's Hospital, Boston, MA, USA,*
[b] *Harvard Medical School, Boston MA, USA*
[c] *Partners HealthCare, Boston, MA, USA*

Abstract

Compared to other laboratory data, microbiology data are a complex mix of quantitative and qualitative results that return iteratively over time. Commercial electronic health records (EHR) frequently have limitations in the manner in which they manage microbiology data, not attempting to codify data but rather displaying it as text. This contributes to time- consuming and error-prone clinical workflows. We developed a microbiology viewer application to aggregate results and implemented it in our EHR.

Keywords:
Electronic Health Records, Manage, Workflows

Introduction

Microbiology data is a critical component for care delivery in Infectious Diseases. Availablity and accuracy of such data facilitates appropriate antimicrobial therapy, guides duration of therapy, can limit the spread of drug-resistant organisms and can enable detection of outbreaks.[1] Comprehensive review of microbiology results is a time-consuming and complex activity made more challenging by a high volume of tests, time delay to test positivity, co-existence of clinically significant and insignificant organisms within a single specimen, and complex antimicrobial susceptibility patterns. [2]

Commercial electronic health records (EHR) generally have limitations with regards to clinical workflows for review and management of microbiology data. EHRs generally do not attempt to codify microbiology data at all, instead displaying it as text. This prevents searching, grouping, and other functions that streamline clinical workflow and instead shifts the cognitive burden of assessment and summary to the clinician. [3] Despite its complexity, microbiology data does have an underlying hierarchical structure of specimens, organisms, tests and susceptibilities that, if preserved, can facilitate electronic summarization and decision support. We developed a tool that reviews text data and displays it in a manner that permits rapid assessment and trending.

Methods

We developed a microbiology viewer application that contains two views; a tabular view, that displays all microbiology and other Infectious Disease-related results and an organism view that displays all results by organism. The tabular view supported the activities of searching, sorting and filtering. It also allowed access to the full text report. The organism view facilitated rapid assessment of all organisms in a patient's history, which allowed a user to see data organized by organism rather than chronologically.

The original prototype of the microbiology viewer was developed as a standalone web-based Java application, outside of our EHR. The prototype viewer used web services to authenticate users, retrieve patient identity from our enterprise master patient index, clinical team assignments from our enterprise patient list, and structured microbiology data from our comprehensive clinical data repository (CDR). After demonstrating the value of the system with usability testing [4], we integrated a hardened version of the application into our EHR (Epic Systems v2018), which permits seamlessly embedding external web applications.

Results

We successfully coded our microbiology viewer to launch within Epic and to coexist within the vendor-supported clinical workflow. Figure 1 shows the tabular view. A key feature of the viewer was the novel organism-specific (i.e. Bacteria, fungus etc.) view (Figure 2) that allowed the user to see data organized by organism rather than chronologically. This view has been particularly useful for patients with complex, multi-year histories, such as those undergoing hematopoeitic stem cell transplant. In addition, we have developed the ability to filter/sort data to permit clinical review to focus on a subset of information, further enhancing user experience. User adoption took place seamlessly and without specific training, further supporting the usability principles that we observed when the product existed outside of the EHR.

Discussion and Conclusion

We have demonstrated successful integration into a commercial EHR of a tool that was initially built entirely outide of our legacy EHR system. Information systems that can aggregate, synthesize and present data are well-suited to improvement of the management of microbiology test results.

Acknowledgements

The authors appreciate the assistant of the Partners HealthCare Clinical Data Repository Team in building and implementing the tool. We also that the Partners eCare Clinical Content Team and Epic Integration Team for designing and supporting the clinical workflow within the Epic Electronic Health Record.

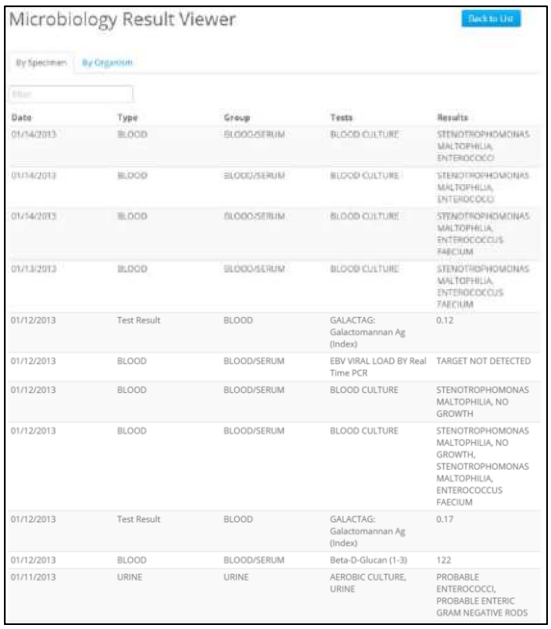


Figure 1. Microbiology Viewer – Results and Filter Function

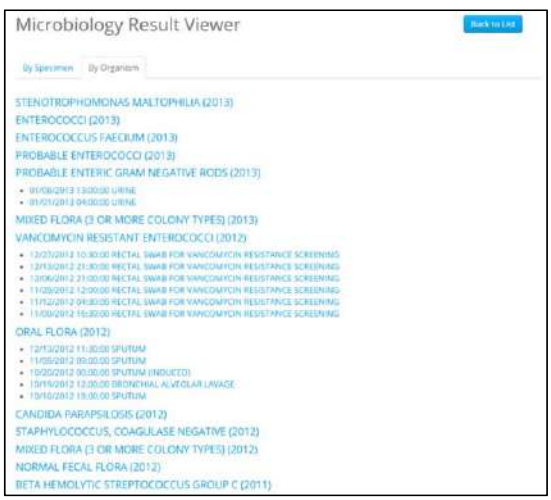


Figure 2. Microbiology Viewer- Organism View

References

[1] D.D. Rhoads, V. Sintchenko, C.A. Rauch, and L. Pantanowitz, Clinical microbiology informatics, *Clin Microbiol Rev* **27** (2014), 1025-1047.

[2] V.P. Ackerman, R.C. Pritchard, D.J. Groot Obbink , R. Bradbury, and A. Lee, Reporting practices of microbiology laboratories, *J Clin Pathol* **33** (1980), 830-835.

[3] J.C. Feblowitz, A. Wright, H. Singh, L. Samal, and D.F. Sittig, Summarization of clinical information: a conceptual model, *J Biomed Inform* **44** (2011) 688-699.

[4] A. Wright, P.M. Neri, S. Aaron, T.T. Hickman , F. Maloney, D. Solomon, D. McEvoy, K. Kron, and G. Zuccotti, Development and evaluation of a novel user interface for reviewing clinical microbiology results, *J Am Med Inform Assoc* ***25*** (2018)1064-1068.

Address for correspondence

Gianna Zuccotti, MD
Partners Healthcare System
399 Revolution Drive,Suite 1065
Somerville, MA 02145-1466
USA
gzuccotti@partners.org

III. Enabling Precision Medicine and Public Health

This page intentionally left blank

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190668

Informatics and Data Science for the Precision in Symptom Self-Management Center

Suzanne Bakken[a,b,c], Adriana Arcia [a], Theresa Koleck[a], Jacqueline A. Merrill[a,b,c], Kathleen T. Hickey [a,d]

[a] *School of Nursing, Columbia University, New York, New York, USA,*
[b] *Department of Biomedical Informatics, Columbia University, New York, New York, USA,*
[c] *Data Science Institute, Columbia University, New York, New York, USA,*
[d] *Department of Medicine, Columbia University, New York, New York, USA*

Abstract

The goal of the Precision in Symptom Self-Management (PriSSM) Center is to advance the science of symptom self-management for Latinos through a social ecological lens that takes into account variability in individual, interpersonal, organizational, and environmental factors across the life course. Informatics and data science methods are foundational to PriSSM's research activities including its pilot studies and research resources. This work highlights three areas: Latino Data Repository, Information Visualization, and Center Evaluation.

Keywords:
Precision Medicine, Nursing Informatics, Self-Management

Introduction

The goal of the Precision in Symptom Self-Management (PriSSM) Center is to advance the science of symptom self-management for Latinos through a social ecological lens that takes into account variability in individual, interpersonal, organizational, and environmental factors across the life course. Informatics and data science methods are foundational to PriSSM's research activities including its pilot studies and research resources. The PriSSM Center is a Center of Excellence funded by the National Institute for Nursing Research (NINR) and a member of its Center Directors Network, which collaboratively works to advance symptom and self-management science. [1,2,3] The purpose of this work is to highlight three areas in which informatics and data science methods have been used to facilitate the goal of the PriSSM Center: Latino Data Repository, Information Visualization, and Center Evaluation.

Methods

Setting

Columbia University Irving Medical Center is an academic medical center in a primarily Latino neighborhood in Northern Manhattan and is affiliated with NewYork-Presbyterian Hospital.

Approaches

Methodological approaches including procedures, data sets, sample size, and analytic plans that where relevant are summarized within each research activity.

Latino Data Repository

All PriSSM Center pilot studies include collection of genomic and phenotype data including a set of NINR-designated common data elements (CDEs) related to patient-reported symptoms. We received supplemental Center funding from NINR to create a Latino Data Repository that integrates genomic data, CDEs, and other phenotypic and environmental data from PriSSM pilot studies as well as our other Latino data sets. We processed, transformed, and integrated all data into a Research Electronic Data Capture (REDCap) database.

Information Visualization

Information visualization is a methodological focus in the PriSSM Center. We convene a monthly interdisciplinary Visualization Design Studio, support a doctoral level course in information visualization for research synthesis, and conduct participatory design research to create information visualizations for returning individual research results.

Center Evaluation

To examine co-authorship patterns and grant and publication topics for PriSSM Center investigators (n=22), we apply network analysis and topic modeling, respectively, to publication and grant corpora created from online sources (PubMed, NIH RePORTER, Federal RePORTER, Columbia University Scientific Profiles).

Results

Latino Data Repository

The Latino Data Repository contains sociodemographc data, NINR CDEs, other patient-reported data such as diet and physical actiivty, Ancestry Informative Markers, and a selected set of common variants and rare mutations. We also store participants consent for data sharing, data linkage, and willingness to be contacted for future research along with medical record number if available. The latter facilitates access to clinical data for particpants who have consented to data linkage. Columbia researchers can request access to both samples and data as demonstrated in a recent Cell publication.[4]

Information Visualization

Through the Visualization Design Studio, we have provided guidance on the design and creation of a broad variety of information visualizations including those for returning

research results, patient-provider communication tools, mHealth apps, and clinical dashboards. We are offering the doctoral-level information visualization class annually. Figures 1 and 2 display information visualizatons for returning results to research participants.

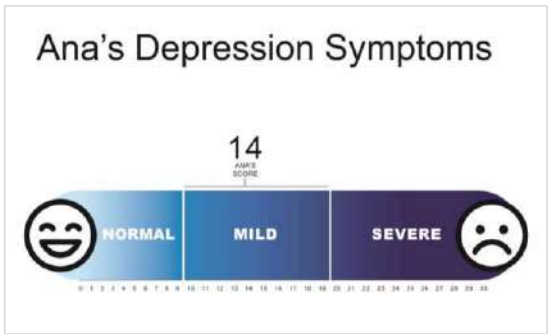


Figure 1. Depression Common Data Element Visualization

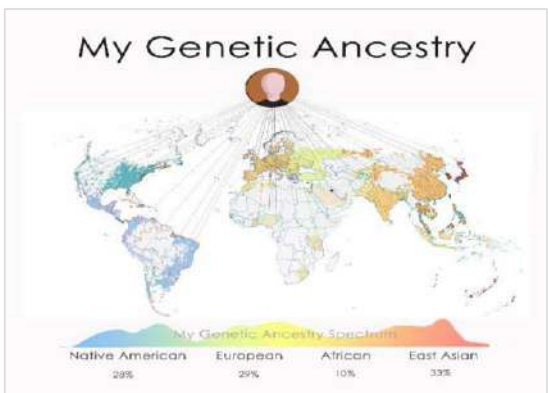


Figure 2. Ancestry Informative Marker Visualization

Center Evaluation

The co-authorship network (Figure 3) shows that Pilot Project principal investigators (PIs) have developed publication networks that distinguish them from the PriSSM Center PIs. The topic model (Figure 4) illustrates the clinical application areas as well as informatics and data science methods in grants (n=44) and publicatons (n=121).

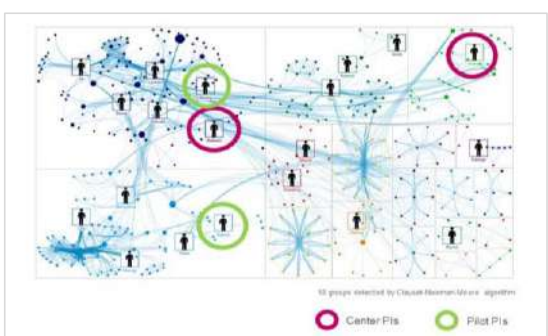


Figure 3. Network Analysis Showing Co-Author Patterns of Center Investigators

Conclusions

Our results demonstrate the important role that informatics and data science play in enabling the PriSSM Center to achieve its scientific goal as well as to evaluate Center performance.

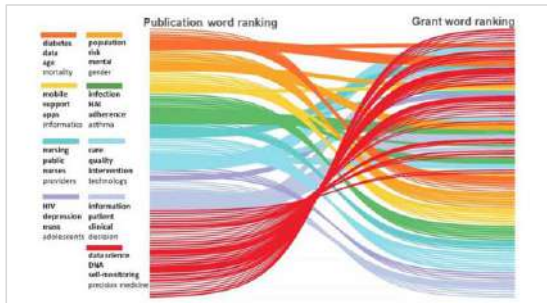


Figure 4. Topic Models of Grants and Publicaitons. Thicker lines mean higher frequency.

Acknowledgements

The PriSSM Center is funded by the National Institutes of Health/National Institute of Nursing Research (P30NR016587, P30NR016587,-02S1, Bakken, Hickey, Merrill, Arcia, Koleck). We thank Dr. Janet Woollen for Figure 2 and Dr. Sunmoo Yoon for the data analysis and visualizations in Figures 3 and 4.

References

[1] N.S. Redeker, R. Anderson, S. Bakken, E. Corwin, S. Dochtery, S.G. Dorsey, M. Heitkemper, D.J. McCloskey, S. Moore, C. Pullen, B. Rapkin, R. Shiffman, D. Waldrop-Valverde, and P.A. Grady. Advancing symptom science through use of common data elements. *Journal of Nursing Scholarship* 47(2015):379-88.

[2] G.G. Page, E.J. Corwin, S. G. Dorsey, N. S. Redeker, D. J. McCloskey, J.K. Austin, B.J. Guthrie BJ, S. M. Moore, D. Barton D, M.T. Kim, S.L. Docherty, D. Waldrop-Valverde, D.E. Bailey Jr., K.T. Hickey KT, S Bakken, and P. Grady. Biomarkers as common data elements for symptom science. *Journal of Nursing Scholarship* 50(2018):276-286.

[3] K.T. Hickey, S. Bakken, M. Byrne, D.E. Bailey Jr., G. Demiris, S.L. Docherty, S.G. Dorsey, M.M. Heitkemper, C.S. Jacelon, T.J. Kelechi, S.M. Moore, N.S. Redeker, A. Starkweather, H. Thompson, T.M. Ward, D.J. McCloskey, J.K. Austin, and P.A. Grady. Precision health: Advancing symptom and self-management science. *Nursing Outlook* (in press).

[4] F. Polubriaginof, R. Vanguri, K. Quinnies, G.M. Belbin, A. Yahi, H. Salmasian, T. Lorberbaum, V. Nwankwo, L. Li, M. Shervey, P. Glowe, I. Ionita-Laza, M. Simmerling, G. Hripcsak, S. Bakken, D.K. Goldstein. Kiryluk, E.E. Kenny, J. Dudley, D.K. Vawdrey, N.P. Tatonetti. Estimate of disease heritability using 7.4 million familial relationships inferred from electronic health records. *Cell* 173(2018):1692-1704.

Address for Correspondence
Suzanne Bakken, sbh22@cumc.columbia.edu.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190669

Information Systems, Statistical Information Availability and Decision Making in the Primary Health Care Level of the City of Buenos Aires

María Belén Islas[a], Cecilia Palermo[a], Micaela Zapata[a], Santiago Esteban[a]

[a] *Health Information Systems Office, Ministry of Health, Buenos Aires City, Argentina*

Abstract

The availability of statistical information is usually associated with 'better decision making', fundamentally within the paradigm of evidence-based decisions. Thus, it is interesting to understand the demands and expectations, interpretations and effective use of statistical information by decision-makers at the primary health care level, considering the implications of the recently implemented Information System in Public Health (Buenos Aires, Argentina).

Keywords:

Attention, Decision Making, Statistics

Introduction

The introduction and development of information systems has an impact on the way in which significant data is generated for decision-making at the clinical, institutional, organizational and resource management levels. The implementation of an Information System (IS) in health -including an Electronic Health Record (EHR) cluster- began, in Buenos Aires City, in the year 2016. This process brought certain transformations in the way in which information is generated, available for and used by decision-makers: previous statistic information was merely descriptive and resulted of hand-filled forms, and the implementation of an EHR lead to the development of new strategies to collect and obtain information.

Within the introduction of an IS, the information sources are unified. This is considered an advantage, because "decision making by health care professionals is often complicated by the need to integrate ill-structured, uncertain, and potentially conflicting information from various sources" [3]. Thus, it is important to understand the way in which decision-makers work with and use the available statistic information. Even though statistical information is referred to be used for either management and clinical aspects, this poster analyze the use of statistical information for decision-making involving the collection, systemization, structuring, storage, combination, distribution and provision of valuable information - considering the impact of the IS introduction in the resulting access to information.

In the field of public health, decision-making is defined by Cediel Becerra as *"the process that implies being able to continuously respond to changing circumstances, anticipating emerging threats and identifying ways to control them"* [1]. It implies decisions that affect the health of entire communities or populations. The decision-making process in managerial roles is a fundamental mean to promote the achievement of objectives set by managers in every institution; in this sense, its exercise cannot depend on the sole experience or intuition of the decision-maker. It should be based on knowledge and interpretation of the best available evidence and local data.

With the implementation of the Electronic Health Record in the first level of care, we aspire to obtain high quality statistical data to support these decision-making processes. Actually, the information resultant of the use of EHR by professionals of the primary health care level is been processed with the aim to develop a dashboard with extensive sets of information. Nevertheless, it is not clear how managers at the primary care level make actual use of the available information.

This work aims to analyze the use of statistical and(or) scientific information in the decision-making process carried out by health center managers of the primary health care level.
Thus, it is necessary to identify the decision-making main topics, the different agents involved through the process, and the difficulties and issues that may arise along it. The main decisions topics and specific demands of information were identified, according to the experiences of Managers of Primary Health Care Level Centers.

As well, it analyzes the impact of EHR concerning the access and availability of information and specific information requirements issued by managers at the primary care level. In light of the little reliable evidence quantifying the extent to which research evidence is used in public health decision making processes [4], in the specific case of Buenos Aires city, the question on which this work relies is: an increased availability of information, necessarily involves an upgrade in decision-making? This question is significant to conduce local analysis, considering background information refers to international settings [5].

Methods

Setting

The Buenos Aires city healthcare network is conformed by a Ministry of Health, 33 Hospitals, 81 Primary Care Centers and it is structured into 12 areas in order to organize health care delivery. The health system employs a total of 41,000 people. Since June 2016 an Electronic Health Record (EHR) is being gradually implemented in the outpatient setting. [2] Until October 2018, 60 healthcare facilities are using EHR and more than 2.5 million of clinical notes had been registered.

Through a qualitative approach, 9 (nine) managers of the Primary Care Centers participated in individual semi-structured interviews. Throughout these interviews, the following dimensions were explored: circuit and current flow of information; personal experience as decision makers; influence of the available statistical data in the decision-making process; and the expectations about information available on dashboards.

Results

During the interviews, managers were asked about their decision-making governance areas. They defined three key dimensions:

Management: this item involves supplies, personnel and infrastructural demands.

Activities: refers to specific activities that are designed according to general prevalence of health problems and characteristics of the population.

Prevention: on one side, this is bound to identify health-trends in order to plan interventions. On the other, it relates to comparative studies between actual and historic information, to identify health-fluctuations in the population.

By processing the content of the interviews conducted, we were able to identify a 'theoretical' and a 'practical' dimension, associated with the importance of statistical information availability. On a theoretical level, the interviewees said that a greater amount of statistical information would contribute to characterize and typify health problems and identify epidemiological tendencies along the city territory. They also mentioned that if typification is not correctly achieved by the use of data, it might be misled by intuition or personal-biases of the healthcare teamwork, and it would generate erroneous planning and diagnoses. On the other hand, on a more practical level, the interviewees revealed that the usefulness of statistical information in decision-making is post-factum, as decisions are not necessarily based on statistical evidence.

Conclusions

The analysis of the decision-makers' perspective points out that there is not a detailed proposal related to the content and availability of statistical information. The demands usually refer to specific topics, such as patient absenteeism, amount of consultations according to each professional, chief complaints and concerns. Even though professionals point out the strong potential of data resulting of EHR, there is still uncertainty over the access and usefulness of information associated to decision-making. Also, there is no specific or structured notion regarding the ideal way to make information available: the electronic format generates a perception of potentiality but at the same time it is not clear to them how data is generated and processed.

Ideally, actors with decision-making power in public health providers should incorporate scientific evidence to make decisions, develop policies and implement programs. However, in reality, these decisions are often based on short-term demands.

Statistic information availability is not sufficient on its own if there is not a clear view of which data is needed, what for, and the how is these data interpreted. Even though the analysis presented on this poster is exploratory, it intends to show the professional's expectations over the use of information regarding decision-making processes.

The importance of the availability of information is not exclusively at the central level or in decision-making positions, but it is also extremely important for professionals (who in turn are the ones who generate it) and the use they can make of it.

References

[1] A.W. Kushniruk, Analysis of Complex Decision-Making Processes in Health Care: Cognitive Approaches to Health Informatics. 2011.

[2] MV. Giussi, A. Baum, F. Plazzotta, P. Muguerza, F. González Bernaldo de Quirós, Change management strategies: transforming a difficult implementation into a successful one, *Stud Health Technol Inform.* **245** (2017), 813-817.

[3] N. Cediel Becerra, Journal of Public Health, Evidence-based public health decision-making tools which can also be used for prioritizing disease, 2009

[4] R.C. Brownson, J.G. Gurney, Evidence-Based Decision Making in Public Health, 1999

[5] L. Orton, F. Lloyd-Williams, D. Taylor-Robinson, M. O'Flaherty, S. Capewell, The Use of Research Evidence in Public Health Decision Making Processes: Systematic Review. *PLoS One* **6(7)** (2011), e21704.

Address for correspondence

María Belén Islas

Ministry of Health, Buenos Aires City, Argentina.

E-mail: mbislas@buenosaires.gob.ar

Phone Number: +54 9 11 66221122

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190670

Standardized Observational Cancer Research Using the OMOP CDM Oncology Module

Rimma Belenkaya[a], Michael Gurley[b], Dmitry Dymshyts[c], Sonia Araujo[d], Andrew Williams[e], RuiJun Chen[f], Christian Reich[g]

[a] OHDSI Oncology Workgroup, Memorial Sloan Kettering Cancer Center, New York City, NY, USA,
[b] Clinical and Translational Sciences Institute, Northwestern University, Chicago, IL, USA,
[c] Odysseus Data Science Inc, Cambridge, MA, USA,
[d] Real World Analytics Solution, IQVIA, London, UK,
[e] Maine Medical Center Research Institute, Center for Outcomes Research and Evaluation, Portland, ME, USA,
[f] Biomedical Informatics Department, Columbia University Medical Center, New York City, NY, USA,
[g] Real World Analytics Solution, IQVIA, Cambridge, MA, USA

Abstract

Observational research in cancer requires substantially more detail than most other therapeutic areas. Cancer conditions are defined through histology, affected anatomical structures, staging and grading, and biomarkers, and are treated with complex therapies. Here, we show a new cancer module as part of the OMOP CDM, allowing manual and automated abstraction and standardized analytics. We tested the model in EHR and registry data against a number of typical use cases.

Keywords:
Oncology, Research, Standardized

Introduction

Observational research utilizes secondary use of observational healthcare databases for pharmacoepidemiology, health outcomes, and health services research. A variety of statistical and epidemiological methods produce estimates of desired or adverse effects caused by medical interventions and associated statistical artifacts such as confidence intervals and p-values. Cancer research is no different from general observational research. However, cancer characterization requires much more details such as: histology of the tumor tissue, anatomical sites affected by the disease, tumor size, affected lymph nodes and metastases, tumor markers (conventional and genomic aberrations), and standardized disease grade and staging.

The treatment of cancer is also more complex than most other conditions since it involves surgery, radiotherapy, chemotherapy, targeted therapy, and immunotherapy. Many of these are administered in cycles. In addition, chemotherapy compounds are applied in predefined regimens.

Cancer diagnosis and treatment interventions are complicated healthcare interactions stretching over longer periods of time. To perform observational cancer research, low-level events need to get abstracted into higher level disease episodes, treatments, and outcomes such as initial diagnosis, drug regimen, radiotherapy, surgical resection, response to treatment, overall and disease-free survival, etc.

Systematic and standardized analytics requires data harmonization, which also enables distributed research networks [1]. Both data format and representation (coding) need to be standardized for true federated analytics. Currently, there is no comprehensive data model or standard semantic terminology system covering the cancer domain available in the public domain to support this approach.

ICD-10 is the most common coding scheme for defining the condition in observational data. However, it details mostly anatomical sites of cancer, while histology and tumor attributes are poorly covered. ICD-O-3 is a classification with explicit topology and histology domains, but it is not connected to other terminologies. The two coding systems are not combined to form meaningful conditions and it lacks detailed tumor characteristics. The CAP Cancer Protocols solves the latter problem by defining a mandatory and specific set of data elements for each cancer type, called synoptic reporting. However, this resource does not have a freely available computer-readable representation and does not have a properly defined terminology. The NAACCR Data Dictionary is addressing this issue and also contains terminologies for treatment. However, its data elements are not conceptualized to carry a distinct meaning independent from context, and for most of its data elements, it does not utilize or map to any external existing coding system. For clinical trials, CDISC developed comprehensive Therapeutic Area Standards, but it covers only four cancer types and also lacks relationships to external standards. HL7 FHIR is developing cancer-specific profiles, but only Breast Cancer Staging is available today. The Nebraska Lexicon addresses all of the above issues, creating a freely available ontology embedded in SNOMED-CT, but it is a work in progress, and currently only covers breast, colorectal, lung cancer, and malignant melanoma.

For cancer treatments, the situation is equally inconsistent. RxNorm is the standard vocabulary for the OMOP drug domain, but it is not specific to oncology and does not contain regimens. Both the NCI List of cancer drugs, the SEER*Rx, and the NCI Metathesaurus contain cancer specific drugs and their indication, but not regimens with constituent ingredients normalized to RxNorm. The NCCN has comprehensive drug and regimen information, but nothing is available in the public domain. The SEER OROT contains a comprehensive cancer drug repository with maps to NDC and HCPCS. Finally, HemOnc contains a rich drug and regimen ontology, including indications and links to RxNorm but requires proper life cycle of their concepts.

Here, we introduce the OMOP Cancer Module. It consists of a comprehensive hierarchy of cancer conditions, including topography, histology and tumor attributes, as well as treatments with regimens. It also contains the data model for abstracted

episodes and a mechanism to link them to the lower level clinical detail. We converted four cancer databases into this model and tested it for a number of typical use cases.

Methods

We used the following standardized vocabularies:

1. Cancer diagnosis: We incorporated all ICD-O-3 histology and topography codes and created equivalent or "uphill" links to SNOMED. We combined all histologies and topologies to create standard conditions. Instead of instantiating the entire Cartesian product, we only used reported combinations derived from ICD-O-3 site, ICD-O-3 SEER Site/Histology Validation List, Columbia University Medical Center (CUMC) Cancer Registry, and Northwestern University (NU) Tumor Registry. We mapped the resulting conditions that existed in SNOMED, and connected the rest of the conditions to their histology and topography and linked them to the higher level SNOMED concept.
2. Modifiers: We incorporated and instantiated both the NAACCR Data Dictionary and the Nebraska Lexicon and de-duplicated the resulting corpus. For ambiguous NAACCR concepts, we pre-coordinated the data elements with its site-specific context.
3. Treatment episodes: We instantiated NAACCR surgery and radiation concepts. We incorporated HemOnc drug regimens and classification. We used OROT to connect drug concepts to the underlying source codes.

Episode model

We introduced new Episode and Episode_Event tables to represent disease and treatment episodes and their connection to lower level events (Figure 1).

Database instantiation

We converted Electronic Health Records (EHRs) and registries from four participating institutions into the new model.

Abstraction

We used an automatic abstraction algorithm to derive disease and treatment episodes [2]. We also performed manual abstraction and compared it to the automatic abstraction.

Empirical Model Testing

We compared the manual to the automatic abstraction and characterized the results. We then tested the following use cases using either abstractions.

- Complete history of cancer disease for patients cohort
- Complete list of modifiers for cohort of patients
- Stratification of therapy by high level classes based on vocabulary hierarchy

For each use case, we distributed a standard query through the network of databases and experts validated the results.

Results

We successfully generated the Standardized Vocabularies and converted the data from four participating institutions: NU, CUMC, Maine Medical Center Research Institute, and IQVIA.

We achieved 95% of coverage for the diagnoses reported in the source data by the Standardized Vocabularies, the remaining 5% represented rare cancers. The positive predictive value (PPV) of the automatic abstraction compared to manual chart review and Cancer Registry was 96% for any cancer diagnosis for four cancers, 100% for chemotherapy, 98% for hormone therapy, 100% for immunotherapy, and 86% for radiotherapy.

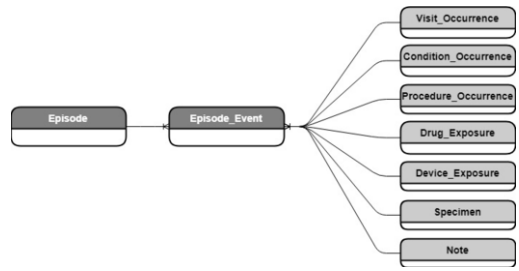


Figure 1– New Episode and Episode_Event tables

Conclusions

We successfully standardized the key aspects of cancer diseases required for observational research. Further work will include: extending Standardized Vocabularies to provide complete coverage of diagnoses and diagnostic modifiers, modeling of genomic data and outcomes, and creating automatic abstraction algorithms that would be utilized in an Open Network of cancer databases.

References

[1] G. Hripcsak, P.B. Ryan, J.D. Duke, N.H. Shah, R.W. Park, V. Huser, M.A. Suchard, M.J. Schuemie, F.J. DeFalco, A. Perotte, J.M. Banda, C.G. Reich, L.M. Schilling, M.E. Matheny, D. Meeker, N. Pratt, and D. Madigan, Characterizing treatment pathways at scale using the OHDSI network, *Proceedings of the National Academy of Sciences of the United States of America* **113** (2016), 7329-7336.

[2] N.M. Carroll, K.M. Burniece, J. Holzman, D.B. McQuillan, A. Plata, and D.P. Ritzwoller, Algorithm to Identify Systemic Cancer Therapy Treatment Using Structured Electronic Data, *JCO Clin. Cancer Informatics*. (2017).

Address for correspondence

Rimma Belenkaya
belenkar@mskcc.org.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190671

Measuring Healthcare-Associated Infection Outcomes: Enhanced Surveillance to Include Process Adherence for Quality Improvement

Simon J. Burrell[a], Ann L. Bull[a], Leon J. Worth[a, b]

[a] Victorian Healthcare Associated Infection Surveillance System (VICNISS) Coordinating Centre, Doherty Institute, Melbourne, Victoria, Australia,
[b] Department of Medicine, University of Melbourne, Melbourne, Victoria, Australia

Abstract

To prevent healthcare-associated infections, there are a range of clinical practices that should be followed. For example, appropriate administration of prophylactic antibiotics [process] is essential to reduce risks of surgical site infections post-operatively [outcome]. Monitoring adherence to these processes provides insights into potential causes of infection. The Victorian Healthcare Associated Infection Surveillance System (VICNISS) captures process data in the same system as outcome data, thereby providing integrated data to support quality improvement within healthcare and reduce the burden of healthcare-associated infections.

Keywords:

Cross Infection, Antibiotic Prophylaxis, Population Surveillance

Introduction

The Victorian Healthcare Associated Infection Surveillance System (VICNISS) was established in 2002 with the primary aim of reducing the incidence of healthcare-associated infections (HAI) in the Australian state of Victoria [1].

While the initial program was targeted towards large acute hospitals (greater than 100 acute beds), subsequently a smaller hospitals surveillance program was established and more recently residential aged care facilities have commenced the HAI surveillance program. There are currently approximately 460 participating facilities.

Surveillance activities through the VICNISS program can be classified into three types: infection outcomes, preventative process, and vaccination compliance/immunity.

When facilities have an increase in infection rate/s, clinicians, quality improvement staff and hospital executives seek to understand how to lower the rate and improve patient outcomes [2]. A sole focus upon monitoring infection outcomes does not enable relevant factors contributing to risk to be identified. It has been demonstrated that clinical processes (or multimodal processes, 'bundles of care') can be implemented to reduce the risk of acquiring a range of infections in healthcare settings [3].

The importance of process adherence in the prevention of surgical site infections is documented in a World Health Organisation (WHO) guideline [4].

Table 1– Examples of processes that are integrated with infection surveillance in the VICNISS program

Infection (outcome)	Associated clinical practice (process)
Surgical site infection	Antibiotic prophylaxis (choice, duration and timing)
Staphylococcus aureus bloodstream infection (s)	Hand hygiene compliance, Central line insertion practices
Central line-associated bloodstream infections	Central line insertion practices
Surgical site infections following colorectal surgery	Colorectal surgery process adherence (multimodal bundle)

Process monitoring requires additional data collection and therefore resource allocation. In our experience, additional data required to monitor the prevention process of antibiotic prophylaxis can be quite easily collected in tandem with surgical site infection data.

This paper describes how we integrated process adherence monitoring, namely surgical antibiotic prophylaxis, with surgical site infection surveillance.

Methods

The VICNISS SSI surveillance program is based on the U.S. Department of Health & Human Services - Centers for Disease Control and Prevention (CDC) National Healthcare Safety Network (NHSN) [1]. In Australia, national medication prescribing guidelines, including recommendations for surgical antibiotic prophylaxis, are widely available and utilised within healthcare [5].

Surveillance for post-operative surgical site infections (SSI) requires the collection of demographic and surgery details for every patient undertaking the surgery under surveillance. Many of these data elements are routinely captured by hospital IT systems and are extracted for infection surveillance use.

Additional data are required to monitor concordance of antibiotic prescribing with the national guidelines [5]: antibiotic agent, time of administration, and if the antibiotic is continued for greater than 24 hours.

Hospital infection prevention staff extract data from their hospital system and either upload a batch file or enter the data onto a web form in the VICNISS secure portal. Validation and

quality checks are processed at time of entry. Data are stored in a relational database structure.

We expanded our web form to allow the collection of additional antibiotic fields at the same time as the surgical procedure data are entered. These fields were also added to the bulk upload process.

To enable reporting of antibiotic prophylaxis compliance the national antibiotic guidelines are transcribed into single business rules that can be applied to each surgical procedure and antibiotic combination. As such, there are four possible results of applied rules: optimal, adequate, inadequate and unknown.

Business rules are stored in a database table and code written to run on a nightly schedule that applies the rules to incoming data. Results are saved into a subsequent table. Once data have been uploaded to VICNISS, individual hospitals are able to run reports showing their infection rates together with process compliance over time [6].

Results

Figure 1 summarises the time-trend for appropriate antibiotic choice for surgical antibiotic prophylaxis. Over time, 'optimal' choice has increased while 'inadequate' choice of agents has diminished. The surgical procedures included in Figure 1 are CAGS, cardiac, hip prosthesis, knee prosthesis, and colorectal surgery.

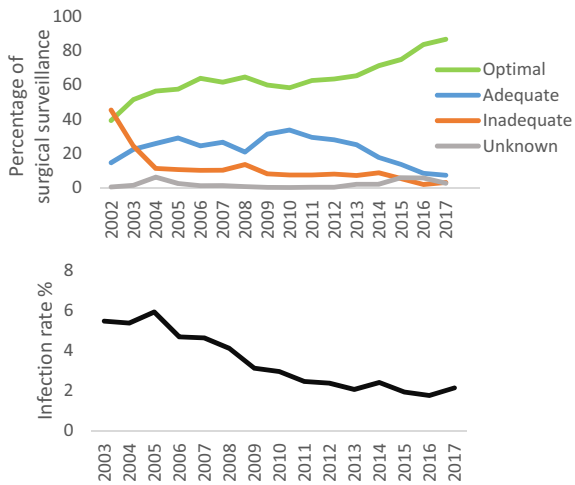


Figure 1– Top: Assessment of appropriateness of prophylactic surgical antibiotic choice over time (high-risk procedures). Observed increase in optimal prescribing. Bottom: Reported surgical site infections following surgery over time (high-risk procedures). Observed reduction in infections outcomes.

In addition to antibiotic choice, the timing of administration and duration of prophylaxis are captured. For prolonged procedures, the administration of repeat doses of antibiotic for prophylaxis, are also recorded.

At a hospital level, these data allow clinicians to identify key areas for quality improvement and education. Antibiotic stewardship teams may opt to focus upon choice, timing or repeat dosing in strategies to enhance quality of prescribing.

Conclusions

While it is well-documented that infection surveillance activities are associated with a reduction in healthcare-associated infections, we demonstrate that a small increase in captured data (to include relevant processes) is an efficient way to gain greater insight into possible opportunities for practice improvement to reduce infection risk[7].

We are achieving our aim of reducing healthcare-associated infections and very efficiently answering the first question that front-line hospital staff ask: 'What should we do when we have an increased infection rate?'

Acknowledgements

VICNISS is funded by the Victorian Department of Health and Human Services, is a Service of the Royal Melbourne Hospital and located at the Peter Doherty Institute for Infection and Immunity. We acknowledge the work of infection prevention and control staff in collecting the surveillance data at each healthcare facility participating in the VICNISS program.

References

[1] Russo PL, Bull A, Bennett N, Boardman C, Burrell S, Motley J, Berry K, Friedman ND, Richards M. The establishment of a statewide surveillance program for hospital-acquired infections in large Victorian public hospitals: a report from the VICNISS Coordinating Centre. Am J Infect Control 2006; 34:430-436.

[2] Thursky K, Lingaratnam S, Jayarajan J, Haeusler GM, Teh B, Tew M, Venn G, Hiong A, Brown C, Leung V, Worth LJ, Dalziel K, Slavin MA. *Implementation of a whole of hospital sepsis clinical pathway in a cancer hospital: impact on sepsis management, outcomes and costs*. BMJ Open Qual 2018; 7:e000355.

[3] Sutherland T, Beloff J, Lightowler M, Liu X, Nascimben L, Urman RD. *Process measures* to improve perioperative prophylactic antibiotic compliance: quality and financial implications. Health Care Manag (Frederick) 2014; 33:289-96.

[4] World Health Organization. Global guidelines for the prevention of surgical site infection. Geneva: World Health Organization; 2016.

[5] Antibiotic Expert Group. Therapeutic guidelines: antibiotic. Version 14. Melbourne: Therapeutic Guidelines Limited; 2010.

[6] Bull AL, Worth LJ, Spelman T, Richards MJ. Antibiotic Prescribing Practices for Prevention of Surgical Site Infections in Australia: Increased Uptake of National Guidelines after Surveillance and Reporting and Impact on Infection Rates. Surg Infect (Larchmt) 2017; 18:834-840.

[7] Bull AL, Thursky KA, Richards MJ, Worth LJ. Monitoring the appropriateness of surgical antibiotic prophylaxis prescribing in Australia: valid and meaningful indicators provide 'data for action'. Anaesth Intensive Care 2016; 44:121-2.

Address for correspondence

Simon Burrell simon.burrell@mh.org.au

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190672

Toward CDSS Benefiting Elderly Patients with Olfactory Disorders

Xiaoqiu Chen[a], Chao Shang[b], Fan Li[c], Zhiwei Cao[a], Yang Gong[d]

[a] *ENT Department of Shengjing Hospital, China Medical University, Shenyang, Liaoning Province, China,*
[b] *Department of Network Center, China Medical University, Shenyang, Liaoning Province, China,*
[c] *School of Medical Informatics, China Medical University, Shenyang, Liaoning Province, China,*
[d] *School of Biomedical Informatics, University of Texas Health Science Center at Houston, Texas, USA*

Abstract

Patients with olfactory disorders are often encountered with a complex workflow and unsatisfied results. This article explores an informatics application to manage olfactory disorders with clinical decision support. Evidence from PubMed was used to analyze and extract the workflow, according to the "five rights" CDS framework. Five focus points and corresponding tips were identified. This work provides disease analysis and life guidance for prediction of the occurrence of more serious diseases.

Keywords:

Olfactory disorders, elderly, clinical decision support

Introduction

Olfactory disorders (OD) are common diseases in ENT clinics. The elderly are the main population affected by OD and are sometimes unaware of risky circumstances. It has been proven that OD are also early signs of neurodegenerative disorders [1].

The current treatment of OD is far from satisfactory. Doctors suffer from only a limited number of methods that can be used for treatment. Patients have to accept unsatisfactory results after complicated examinations and treatment.

The clinical decision support system (CDSS) is typically an effective application of electronic health records (EHR) [2]. Osheroff's "five rights" of clinical decision support are a good summary of what needs to be done in design, development, and implementation of CDSS application [3].

This poster explores how we apply the "five rights" CDS framework to effectively manage OD, from the perspective of all spectrum of OD participants including doctors, patients, patient family members or caregivers, aiming to create personalized informatics prescriptions, and thus improve patients' quality of life.

Methods

Articles related to senile OD were captured from the PubMed database between 2008~2018. MeSH terms "olfactory disorders" and "age 45+ years" were used in the literature search strategy. A total of 1047 articles were retrieved. 73 articles were included and further reviewed based on the titles and abstracts, once the following information was extracted from each article: the morbidity, onset characteristics, inducing and prognostic factors, hazardous events, treatment methods and cure rates of OD. Using the "five rights" as a CDSS theoretical framework, three professors of otolaryngology were consulted, and combined with the application data of the EHR system in the nasal clinic where the researcher worked (more than 40,000 outpatient visited per year), the entire OD workflow was analyzed.

Results

In the EHR system, there are five focus points found to intervene with the treatment workflow of OD as the workflow progressed.

The 1st focus point: age over 50

The incidence of OD in the elderly is extremely high (*Table 1)*. The growing aging population will inevitably surge the OD incidence accordingly.

Table 1— Epidemiologic Surveys of OD Patients in Different Countries

	Age Group			
Country	**≥50ys**	**≥60ys**	**65-80ys**	**≥ 80ys**
2002 U.S	25%[4]			
2004 SWE	19.1%[5]			
2009 U.S			50%[6]	75%[7]
2011 U.S		12.5%[8]		

The 2nd focus point: lasting more than 1 year

The basis for OD treatments relies on the identification of the cause. In a latest systematic review, OD commonly origin from five causes: naso-sinal disease, head trauma, upper respiratory infection [URI], age-related loss, and congenital disorders [9]. Within a year, all of them can be treated accordingly and get a stable therapeutic effect.

Other than the above-mentioned causes, the National Health and Nutrition Examination Survey (NHANES) self-reported olfactory function index appears to offer 12 months to screen for anosmia (loss of olfaction) in the community [10].

The 3rd focus point: hazardousevents caused by OD

The hazardous events caused by OD are not rare. In 2004 and 2014, two scientists separately published the results that the incidence of experienced any hazardous event (*Table 2*) is progressively increased with degree of impairment and there is significant differences based on age [11][12]. The questions about hazardous events are simple indicators for assessing the impact on patients' previous life, and are warnings and reminders for their future life.

Table 2 Ratio of hazardous events among OD patients

OD pts	**Pts with events**	**events**	**Cooking related**	**Spoiled food**	**Neglect gas leak**	**Neglect fire**
340	124(37%)	165	45%	25%	23%	7%
543	179(33%)	266	47%	30%	12%	11%

Figure 1- Five Focus point intervention towards CDSS

The 4th focus point: questionnaire of OD (QOD)

The QOD is widely used in clinical and field surveys since it was developed by Thomas Hummel and Johannes Frasnelli in 2005 [13]. Due to its high degrees of reliability and validity, the QOD is a self-reporting and systematic questionnaire implemented at the national level in the United States. Meanwhile the QOD score comparison before and after treatment can be used to evaluate the patient's life status, ease communication barriers caused by information asymmetry, and resolve the contradiction between doctors and patients.

The 5th focus point: Family history and early manifestations of neurodegenerative diseases.

About 20 years ago, neurologists observed that OD are a cardinal feature of several neurodegenerative diseases such as Alzheimer's Disease (AD) and Parkinson's Disease (PD). In both diseases, the deficit is present in 85~ 90% of early-stage patients. Subsequent research suggests that OD in first-degree relatives of PD patients is associated with at least a 10% increased risk for developing clinically defined PD within 2 years [14]. The EHR is an ideal location where the above five points can be found easily.

During the regular workflow, how the 5 focus points to intervene in the follow-up treatment of OD patients intervened is shown in Figure 1.

Conclusion

The current OD workflow is often incomplete for the following reasons: (1) No one has explained the correlation between age and course with prognosis; (2) No one pays attention to the family history of neurodegenerative diseases and whether dangerous events occur in the patient's life; (3) The main purpose of current OD diagnosis and treatment is still to find the cause, often without accurate quantitive indicators; and (4) Almost no one gives a valid informatics prescription. With the help of EHR, the OD diagnosis and treatment process under the intervention of CDSS can help reduce medical errors, clear referral, predict the possibility of ND, and improve patient life treatment.

References

[1] J. Hidalgo, G. Chopard, J. Galmiche, L. Jacquot, and G. Brand, Just noticeable difference in olfaction: a discriminative tool between healthy elderly and patients with cognitive disorders associated with dementia, *Rhinology* **49** (2011), 513-518.

[2] D. Blumenthal and M. Tavenner, The "meaningful use" regulation for **electronic** health records, *The New England journal of medicine* **363** (2010), 501.

[3] J.A. Osheroff, I. Healthcare, and S. Management Systems, *Improving outcomes with clinical decision support: an implementer's guide*, HIMSS, Chicago, IL, 2012.

[4] C. Murphy, C.R. Schubert, K.J. Cruickshanks, B.E.K. Klein, R. Klein, and D.M. Nondahl, Prevalence of Olfactory Impairment in Older Adults, *JAMA* **288** (2002), 2307-2312.

[5] A. Brämerson, L. Johansson, L. Ek, S. Nordin, M. Bende, f. Medicinska, v. Institutionen för klinisk, p. Institutionen för, f. Samhällsvetenskapliga, u. Umeå, and h. Öron- näs- och, Prevalence of olfactory dysfunction: the skövde population-based study, *The Laryngoscope* **114** (2004), 733-737.

[6] D. Lafreniere and N. Mann, Anosmia: Loss of Smell in the Elderly, *Otolaryngologic Clinics of North America, The* **42** (2009), 123-131.

[7] R.L. Doty, The olfactory system and its disorders, *Seminars in neurology* **29** (2009), 74.

[8] C.R. Schubert, K.J. Cruickshanks, B.E.K. Klein, R. Klein, and D.M. Nondahl, Olfactory impairment in older adults: Five-year incidence and risk factors, *The Laryngoscope* **121** (2011), 873-878.

[9] G.A. Scangas and B.S. Bleier, Anosmia: Differential diagnosis, evaluation, and management, *Am J Rhinol Allergy* **31** (2017), 3-7.

[10] H.J. Hoffman, S. Rawal, C.-M. Li, and V.B. Duffy, New chemosensory component in the U.S. National Health and Nutrition Examination Survey (NHANES): first-year results for measured olfactory dysfunction, *Reviews in Endocrine and Metabolic Disorders* **17** (2016), 221-240.

[11] D.V. Santos, E.R. Reiter, L.J. DiNardo, and R.M. Costanzo, Hazardous Events Associated With Impaired Olfactory Function, *Archives of Otolaryngology–Head & Neck Surgery* **130** (2004), 317-319.

[12] T.S. Pence, E.R. Reiter, L.J. DiNardo, and R.M. Costanzo, Risk factors for hazardous events in olfactory-impaired patients, *JAMA Otolaryngol Head Neck Surg* **140** (2014), 951-955.

[13] J. Frasnelli and T. Hummel, Olfactory dysfunction and daily life, *European Archives of Oto-Rhino-Laryngology* **262** (2005), 231-235.

[14] R.L. Doty, The olfactory system and its disorders, *Semin Neurol* **29** (2009), 74-81.

Address for correspondence

Yang Gong, MD, PhD

7000 Fannin St Suite 600, Houston, TX 77030, USA

Yang.Gong@uth.tmc.edu

Phone: +1 713 500 3547, Fax: +1 713 500 3907

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190673

A Generic IT Infrastructure for Identity Management and Pseudonymization in Small Research Projects with Heterogeneous and Distributed Data Sources Under Consideration of the GDPR

Hauke Fischer[a], Rainer Röhrig[a], Volker Sebastian Thiemann[a]

[a] *Division of Medical Informatics, Carl von Ossietzky University, Oldenburg, Germany*

Abstract

Each year, millions of rescue operations occur in Germany. Despite rising numbers, there is a shortfall of systematic quality management (QM). Focusing on finding the required quality parameters, this work aims at linking heterogeneous and distributed data sources on the patient-level. Presented here is necessary IT infrastructure for linking patients' data records properly and efficiently. The focus of the IT infrastructure is to provide a lightweight and easy applicable solution with an utmost generic approach.

Keywords:

Privacy, Health Services Research, Emergency Medical Services

Introduction

"*About 14 million emergency runs take place annually in Germany. However, little systematic quality assurance (QA) that includes the course of the patient, medical outcomes and interregional comparisons is conducted*" [1]. Founded to address this shortcoming, the *Inno_RD* [2] research project investigates parameters necessary to establish a systematic quality management for German emergency medical service operations. In order to evaluate the medical outcome of these, medical data of patients from different sources needs to be linked. In Germany, health service providers and health insurances share specific data sets, consisting of billing information needed for later reimbursement. For each case, the data sets contain the health insurance number of the respective patient. Connecting the data sets from multiple service providers and health insurances allows for the tracking of the patient course beyond single contacts to a specific service provider. Thus, there is the need for an IT infrastructure to link the data on a patient-level.

One must consider the recent changes in regulation induced by the general data protection regulation (GDPR), especially regarding sensitive information such as health data. When following the GDPR, it is a requirement to link data without directly sending and using the patients' names or identifying information, as the use of pseudonymization is encouraged if not mandatory when processing personalized (health) data. The GDPR defines pseudonymization as "*the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person*" (Article 4 (5)).

We developed a data protection concept, determining the technical requirements for the implementation of the IT-infrastructure. Our goal was to deliver a lightweight and easy-to-handle IT infrastructure for identity management and data linkage in small, distributed research networks. This IT infrastructure must meet the needs of offering local file processing and batch processing of data.

State of the Art

Different approaches for the given challenge can be found in literature. The *ToolPool Gesundheitsforschung* [3] has provided an overview of established solutions for the German health care system. The *ToolPool* contains approaches like the *Mainzelliste* [4] or *MOSAIC* [5]. The *Mainzelliste* is an identity management system based on pseudonymization and REST-Services for distributed data warehouses developed by the University of Mainz. *MOSAIC* is a modular approach for data management in epidemiological studies, which provides identity management as well as pseudonymization tools. The University of Greifswald developed MOSAIC. Most of the existing solutions process identifying patient data consisting of names and additional personal information for identity management. These systems employ complex algorithms to convert strings into phonetic codes or use word distance techniques to ensure proper linkage [6]. The latter approaches provide identity management in the form of federated IT infrastructures. Medical research registers or data warehouses with continuous data flow often use these techniques. The investigated tools are not designed for local file processing and fast batch processing of large data sets.

Concept

Based on the data protection concept and the project's context, we performed a requirement analysis for the IT infrastructure. We prioritized these requirements according to the MoSCoW prioritization [7]. We base the core concept for data linkage within the project on a cryptographically secure identifier generator and pseudonymization service that is supposed to be able to translate the health insurance number used as a pseudonym by health service providers and health insurances into a new pseudonym. The direct usage of the health insurance number is not possible. A pseudonym should minimize the possibility to identify the respective patient, and the health insurance number identifies the patient on each provider site. Thus, we need to build cryptographically a new pseudonym. One can assume, that the quality of the data of the health insurance

number is quite high due to its purpose as key value in the billing process.

Using these concepts the data is processed based on a two level pseudonymization approach in our IT infrastructure. On the first level of the pseudonymization process, we generate a permanent identifier (PID). For generating the PID, we used the personal identity data (IDAT, being the health insurance number) of each patient. The PID is sent to a pseudonymization service (PSS). The PSS performs the second level of the pseudonymization process. For each deposited PID the PSS generates a unique permanent pseudonym (PSN).The data analyst receives this PSN in place of the original IDAT, enabling the linkage of different datasets. Each IDAT and thus PSN is linked to specific medical data (MDAT).The MDAT is not touched as it routed through to the PSS in encrypted form [8]. With the PSN, the data analyst can link data from different data providers on a patient-level. With this approach, none of the three participants hold enough information to single-handedly trace the research data back to a patient or vice versa (c.f. Fig. 1).

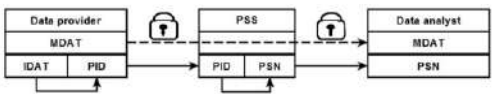


Figure 1: Pseudonymization and transmission of encrypted data

We base the PID generation on a pepper-based hash method. Hashing with pepper is a cryptographic process that adds a secret phrase serving as a password to any input. This method shares the phrase between all data providers and keeps it away from identity management, pseudonymization and the data receiver. The aim of this approach was to minimize the possibility of reconstructing the health insurance number of a patient. Using our IT infrastructure, we then send the PID to the identity management, which generates a PSN for each PID and a temporary ID (tempID). Identity management sends the tempID to the data provider, to replace the health insurance number in the original data set. The data receiver then collects the data set with the tempIDs. The receiver of the data set sends the tempID to receive the respective PSN replacing the tempID.

We base the IT infrastructure on a client-server-architecture where the PSS functions as a server. It provides and manages the allocation of PIDs, tempIDs and PSNs. A software-application, or the so-called pseudonymization software, manages the client-side. The PSS acts as a passive service provider reacting only to send requests from the pseudonymization software. Designed to be firewall-friendly, this approach gives full control of all connections to the client software. Also based on the GDPR principle of data minimization (Article 5 (1) (c)), this system offers full control over the data delivery process to each data provider. The pseudonymization software contains a module for data receivers and a module for the data provider. An access management via authentication keys decouples the different functionalities of the two user groups.

Implementation

On the server-side of the IT infrastructure we are using Java Enterprise Edition for a lightweight server application. On the client-side, we employ an executable Java desktop application with a graphical user interface able to process data in CSV format. The desktop application offers encryption of the pseudonymized data based on a hybrid encryption process using AES and RSA encryption. The software also supports the transfer of files via a separate FTP file server. This option simplifies the process of pseudonymization and sending the data in an encrypted way for the user to a user interface that is similar to a simple file upload graphical user interface.

Lessons learned

The designed and implemented IT infrastructure delivers a generic tool to link data records for small research projects with distributed data sources. The main prerequisite is a unique identifier that identifies the subjects across all data sources. The approach encapsulates the data provider, the data analyst and the trusty agency through our pseudonymization concept. This is reached through the use of the client-server architecture and locally executable client software.

During the development process, the need for a toolbox combining multiple methods needed for small research projects became obvious. This toolbox needs to integrate the developed tools and offer further instruments, which we need to identify. There is a need for potential solutions inter alia for encryption, pseudonymization, record linkage, and data preparation as well as legal requirements, such as the data protection concept. These could save research projects time and gives the possibility to focus more on research itself.

Acknowledgements

Funding by Innovationsfond, No. 01VSF17032.

References

[1] S. Piedmont., D. Brammen, D. Branse, K. Focke, W. Kast, B.-P. Robra, Auf dem Weg zur integrierten Qualitätssicherung im Rettungsdienst, *Notfall Rettungsmed* **4** (2018), 261.

[2] OVGU - Herzlich Willkommen zu Inno_RD!, (2018). http://rettungsdienst-im-fokus.ovgu.de/ [Accessed October 20, 2018]

[3] Startseite. https://www.toolpool-gesundheitsforschung.de/ [Accessed October 20, 2018]

[4] M. Lablans, A. Borg, F. Ückert, A RESTful interface to pseudonymization services in modern web applications, *BMC Med Inform Decis Mak* **15** (2015), 2.

[5] M. Bialke, T. Bahls, C. Havemann, J. Piegsa, K. Weitmann, T. Wegner, W. Hoffmann MOSAIC--A Modular Approach to Data Management in Epidemiological Studies, *Methods Inf Med* **54** (2015), 364–371.

[6] A.K. Elmagarmid, P.G. Ipeirotis, V.S. Verykios, Duplicate Record Detection: A Survey. *IEEE Transactions on Knowledge and Data Engineering* **19** (2007), 1–16.

[7] S. Hatton, Early Prioritisation of Goals, *Proceedings of the 2007 conference on Advances in conceptual modeling foundations and applications,* (2007), 235–244.

[8] K. Pommerening, T. Müller, Leitfaden zum Datenschutz in medizinischen Forschungsprojekten: Generische Lösungen der TMF 2.0*,* MWV Med. Wiss. Verl.-Ges, Berlin (2014).

Address for correspondence

Hauke Fischer, Division of Medical Informatics, Carl von Ossietzky Unviersity Oldenburg, Germany – email: hauke.fischer1@uol.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190674

PRO (Patient Reported Outcomes) Implementation: From Vision to Reality

Alex Galper[a], Ora Shamai-Rosler[a], Dr. Varda Stanger[a], Dr. Eyal Zimlichman[a]

[a] Department of Medical Quality, Sheba Medical Center, Ramat Gan, Israel

Abstract

Patient-reported outcomes measure patients' views of their health-related quality of life. With the increasing interest and usage of PRO's in the world, Sheba Medical Center is the pioneer in the field in Israel looking at the potential value in PRO implementation in routine practice as a quality improvement tool together with the opportunity to help drive changes in healthcare delivery.

The PRO implementation process involves the multi-disciplinary ward or clinic staff in order to engage them and transform the process into a work routine. The collection methods and usage differ from one clinical area to another and require cultural adaptation in each area. Adjusted reporting and usage during patient encounter assist to improve patient enrolment and provide better and more accurate outcomes for the patient and the medical center.

Keywords:
Patient Reported Outcome Measures, Health-related Quality Of Life, Quality of Health Care

Introduction

Objective clinical outcomes are highly valued by healthcare professionals, but may not reflect the disease impact on the patient's life. Patient-reported outcomes (PROs) are of any health report or indicator given directly by the patient, without an external interpretation by a physician or researcher. This tool can be especially useful following medical procedures and interventions.

PROs were initially developed for research use, which has culminated in some regulatory bodies mandating their use. From there, PROs were adopted by a limited number of doctors to enhance the clinical management of individual patients. In recent years they have been used in order to assess and compare the outcomes achieved by healthcare providers.

Patient-reported outcomes (PROs) ask patients to evaluate the features of their health, including symptoms, quality of life, mental and functional status following healthcare treatment. The resulting data provides valuable information on how healthcare interventions affect critical aspects of an individual's life[1].

An ongoing PRO program launched at the Sheba Medical Center in 2015, based on the pioneering work at the Boston-based Partners HealthCare network. The pilot stage began with three clinical areas – prostate cancer, acute coronary disease, and cataract surgery.

The project aims to improve the quality of medical care and service by identifying and collecting outcomes that are significant for treating patients by reporting the results back to clinicians and the patients and by thus enhance patient engagement, adherence and shared decision making [2].

Methods

The PRO process begins with an initiative that usually comes from the ward head or clinician inside the ward or clinic. At the initial meeting, a clinical and patient profile is built in order to later adopt the required questionnaires, implementation method and work routine to the unique clinic culture and regime. Primary work with the clinical staff is required to choose the correct tools and questionnaires according to the clinician, patient preferences, language barriers, and relevant literature. In Sheba Medical Center each adult questionnaire contains the PROMIS Global Tool in addition to disease-specific reliable and validated questionnaires.

The information is gathered from patients during the waiting period in outpatient clinics when the patient is asked to fill out a questionnaire consisting of questions regarding his quality of life, functional and mental status and symptoms related to his medical condition. The set is completed on a tablet using a dedicated applet or with a manual questionnaire, which is later transferred to the PRO dedicated database. Then, on specific periods, the patients are tracked, using the same toolset, during the follow-up meeting, directly via the patient portal (Sheba Connect) or using a phone operator. Each tool set is processed by the dedicated PRO system and its results are presented in an easy to read PDF format that was developed using both clinicians and patients. While the patients can access the report via the patient portal, the report is embedded into the EMR and accessible to the clinicians.

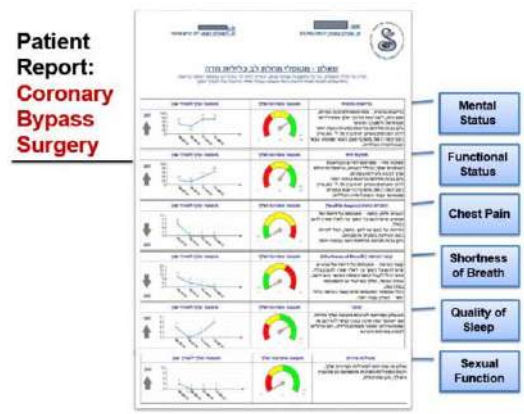


Figure 1 – patient report

In the initial implementation period, the primary focus is put on staff involvement and education. Clinical staff is educated about the collection process method and outcomes, especially report structure and benefits in routine work. The administrative staff is involved in adopting the process into their work routine.

Results

The program began in 2015 in three areas, developed into 20 active clinical areas and 40 additional areas in different stages of information gathering and implementation. From an active managerial approach focused on area recruiting it transformed to dealing with ward and clinic request for PRO program enrollment. The main volume is relying on the early areas, cataract (30%) and acute coronary disease (26%) together with oncology (14%) and orthopedics (21%) while other disciplines, including chronic illness, are being integrated into the program.

Program growth and collection methods improvement intend to double the number of questionnaires collected between 2017 and 2018 and almost triple in relation to 2016. Similarly, the number of participating patients raised in 63% between 2017 and 2018 and rose by 115% against 2018. The use of the patient portal had a significant impact and involved 11% of all complete tool sets and 30% of the follow-up questionnaires.

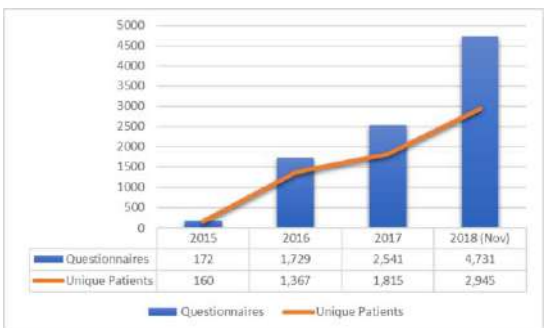


Figure 2 – PRO collection

Discussion

Current basic collection tools and methods that depend on man-power should be replaced with electronic collections methods such as IVR and automated messaging.

Constant project growth and deep clinical implementation, as part of medical quality improvement, require constant efforts and increasing resources as the foundation for a much elaborate system for Sheba Medical Center and the future of medicine using international collaboration.

Shared big data combined with socio-demographic, genomic and other EMR existing data can be used to create and measure outcome benchmarking as well as construct patient predictive tools for patients to make informative treatment decisions and to perform as clinical decision support tool for both patients and clinicians, at a hospital or home setting.

Conclusions

The PROs program expansion in the past three years shows success in patients, clinicians and managers recruitment to the program. Understatement of the process and its benefits of all involved parties increases the patient responsiveness and continues to generate interest among wards and clinics towards an accelerated expansion expected in 2019. However, many challenges still exist, such as educating clinicians to use the tools in routine clinical encounters, collecting and reporting processes as part of the routine clinical environment and implementing the use of tools by the online portal system.

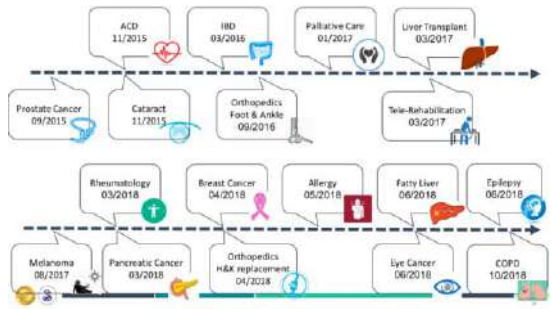


Figure 3 – PRO areas

Acknowledgements

The Gertner Institute for epidemiology and healthcare policy, Chaim Sheba Medical Center, Tel Hashomer.

References

[1] Weldring, T., Smith, S. M., Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health Services Insights* 6 (2013) 61–68. doi:10.4137/HSI.S11093

[2] Lavallee, D. C., Chenok, K. E., Love, R. M., Petersen, C., Holve, E., Segal, C. D., Franklin, P. D., Incorporating Patient-Reported Outcomes Into Health Care To Engage Patients And Enhance Care. *Health Affairs* 35 (2016): Patients' & Consumers' Use Of Evidence.

Address for correspondence

Alex.Galper@sheba.health.gov.il

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190675

Ethnic Difference in Skin Infection Rates: An Analysis of Electronic Health Records

Yulong Gu[a], John Kennelly[b]

[a] School of Health Sciences, Stockton University, Galloway, New Jersey, United States of America
[b] Department of General Practice and Primary Healthcare, University of Auckland, Auckland, New Zealand

Abstract

Diagnosis and prescription data in the electronic health records (EHR) of a New Zealand primary care clinic are analyzed to identify skin infection occurrences in support of a public health initiative. High prevalence of this disease and substantial differences in the prevalence and occurrence rates among ethnic groups are identified. The findings indicate application potential of the analysis algorithm in public health for identifying the population groups with high needs and assessing intervention impact.

Keywords:

Electronic Health Records, Infectious Skin Diseases, Ethnic Groups, Public Health

Introduction

Skin and Soft tissue infections (SSTI) are highly infectious diseases caused by *Staphylococcus aureus* and/or *Streptococcus pyogenes* bacteria. The SSTI rates are rising in New Zealand, with significant health disparity among ethnic groups [1]. SSTI are easily treated in individuals but the blood borne spread of the infection carries a serious threat to survival with the development of osteomyelitis, septic arthritis, pneumonia, and cerebral abscesses. This suggests that SSTI treatment should be considered from a public health perspective. As a part of the public health initiative (The Whakakotahi Initiative), The Happy Skin Project aims to improve SSTI management in the community, particularly among Pacific people. In support of this project, we analyzed baseline electronic health records (EHR) from an Auckland primary care clinic to identify SSTI occurrences.

Methods

The participating clinic serves predominantly Pacific people in West Auckland. Anonymized EHR data of recent 12 months (September 2017 – August 2018) were extracted from this clinic using Structured Query Language (SQL) queries on demographic, diagnosis, and prescription data. The patients of 24 years of age or younger who are registered at this clinic are included in this study. Three ethnic groups are considered in the analysis as Tuvaluans, other Pacific people (including Samoan, Tongan, Cook Island Maori, Niuean, Tokelauan, Fijian, and Pacific people not further defined), and non-Pacific people.

The SSTI identification algorithm was developed based on previous research that validated the use of EHR data in automatic detection of SSTI occurrences [2]. In this study, the diagnosis records of cellulitis, abscess, boil, impetigo, infected eczema, folliculitis, furuncle, and other skin infections were identified by examining the EHR-recorded READ codes (READ Version 2 [3], as implemented in the clinic's EHR systems). Prescription records of the antibiotic medications commonly used to treat SSTI such as fusidic acid were also examined to identify those SSTI occurrences in which the diagnosis was not coded in EHR. (The identification algorithm, including the list of READ codes and the drug names used in this study are available on request.) To verify the algorithm accuracy, in terms of the positive predictive value, the EHR records (diagnosis, prescription, notes, and laboratory testing results) of a small sample of algorithm-identified SSTI occurrences (n=135) were manually reviewed by a clinician.

The SSTI prevalence rate is calculated based on the algorithm-identified SSTI occurrences within the 12 months and is presented in percentage. Chi-square tests are performed to examine ethnicity differences in SSTI prevalence. The monthly and yearly occurrence rates per 1000 patients are calculated. An example of the monthly rate equation is:

$$\text{Monthly occurrence rate} = \frac{1000 * (\#\text{ of SSTI occurrences in the month})}{\text{Total \# registered patients}}$$

The monthly SSTI occurrence rates by ethnic group are also tracked over the 12 months. Microsoft Excel and SAS package are used in the data analysis (with significance at <0.05 level).

Results

All registered patients aged 24 or younger (N=4644) are included in this analysis. Among these patients, 81% are Pacific people (including 15% Tuvaluan and 66% other Pacific people), 19% are non-Pacific people.

High SSTI Prevalence and Significant Ethnic Differences

As shown in Table 1, the overall SSTI prevalence rate for the 12 months in the sample is 12%, with significant ethnic difference observed (X^2=39.6, p<0.01).

Table 2 shows the total number of SSTI occurrences over the 12-month period and the occurrence rate (per 1000 patients) in this period for each ethnic group. The substantial ethnic differences appear persistent over time, as shown in Figure 1.

Table 1 – SSTI Patient Counts by Ethnic Group

Ethnic group	SSTI patient number	Total patient number	SSTI prevalence
Tuvaluan	125	717	17%
Other Pacific people	349	3053	11%
Non-Pacific people	61	874	7%
Total	**535**	**4644**	**12%**

Table 2 – SSTI Occurrences over 12 Months by Ethnic Group

Ethnic group	SSTI occurrence number	Total patient number	Yearly occurrence rate per 1000 patients
Tuvaluan	201	717	280
Other Pacific people	501	3053	164
Non-Pacific people	79	874	90
Total	**781**	**4644**	**168**

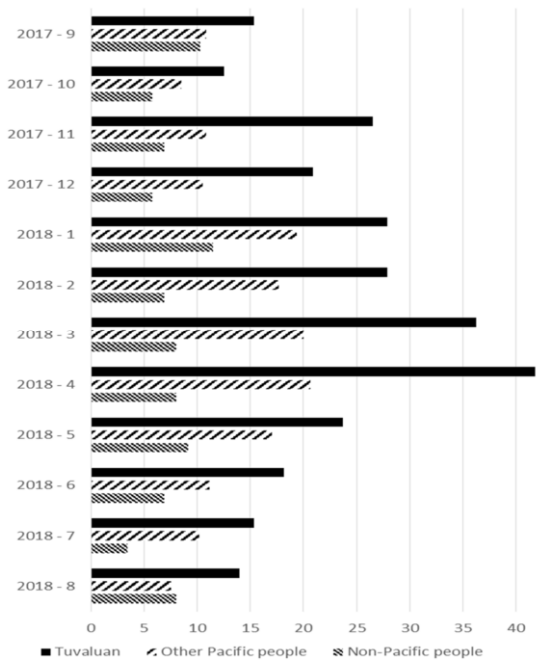


Figure 1 – Monthly SSTI Occurrence Rate per 1000 Patients by Ethnic Group (Sep. 2017 – Aug. 2018)

High Level of Algorithm Accuracy

High level of accuracy is indicated via the manual review of a small sample of the algorithm identified SSTI occurrences (positive predictive value = 100%). Within the ongoing Happy Skin Project, we plan to continue testing for algorithm accuracy and track the intervention impact on SSTI rates.

Discussion

Pacific people living in New Zealand are overrepresented in infectious skin diseases [1]. This study confirms this disparity and finds even higher SSTI rates among a subgroup of the Pacific people, Tuvaluan people. This identifies an urgent need for public health intervention in this population. This study also demonstrates meaningful use of EHR technology in public health initiatives that aims to improve health outcome and reduce health disparity. The seasonal variation of SSTI rates observed in the study (Figure 1) is consistent with previous research that finds higher SSTI rates are associated with higher temperature and humidity [4]. This seasonal variation will be considered when assessing the Happy Skin Project impact on the disease occurrence rate. The outcome analysis may also examine the temporal patterns of the disease and the causes for the variation.

A limitation of this study is that the convenience sampling of one clinic may not represent the clinics or the population in other settings. Another limitation is that we have only tested positive predictive value in the algorithm accuracy evaluation, but further testing with more cases and more measures is planned. Ongoing efforts are also made to support continuous refinement of the SSTI identification algorithm while we monitor the ongoing public health initiative.

Conclusions

High SSTI prevalence and substantial ethnic differences in the disease rates are identified using diagnosis and prescription data recorded in EHR. The current baseline analysis indicates feasibility of EHR data analysis in public health initiatives to identify the population groups with high healthcare needs and to assess the intervention impact on disease rates.

Acknowledgements

Funding for The Happy Skin Project is provided by the Health Quality & Safety Commission New Zealand (HQSC). We also thank the clinical and community team members, including Sally Dalhousie, Temasi Kitara, Edlyn Hetutu, Fala Haulagi, Minister Teatu Fusi, Uelese Malaga, Tauliani Monise, Marama Papau, Ana Aitcheson, and Hira Harema.

References

[1] Williamson DA, Ritchie SR, Lennon D, Roberts SA, Stewart J, Thomas MG, et al. Increasing incidence and sociodemographic variation in community-onset Staphylococcus aureus skin and soft tissue infections in New Zealand children. The Pediatric infectious disease journal. 2013;32(8):923-5.

[2] Gu Y, Kennelly J, Warren J, Nathani P, Boyce T. Automatic Detection of Skin and Subcutaneous Tissue Infections from Primary Care Electronic Medical Records. Stud Health Technol Inform. 2015;214:74-80.

[3] Health and Scocial Care Information Centre. Read Codes Leeds, England: Health and Social Care Information Centre; 2015 [cited 2015 Jan 30]. Available from: http://systems.hscic.gov.uk/data/uktc/readcodes

[4] Wang X, Towers S, Panchanathan S, Chowell G. A Population Based Study of Seasonality of Skin and Soft Tissue Infections: Implications for the Spread of CA-MRSA. PLoS ONE. 2013;8(4):e60872.

Address for correspondence

Yulong Gu, School of Health Sciences, Stockton University, 101 Vera King Farris Drive, Galloway, NJ 08205; Email: yulong.gu@stockton.edu.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190676

Genomic Common Data Model for Biomedical Data in Clinical Practice

Seo Jeong Shin[a], Seng Chan You[b], Jin Roh[c], Yu Rang Park [d], Rae Woong Park[a,b]

[a] Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, South Korea;
[b] Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea;
[c] Department of Pathology, Ajou University Hospital, Suwon, South Korea;
[d] Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, South Korea

Abstract

A common data model for clinical NGS panel data that is used in a distributed research network to achieve large scale to make evidence for improving patient care should be developed. This study developed OMOP-CDM extension for NGS panel data and confirmed the feasibility of the model by finding the differences between a database generated by research-purpose and clinical practice. We believe this data model can be used in distributed research model and will facilitate the usage of the clinical NGS data in patient care.

Keywords:

High-Throughput Nucleotide Sequencing; Data Analysis; Observational Study

Introduction

An impact on improving the medical care of cancer has increased the role of clinical next-generation sequencing (NGS) in precision medicine. The patient's NGS data are already being generated by large-scale projects that finding the role of clinical NGS in precision medicine [1]. However, the NGS panel data is relatively sensitive in privacy, which raises worries about exporting the data out of the center and collecting it in a silo. Also, the NGS panel data from the study cohort cannot be applied directly to the whole population in the point-of-care due to low diversity.

There has been a global trend to collect genomic data of patients in clinical practice through routine tests [2]. Moreover, since March 2017, the conditional insurance for NGS-based cancer panel has been provided in South Korea. The Distributed Research Network (DRN) uses the Common Data Model (CDM) and allows researchers to share the results of analysis rather than the data of multi-center without exporting sensitive data. The DRN is being adopted by global research consortium, including the Observational Health Data Sciences and Informatics (OHDSI). The Informatics for Integrating Biology and the Bedside (i2b2), a warehouse platform for clinical and genomic data, only covers health centers in the United States, which is resulting in the limitation.

To support diversity, such as ethnicity, and address privacy issues of genomic data, the Observational Medical Outcomes Partnership (OMOP) CDM used in OHDSI was chosen in this study. The OMOP-CDM is developed by OHDSI and their clinical data has been transformed into OMOP-CDM over twenty countries, 1.5 billion patients. However, OMOP-CDM shows low coverage of NGS panel data and is hard to reflect the stream of precision medicine. Although a few standard models have been introduced for genomic data, they had limitations on application.

The aim of this study is to design an OMOP-CDM extension model for NGS panel data that could be used in the DRN. Moreover, we purposed to confirm the feasibility of the data model through conducting multi-center analysis and describing the difference of NGS panel data harvested from the research project and clinical practice.

Methods

Building Genomic CDM

Existing data model and NGS report format, such as COSMIC and ISO20428, were reviewed to design OMOP-CDM extension to cover clinical NGS panel data. Standard nomenclature, such as HGVS (Human Genome Variation Society) and HGNC (HUGO Gene Nomenclature Committee), was adopted for a controlled vocabulary system.

Genomic Data Transformation

Clinical NGS panel data of Ajou University Hospital (AUSOM) were transformed by Ajou University to the genomic CDM. Yonsei University downloaded NGS data from TCGA database and transformed the data to the genomic CDM. R version 3.5.1 was used in the data processing. For storing and querying the sequencing data during data conversion, SQL script on Microsoft SQL Server 2017 was used as the relational database backend.

Data Description

In AUSOM, 92 lung adenocarcinoma (LUAD) and 22 lung squamous cell carcinoma (LUSC) patients had NGS cancer panel test. Public NGS data of the lung cancer cohort studies, Pan-Lung Cancer (TCGA, Nat Genet 2016) [3] were downloaded through the cbioportal of Memorial Sloan-Kettering Cancer Center (MSKCC). The TCGA database (TCGA) contained 603 LUAD patients and 457 LUSC patients.

Study Design

The data in G-CDM, such as gene name, variant type, and actionable variants, was compared between institutions for model validation. Comparison between two different databases is conducted using a Chi-squared test and results with $P<0.05$ were regarded as statistically significant. T-test was computed for the group differentiation. The waterfall plot was generated using an R package "GenVisR" that available via Bioconductor. R Shiny was used in implementing the data explorer as a pilot application of the genomic CDM.

Results

Data Structure of Genomic CDM

For linking clinical data of an OMOP-CDM, the information about a person, specimen, procedure, condition, and care site in NGS panel results was stored in a PERSON, SPECIMEN, PROCEDURE_OCCURRENCE, CONDITION_OCCURRENCE, and CARE_SITE table respectively. The data about NGS panel and pipeline were stored into four tables; GENOMIC_TEST, TARGET_GENE, VAIRNAT_OCCURRENCE, VARINT_ANNOTATION. Relationship between tables can be identified in Figure 1. Specification of a genomic CDM can be identified in the 'CDM Builder' category of the OHDSI forum site (http://forums.ohdsi.org/t/genomic-data-in-the-cdm).

Data Comparison for Model Validation

AUSOM and TCGA databases had differences in variant frequency. The 15 genes as a union of top 10 genes in each database were the targets for profiling. On AUSOM, top 10 genes had a frequency over 75% out of lung cancer patients, while only TP53 gene had a frequency over 25% in TCGA. Especially, EGFR variants had 89.47% frequency in AUSOM, while 6.96% frequency in TCGA. The mean proportion was different between databases (AUSOM: 69.59, TCGA: 12.86, P<0.00, 95%CI: 33.99-79.47). All 15 genes had different proportions between databases (P<0.05). The gene order was also different between databases, but the variant type of most frequent was a 'missense' in both databases (Figure 2).

An Application Based on Genomic CDM

Data exploring application 'GeneProfiler' was developed to generate the results of the descriptive analysis, such as variant types, pathogeny distribution, gene/variant proportion, and it can service patient proportion of input queries such as gene symbol and HGVSp (Figure 3). The 'GeneProfiler' code was uploaded and uploaded to the public in OHDSI Github (https://github.com/ABMI/GeneProfiler).

Conclusions

In this study, we propose a genomic CDM, as an expansion point for a broader discussion. During data comparison, differences in clinical NGS data between institutions were found. This study conducted distributed research using genomic CDM between institutions minimizing privacy issue. This is meaningful in that this study was done in an area of the genomic data that was not previously performed. We believe that the construction and adoption of the standard data model will facilitate the usage of the clinical NGS panel data.

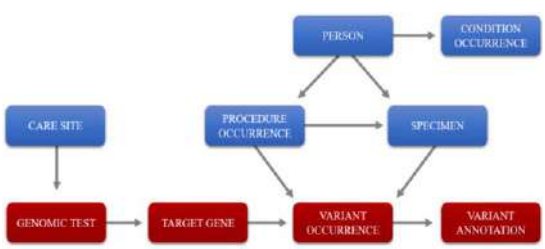


Figure 1– Schematic Diagram of the Relationship between Tables that Make Up the GCDM.

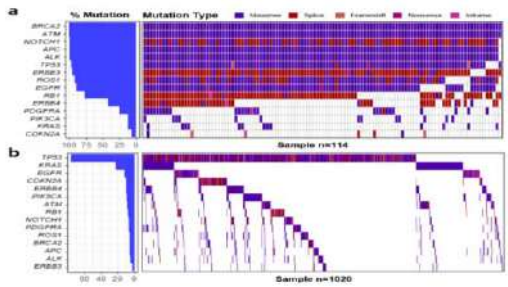


Figure 2– Waterfall Plot Describing the Variant Profile of the top 10 Genes in AUSOM (a) and TCGA (b) Databases.

Figure 3– Data Exploring tool 'GeneProfiler' for G-CDM.

Acknowledgements

This work was supported by the Bio Industrial Strategic Technology Development Program (20001234) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea) and a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea [grant number: HI16C0992].

References

[1] Consortium APG, AACR Project GENIE: Powering Precision Medicine through an International Consortium, *Cancer Discovery*, **7** (2017), 818-831.

[2] C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, and M. Landray, UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age, *PLoS Medicine* **12** (2015), e1001779.

[3] J.D. Campbell, A. Alexandrov, J. Kim, J. Wala, A.H. Berger, C.S. Pedamallu, S.A. Shukla, G. Guo, A.N. Brooks, and B.A. Murray, Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas, *Nature genetics* **48** (2016), 607..

Address for correspondence

Rae Woong Park, MD, PhD

Department of Biomedical Informatics, Ajou University School of Medicine, 164, World cup-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 16499, Republic of Korea

Tel : +82-31-219-4471, E-mail : veritas@ajou.ac.kr

Yu Rang Park, PhD

Department of Biomedical Systems Informatics, Yonsei University College of Medicine, 50-1, Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea

Tel : +82-2-2228-2493, E-mail : yurangpark@yuhs.ac

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190677

Linking Exome Sequencing Data with Drug Response Aberrations

Konstantinos Kyriakidis[a], Alexandra Charalampidou[b], Pantelis Natsiavas[c,d], Ioannis S. Vizirianakis[e], Andigoni Malousi[a]

[a] Lab. of Biological Chemistry, School of Medicine, Aristotle University of Thessaloniki, Greece
[b] Scientific Computing Office, Information Technology Centre, Aristotle University of Thessaloniki, Greece
[c] Sorbonne Université, INSERM, Univ Paris 13, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances pour la e-Santé, LIMICS, F-75006 Paris, France
[d] Institute of Applied Biosciences, Centre for Research & Technology Hellas, Thermi, Thessaloniki, Greece
[e] Lab. of Pharmacology, School of Pharmacy, Aristotle University of Thessaloniki, GREECE

Abstract

Next-generation sequencing has prompted the development of numerous -omics applications. Along with experimental procedures, various computational pipelines became available to address the inherent complexities concerning the volume and quality of data. These pipelines are effective and routinely applied; however, interpreting their outcomes into actionable evidence is still poorly addressed. In this context, this work proposes a method for translating patient genomic profiles to drug response aberrations by integrating pharmacogenomic data into sequencing data analysis pipelines.

Keywords:

Polymorphism, Pharmacogenomic Variants, High-Throughput Nucleotide Sequencing

Introduction

During the last decade, next-generation sequencing (NGS) technologies coupled with bioinformatics advancements have proved their potential to address previously unmet needs, such as improved personalized diagnosis, disease management and stratified treatment. Exome sequencing of patient samples is among the most popular and cost-effective NGS applications that enables the detection of single-nucleotide variants and small insertions/deletions with potential clinical significance. Transforming sequencing data, called reads, to genomic variants is a complex computational problem requiring the execution of several tools that are usually combined into workflows [1-3]. Most of these workflows end up with a set of high-quality variant calls that is annotated with respect to their functional consequence, while the therapeutic importance of the detected variants is rarely evaluated [4].

To cope with this issue, the present work proposes a method that besides the main processing steps it elaborates further on the interpretation of the called variants by identifying those with pharmacogenomic significance and by evaluating the type of interactions and level of evidence that are associated with known drug response aberrations.

Methods

The proposed method comprises of three modules (Figure 1). Module 1 contains a set of chained processes that transforms raw exome sequencing reads to high-quality mappings against reference human genome. Specifically, the reads are first checked for their quality and trimmed to remove low-quality base calls, adapter sequences, etc. Then, BWA [5] is used to align high-quality reads against the indexed human reference genome, followed by Picard [6] and GATK [7], which remove PCR duplicates and detect systematic errors in base quality scores, respectively. Module 1 exports quality reports of the reads and alignments, as well as the alignment files that will be subsequently used to identify variants.

Module 2 performs variant calling for each sample and joint genotyping on multi-sample variant files, using the GATK workflow [7]. Variants are then processed, including quality control, filtering and functional effect analysis. The latter functionality is performed by GEMINI [8], which also annotates variants, with respect to their functional pathways and protein-protein interactions. Finally, Module 3 automates the identification of pharmacogenomically important variants in a two-step procedure that includes: a) parsing of local distributions of curated clinical annotations, and b) identification of interacting drug labels, along with the type of interaction (toxicity, efficacy, metabolism dosage), the level of evidence, the associated phenotype etc., that are available through PharmGKB [9].

Results

The methodology is implemented in an automated workflow that combines freely accessible tools and public data repositories for variants annotation and variant-drug interactions. Currently, the workflow is able to analyse both whole-exome and targeted gene sequencing data for the detection of germline variations. The workflow is built to run on high-performance computing infrastructures, enabling parallelization, that is particularly important for whole-exome data analyses.

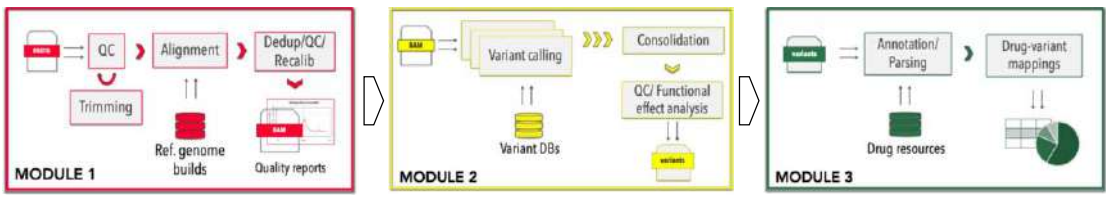


Figure 1 – Modules of the proposed methodology to compute drug-variant interactions from sequencing reads. The modules are chained in a pipeline that exports drug response aberrations together with intermediate data and quality control (QC) reports. FASTQ is the file format of the sequencing reads and BAM is the file format of the alignment files.

Besides the variants and alignment files generated during the execution of Modules 1 and 2, the workflow exports user-friendly reports of the drug-rankings and various visualizations that are built on the variant-drug associations. An example visualization is shown in Figure 2, where variant-drug interactions are depicted with respect to their type of interaction (axis y) and level of evidence (size of concentric circles).

Conclusions

During the last few years pharmacogenomic research has revealed strong associations between genomic profiles of patients and their response to drug treatment. The incorporation of these associations to downstream analysis of high-throughput data sequencing pipelines can be particularly important for ranking the pharmacogenomically actionable variants and for enabling stratified therapeutic decisions. In this context, the proposed method is designed and implemented in order to facilitate targeted treatment decisions and to provide a means to identify biochemical mechanisms underlying variant-drug interactions.

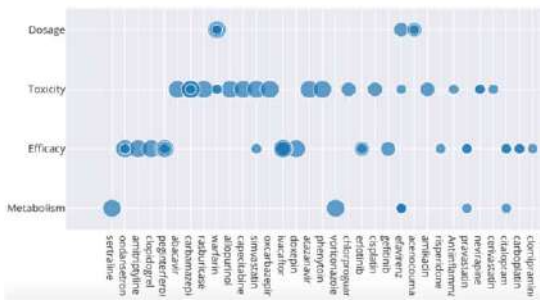


Figure 2 – Example drug-ranking visualization with respect to the type of variant-drug interaction. Concentric circles denote multiple evidence levels. Bigger circles correspond to higher evidence levels (Level 1A/1B), according to PharmGKB [9].

References

[1] J.L. Causey, C. Ashby, K. Walker, Z.P. Wang, M. Yang, Y. Guan, J.H. Moore, and X. Huang, DNAp: A Pipeline for DNA-seq Data Analysis, *Sci. Rep.* **8** (2018).

[2] Y. Guo, X. Ding, Y. Shen, G.J. Lyon, and K. Wang, SeqMule: Automated pipeline for analysis of human exome/genome sequencing data, *Sci. Rep.* **5** (2015).

[3] Z.K. Liu, Y.K. Shang, Z.N. Chen, and H. Bian, A three-caller pipeline for variant analysis of cancer whole-exome sequencing data, *Mol. Med. Rep.* **15** (2017) 2489–2494.

[4] J. Hintzsche, J. Kim, V. Yadav, C. Amato, S.E. Robinson, E. Seelenfreund, Y. Shellman, J. Wisell, A. Applegate, M. McCarter, N. Box, J. Tentler, S. De, W.A. Robinson, and A.C. Tan, IMPACT: A whole-exome sequencing analysis pipeline for integrating molecular profiles with actionable therapeutics in clinical samples, *J. Am. Med. Informatics Assoc.* **23** (2016) 721–730.

[5] H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, *{arXiv} Prepr. ArXiv1303.3997* (2013).

[6] Broad Institute, Picard Tools (https://broadinstitute.github.io/picard/).

[7] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and others, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, *Genome Res.* **20** (2010) 1297–1303.

[8] U. Paila, B.A. Chapman, R. Kirchner, and A.R. Quinlan, GEMINI: Integrative Exploration of Genetic Variation and Genome Annotations, *PLoS Comput. Biol.* **9** (2013).

[9] M. Whirl-Carrillo, E.M. McDonagh, J.M. Hebert, L. Gong, K. Sangkuhl, C.F. Thorn, R.B. Altman, and T.E. Klein, Pharmacogenomics knowledge for personalized medicine, *Clin. Pharmacol. Ther.* **92** (2012) 414–417.

Address for correspondence

Dr. Andigoni Malousi
Lab. of Biological Chemistry, Medical School
Aristotle University of Thessaloniki, Greece
e-mail: andigoni@auth.gr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190678

The Preliminary Outcome of Applying a Patient Transportation Management System for Non-Emergency Intra-Hospital Transportation of Patients

Ying-Li Lee[a,b], Tsai-Feng Chien[a], Yu-Wen, Guo[c], Mei-Hui Lin[a]

[a] *Department of Nursing, Chi Mei Medical Center, Tainan, Taiwan (ROC)*
[b] *Department of Nursing, Chung Hwa University of Medical Technology, Tainan, Taiwan (ROC)*
[c] *Department of Information, Chi Mei Medical Center, Tainan, Taiwan (ROC)*

Abstract

In Taiwan, the safety of intra-hospital patient transportation (IHT) is an important issue of patient safety. However, the effects on the quality of patient transportation and the results of patient safety in applying informatics and communication technology were less discussed. The purpose of this study is aimed to understand the current status of IHT events through the patient transportation management system as a reference for further improving the IHT quality.

Keywords:

Transportation of Patients,, Information Technology

Introduction

In hospitals, patients are frequently transported to other professional units in the hospital for further diagnosis or treatment, called intra-hospital transportation (IHT). Over the years, risks and adverse events of IHT have been extensively studied and identified as risky processes, especially for critically ill patients. However, IHT of non-emergency patients is the most frequent event in the hospital, but it is rarely discussed [1].

The goal of IHT is efficiently stated as the 'right patients' to get to the 'right place' at the 'right time'. Yang et al observed 102.5 patient safety events per 100,000 IHTs in Taiwan. 71.8% of the patient safety events were no harm events, and 71.4% were related to the process of transportation. The most common safety events were: delayed departure (13.1%), error in the process (12.1%) and prolonged waiting after arrival (9.2%). 58 (28.2%) events were associated with patient harm, including mild harm in 35 (17.0%) and moderate harm in 23 (11.2%) due to typical acute change in physiological status [2]. Furthermore, poor efficiency of IHT can lead to increased costs for the hospital, longer hospital stays and patient anxiety [1].

The pre-transport communication, the appropriateness of the transport personnel, the timing of the transportation, and the safety and efficiency of the transport processes are key factors that affect the quality of transportation and safety of the non-emergency patient transportation. Non-emergency patient transport processes should be standardized to reduce risks of IHTs and increase the efficiency of patient transportation. The application of information and communication technologies (ICT) may help to improve the quality of transport services for patients [1]. However, related research is rare.

To improve transportation quality of the non-emergency patient, the study hospital has developed a patient transportation management system which was developed by applying ICT and combined with nurses' and patient transport team's participation to transport non-emergency patients. The purpose of this study is to understand the current status of IHT events through the patient transportation management system as a reference for improving the IHT quality.

Methods

The study was conducted at a medical center in southern Taiwan. The hospital has 1,278 beds. Deducting emergency department, intensive care units, and special purpose beds, there are 843 beds in the general ward. There are 5 medical-related buildings in the hospital. All patients have to be graded for transportation safety, according to their condition, by their doctors after hospitalization. According to the hospital's policy, only non-emergency patients can be transported by the patient transport team members in the hospital. The hospital's transport team consists of 2 senior nurses (team leaders) and 43 members.

One or two senior transport team members were in charge of the assignments during each shift before the system implementation. They had to print the transportation schedule from the computer every 4 hours, receive the instant telephone notifications, write down the patient information and assign the task to team members with radio pagers. It was difficult to measure the time interval between each step. Therefore, we fail to obtain data prior to system implementation.

According to the transportation process of the hospital, a patient transportation management system was developed by applying mobile phone and computer network communication technology. The system can receive scheduling appointments for all non-emergency patient transportation every day from the appointment scheduling systems of other departments in the hospital, as well as immediate patient transport requests. Each transport team member must carry a smartphone with him or her during work to receive the assigned task instantly through the mobile application. The detail patient and transportation task information will be presented in the app.

All transport team members followed standard operating specifications. The working status of transport personnel was divided into available, on duty, resting, and at meals. The action flow of each transport event included assigned, took over, transporting, and completed.

We recorded the time of each step performed by each member in each transport event. Variables included time of appointment, time of assignment, transport personnel, the location of take over, time of take over, destination, time of arrival, transport tool, main transport purpose, etc. After 2 months of system implementation, all data was remitted and descriptive statistical analysis was performed through Tableau Desktop public edition 2018.3.0.

Results

Two months after system implementation, the transport team members were familiar with the system operation. We conducted a retrospective analysis on 24,457 patient transport requests from August to September 2018. We excluded 922 (3.77%) requests canceled before transport assignment and 546 (2.23%) uncompleted requests due to patient not ready for transport. A total of 22,913 (93.69%) patients were taken over and completed transportation smoothly, including 76 (0.31%) transportations without recorded completion time. The most frequent patient transportation happened between the general ward and radiology (8,138), followed by general ward and operating room (4,242), and general ward and exam room (4,002) (Figure 1).

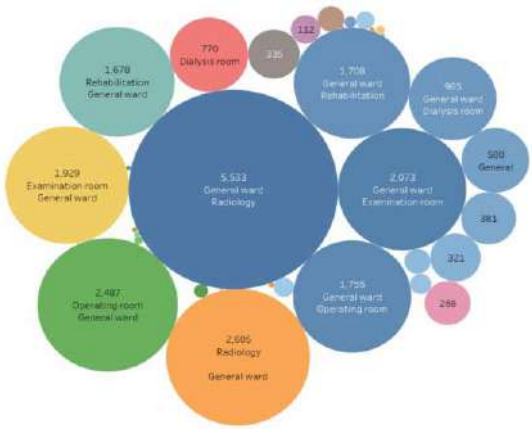


Figure 1- The frequency analysis between the transport units.

The frequency of transportation purpose was the highest for various examinations (12,456, 54.18%), followed by various treatments (6,037, 26.26%), surgery (4,101, 17.84%) and consultation (395, 1.72%). The main purpose and classification of transportations were analyzed and shown in Figure 2.

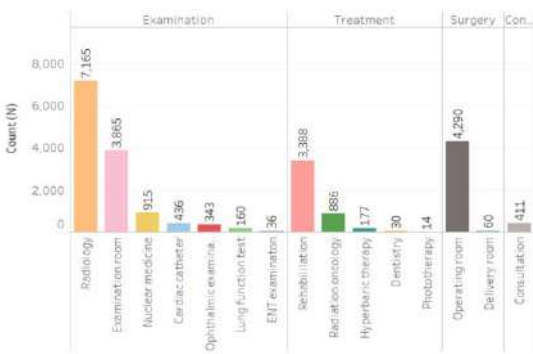


Figure 2- The histogram of the main purpose and classification of transportations.

The duration between the stages of each IHT event was analyzed. After the patient transportation management system implementation, most patients were taken over within 10 minutes of assignment (76.82%). After take over, 62.41% of the patients arrived at the destination within 10 minutes. The rate of transportation completion within 20 minutes was 93.71% (Table 1).

Table 1- The analysis of the execution duration between each stage of each IHT event.

Time interval (minutes)	From assignment to take over patients (n, %)	From taking over patients to mission completed (n, %)
No assign	245 (1.07)	76 (0.33)
> 30	888 (3.86)	385 (1.67)
20~30	714 (3.11)	986 (4.29)
10~20	3,481 (15.14)	7,195 (31.3)
3~10	16,238 (70.63)	10,809 (47.02)
< 3	1,423 (6.19)	3,538 (15.39)

Discussion

Communication, efficiency, and appropriateness are regarded as important issues affecting the quality and safety of non-emergency hospital transport [1]. The implementation of information systems has been widely recognized in improving communication and administrative efficiency of hospitals. One limitation of this study is the inability to obtain data prior to system implementation, but the results of this study can be used as a status analysis before IHT quality improvement. We found that most IHTs were concentrated in transportations between the general wards and the radiology department, operating room or the rehabilitation department. In addition, 6.29% IHTs were not completed within 20 minutes. The results of this study will help us further explore the appropriateness and efficiency of the current examination and treatment process and may prove helpful for designing automatic patient transportation management systems.

Conclusions

In Taiwan, the communication and safety of handing over patients and the risk management of patient transportation are important issues in patient safety. Safe and efficient patient transportation is a common expectation for patients, families, and health providers. Through the implementation of a patient transportation management system, we have gathered baseline data on the timeliness of non-emergency patient transportation. In the future, the impact of this system on the quality and safety of non-emergency patients will be further evaluated.

References

[1] I.M. Hains, A. Marks, A. Georgiou, and J.I. Westbrook, Non-emergency patient transport: what are the quality and safety issues? A systematic review, *Int J Qual Health Care* **23** (2011), 68-75.

[2] S.H. Yang, J.S. Jerng, L.C. Chen, Y.T. Li, H.F. Huang, C.L. Wu, J.Y. Chan, S.F. Huang, H.W. Liang, and J.S. Sun, Incidence of patient safety events and process-related human failures during intra-hospital transportation of patients: retrospective exploration from the institutional incident reporting system, *BMJ Open* **7** (2017), e017932.

Address for correspondence

Ying-Li Lee, RN, PhD
Chi Mei Medical Center; No.901, Chung-Hwa Rd., Yung-Kang Dist., Tainan City 701, Taiwan (ROC)
Email: n930969@mail.chimei.org.tw
Phone: 886-6-281-2811 Ext.53008

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190679

Requirement Analysis for Developing a Patient Participation Program in Patient Safety

Nam-Ju Lee[a], Shinae Ahn[a], Miseon Lee[a]

[a] *College of Nursing, Seoul National University, Seoul, Republic of Korea*

Abstract

Patients' roles in preventing errors has been emphasized. Patients' and families' participation is one of the important strategies to improve patient safety. Therefore, it is necessary to educate patients and families who visit hospitals. Web-based educational programs can be useful tools to provide patient safety information easily and enhance patients' and families' knowledge. This study analyzed requirements for developing a web-based program for patient participation in patient safety.

Keywords:
patient safety, patient participation, education

Introduction

Patient participation has been recognized as an important factor to improve quality of care, which is associated with positive health outcomes [1]. Ensuring that patients have the information needed to act as advocates for themselves may help to decrease patient safety incidents. The concept of 'patient participation' and 'patient-centeredness' implies that the patient can choose realistic goals independently and engage voluntarily in health care during the care process [2].

Patients and their families who have more knowledge about safety issues and information are more likely to participate in patient safety practices in their health care [3]. To promote patient participation in health care, appropriate interventions with educational materials and health information have been considered a crucial component of the methodology [4]. Therefore, healthcare professionals need to develop techniques and educational tools that increase patients' knowledge in preventing errors. However, there are not sufficient educational programs focusing on patient participation in patient safety activities in South Korea.

The purpose of this study was to assess patients' and families' educational needs for improving participation in safety practices and to identify the most important areas to be addressed in a web-based program for patient participation.

Methods

To construct the educational contents of a web-based program for patient participation, we conducted a survey, focus group interview, and literature review.

Previously, we conducted a online survey of health consumer's willingness to participate, their recognition of the importance of their participation, and their experience of engaging in patient safety practices. Our population was adults aged 19 years or older who had visited hospitals in South Korea within the previous one year [5].

Two-hour focus group interviews were held with two groups of six participants who had completed the survey and volunteered for the focus group interview. Data were collected during March 2018. The key interview question was as follows: "How was your experience with visiting medical institutions as a patient concerning patient safety?" All interviews were recorded and transcribed. After each interview, the research team listened to the recorded data, reviewed the transcribed data, and then coded the data. The qualitative data were analyzed using content analysis.

In addition, we conducted a literature review related to patient participation, patient safety frameworks, and evidence-based strategies for patient involvement in healthcare.

Results

A total of 493 respondents completed the online survey and 65% of the particpants reported that they had experience of patient safety incidents. The survey results demonstrated that their experience of participating in patient safety was relatively low compared to their perceived importance to patient safety and their willingness to participate in patient safety activities. They were reluctant to engage in confrontational situations, such as asking healthcare professionals to wash their hands [5].

Eight interview participants were female and four were male. Respondents' educational needs in patient safety practices varied considerably. Based on the focus group interviews, we analyzed interview data and categorized what patients wanted to know as 'patient rights', 'disease and diagnosis', 'treatment', 'surgery and medical testing', 'medication', 'patient advocacy', 'error disclosure', 'patient question checklists', and 'manuals for empowered patients'.

Through a review of the literature based on the patient safety frameworks and patient participation, we identified four competencies needed to improve patient participation in healthcare; 'speak up', 'ask questions', 'find health information', and 'engage with healthcare'.

According to the survey results, educational needs assessment, and the literature review of previous research on patient participation in healthcare, we identified thirteen priority areas to construct educational contents based on the most effective safety practices in reducing patient safety incidents: 'introduction to patient participation', 'patient rights', 'error prevention', 'patient advocacy', 'medical testing', 'surgical care safety', 'medication safety', 'infection prevention and control', 'fall prevention', 'patient-provider communication', 'patient identification', 'preparation for admission', and 'transitions of care'.

After initial development, the educational contents on patient participation were reviewed by four experts (one doctor, two

nurses, and one pharmacist) for content validity test. On the basis of their comments, some contents were too complex or difficult to understand were modified.

Conclusions

Health consumers are not active in particiting in patient safety activities in South Korea. They need to be informed of more information and resources for patient safety to enhance their engagement in safety practice. We are implementing a web-based program for patient participation in patient safety practice through content development based on the results of this requirement assessment.

Acknowledgements

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1D1A1B03034406).

References

[1] Y. Longtin, H. Sax, L.L. Leape, S.E. Sheridan, L. Donaldson, and D. Pittet, Patient participation: current knowledge and applicability to patient safety, *Mayo Clin Proc* **85** (2010), 53-62.

[2] J. An, S.J. Kim, S. Park, K.T. Moon, and E-C. Park, The effects of patient education on patient safety: can we change patient perceptions and attitudes? Lessons from the Armed Forces Capital Hospital in Korea, *Int J Qual Health Care* **29** (2017), 392-398.

[3] S. Cho, E. Lee, Effects of the smartphone application "Safe Patients" on knowledge of patient safety issues among surgical patients. *CIN: Computers, Informatics, Nursing* **35** (2017), 639-646.

[4] S.N. Weingart, M. Toth, J. Eneman, M.D. Aronson, D.Z. Sands, A.N. Ship, R.B. Davis, and R.S. Phillips, Lessons from a patient partnership intervention to prevent adverse drug events, *Int J Qual Health Care* **16** (2004), 499–507.

[5] N-J. Lee, S. Ahn, M. Lee, and H. Jang, Health consumers' self-reported willingness, recognition of importance, and experience of participation in patient safety practices, ISQUA 2018.

Address for correspondence

Nam-Ju, Lee,
College of Nursing, Seoul National University
103 Daehak-ro, Jongno-gu,
Seoul, 03080, Republic of Korea
Email: njlee@snu.ac.kr
Phone: 82-2-740-8812

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190680

HLA Allele Distribution Associated with Adverse Drug Reactions in Organ Transplant Patients

KyeHwa Lee[a], Yi-Jun Kim[a], Hyo-Jung Kim[b], and Ju Han Kim[a,b]

[a] Center for Precision Medicine, Seoul National University Hospital, Seoul 03082, South Korea,
[b]Seoul National University Biomedical Informatics (SNUBI) and Systems Biomedical Informatics Research Center, Division of Biomedical Informatics, Seoul National University College of Medicine, Seoul 110799, South Korea

Abstract

The results of Human Leukocyte Antigen (HLA) antigen testing in transplant patients are not generally used to predict future adverse events. In this study, free-text HLA screening results of transplant patients were analyzed and stored in a database, and the frequencies of patients with adverse events according to HLA allele were extracted. Approximately 25% of patients had HLA alleles associated with serious drug side effects.

Keywords:

HLA Antigens, Pharmacogenomic Testing, Drug-Related Side Effects and Adverse Reactions

Introduction

Human Leukocyte Antigen (HLA) is a protein that plays an essential role in the immune function of our body with a wide variety of allele types [1]. HLA diversity is particularly important in organ transplantation because transplant recipients and donors with different serological HLA proteins will exhibit organ transplant rejection [2]. Therefore, transplant recipients need HLA screening before transplantation. Recently, HLA diversity has been reported to cause severe drug hypersensitivity as well as organ transplantation rejection [3]. According to the PharmGKB database, there are about 40 drugs known to cause HLA-related adverse reactions, commonly prescribed drugs such as statins and NSAIDs were involved as well [4]. However, the HLA results of transplant patients and donors have not been used to predict future adverse drug reactions. This is because the HLA test is performed in different ways, ranging from a simple serological test to next-generation sequencing (NGS) test. Also, because the nomenclature for the representation of the HLA results is continuously updated, the test results simply stored as free text in the electronic medical record (EMR) [5]. In this study, we designed a data model, GDM-STAR, which expresses HLA test results according to HLA nomenclature and stores the test results in a database. We then used this database to derive predictable drug adverse events using HLA testing.

Methods

We extracted HLA test results performed between February 2002 and June 2018 including basic clinical data of the patients using SUPREME®, a clinical data warehouse at the Seoul National University Hospital. The research design and analysis methods were approved by the Institutional Review Board of Seoul National University Hospital (Seoul, South Korea). We parsed free text in clinical notes to extract HLA test results using the Python 3 regular expression function and loaded them to a database according to the allele type of HLA-A, B, C, DR, and DQ genes. Incomplete test results were filled in by a manual review by a research nurse (JY Lee). We then matched the HLA allele type loaded in the database with the HLA genotype in the PharmGKB clinical variants table. Finally, we derived a patient-drug table associated with a specific allele. All descriptive statistics and graphs were generated using R software [6].

Results

As a result of HLA parsing, a total of 16 HLA-A, 43 HLA-B, 15 HLA-DR, 19 HLA-DR, and 12 HLA-DQ genes were identified respectively. Of the 11,289 total transplant recipients, the HLA allele type of the patient matched the PharmGKB drug adverse event related genotype in 2,813 patients, accounting for 25% of the total patients. A total of 34 PharmGKB drugs or classes were matched to these patients. A total of 258 (9.4%) patients had drug-related HLA alleles in all five gene families (Figure 1). The most common drug associated with the patient was carbamazepine, and a total of 1,975 patients were found to have a side effect-associated gene for this drug. Table 1 shows drugs and frequencies of patients related to that drug with a patient frequency of 1% or more.

Conclusions

In this study, we designed a parser to extract HLA test results contained in free text in EMR and transform them into a tabular form according to the the five HLA gene families. Assuming that the organ transplant recipients are representative of the population, the database of HLA results shows that we could prevent adverse drug reaction in about 25% of the patients by building an HLA database.

It is essential to integrate such a clinical decision support (CDS) system within the EMR to achieve real-time precision medicine. Developing and testing of the CDS system using the HLA database is subject to further study. It is also important to increase physicians' understanding of pharmacogenomics to efficiently utilize the pharmacogenetics information derived from existing NGS tests.

Table 1- Distribution of patients by HLA allele- matching drugs

No	Drugs	#Samples	%
1	Carbamazepine	1975	17.5
2	Nevirapine	1907	16.9
3	Allopurinol	1330	11.8
4	Oxcarbazepine	1119	9.9
5	Lamotrigine	827	7.3
6	Peginterferon alfa-2b,ribavirin	696	6.2
7	Clindamycin	510	4.5
8	Phenobarbital	498	4.4
9	Interferon beta-1a	428	3.8
10	Ticlopidine	400	3.5
11	Statins	308	2.7
12	Antithyroid Preparations	307	2.7
13	Antiepileptics	295	2.6
14	Phenytoin	257	2.3
15	Azathioprine,mercaptopurine	199	1.8
16	Lapatinib	199	1.8
17	Influenza vaccines	173	1.5
18	Sulfasalazine	152	1.3
19	Methazolamide	120	1.1
20	Acetazolamide	120	1.1

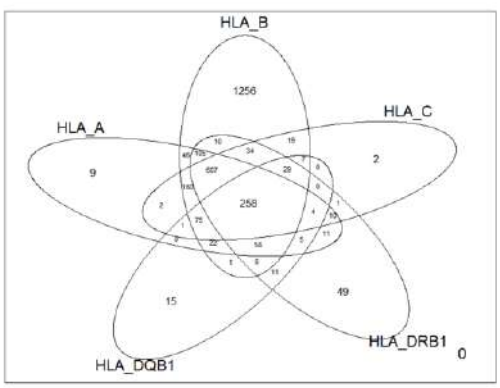


Figure 1– Distribution of patients with HLA allele belonging to a specific HLA gene family

Acknowledgements

This research was supported by a grant (16183MFDS541) from Ministry of Food and Drug Safety in 2018.

References

[1] Clinical Role of Human Leukocyte Antigen in Health and Disease - Mosaad - 2015 - Scandinavian Journal of Immunology - Wiley Online Library. https://onlinelibrary.wiley.com/doi/full/10.1111/sji.12329. Accessed 25 Nov 2018.

[2] Mahdi BM. A glow of HLA typing in organ transplantation. Clin Transl Med. 2013;2:6.

[3] Fan W-L, Shiao M-S, Hui RC-Y, Su S-C, Wang C-W, Chang Y-C, et al. HLA Association with Drug-Induced Adverse Reactions. Journal of Immunology Research. 2017. doi:10.1155/2017/3186328.

[4] Thorn CF, Klein TE, Altman RB. PharmGKB: The Pharmacogenomics Knowledge Base. Methods Mol Biol. 2013;1015:311–20.

[5] Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in Medicine. 2015;17:405–23.

[6] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2018. https://www.R-project.org/.

Address for correspondence

Ju Han Kim

juhan@snu.ac.kr (preferred method of contact)

82-2-740-8320

Seoul National University Biomedical Informatics (SNUBI) and Systems Biomedical Informatics Research Center, Division of Biomedical Informatics, Seoul National University College of Medicine, Seoul 110799, South Korea

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190681

Extending CQL with openEHR to Express Clinical Quality Indicators

Mengyang Li[a,b], Yunlong Zhi[a,b], Xudong Lu[a,b,c], Hailing Cai[a,b]

[a]*College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, Zhejiang, China*
[b]*Key Laboratory for Biomedical Engineering, Ministry of Education, China*
[c]*School of Industrial Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands*

Abstract

Clinical Quality Language (CQL), a HL7 authoring language to express clinical quality indicators, provides the capability to express logic that is human readable yet structured enough for processing a query electronically. OpenEHR is a widely used, modeling methodology, but currently CQL cannot support openEHR archetypes which hinders its usage in openEHR environment. This paper presents a method to express and compute clinical quality indicators by extending CQL with openEHR archetypes. To verify the feasibility of this method, 64 indicators from the Centers for Medicare & Medicaid Services (CMS) and 118 indicators from local environment in China were utilized. The results show that those indicators can be well represented and computed in openEHR environment.

Keywords:

Language, Quality Improvement, Quality of Healthcare Care

Introduction

It is a challenge to assess the validity of healthcare services to improve the patients' health. Clinical quality measures are important tools that help achieve the goal. Typically, clinical quality information can be represented in the form of indicators [1]. It can make information about clinical quality comparable across different healthcare service providers.

There exist some problems with the development of indicators. First, they are conducted by different providers for specific purposes, like government and large medical institutions. The absence of a unified and well-defined specification makes it difficult to share indicators among different organizations and cover the need for regional medical organizations. Second, the fuzzy expression of the natural language may lead to inaccurate computation.

Clinical Quality Language (CQL) is a HL7 authoring, language standard and provides the ability to express logic that is human readable yet structured enough for processing a query electronically. It acts as a mature specification and is widely used in the domains of quality measurement and decision support. As a well-known EHR modeling specification, openEHR is designed to ensure universal interoperability among all forms of electronic data. However, currently CQL cannot support openEHR archetypes which hinders its usage in openEHR environment.

The objective of this research is to explore an approach to express and compute clinical quality indicators through extending CQL with openEHR to facilitate the share of clinical quality information in openEHR environment.

Methods

In order to apply CQL directly to archetype-based context, we have proposed three extensions including archetype binding, terminology binding, and metadata.

Archetype Binding

The openEHR specification is used as the information model to express the medical concepts in clinical quality indicators. The archetype binding mechanism in Guideline Definition Language (GDL) [2] has been taken as a reference. It helps design indicators in a unified form. The mechanism established a relationship between medical concepts and data elements in archetypes which can lay the foundation for understanding meaning of indicators and binding terminologies.

Terminology Binding

Terminology can be used to disambiguate the expression of indicators. For the sake of including both public and private terminology sets, terminology binding mechanism was brought in to the extension. Two mechanisms used in Archetype Definition Language (ADL) [3] has been taken as a reference. One is to use the external terminologies directly and the other is defining a set of dictionaries about terms within the archetypes.

Metadata

Metadata was designed to help manage indicators easily and share indicators among different organizations. A series of metadata data elements were defined, including the name, version, purpose and so on.

Results

64 clinical quality indicators defined and revised by CMS in 2016 and 118 indicators from National Health Commission of the People's Republic of China were expressed successfully. A complicated indicator about controlling high blood pressure [4] was chosen to describe the detailed process of expression.

The complete expression is in Figure 1. The overall operational process is shown in Figure 2.

Figure 1. Complete Expression for Controlling High Blood Pressure

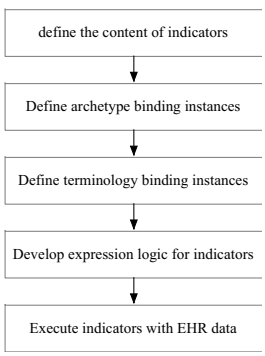


Figure 2. The Overall Operational Process

Discussion

By introducing archetype binding, terminology binding and metadata to formalism of quality indicators, it is helpful to those developers in openEHR community who works on assessment of indicators. Compared to CLIF[5], the method can take advantages of archetypes directly rather than define information models, related terminology and operations. On account of extending from CQL, the method can also support clinical decision support and clinical quality measurement relative to LERM[6]. The limitation of the method is that a relatively small number of indicators were expressed for validation.

Conclusions

We have explored a method to extend CQL with openEHR specification. Metadata, terminology binding and archetype binding have been refined and redesigned to satisfy the need of formalism of clinical quality indicators with the help of unified information models. These extensions can eliminate the inaccuracies of indicators and make them comparable among different organizations.

Acknowledgements

This work was supported by grant National Key R&D Program of China [grant number 2016YFC1300300].

References

[1] Gibberd, R., et al. Using indicators to quantify the potential to improve the quality of health care. *International Journal for Quality in Health Care* **16 suppl_1** (2004): i37-i43.

[2] Guideline Definition Language. http://www.openehr.org/releases/CDS/latest/docs/GDL/GDL.html. Accessed 19 Dec 2016.

[3] Archetype Definition Language 1.4 specification [EB/OL]. https://www.openehr.org/releases/AM/latest/docs/ADL1.4/ADL1.4.html.

[4] Controlling High Blood Pressure [EB/OL]. https://ecqi.healthit.gov/ecqm/measures/cms165v4.

[5] Dentler, K., Teije, ten, A., Cornet, R., & de Keizer, N. (2012). Towards the Automated Calculation of Clinical Quality Indicators. In D. Riaño, A. ten Teije, & S. Miksch (Eds.), Knowledge Representation for Health-Care (pp. 51–64). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-27697-2_4

[6] Medlock, S., Opondo, D., Eslami, S., Askari, M., Wierenga, P., de Rooij, S. E., & Abu-Hanna, A. LERM (Logical Elements Rule Method): A method for assessing and formalizing clinical rules for decision support. *International Journal of Medical Informatics*, **80(4),** (2011), 286–295.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190682

Near-Real Time Monitoring of Vaccine Uptake of Pregnant Women in a Primary Care Sentinel Network: Ontological Case Definition Across Heterogeneous Data Sources

Harshana Liyanage[a], John Williams[a], Rachel Byford[a], Sameera Pathirannehelage[a], Simon de Lusignan[a]

[a] Department of Clinical & Experimental Medicine, University of Surrey, UK

Abstract

Vaccination against influenza is important in pregnancy for the health of both mother and unborn baby. Influenza introduces risks to pregnancy and to the baby who relies on maternal antibodies for protection. Because the data associated with pregnancy is fragmented across multiple providers of health care, it is challenging to conduct pregnancy-related public health surveillance using a single data source. We report the integration of a novel ontological approach to identifying pregnancies in routine data with a web-based dashboard that feeds back information to general practices in a sentinel network. As a result, practices receive information about how well they are performing influenza vaccination in pregnancy in near-real-time.

Keywords:

Electronic Health Records, Pregnancy, Vaccines;

Introduction

Uptake of influenza vaccine in pregnancy is very important because there are considerable associated risks, both to mother and unborn baby. Influenza in pregnancy can result in preterm delivery and also morbidity and mortality for the mother. This was apparent in the 2009 pandemic. In the UK hospital and maternity data give a complete retrospective overview of data, but this is not available for many months after collection. Public health bodies need near real-time monitoring of vaccine uptake in pregnancy and also to be able to study influenza vaccine effectiveness.

Monitoring influenza vaccination during pregnancy is challenging due to the suboptimal recording of pregnancy data in computerised medical record (CMR) systems [1]. Pregnant women are considered to be a risk group in influenza immunisation programmes and therefore require to be monitored by sentinel networks. The Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) works in collaboration with Public Health England (PHE) to monitor influenza, other infections, and vaccine effectiveness – particularly influenza - in England.

Pregnancy related data is sparse are often scattered across the health care providers. Obstetric care is frequently shared between midwives, the family physician, and specialist obstetric units. Therefore data is stored in different formats and locations.

Dashboards that visualise health data are increasingly being used as decision support tools to make health care delivery more dynamic. This paper describes how a pregnancy case identification method was incorporated into an updated version of the dashboard that is currently being used to monitor the vaccine uptake of pregnant women in the sentinel network.

Methods

Identification of pregnant women from routine data

We developed an algorithm that accurately inferred pregnancies even when the availability of pregnancy-related data was limited. This was achieved by adopting an ontological approach for case finding [2]. Ontological approaches use technology agnostic information structures and can make use of existing technology stacks that assist with semantic reasoning and formal validation of the relationship between concepts used. An ontological approach to identifying pregnancies and associated complications uses a systematic approach to derive this information from routine data. Both direct pregnancy codes and surrogate markers, such as therapeutic data, can be used to identify cases. The ontological concepts were identified from a combination of literature searches and input from experienced General Practitioners. The concepts were annotated using two coding systems used in computerised medical record systems within the sentinel network (i.e. Clinical Terms version 3 (CTV3) and Read V2). The pregnancy ontology was implemented using the OWL (web ontology language) standard using Protégé ontology authoring environment. The pregnancy ontology is currently hosted on the BioPortal ontology repository. We implemented the pregnancy case identification algorithm using Structured Query Language (SQL).

Integration of the pregnancy algorithm outputs with the winter wellness dashboard

The dashboard was developed as part of a "Winter Wellness" dashboard. The purposes of the dashboard were to feedback to the practices about their clinical care. The dashboard interface contained a combination of tabular and graph displays [3].

The RCGP RSC database is refreshed on a weekly basis using increment data extracted from contributing general practices in the sentinel network. The weekly increment update is followed by the execution of the pregnancy case identification algorithm on the weekly data. This allows the identification of additional pregnancy outcomes recorded during the week. This information is processed with the vaccination data received through the same weekly increment data to ascertain the status of vaccination of pregnant women in the sentinel network.

Results

We were able successfully to integrate the vaccination data of pregnancy women in the winter wellness dashboard. The dashboard allowed the user to view the number of pregnant women eligible for influenza vaccination and the number of pregnant women that were administered with the influenza vaccination. RCGP RSC has over 200 practices currently uploading over 2 million patient records each week (www.rcgp.org.uk/rsc) into a freely available national report about influenza and over thirty other monitored conditions. We have now added the ability to report on the uptake of influenza vaccine in pregnancy.

We seem to have a similar cumulative vaccination rate in pregnant women as non-pregnant women in the same age-group The vaccination rates shown in Figures 1 and 2 are the vaccination performance of the sentinel network for the current season from ISO week 35 (2018) to 11 (2019). This information can also be obtained on the dashboard for each practice by entering a practice specific code.

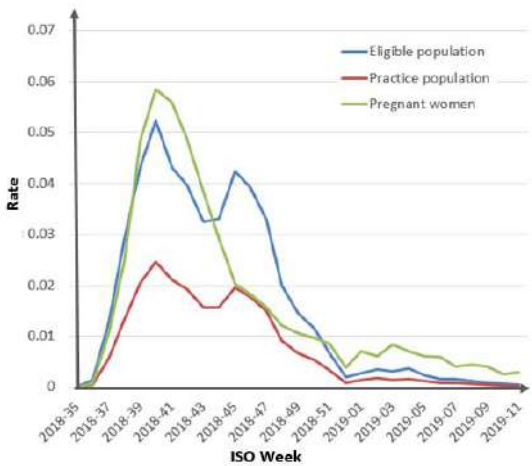


Figure 1 – The graph indicates the rate of practice population, rate of population eligible for vaccination and rate of pregnant women vaccinated at each ISO week (from week 35(2018) to week 11(2019))

Figure 1 indicates two peaks in weekly vaccination (week 40 and 44) when most influenza clinics are held at general practices for the current season. Figure 2 provides a cumulative rate of vaccination for the three groups described above. This information is particularly useful to understand if the vaccination programmes within the general practices are in a trajectory to meet the seasonal vaccination targets set.

Discussion

The use of the ontological approach also allowed us to consistently ascertain pregnancies throughout a network that uses and two different types of clinical coding systems for recording pregnancy events. We anticipate that this method will also make it easy to adapt the case identification in the system to the needs of SNOMED CT once this is fully adopted in English primary care.

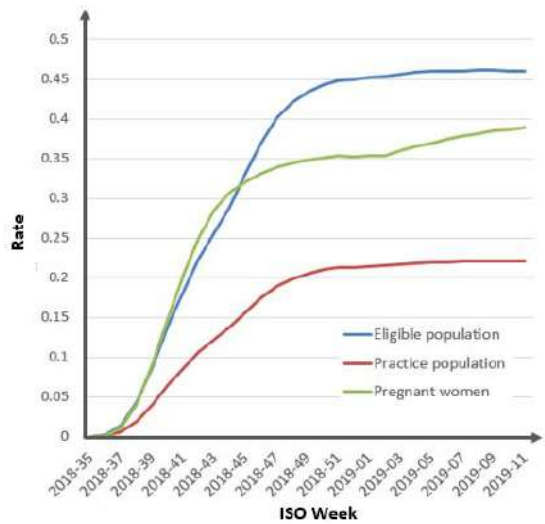


Figure 2- The graph indicates the cumulative rate of practice population, rate of population eligible for vaccination and rate of pregnant women that were vaccinated each ISO week (from week 35(2018) to week 11(2019)) from the beginning of the influenza season

Conclusions

To our knowledge this poster reports the first near-real time monitoring dashboard for influenza vaccine uptake in primary care. Realising this goal was particularly challenging due to suboptimal data quality of pregnancy events in primary care and the heterogeneity of the CMR systems we collect data from and coding systems in use. A web-based dashboard is now operationalised and providing weekly feedback of vaccine uptake to general practices in our sentinel network and will deliver reports about the level of uptake in our nationally representative network of practices to public health colleagues.

Acknowledgements

GPs and patients in RCGP RSC practices for sharing pseudonymised data. Public Health England is RCGP RSC's principal funder. EMIS, TPP Systm One, In-Practice Vision and Wellbeing software for collaboration with data extraction.

References

[1] M.C. Steinhoff, N. MacDonald, D. Pfeifer, and L.J. Muglia, Influenza vaccine in pregnancy: policy and research strategies, *Lancet* 383 (2014), 1611-1613

[2] S. de Lusignan, In this issue: Ontologies a key concept in informatics and key for open definitions of cases, exposures, and outcome measures, J Innov Health Inform 22 (2015), 170.

[3] S. Pathirannehelage, P. Kumarapeli, R. Byford, I. Yonova, F. Ferreira, and S. de Lusignan, Uptake of a Dashboard Designed to Give Realtime Feedback to a Sentinel Network About Key Data Required for Influenza Vaccine Effectiveness Studies, Stud Health Technol Inform 247 (2018), 161-165.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190683

Effects of Enterprise Digital Assistants in Medication Dispensing Operations: Case HUS Hospital Pharmacy Meilahti 2014-2018

Teressa Lyly, Santeri Palomäki, Paulus Torkki, Tomi Malmström

Aalto University School of Science, HEMA Institute, Finland
Department of Industrial Engineering and Management

Abstract

Adverse drug effects are a serious problem in hospital patient care [1]. The use of Enterprise Digital Assistants (=EDA) proved succesfull in HUS hospital pharmacy. The quality of order picking was better with EDAs. In 2018 75% of order rows and over 1 milloin packages in are picked with EDAs. There is a growing need for more detailed information of the medicines given to each patient. With EDAs this goal can be achieved cost effectively.

Keywords:

Medication Safety, Barcode Technology, Hospital Pharmacy

Introduction

Medication errors are common errors in hospital treatment and care operations [1]. In order to minimize and prevent the number of adverse drug effect to patients hospitals must find out new solutions to ensure cost effective and reliable medication to their patients. New technologies can help to solve these problems. In HUS the cost of adverse drug event attributed to € 1.8 million in additional costs [2]. Hospital pharmacies supply hospital wards with their daily medication needs. The accuracy of medication picking in hospital pharmacy is an important step in hospital medication safety chain. In this research the hospital pharmacy changed the way the medicines were recognized during medicine picking. The traditional way is to print the medicine order into a paper sheet. The personnel would collect the right medicine with the help of this paper. In the paper there is the name of the medicine, the strength of the medicine and the number of packages ordered, the place of the medicine and the code of the medicine and the details of the ordering ward. It is labourous to read and recognize all this by eye. All these hand picked medicines have to be checked by a pharmacist before they can be delivered to the wards. An enterprice digital assistant (EDA) was introduced to the hospital pharmacy in order to recognize the medicine ordered also with the help of a bar code recognition technique, Picture 1. Also there was no longer need for paper lists as all the information of the medicine order could be found in the screen of the EDA with the help of the new pharmacy software program.

Methods

This study was carried out by analysing the times used in order picking in a hospital pharmacy medicine picking in HUS Pharmacy Meilahti in Finland. First the manual order picking process was analysed by 8843 order lines. Later the order picking of selected hospital wards (n=8) were carried out with the help of EDA equipment in 2014. This sample consisted of 9286 order rows. In 2016 and 2017 a second phase of the pharmacy software program was applied and the EDAs were taken into use to all hospital wards (n = >500 in Meilahti) and all types of orders lines, except the narcotics were the traditional order picking is still in use. In 2018 32000 order rows picked with EDA in Meilahti were analysed.

Results

The number of picking errors before phaumaceutical inspection was smaller with the use of EDAs (0.136%) as compared with the traditional picking, 0.938% in 2014, Table 1 [4].

Table 1 – Comparison of Traditional order picking and EDA picking in 2014

Picking technique, n	Picking speed	Picking errors
Picking with paper, n=8843 order rows	1.4 rows per minute	0.98 %
Picking with EDA, n=9286 order rows	2.29 rows per minute	0.14 %

With the phaurmaceutical inspection the picking error rate reported by the hospital pharmacists was 0.131%. In 2017 an analysis of the order picking errors resulted good as the wrong strengteh and wrong medicine errors dimished with the use of EDAs [3]. Also the speed of order picking was faster with EDAs than with traditional picking in 2014. With traditional order picking the result was 1.4 order rows per minute. With an EDA the order picking rate was faster, 2.29 rows per minute. A secondary check out of the hand picked medicines by a pharmacist was found out unnecessary as the picking result was better with the use of EDAs than with traditional hand picking with pharmaceutical inspection. In 2017 the number of order lines picked with the use of EDAs exceeded 130 000 order rows in HUS Pharmacy Meilahti. In 2018 already 75% of the order rows are now picked with the help of EDAs in HUS Pharmacy Meilahti, Table 2. With the help of the second phase of pharmacy software the amount of medicine packages recognized with the EDAs was raised from 44% to 100%. During 2014 to 2018 the EDAs also developed and the time used for order picking was even faster, 2.1 rows per minute, n=32000. The picking process was also more reliable as in 2014 only one medicine package / order was recognized with the help

of an EDA (57%) but in 2018 all the picked medicine packages were forced to be recognized by the settings of the hospital software program by their bar code (100%) with the help of an EDA. This change in the hospital pharmacy sorfware settings helps in preventing the picking errors very efficiently. Also the EDA equipment developed during 2014-2018 and the third generation of these equipment are faster to work with. With these developments both a faster and a more precise picking process could be achieved.

Table 2 - The development of EDA order picking in Meilahti

Year	Medicine packages recognized with a bar code	% recognized with a bar code	All order rows	Order rows picked with EDA	% of order rows picked with EDA
2014: 4 months	1046	57.75	233725	571	0.24
2015: 12 months	21418	44.17	260698	15394	5.9
2016: 9 months	11469	71.93	262210	5754	2.19
2017: 12 months	5E+05	96.02	251798	1E+05	52.03
2018: 5 months	3E+05	100	113457	86072	75.86

Conclusions

The HUS hospital pharmacy is very satisfied with the new technique of picking medicine packages with the use of EDAs. The number of medicine picking errors is smaller and the rate of order picking is faster with the new EDA models. More pharmaceutical personnel nowadays work at the hospital ward directly with patient medication processes instead of the hospital pharmacy warehouse logistics chain. Therefore HUS hospital phacmacy has now applied this technique of order picking in all their 6 hospital pharmacies during 2016-2018. In year 2018 the number of order lines picked with the help of EDAs in these 6 hospital pharmacies will exceed 1 milloin order rows. HUS Pharmacy will further develop the process by adding more information into the bar coding of each medicine package as there is a growing need for more detailed information of the medicines given to each patient.

Acknowledgements

This research was funded by Business Finland, former Tekes and HUS Pharmacy. This is a part of a research program SalWe-SHOK Get It Done - Personalized Diagnostics and Care 2014-2018.

References

[1] Institute of Medicine. To Err is Human: Building a Safer Health System. National Academy Press. Washington D.C 2000.

[2] L. Tyynismaa, S. Isokirmo, M. Saanila-Sotamaa, K. Carlsson, 2+13. Lääkityspoikkeamat HUSin HaiPro-aineistossa 2012. HUS internal documents unpublished 2012.

[3] E. Ahomäki, Kämmenmikrolaitteiden käyttöönotto HUS-Apteekin keräilyssä. HUS internal document unpublished. 2018.

[4] S. Palomäki, Effect of Enterprise Digital Assistants in Medication Dispensing Operations – Case Hospital Pharmacy. Master's Thesis. Department of Information and Service Economy, Aalto University 2015.

Address for correspondence

Teressa Lyly,
Maarintie 8 Espoo Finland.
Teressa.Lyly@aalto.fi,
Tel. +358503299599,

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190684

Development of a Practical Course to Assist Elementary School Students in Acquiring the Ability to Support for Elderly People with Dementia

Seiko Masuda[a], Kotoka Murashima[a], Yukie Majima[a], Yumiko Nakamura[a]

[a] *Graduate of Humanities and Sustainable System Sciences, Osaka Prefecture University, Sakai, Osaka, Japan*

Abstract

Support for elderly people with dementia is required in the community. Many elementary school students are taking dementia kids supporter training courses, but it is difficult to nurture the ability to correctly understand and respond to dementia.. By using multiple information communication technologies, we developed a practical program with concrete response method to convey knowledge on supporting of elderly people with dementia..

Keywords:

Dementia; primary schools, robotics

Introduction

Supporting with elderly people with dementia is becoming an imminent issue in Japan that is faced with the highest elderly people with dementia prevalence rate of 2.33% among the OECD member countries with its population aging at a rapid pace like none in the world [1]. Neighborhood connections are declining, and more elderly are living alone. Under such circumstances, however, elderly people in need of assistance tend to hesitate to seek help from people around them. Many elderly are also living with dementia [2].The isolation of the elderly leads to abuse, unattended death, suicide of their caregivers from fatigue, and other threats to health and safety[3].

Therefore, existing organizations, such as neighborhood associations, elderly people groups, and parents and teachers associations, are keeping their watchful eye on senior citizens. But young generations know little about how to associate with elderly persons with dementia. Thus, public awareness campaigns about elderly with dementia are necessary among young people.

The elderly with dementia supporters training course, has been developed in Japan, with the aim of nurturing the ability to understand the elderly with elderly without prejudice and treat them with respect. The organizers also established a framework to educate supporters by using learning materials that are commonly applicable in Japan.

The current elderly with dementia supporter training course promotes an understanding of dementia for members of society. Participants apply the knowledge and skills they have learned in the course wtih practical experiences of engagining with elderly. .Elementary school students particpated in the course to increase their enagegement with elderly persons.. Organizers provided sought to create enjoyable learning material that would inspire students' interest in the disease. The developer utilized information communication technologies that are being introduced into school education for its highly expected learning efficacy. The developer combined a set of slides that aimed at raising the awareness of elderly with dementia among primary school students with the communication robot named Pepper and a clicker. The program enabled the robot to play multiple roles, to assist students in gaining knowledge, and learn specific countermeasures related to dementia. The developer is now examining the efficacy of the program.

Methods

We recruited elementary school students through posters placed in three districts of H prefecture from 2016 to 2017. Developed Softbank Robotics, Pepper was used for our study. Pepper is a humanoid robot measuring 121 centimeters high, and was designed to recognize faces and voices and strike a conversation. A special program has been built into Pepper that enables Pepper to play the roles of a tutor and as an elderly person with dementia. The developer has given the humanoid robot multiple roles that target the robotic process of acquiring knowledge and subsequently learning practical skills.. Ethics approval was obtained by the research ethics committee of the affiliated institution of the author.

Materials for Acquiring Knowledge

In the earlier part of the training course, the slides rely on illustrations and stories to encourage participants to draw an affirmative image of elderly people. The content was designed to help participants understand that dementia is a cognitive disease that makes it difficult to remember, think and learn.

To maintain the interest of participants, quizzes about the elderly with dementia were given by Pepper at the beginning, the middle, and near the end of the course. Participants answered the quiz using the clicker. The clicker enabled participants to confirm the knowledge they had learned in real-time and to maintain their motivation. To comparie their degree of understanding of the elderly and the disease, answers from all participants were instantly displayed. Anonymity secured by the clicker encouraged all participants to answer the quiz and compare their answer with other participant's answers.

Materials for Mastering Skills

To effectively achieve a successful learning environment we applied the M. D. Merrill First Principles of Instructional Design (ID) to the design of the teaching material [4]. Figure 1 shows the timing to apply the first principles of Merrill's ID to teaching materials. Specifically, when Pepper showed the symptoms of dementia, participants face ① a problem (① problem) with their knowledge (② activation). After that, when participants did not do the right correspondence, Pepper gentlely provided an example (③ example). Next, participants tried to apply their understanding to different challenging scenarios (④ application), and and and looked back on

Pepper's reaction when answering questions (⑤ integration). Through this process, participants could effectively integrate learning.

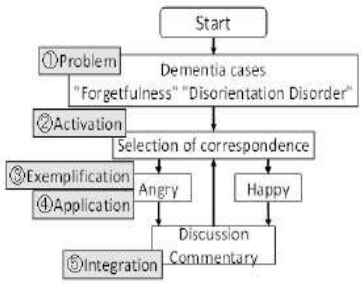


Figure 1-Application Method of Merrill's First Principles of Instructional Design

Results

We recruited 38 elementary school students through the posters. To the question: "Was slide teaching materals easy to understand?", 35/38 people (92%) responded "It was easy to understand". In the free description, there was a statement that "It was easy to understand with colorful pictures and figures, large letters". 36/38 people (95%) answered "It was easy to understand" about "Pepper's talk was easy to understand". Additional statements in the free description, included "the talk stopped suddenly, and it was hard to hear", "it was a one word". Regarding "Was it easy to understand using Clicker", all responded "easy to understand" The following was noted on the questions when comparing the correct responses before and after the course: "Dementia can Affect anyone" ($p < 0.001$), "Dementia patient and trouble family" ($p = 0.002$), "I know about elderly support networks "($p = 0.004$). (see Table 1).

About the learning effect of the modeling of the role of elderly with dementia questionnaires, we conducted a questionnaire by five level evaluations to five experts in community welfare. The average points of the results were 4.2 points for "① problem", 3.6 points for "② activation", 4.6 points for "③ exemplification", 4.4 points for "④ application", and 4.2 points for "⑤ integration". On the whole, we found the course to be effective in showing how elementary school students' behavior affects elderly people with dementia; they can see the influence of their own action and enhance the educational effects.

*Table 1-Comparision of Understanding of Dementia **b**efore and After the Workshop (n=38)*

	Before			After			p
	Yes n(%)	No n(%)	Do not know n(%)	Yes n(%)	No n(%)	Do not know n(%)	
Dementia can Affect anyone	15 (39.5)	12 (31.6)	11 (28.9)	35 (92.1)	2 (5.3)	1 (2.6)	<0.001 *
Dementia patient and troubled family	22 (57.9)	3 (7.9)	13 (34.2)	35 (92.1)	1 (2.6)	2 (5.3)	0.002 *
Everyone is easy to live in the town that will help people in need	34 (89.5)	1 (2.6)	3 (7.9)	37 (97.4)	1 (2.6)	0 (0.0)	0.102
If there is a person who is in trouble, I want to help	33 (86.8)	1 (2.6)	4 (10.5)	35 (92.1)	0 (0.0)	3 (7.9)	0.276
I know about elderly support networks	5 (13.2)	24 (63.2)	9 (23.6)	18 (63.6)	11 (28.9)	8 (21.1)	0.004 *

Signed rank test of Wilcoxon * P <0.05

Discussion

The use of multiple media encouraged participants to be actively involved in the elderly with dementia supporter training course and maintained their motivation. The course helped to build an affirmative appreciation of the elderly, and improve participants' understanding of elderly with dementia. Acting as a tutor, Pepper raised the interest and concern of participants, explaind important parts of the slides and gave a quiz. By making a robot play the role of a tutor, the program created a trilateral relationship that advanced proactive learning, instead of a bilateral confrontational relationship between teacher and learner. In order to strengthen the understanding of dementia among all participants. future development should focus on creating content that will encourage participants to exchange their knowledge with each other and subsequently build such a content into the program. Participants will be able to respond appropriately to elders with dementia through their role-playing interactions with Pepper.

Conclusions

We are incorporating Pepper to serve as tutor and role-play an elderly person with dementia. By making the humanoid robot take multiple roles, will allow for particpants to acquire and practice their new knowledge. Future work will focus on a practical application in the program where the robot will play the role of elderly person with dementia and incorporates the primary school grade viewpoint.

Acknowledgements

This research was conducted with the aid of JSPS 16 K 12355.

References

[1] OECD, *Health at a Glance2017: OECD Indicators,* http://www.oecd.org/health/health-systems/health-at-a-glance-19991312.htm (see.2018.10.20)

[2] Ikuko Kitamura, et.al., A study on the effectiveness of specialized collaboration, centered on Regional Comprehensive Support Center for Continuing Home - living of Elderly with Dementia, *Nihon Fukushi University social welfare studies* **130** (2014), 191 - 208.

[3] J.M. Keller, translated to Japanese by Katsuaki Suzuki. *Instructional design by ARCS model for designing motivation to learn*, Kitaoji Publishing (2010), 45-78.

[4] M.D.Merril, First principles of instruction, *Educational Technology Research and Development* **50** (2006),43-59.

Address for correspondence

Seiko Masuda

Addres:1-1, Gakuen-Cho, Naka-Ku, Sakai-Shi, Osaka 599-8531, Japan. Telephone: +81-72-254-9615

E-mail: s-masuda@kis.osakafu-u.ac.jp

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190685

A Cost-Effectiveness Simulation of Specialist Dispatching System in Japan for Treatments of Patients with Acute Ischemic Stroke Using a Geographic Information System

Yasuhiro Morii[a], Toshiya Osanai[b], Tomoki Ishikawa[c], Kensuke Fujiwara[a], Takumi Tanikawa[d], Eiichi Kobayashi[a], Katsuhiko Ogasawara[a]

[a] Graduate School of Health Science, Hokkaido University, Sapporo, Hokkaido, Japan
[b] Department of Neurosurgery, Graduate School of Medicine, Hokkaido University, Sapporo, Hokkaido, Japan
[c] Institute for Health Policy and Economy, Minato-ku, Tokyo, Japan
[d] Faculty of Health Sciences, Hokkaido University of Science, Sapporo, Hokkaido, Japan

Abstract

Regional disparities in the implementation rates of recombinant tissue-type plasminogen activator and endovascular thrombectomy treatments have been reported in Japan. We simulated the cost-effectiveness of specialist dispatching system in Hokkaido, Japan using Geographic Information System. In the system a qualified specialist is dispatched to another hospital for endovascular thrombectomy. Since the system improved patient accessibility, and the cost-effectiveness was excellent, the system could help enhance the equality and cost-effectiveness of ischemic stroke treatments in Hokkaido.

Keywords:

Cost-Benefit Analysis, Geographic Information System, Ischemic Stroke

Introduction

The proportions of stroke in the medical and nursing care expenditures are not ignorable. Therefore, it is necessary to consider the cost-effective management of stroke treatments. Endovascular thrombectomy (EVT) combined with recombinant tissue-type plasminogen activator (rt-PA) has been known as effective for acute ischemic stroke (AIS) patients [1]. However, due to the time limits, regional disparities in accessibility to the treatments have been reported. Specialist dispatching system (SDS), in which a qualified specialist for EVT is dispatched to another hospital for EVT, has been partly in practice in Hokkaido. SDS can be a solution to the regional disparities that will be achievable without allocating any further resources. However, there is no research that evaluates the effect of SDS on patient accessibility to rt-PA and EVT, and the effect on the cost-effectiveness of SDS. To suggest a solution to enhance the equality and the efficiency of AIS treatments in Hokkaido, this study simulates the effect of SDS on accessibility to rt-PA and EVT, and simulates the cost-effectiveness of SDS from the public perspective.

Methods

Subjects and Outcomes

The subjects in this study are a virtual cohort of patients with AIS in Hokkaido, Japan, at the time setting of 2015. Solitary island areas are excluded. There are 2 scenarios: (1) current scenario in which patients are taken to the closest of the 29 facilities where qualified specialists for EVT work (Hub hospitals); (2) SDS scenario in which a qualified specialist for EVT is dispatched from a hub hospital to another hospital without a specialist that is equipped for rt-PA and EVT, and that is within 60-minute drive distance. The primary outcome is the Incremental Cost-effectiveness Ratio (ICER) from the public perspective, which is calculated by dividing incremental costs by incremental Quality Adjusted Life Years (QALY) (Equation1). In this study, ICER is evaluated using the threshold of 5,000,000JPY per QALY (approximately US$44.154 at the exchange rate on November 7, 2018), which has been suggested for cost-effectiveness evaluation in Japan [2]. Cost gain (reduction) by the system is also analyzed.

$$\mathrm{ICER} = \frac{\text{Cost with SDS} - \text{Cost without SDS}}{\text{QALYs with SDS} - \text{QALYs without SDS}} \quad \text{(Equation1)}$$

Patient Generation

The number of AIS patients in Hokkaido in 2015 is estimated from the patient's behavior survey 2014 by the Ministry of Health Labour and Welfare [3], and are generated on Geographic Information System (GIS) based on the population distribution (Arc GIS Desktop 10-ESRI Inc.), using the "Random Point Generation" function.

Transport Time Analysis

Patient transport time is analyzed with the "Closest Facility Analysis" function of GIS. The sum of ambulance arrival time at the patient's location and the time from the arrival up to the hospital is defined as "transport time". The patients are grouped into four by the transport time windows, 0-1hour, 1-2 hours, 2-3 hours, and 3 > hours.

Patient Outcome Estimation

Patient severity is estimated in the form of modified Rankin Scale (mRS) at 90 days after onset, based on the transport time windows [4] (Table 1). Our analysis attempts to reflect the nature of rt-PA and EVT that the later a treatment start is, the worse outcome will be expected. The data of treatment time and patient outcome from the HERMES research is used for the analysis [4]. It is assumed that 16.9% of patients transported within 3 hours are eligible for rt-PA and EVT, according to the data from the Stroke Databank which is a nationwide database of stroke cases in Japan [5].

Cost and QALY Estimation

Total costs and QALYs are calculated by multiplying the per-patient cost at each time window by the number of patients at the time window, and summing the total costs of all the time window. Medical costs, as well as nursing care costs, are included in the cost analysis since stroke patients are likely to suffer post-stroke disability lifetime. Those costs are estimated

based on the estimated patient mRS stages. Data on the medical cost for patients at each mRS stage is obtained from the Stroke Databank [6] (Table 1). Nursing care costs are estimated from patients' Care Need levels (Table1), which is the official classification of patients with nursing care. (The Care need levels are divided into 7 categories (Support level 1 to 2, and Care Level 1 to 5), with Care Level patients severer than Support Level patients, and larger numbers mean that patients tend to have a severer disability). The estimated mRS stage is changed into the care needs and the nursing costs at each care need level are calculated, according to Yamaga et al. [7]. It is assumed that patients utilize nursing care services at the rates shown in Table2 [8]. QOL utility data at mRS stages are obtained from the study by Hattori [8], which collected data from a cohort of Japanese stroke patients.

Results

The results of patient transport time are shown in Table 4. The rates of patients transported within 3 hours would increase from 98.5% to 99.0% with SDS (45 increase in number). Other 203 patients (2.3% of all the patients) could be benefited from earlier treatment windows. Estimated Per-patient costs at each time window are shown in Table 5. The results showed that shorter transport time led to cheaper cost expectations. The ICER by implementing SDS was dominant since QALY gain, and $214,710 cost reduction would be achieved with SDS.

Table 1 – The expectations of mRS stage distributions at 90 days at the transport time windows and treatment eligibility

Time window	Treat eligiblity	nRS stage (%)						
		0	1	2	3	4	5	6
0-1 hour after onset	Eligible for rt-PA and EVT	0.11	0.216	0.245	0.194	0.116	0.05	0.07
1-2 hour after onset		0.096	0.198	0.24	0.202	0.127	0.057	0.08
2-3 hour after onset		0.084	0.181	0.213	0.209	0.138	0.063	0.092
All time windows	Not eligible for rt-PA and EVT	0.036	0.062	0.125	0.087	0.312	0.15	0.225

Table 2 – Per-patient medical costs, Care need levels, nursing care service utilization rates, QOL utilities for each mRS stage

	mRS stage						
	0	1	2	3	4	5	6
Total medical costs (US$)	7,594	10,244	14,570	16,780	20,130	28,520	23,930
Care need level	/	Support level 1	Support level 2	Care level 1	Care level 2 or 3*	Care level 4 or 5*	/
Nursing care service utilization rate	0.0%	26.1%	91.8%	98.5%	100.0%	100.0%	/
QOL utility	0.89	0.797	0.65	0.588	0.363	0.092	0

Conclusions

This study simulated the effect of SDS on accessibility to rt-PA and EVT, and cost-effectiveness of SDS from the public perspective. Excellent cost-effectiveness, or, cost reduction effect of SDS was revealed. Therefore, even though the potential patients who would be benefited don't account for a large part of all patients in Hokkaido, SDS will enhance the equality and cost-effectiveness of AIS treatments in Japan without allocating any further medical resources. Further analysis is required on each secondary medical area, which is a regional unit to provide in-hospital medicine, for the purpose of evaluating the areas where SDS will be the most effective.

Table 3 – Per-patient nursing costs for each Care need level

	Support level 1	Support level 2	Care level 1	Care level 2	Care level 3	Care level 4	Care level 5
Nursing care costs (US$/year)	3,079	5,263	11,264	15,224	22,296	26,757	30,811

Table 4 – The number of patients in each time windows

Scenario	Transport Time Window					
	Total	0-1 hour	1-2 hour	2-3 hour	Over 3 hours	Within 3 hours
Current	8,528	7,085	8,037	8,397	131	98.5%
SDS	8,528	7,270	8,104	8,442	86	99.0%

Table 5 – Expected per-patient costs at each time window

Transport time	Medical cost	Nursing cost	Total cost
0-1 hours	$14,459	$18,730	$33,189
1-2 hours	$14,877	$20,055	$34,932
2-3 hours	$14,418	$21,242	$35,660
3 hours	$20,177	$35,547	$55,724

References

[1] M. Goyal, B.K. Menon, W.H. Van Zwam, D.W. Dippel, P.J. Mitchell, A.M. Demchuk, A. Dávalos, C.B. Majoie, A. Van Der Lugt, and M.A. De Miquel, Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials, *The Lancet* **387** (2016), 1723-1731.

[2] T. Fukuda, T. Shiroiwa, S. Ikeda, A. Igarashi, M. Akazawa, H. Ishida, S. Noto, S. Saito, H. Sakamaki, and K. Shimozuma, Guideline for economic evaluation of healthcare technologies in Japan, *J Natl Inst Public Health* **62** (2013), 625-640.

[3] The Ministry of Labor, Health and Economy "Patient Survey 2014".

[4] J.L. Saver, M. Goyal, A. Van der Lugt, B.K. Menon, C.B. Majoie, D.W. Dippel, B.C. Campbell, R.G. Nogueira, A.M. Demchuk, and A. Tomasello, Time to treatment with endovascular thrombectomy and outcomes from ischemic stroke: a meta-analysis, *Jama* **316** (2016), 1279-1289.

[5] STROKE 2013. The epidemiological trend of stroke observed from the Japanese Stroke Databank. 2013. Available at http://att.ebm-library.jp/conferences/2013/stroke/08.html (Accessed on September 26, 2018)

[6] Kobayashi H. The Japanese Stroke Databank 2015 Total Medical Fees. Nakayama Shoten.

[7] Yamaga M, Ikeda S. Cost of illness after stroke. *Journal of International University of Health and Welfare*. 2016, 21:1, 82-92

[8] N. Hattori, T. Hirayama, and Y. Katayama, Medical care for chronic-phase stroke in Japan, *Neurologia medico-chirurgica* **52** (2012), 175-180.

Address for Correspondence

Corresponding Author: Katsuhiko Ogasawara

Email : oga@hs.hokudai.ac.jp

Mailing Address: Kita 12, Nishi 5, Kita-ku, Sapporo, Hokkaido, Japan. Phone Number : +81 011-706-3409

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190686

Developing a Model for Using Clinical Routine Data to Analyze Nursing Sensitive Patient Outcome Indicators

Renate Nantschev[a], Werner O. Hackl[a], Elske Ammenwerth[a]

[a] *Institute of Medical Informatics, UMIT – University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria*

Abstract

Under the term "big data", the secondary use of available IT-based clinical routine data promises to easily support quality management and research. Available clinical data can, for example, be used to measure care quality, as this is an important aspect in the evaluation of health interventions. Different indicators are available for measuring care quality, e.g. nursing sensitive outcome indicators. However, a specific methodology for using clinical routine data to measure quality of nursing care is not available at the moment. We present first ideas for a model to use clinical routine data to derive nursing outcome indicators measuring quality of care. In particular, we are interested in the question of which care indicators can be extracted to predict care outcome.

Keywords:

health care quality indicators, outcome assessment, patient safety

Introduction

Measuring care quality is an important aspect in the evaluation of health interventions due to increasing economic pressure and legal requirements on health care facilities. For this purpose, meaningful key indicators are necessary.

In many countries, basic indicators for care quality are defined by law. For example, in Austria, the Inpatient Quality Indicators (A-IQI) are already established as a legal instrument for measuring the quality of outcomes [1, 2].

However, this instrument ignores indicators that measure the quality of nursing care. Nursing sensitive outcome indicators can be used to measure the quality in nursing. Nursing-sensitive outcomes are measurable patient conditions based on interventions for which nurses are responsible [1]. These indicators can be subdivided into subjective and objectively measurable outcomes. The subjective outcomes are e.g. patient satisfaction, well-being and quality of life. The objective outcomes are e.g. patient safety, symptom management, functional status and self-management.

Nursing-sensitive outcomes may also focus on patient safety. Six types of nurses-reported adverse events are often described. These are: medication - administration errors, pressure ulcers, patient falls with injury and three types of healthcare-associated infections (urinary tract, catheter related bloodstream infections and pneumonia). All are considered sensitive to nursing care [3].

Currently a standardised collection of data with scientifically supported instruments for assessing the quality of nursing care outcomes does not exist in Austria [1].

In the USA, for example, data on nursing outcome indicators are collected in a standardised way in national databases such as the collaborative alliance for nursing outcomes (CALNOC) or the national database of nursing quality indicators (NDNQI).

This data is used as a basis for analysing Nursing Outcome. In Europe the currently largest study on nursing outcomes was conducted in 12 countries and with a total of 30 hospitals. Data was collected on structure, process and outcome quality. With this data the research group was able to perform nursing specific outcome analyses [4].

In practice, models of this kind require a considerable amount of additional documentation, and the collection of data is always carried out on a selective basis, e.g. once a quarter. However, routine clinical documentation is a key task in nursing practice and produces a great amount of electronically available data. Standardised methods for using clinical routine data to measure nursing outcomes are currently not available. A further challenge is to structure the clinical routine data in order to make it usable for analysing quality of care.

Objective

To develop a model for using clinical routine data to analyze quality of care for selected patient safety nursing sensitive outcome indicators.

Methods

The development of the model is based on the SPIRIT (Systematic Planning of Intelligent Reuse of Integrated Clinical Data) methodology. SPIRIT provides a general framework to plan and design solutions for the reuse of clinical routine data [5].

The development of the conceptual data model follows three steps (cf. Fig. 1):

1. Expert interviews

Experts from nursing quality control divisions from different hospitals are sampled for interviews. A semi-structured interview guide is used. The interviews cover four areas: (1) Relevance of nursing outcome indicators; (2) Currently used measurement methods; (3) Process of creating and structuring care data (4) Needs for care outcome analysis. The interviews are analysed using the qualitative content analysis according to Mayring [6]. Based on the results, questions for the data model are developed.

2. Identification of influencing factors for outcome.

In a second step, evidence for factors influencing nursing outcome indicators (e.g. patient falls) are collected from scientific literature. A systematic literature research is conducted in scientific databases such as Pub Med and CINAHL. All influencing factors from studies and guidelines considered to be relevant are included as objects in the conceptual data model.

3. Identification of clinical processes

Scenarios are created in order to obtain information about where routine clinical data is created in the process of care. Based on this, the relationships between the objects in the data model are detailed. The scenarios are evolved using five phases as described in Kosow (2008). First is the identification of the scenario field, second is the identification of the key factors, third key factors are analysed, then the scenario is generated and finally transferred [7].

After the development of the conceptual data model, a case study is conducted to match the objects of the data model with the available clinical routine data in a health care facility. Statistical analyses are carried out with the available data and a procedure for the analysis is described. Based on the results, a framework for using clinical routine data to analyse nursing sensitive patient outcomes is created.

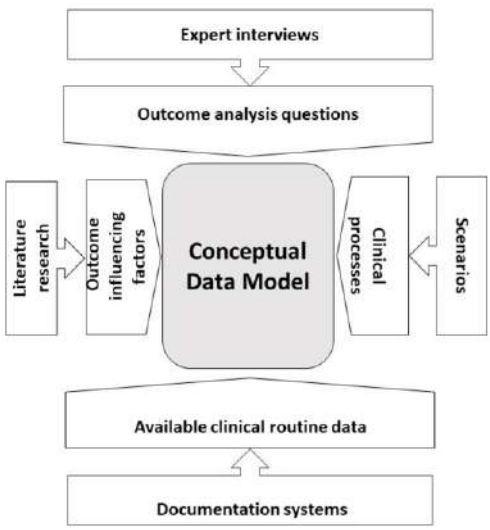


Figure 1: Overview of the conceptual nursing sensitive outcome data model with the different steps of development.

First Results

Expert interviews and literature research are currently under way. Nursing outcome indicators and quality measurement approaches have been identified based on a broad literature research.

Discussion

This study will offer a way to detect quality of care by using routinely collected patient data. The conceptual nursing sensitive outcome data model is developed as a best-practice framework, which permits the use of clinical data to measure nursing outcomes. This study will reveal if routinely collected patient data is suitable for measuring quality of care and patient outcome. If it fails to do so, it can produce input into how clinical documentation must be modified in order to achieve this goal.

Acknowledgements

This work is part of the PATIS-Project (A Patient safety Intelligence System and framework for the secondary use of multimodal clinical data to assess and improve patient safety) which is funded by the Austrian Science Fund (FWF-P 29076).

References

[1] S. Türk, Austrian Inpatient Quality Indicators (A-IQI), *Bundesministerium für Gesundheit.* (2015).
[2] F. Stewig, P. Schuchter, B. Domittner, and E. Breyer, Pflege-Ergebnisqualität, Gesundheit Österreich GmbH, Wien, 2013.
[3] D. Ausserhofer, M. Schubert, M. Desmedt, M.A. Blegen, S. De Geest, and R. Schwendimann, The association of patient safety climate and nurse-related organizational factors with selected patient outcomes: A cross-sectional survey, *Int. J. Nurs. Stud.* **50** (2013) 240–252. doi:10.1016/j.ijnurstu.2012.04.007.
[4] D. Ausserhofer, B. Gehri, S. De Geest, K. Fierz, and R. Schwendimann, [Quality and general practice development in Swiss acute hospitals as a result of study results - a follow-up survey of nursing directors of RN4CAST hospitals], *Pflege*. **28** (2015) 49–56. doi:10.1024/1012-5302/a000404.
[5] W.O. Hackl, and E. Ammenwerth, SPIRIT: Systematic Planning of Intelligent Reuse of Integrated Clinical Routine Data, *Methods Inf. Med.* **55** (2016) 114–124. doi:10.3414/ME15-01-0045.
[6] P. Mayring, Qualitative Inhaltsanalyse, 10.Aufl., Beltz, Weinheim Basel, 2008.
[7] H. Kosow, and R. Gaßner, Studies Methods of Future and Scenario Analysis, *Dtsch. Inst. für Entwicklungspolitik*. (2008).

Address for correspondence

Renate Nantschev, UMIT, Institute of Medical Informatics, Eduard Wallnöfer Zentrum 1, 6060 Hall in Tirol, Austria, renate.nantschev@umit.at.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190687

The Survey for Determining Knowledge-Related Problems in the Dissemination of ICD-11

Akemi Nishio[a], Eizen Kimura[b], Ryoma Seto[a], Yoko Sato[c], Hiroshi Mizushima[b]

[a] *Tokyo Healthcare University Setagaya, Tokyo, Japan*
[b] *National Institute of Public Health, Wako, Saitama, Japan*
[c] *National Defense Medical College, Tokorozawa, Saitama, Japan*

Abstract

Changes in the 11th revision of the International Classification of Diseases (ICD-11) from ICD-10 may significantly impact coding quality. We conducted a field trial in 2017 to evaluate the line coding quality of 19 cases coded using the coding method of the World Health Organization. The cases with low agreement between the accuracy rates of ICD-10 and ICD-11 were cases that required the extension code. We should prepare effective educational content about how to use the extension code for proper coding in ICD-11.

Keywords:

International Classification of Diseases, World Health Organization, Clinical Coding

Introduction

The World Health Organization (WHO) requires that investigations of the effects of the revisions through field trials assess applicability, reliability, and usefulness so that the WHO may fix issues before the rollout of the final version of ICD-11[1]. Field trials are conducted by experts who perform ICD coding in actual medical settings so that they not only provide useful information on the effects of revision on statistical data but also measure the effectiveness of the revision itself[2; 3]. The WHO provided the ICD-11 Reference Guide and the ICD-FiT (a web-based data entry tool to support the protocols of implementation of ICD-11 field trial core study), case scenarios, and diagnostic terms for the field trial[4]. We conducted a field trial in 2017 and evaluated line coding and case coding. In this study, we investigated the relationship between the gold standard code provided by WHO and the codes entered by the participants to determine what education and training are necessary for ICD-11 dissemination.

Methods

We conducted a field trial from August 2017 to September 2017, in collaboration with the Japan Society of Health Information Management and the Japan Hospital Association. Participants were recruited online. During the trial period, they logged into the field trial system (ICD-FiT) and entered the answers at their convenience time. We translated the ICD-11 reference guidelines and some of the instructions in ICD-FiT into Japanese. However, the overall interfaces and messages in ICD-FiT were left in English. We examined whether the ICD-10 and ICD-11 codes coded by participants were consistent with 19 gold standard diagnostic terms provided by WHO.

Agreement analysis using Gwet's AC1[5] was carried out to confirm whether the answers of ICD-10 and ICD-11 were consistent with one another. If divergence observed in the accuracy rates of ICD-10 and ICD-11, we assume that there is some gap of knowledge for coding. We used the R Statistical Software v.3.5.1 (R Foundation for Statistical Computing, Vienna, Austria) and 'sirt' package (supplementary item response theory models).

We classified comments of respondents who found difficulty during coding into eight categories and tabulated the proportions of each category for ICD-10 and ICD-11 (Table 2): (1) the classification code was ambiguous; (2) the inclusion term was insufficient or missing; (3) the classification code was indistinguishable from the others; (4) the coding guide was incomplete or unclear; (5) the description of a diagnostic term was ambiguous; (6) lack of knowledge about the code; (7) the concept of the code was too broad; and (8) other. After the field trial, we administered a questionnaire on English proficiency to the participants.

Of note, the ICD-11 provided by WHO in this trial was the development version as of August 2017 and is not identical to the ICD-11 that is currently published.

Results

Three hundred seventy-eight health information managers participated as evaluators of ICD-Fit. The total number of responses for line coding was 3,959 for 19 diagnostic cases. The accuracy rates of ICD-10/11 on Central line associated Escherichia coli (E. coli) sepsis and Kaposi's sarcoma of the soft palate with AIDS were 0%. The results of the other 17 cases are shown in Table 1. Cytomegalovirus Colitis, Right breast angiolipoma, Pleomorphic adenoma of right parotid gland showed extremely low Gwet's AC1 value; however, Cytomegalovirus Colitis had higher accuracy rate for ICD-11 than that for ICD-10. There were 8 cases where the accuracy rate of ICD-11 exceeded the ones of ICD-10, and 9 cases were lower than the ones of ICD-10. In the case where an extension code was required, the accuracy rate of the ICD-11 was meager.

The numbers of participants who experienced difficulty in coding were 91 and 115 for ICD-10 and ICD-11, respectively (Table 2; *n* represents the number of responses and *%* the proportion). Lack of knowledg of ICD-10 and that the concept of code was too broad for ICD-11 were the top reasons for causing difficulty in coding. In the cases with 0% accuracy rate in ICD-11, too broad concept of the code was the most common reason. There were 93 respondents to the questionnaire on English proficiency; 25 people chose "understand to a certain extent," 67 people chose "do not understand well," and 1 person chose "other.".

Table 1. The agreements and accuracy rate of the coding results of ICD-10 and ICD-11

ID	Disease Name	n	Gwets'AC1	ICD10 Accuracry Rate(%)	ICD11 Accuracy Rate(%)	w Extention
1192	Puerperal sepsis	291	0.8968	93.47	95.88	
1193	Sepsis due to Escherichia coli (E. coli)	187	0.6715	50.36	87.77	
1194	Staphylococcus sepsis	151	0.6865	50.91	80.36	
1195	Methicillin-Resistant Staphylococcus Aureus (MRSA) Septicemia	253	0.6336	11.86	30.04	
1196	Sepsis due to urinary tract infection	238	0.6328	26.89	1.68	
1199	Chronic obstructive pulmonary diease with pneumonia and HIV disease	194	0.9948	0.52	0.00	
1201	AIDS-related dementia	82	0.8925	4.10	0.00	
1202	Pneumoncytis pneumonia (PCP) with AIDS	180	0.8759	0.00	11.11	
1203	Wasting syndrome due to HIV	178	0.7362	84.27	88.76	
1204	Clostridium difficile diarrhea	198	0.7013	93.43	78.79	
1205	Acute gastroenterotos amd dehydration	185	0.7561	18.92	1.08	○
1206	Cytomegalovirus Colitis	197	-0.2054	36.55	89.85	
1207	Right breast angiolipoma	186	-0.173	65.59	0.00	○
1208	Moderately differentiated invasive adencarinoma of the duodenum	173	0.5652	94.80	67.05	
1209	Pleomophic adenoma of right parotid gland	168	-0.5114	80.89	3.57	○
1210	Neurofibromatosis	182	0.478	70.88	82.42	
1212	Mature teratoma of ovary	175	0.9331	6.29	0.00	○

Table 2. Causes of difficulty in line coding

Choice Number	ICD10		ICD11	
	n	%	n	%
1	6	6.59	3	2.61
2	4	4.40	5	4.35
3	10	10.99	5	4.35
4	3	3.30	5	4.35
5	5	5.49	15	13.04
6	28	30.77	33	28.70
7	18	19.78	43	37.39
8	17	18.68	6	5.22
Total	91		115	

Discussion

In most of the cases that require an extension code, the accuracy rate of ICD-11 was low. The current study was conducted only on infectious diseases; therefore, we could not determine the critical factors influencing whether the ICD-11 codes by participants matched the gold standard. We should investigate the cases requiring extensions with other types of diseases and the details in errors. Since this study was generally conducted in English with only a few materials translated into japanesse, its influence on the low accuracy rate in the complex code was considered to have been great. For a short diagnostic term, ICD codes can be easily retrieved by copy and paste in the coding tool. However, for the longer ones, it's hard for evaluators to choose the appropriate word to search. Linguistic problems might have inhibited non-English native participants from acquiring the appropriate search terms. We plan to confirm how the accuracy rate will change after all materials in ICD-FiT translated into Japanese.

For future work, we will investigate sections other than infectious disease, identify common causes of gaps between ICD-10 and ICD-11 in accuracy rates, and develop an effective curriculum in which participants can learn new features of ICD-11 efficiently.

ConclusionIn the cases that the extension code was included, the coincidence rate between ICD-11 respondents and the gold standard was very low. The low accuracy in the cases with extension code might be also due to the language barrier. We think it is necessary to translate the materials into Japanese and to have more field trials after teaching the notions of extension code to take ICD-11 to the practical stage.

Acknowledgements

This work was supported by Health and Labor Sciences Research Grant Number 201701006A.

References

[1] P.M. Ellis, G. Welch, G.L. Purdie, and G.W. Mellsop, Australasian field trials of the mental and behavioural disorders section of the draft ICD-10, *Australian and New Zealand Journal of Psychiatry* **24** (1990), 313-321.

[2] R. Raghuram and C. Shamasundar, ICD-10 Field Trials in India-A Report, *Indian journal of psychiatry* **34** (1992), 198.

[3] M. Wamboldt, A. Cordaro Jr, and D. Clarke, Parent‐Child Relational Problem: Field Trial Results, Changes in DSM‐5, and Proposed Changes for ICD‐11, *Family process* **54** (2015), 33-47.

[4] M. Donada, N. Kostanjsek, V. Della Mea, C. Celik, and R. Jakob, Piloting a collaborative web-based system for testing ICD-11, *Stud Health Technol Inform* **235** (2017), 466-470.

[5] K.L. Gwet, Computing inter‐rater reliability and its variance in the presence of high agreement, *British Journal of Mathematical and Statistical Psychology* **61** (2008), 29-48.

Address for correspondence

Akemi Nishio a-nishio@thcu.ac.jp Tokyo Healthcare University 3-11-3 Setagaya Setagaya ku, Tokyo,154-8568 ,Japan

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190688

Detection Algorithm for Inadequate Blood Specimens Due to Contamination with an Infusion Solution in the Clinical Chemistry Tests: Prevention of Incidents by Blood Draw Error

H Nozaka[a], H Sazawa[b], E Ozaki[a], M Nakano[a], H Takami[a]

[a] *Hirosaki University Graduate School of Health Sciences, Hirosaki City, Aomori, Japan*
[b] *Department of Clinical Labolatory, Hokkaido Cancer Center, Sapporo City, Hokkaido, Japan*

Abstract

The aim of this work is to study the influence of contamination with infusion in clinical chemistry tests and to design an algorithm for detection of inadequate blood specimen. We show that panic value of postassium (K+)/ glucose (GLU) or decrease of total protein (TP), albumin (ALB), urea nitrogen (BUN), uric acid (UA), high-density lipoprotein cholesterol (HDL-C), total cholesterol (T-CHO), and calcium (Ca) is an index of contamination of drip infusion solution. Through a clinical study, we show that our algorithm is useful for preventing adverse medical errors.

Keywords:
Patient Safety, Medical Errors, Clinical Chemistry Tests

Introduction

Panic value due to inadequate blood specimen is often reported in hospitals as one of the causes of contamination with drip infusion solution into the blood [1 – 2]. It is shown that contamination with drip infusion solution is caused by incorrect blood drawing techniques or positions. Therefore, drawing blood guidelines are determined by Japan Committee for Clinical Laboratory Standards (JCCLS) or National Committee for Clinical Laboratory Standards (NCCLS). Panic values due to incorrect blood drawing may be detected and confirmed using conventional quality control methods. However, it is difficult to detect contamination with drip infusion solution, where tests may not reach the panic value and may go unnoticed, leading to adverse incidents in hospital. The aim of this work is to study the influence of contamination with drip infusion in clinical chemistry tests, and to design an algorithm for detection of inadequate blood specimen that is contaminated with drip infusion solution.

Materials and Methods

Materials

The subject of research is a healthy adult person. We used five types of drips with different contents of electrolytes and glucose isotonic electrolyte infusion solution (Lactec) or hypotonic electrolyte infusion solution (T1 /T2 /T3 /T4)) for this study.

Methods

The infusion needle was injected into median antebrachial vein and the drip solution was then infused. The flow speed of infusion was controlled from 0 ml/h to 150 ml/h with infusion pump. Blood was drawn from up/down stream position of same side vein as well as from the opposite side vein (see Figures 1 and 2). After blood centrifugation, serum was collected and analyzed using a dry-chemistry auto-analyzer (FUJI DRI-CHEM 4000). We then extracted the biochemical test items that changed due to the contamination and determined the cutoff value of infusion contamination. We designed a detection algorithm for inadequate blood specimens due to contamination with infusion solution in the clinical chemistry tests. The sensitivity and specificity of the algorithm was evaluated using contaminated or non-contaminated (blineded) blood specimen from 20 patients.

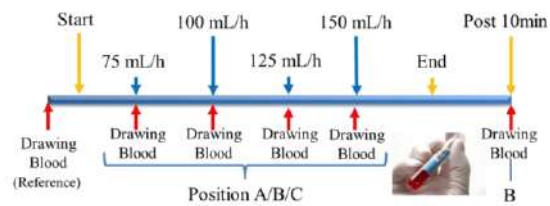


Figure 1 - Flow Chart

Figure 2 - Blood drawing position

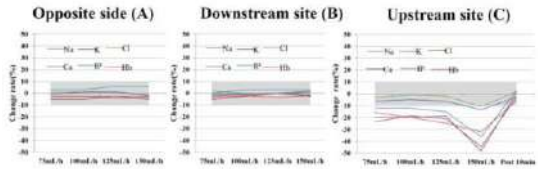
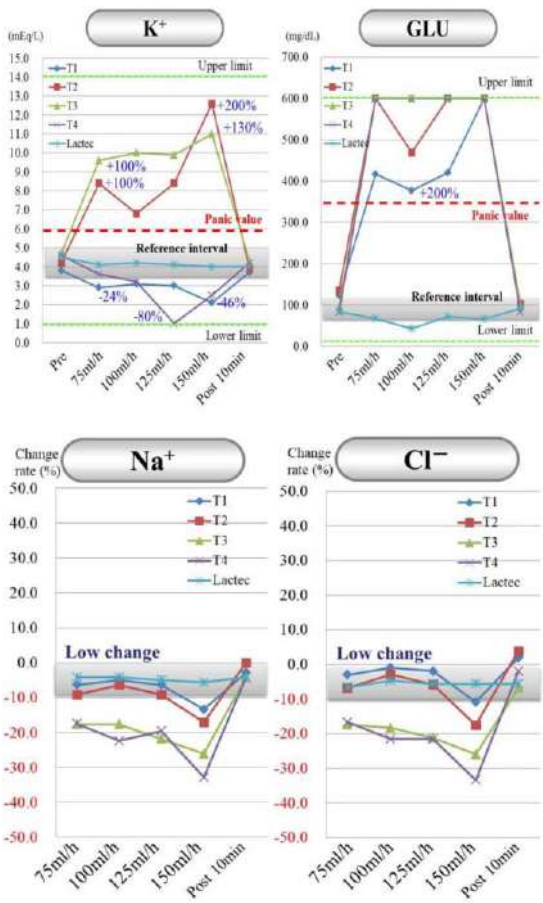


Figure 3 - Influence at each blood drawing site by T3 solution

Results

The influence at each blood drawing site is shown in Figure 3. The change of value was not observed at the downstream site and the opposite side. Conversely, significant change was observed at the upstream site. The influence of infusion speed by infusion type at the upstream site is shown in Figure 4. Potassium (K+) and glucose (GLU) values changed significantly (over +200%:75ml/h), reaching panic values.



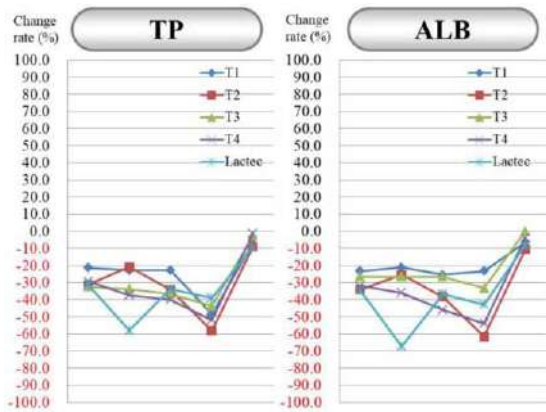


Figure 4 - Influence of infusion speed by infusion type

Total protein (TP), albumin (ALB), urea nitrogen (BUN), uric acid (UA), high-density lipoprotein cholesterol (HDL-C), total cholesterol (T-CHO), and calcium (Ca) decreased in proportion to flow speed (-20%:75m/h, -60%:150m/h). Other biochemical tests showed slight changes (within -20%), trending towards the lower limit level of analysis.The detection algorithm for inadequate blood specimens is shown in Figure 5. We designed two kinds of algorithm based on the result. Type A showed 66.7% sensitivity and 100% specificity. Type B showed 83.3% sensitivity and 75.0% specificity.

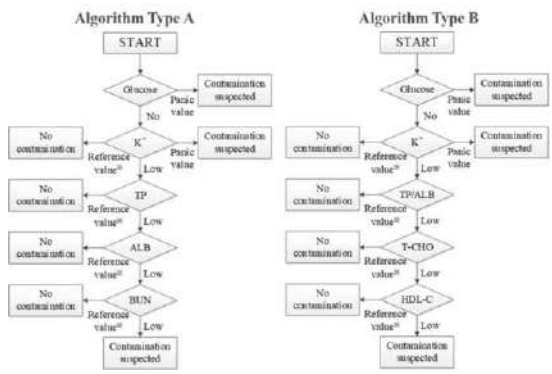


Figure 5 - Detection algorithm for inadequate blood specimens

Conclusions

We showed that increase in K+/ GLU to panic values or decrease in of TP, ALB, BUN, UA, HDL-C, T-CHO, Ca is an index of contamination of infusion solution. For the latter case, even if each of these biochemical values do not to reach the panic value, we should suspect contamination with infusion solution, when the tests show a downward tread. Our detection algorithm seems to imply that confirming the reduction of red blood cell related values may prevent medical errors. This algorithm is also effective for detection of contamination of infusion at low speed. The sensitivity of the algorithm, however, needs improvement. Our future study will be focused on quality control methods for auto analyzer in contamination of infusion.

Acknowledgements

This work was supported by SCOPE of the Japan Ministry of Internal Affairs and Communications.

References

[1] Samaranayake NR1, Cheung ST, Chui WC, Cheung BM, Technology-related medication errors in a tertiary hospital: a 5-year analysis of reported medication incidents, *Int J Med Inform*, 81(12), 828-33.
[2] Shimazu C, Miyazawa Y, Utilization of panic values in our institution: hematological testing, *Rinsho Byori*, 57(7), 665-70.

Address for Correspondence

Hiroyuki Nozaka
Hirosaki University Graduate School of Health Sciences
66-1, Honcho, Hirosaki-shi, Aomori 036-8564, Japan
E-mail: hnozaka@hirosaki-u.ac.jp

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190689

Development and Use of a Cancer Research Funding Database: Promoting Strategic Global Cancer Research Using the International Cancer Research Partnership Database

Toshio Ogawa[a], Lynne Davis[b], Teruhiko Yoshida[c], Yuri Kitamura[d], Tomotaka Sobue[d]

[a] International University of Health and Welfare Graduate School, Tokyo, Japan,
[b] International Cancer Research Partnership, London, UK
[c] National Cancer Center, Tokyo, Japan
[d] Osaka University Graduate School of Medicine, Osaka, Japan

Abstract

The International Cancer Research Partnership (ICRP) has developed a cancer research funding database since its establishment, with data gathered from the participating funnding organizations. We estimated and compared the total amount of cancer research funding from governmental organizations in the USA, the UK and Japan using ICRP and publicly available databases. We also discussed use of the ICRP database as a tool to consider the cancer research funding allocation at a national level.

Keywords:

Cancer research, Database, Government organizations

Introduction

Strategic cancer research funding is needed to maximize the impact of the funds at both national and global levels. The International Cancer Research Partnership (ICRP) was established in 2001 [1], and is an alliance of governmental and charitable organizations funding regional, national, and international cancer research grants and awards. There are 24 Partners, representing 129 cancer research funding organizations in 2018 from Australia, Belgium, Canada, Denmark, France, Hong Kong, Japan, the Netherlands, the United Kingdom, and the United States.

ICRP partners code their annual funding portfolios to a common format and submit the coded portfolios to a publicly-available database on the ICRP website (hereinafter, ICRP database). The ICRP database currently contains over 78,000 awards from 1990, and estimated roughly 60% of cancer research funding at global level.

The ICRP database contains a classification system, "Common Scientific Outline" (hereinafter, CSO) [2]. CSO was originally developed by the National Cancer Institute and the US Department of Defense to analyse cancer research funding comprehensively. The CSO has been used by ICRP and various funding agencies and countries to analyze and achieve appropriate allocation of cancer research funding.

The current version of the CSO (CSO v2) was adopted by the ICRP in April 2015 and all awards in the ICRP database are coded to this version. CSO v2 is organized into six broad areas of scientific interest in cancer research (Table 1) [1] .

Table 1– CSO codes [1]

CSO 1	Biology
CSO 2	Etiology
CSO 3	Prevention
CSO 4	Early Detection, Diagnosis, and Prognosis
CSO 5	Treatment
CSO 6	Cancer Control, Survivorship, and Outcomes Research

The ICRP database also contains a standard cancer type coding scheme. CSO and site coding lay a framework to improve coordination among research organizations, making it possible to compare and contrast the research portfolios of public, non-profit, and governmental research agencies.

ICRP publishes regular data reports using ICRP database, e.g., 2005-2008 benchmark analysis of cancer research funding, Disparities in cancer research, and Childhood cancer [1].

The governments fund a considerable amount of medical research to cancer care including biology, aetiology, prevention, treatment and public health [3]. Given that the total amount of money devoted to cancer research is limited, it is important to maximize allocation of cancer research funding as a part of national caner control programmes (NCCPs). As the ICRP database contains most of the cancer research funding from governmental organizations in some countries, such as the US and UK, it could be a useful tool to consider the cancer research funding allocation at national level.

The purpose of this study is to estimate and compare the total amount of cancer research funding from governmental organizations (hereinafter public cancer research funding) in the USA, the UK and Japan using ICRP database and publicly available databases, and to discuss effective use of the ICRP database as a part of national cancer control programmes.

Methods

We extracted data of public cancer research funding between 2011 and 2015 from the USA and the UK from the ICRP database.

As most of Japan's public cancer research grants data has not yet been included in the ICRP database, we extracted public cancer research grants data in Japan between 2011 and 2015 from publicly available databases of three ministries of Japan; namely the Ministry of Health, Labour and Welfare (MHLW), the Ministry of Education, Culture, Sports, Science and Technology (MEXT) and the Ministry of Economy, Trade and Industry (METI). We allocated at least one CSO and one site

code per grant. Grants with more than one CSO or site codes were divided equally by the number of CSO or site codes for the analyses. All Japanese grants were converted from Japanese Yen to US$ using the exchange rates of 1 JPY = US$ 0.0114.

We estimated the total amount of public research funding in the USA, the UK and Japan. We conducted a comparative analysis of public cancer research funding among the three countries, CSO and cancer site. All data was analyzed using SPSS ver. 25 (IBM, NYC).

Results

Annual public cancer research funding between 2001 and 2015 was estimated at US$ 4,407 – 5,814 million in the USA, US$ 289 – 324 million in the UK, and US$ 244 – 333 million in Japan.

Public cancer research funding was distributed primarily to "CSO5 Treatment" in all three countries, followed by "CSO1 Biology". The smallest proportion was estimated as "CSO3 Prevention" in the USA and Japan, and "CSO2 Etiology" in the UK (Table 2).

Table 2– Total Amount of Public Cancer Research Funding in the USA, UK and Japan in 2011-2015

		Total Amount (Million US$, [%])				
		2011	2012	2013	2014	2015
US	CSO1	1,274.3 (24.5%)	1,155.3 (24.6%)	977.5 (22.2%)	1,089.4 (23.3%)	1,477.1 (25.4%)
	CSO2	738.7 (14.2%)	630.5 (13.4%)	617.0 (14%)	623.7 (13.4%)	840.4 (14.5%)
	CSO3	424.8 (8.2%)	438.7 (9.3%)	440.5 (10%)	416.1 (8.9%)	469.3 (8.1%)
	CSO4	690.4 (13.3%)	640.9 (13.6%)	651.3 (14.8%)	659.2 (14.1%)	775.3 (13.3%)
	CSO5	1,426.6 (27.4%)	1,269.1 (27%)	1,222.2 (27.7%)	1,376.4 (29.5%)	1,661.7 (28.6%)
	CSO6	646.1 (12.4%)	569.0 (12.1%)	498.8 (11.3%)	502.6 (10.8%)	590.4 (10.2%)
	Total	5,200.9 (100%)	4,703.5 (100%)	4,407.3 (100%)	4,667.3 (100%)	5,814.1 (100%)
UK	CSO1	73.8 (25.6%)	62.9 (20.8%)	60.2 (20.8%)	64.4 (21.5%)	60.0 (18.5%)
	CSO2	21.5 (7.4%)	14.7 (4.9%)	11.5 (4.0%)	13.7 (4.6%)	17.6 (5.4%)
	CSO3	22.8 (7.9%)	28.8 (9.5%)	26.8 (9.3%)	27.1 (9.0%)	29.1 (9.0%)
	CSO4	45.9 (15.9%)	51.9 (17.2%)	51.4 (17.8%)	51.9 (17.3%)	56.4 (17.4%)
	CSO5	91.9 (31.8%)	106.6 (35.3%)	101.9 (35.3%)	104.1 (34.7%)	122.1 (37.7%)
	CSO6	32.7 (11.3%)	37.6 (12.4%)	37.2 (12.9%)	38.8 (12.9%)	39.0 (12%)
	Total	288.6 (100%)	302.4 (100%)	289.1 (100%)	300.1 (100%)	324.2 (100%)
JP	CSO1	56.7 (21.9%)	63.6 (23.5%)	59.8 (24.5%)	50.6 (13.1%)	70.6 (21.2%)
	CSO2	23.7 (9.1%)	24.2 (8.9%)	24.9 (10.2%)	36.3 (9.4%)	31.1 (9.4%)
	CSO3	9.7 (3.8%)	12.3 (4.5%)	9.2 (3.8%)	7.8 (2.0%)	4.8 (1.4%)
	CSO4	40.1 (15.5%)	31.8 (11.7%)	31.9 (13.1%)	54.6 (14.1%)	61.8 (18.6%)
	CSO5	100.1 (38.7%)	112.2 (41.4%)	97.1 (39.8%)	213.1 (55.2%)	144.6 (43.5%)
	CSO6	28.7 (11.1%)	26.9 (9.9%)	21.1 (8.7%)	24.0 (6.2%)	19.6 (5.9%)
	Total	258.9 (100%)	271.0 (100%)	243.9 (100%)	386.3 (100%)	332.5 (100%)

Among the cancer sites evaluated, "Not Site-Specific Cancer" received the highest amount of funding in the three countries. There were differences in the allocation of public cancer research funding per cancer site among the three countries. In the past 5 years, breast cancer, prostate cancer and lung cancer were funded as the top three cancer sites in the USA, while leukemia, breast cancer, colon and rectal cancer, and prostate cancer in the UK, and lung cancer, colon and rectal cancer, leukaemia and liver cancer in Japan. The differences may be partly explained by the cancer mortality and incidence of each country.

Table 3– Top Three Publicly Funded Cancer Research by Cancer Site in the USA, UK and Japan in 2011-2015

		2011	2012	2013	2014	2015
US	1	Not Site-Specific Cancer	Not Site-Specific Cancer	Not Site-Specific Cancer	Not Site-Specific Cancer	Not Site-Specific Cancer
	2	Breast Cancer	Breast Cancer	Breast Cancer	Breast Cancer	Breast Cancer
	3	Prostate Cancer	Prostate Cancer	Lung Cancer	Lung Cancer	Lung Cancer
UK	1	Not Site-Specific Cancer	Not Site-Specific Cancer	Not Site-Specific Cancer	Not Site-Specific Cancer	Not Site-Specific Cancer
	2	Leukemia	Breast Cancer	Breast Cancer	Breast Cancer	Prostate Cancer
	3	Breast Cancer	Leukemia	Colon and Rectal Cancer	Prostate Cancer	Breast Cancer
JP	1	Not Site-Specific Cancer	Not Site-Specific Cancer	Not Site-Specific Cancer	Not Site-Specific Cancer	Not Site-Specific Cancer
	2	Lung Cancer	Lung Cancer	Lung Cancer	Liver Cancer	Lung Cancer
	3	Colon and Rectal Cancer	Leukemia	Liver Cancer	Lung Cancer	Leukemia

Conclusions

This study indicated that public cancer research funding can be analysed using the ICRP database. The USA, the UK and Japan have common features of public cancer research funding in general, i.e., relatively higher investments in treatment and biology (CSO), and breast cancer (cancer sites). However, some differences have been revealed from our analysis particularly in the allocation of cancer research funding by cancer sites.

Acknowledgements

This research supported by Japan Agency for Medical Research and Development (AMED).

References

[1] ICRP website: https://www.icrpartnership.org/ (November 2018 Access)
[2] NCI: NCI Research Coding and Reporting. https://www.cancer.gov/about-nci/budget/nci-coding (November 2018 Access)
[3] S. Eckhouse, R. Sullivan, A survey of public funding of cancer research in the European union, *PLoS Med* **3** 2006, e267.

Address for correspondence
Toshio Ogawa, MSc., PhD
Associate Professor, International University of Health and Welfare Graduate School, 4-1-26 Akasaka Minato-ku, Tokyo 107-8402 JAPAN.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190690

Use of Alternative Currencies, Blockchain Technology, and Predictive Analytics for Chronic Disease Prevention: A Conceptual Model

Alessia Paglialonga[a], Karim Keshavjee[b]

[a] *Consiglio Nazionale delle Ricerche (CNR), Istituto di Elettronica e di Ingegneria dell'Informazione e delle Telecomunicazioni (IEIIT), Milan, Italy*
[b] *Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada,*

Abstract

This contribution outlines a conceptual model of a novel approach to chronic disease prevention and management: Preventative Health Account Supported by Electronic Medical Records (PHASE). The PHASE model combines predictive computation of risk scores for chronic disease and use of alternative currencies on blockchain technology to build a preventative health account to help individuals invest in a healthier lifestyle, reduce their risk scores and, in turn, reduce the related potential cost to the healthcare system.

Keywords:
Electronic Health Records, Preventive Medicine, Social Determinants of Health.

Introduction

The financial and societal burden of chronic diseases (CD) is very high worldwide. Globally, more than 70% of deaths are due to CD. CD directly affects health care budgets, employee productivity, and economies [1-2]. CD are highly preventable as health promotion and prevention policies can limit the impact of modifiable risk factors. One way to limit risk factors is to focus on behavior and lifestyle as typically what an individual does (or fails to do) along with general well-being, play a key role in CD development [3-4]. In general, the concept of individual health extends beyond the absence of disease as it encompasses the state of complete physical, mental, and social well-being [5]. In the past decades there has been a growing interest in health promotion and prevention strategies but they typically follow a generalized approach of CD prevention which does not consider the individual profile and context nor the fact there are differences in CD risk patterns and incidence related to socioeconomic status (SES). For example, the fact that people from lower SES are disproportionately affected by CD and have difficulty affording the cost of preventative care [6]. A variety of financial incentives have been proposed to promote healthy lifestyle and CD prevention, for example cash rewards, shopping vouchers, discounts for goods, or in-kind benefits, as well as non-guaranteed rewards (e.g. the chance of winning a lottery or raffle) [7]. Similarly, the concept of health spending accounts (HSAs) has been introduced as an alternate method to decrease healthcare costs for CD. HSAs work by providing the insured individual a specific amount of funds to be used exclusively for healthcare, typically for standard medical services [8]. However, financial incentives and HSAs have not been explored systematically for prevention in any country yet.

In general, the current systems largely focus on treating diseases rather than providing sufficient funding for services or products to improve mental and social well-being, and they do not compensate or reward disease prevention [9]. There is a need for health promotion and disease prevention programs that take into account SES as well as risk profiles for CD and that are able to focus on the individual behavior, lifestyle, and well-being. In this contribution, we outline the conceptual model of a novel approach to CD prevention that combines alternative currencies (ACs) on blockchain technology and predictive analytics supported by Electronic Medical Records (EMRs) data for individualized intervention.

Conceptual Model

Figure 1 outlines the conceptual model of the proposed approach, the Preventative Health Account Supported by EMRs (PHASE).

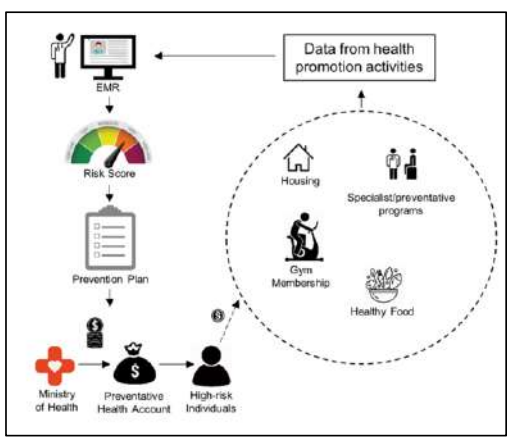


Figure 1 – Conceptual Model of PHASE

In the PHASE model the government or, in general, the organization overseeing healthcare expenditures, determines an allotted amount of money to provide to pre-identified individuals. Each individual has an associated risk score that identifies the potential high-cost users of the healthcare system. The risk score is computed by using predictive analytics on data collected from physician EMRs including, e.g., specific risks for CD, SES, and other variables. The risk score determines the individual prevention plan and the related amount of funds

deposited in the individual's preventative health account. The risk score is evaluated regularly by a family physician to determine any changes and update the individual's prevention plan and account accordingly. Data are collected within the physicians' EMR which is sent to the organization overseeing healthcare expenditures.

In the proposed PHASE model two different ACs can be deposited into the individual's account, namely the Healthcare AC (hAC) and the Reward AC (rAC). ACs are supported on a blockchain to automate transactions following user-defined criteria and logic through the use of its smart contract functionality.

The hAC can be used to purchase services or products that would help individuals to lower their risk scores. For example, the hAC can be spent at pre-approved businesses/organizations, such as healthcare providers or participating stores, e.g. to purchase fresh food, to sign up for a gym membership, to pay for housing, and so on. As the aim of hAC is to decrease the risk scores, allotment will be revised based on changes to risk scores and healthcare costs avoided. Since hAC can only be used at pre-approved organizations it will help to prevent any misuse of funds. The PHASE model also includes incentives to businesses/organizations to participate in such a program, for example a tax break or an annual credit related to the amount of hAC spent. The rAC is introduced in the PHASE model, in parallel to the hAC, to reward individuals who successfully lower their risk scores. The rAC acts as a rewards-based incentive for individuals to actively improve their health. It is deposited in the individuals' account based on their progress and how much they have saved the healthcare system. The rAC may be a combination of loyalty points and monetary incentives to be used on a wider variety of products or services compared to hAC. The goal of the rAC is to encourage individuals to continue maintaining or lowering their risk scores.

The proposed PHASE model is supported on a private, permissioned Ethereum blockchain to capture hAC and rAC transactions between government, physicians, individuals, and businesses/organizations. The solution will require two types of smart contracts built within the blockchain, owned and updated by the government or agency responsible for ACs:

1. The Complete Data Contract, holding a data structure containing the logic requiring all necessary patient information to be entered before the information can be stored in the data structure and used to assess their eligibility for hAC and rAC.
2. The AC Distribution Contract, containing the logic outlining how much hAC and rAC can be distributed based on data provided by the physician's EMR.

For the proposed model to be successful, key components such as the correct balance of hAC and rAC (amount and timing), which CD to focus on, as well as how to compute the risk scores, need to be optimized. The amount of the incentive acts as a motivator or demotivator, shaping the individual's perspective of the behaviour whereas the timing of the incentive influences reinforcement of the desired behaviour.

In general, the program would be considered effective and sustainable to the extent that allocation of AC to individuals with selected conditions can induce behavior change, lower risk scores, and help individuals maintain a healthy lifestyle. For these reasons, the program must have stringent eligibility criteria that take into account factors such as genomic predisposition and social determinants of health in order to serve its intended population and help reduce the related healthcare costs.

Conclusions

We have presented the conceptual model of PHASE, a disease prevention solution drawing on components of HSAs and incorporating two types of ACs as a way to provide necessary financial assistance to people from lower SES to prevent and/or better manage their CD. The use of blockchain technology can help reduce transactional costs between stakeholders and ensure accountability, as well as record a history of transactions. Further investigations are needed to determine how to correctly implement the model (identification of selected CD, computation of risk scores, allottment of AC, monitoring of healthcare costs, and role of the various stakeholders) in a way that it can contribute to sustained health behaviour change, decreased CD risk scores, and reduced financial burden on the healthcare system.

Acknowledgements

We acknowledge the work of Anson Li and Lauren Ribeiro on the initial concept design.

References

[1] World Health Organization (WHO), *Noncommunicable diseases country profiles.* WHO, Geneva, 2011.

[2] World Bank. *The growing danger of non-communicable diseases: Acting now to reverse course*. World Bank Human Development Network, Washington, DC, 2011.

[3] H. Schmidt, Chronic Disease Prevention and Health Promotion. In: *Public Health Ethics: Cases Spanning the Globe*. H. Barrett, et al. (eds). Springer, Cham, CH, 2016.

[4] J.A. Okely, C.R. Gale. Well-Being and Chronic Disease Incidence: The English Longitudinal Study of Ageing. *Psychosom Med* 78 (2016), 335–344.

[5] WHO, *Constitution of The World Health Organization*. Basic documents, 48th Ed. WHO, Geneva, 2014.

[6] S. Kim, B. Lee, M. Park, S: Oh, H.J. Chin, H. Koo. Prevalence of chronic disease and its controlled status according to income level. *Medicine (Baltimore)* 95 (2016), e5286.

[7] M. Tambor, M. Pavlova, S. Golinowska, J. Arsenijevic, W. Groot. Financial incentives for a healthy life style and disease prevention among older people: a systematic literature review. *BMC Health Serv Res.* 16 Suppl 5 (2016), 426.

[8] J.B. Christianson, S.T. Parente, R. Taylor. Defined-contribution health insurance products: development and prospects. *Health Aff (Millwood).* 21(1) (2002), 49-64.

[9] S. Peckham, A. Hann, S. Kendall, S. Gillam. Health promotion and disease prevention in general practice and primary care: a scoping study. *Prim Health Care Res Dev.* 18 (2017), 529-540.

Address for correspondence

Alessia Paglialonga: Consiglio Nazionale delle Ricerche (CNR), Istituto di Elettronica e di Ingegneria dell'Informazione e delle Telecomunicazioni (IEIIT), Piazza Leonardo da Vinci 32, I-20133 Milano, Italy. e-mail alessia.paglialonga@ieiit.cnr.it

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190691

Understanding U.S. Adults' Zika Virus Risk Perceptions and Mitigation Behaviors to Improve Technology-Supported Risk Communication

Tera L. Reynolds[a], Xinning Gui[a], Yunan Chen[a], Kai Zheng[a]

[a] Department of Informatics, University of California Irvine, Irvine, CA, USA

Abstract

Understanding public risk perceptions, and how these affect behavior, is critical to public health's ability to leverage technology for risk communications. However, little is known about Zika virus risk perceptions. We addressed this gap by analyzing nationally representative (U.S.) survey data. Our results suggest that a minority of U.S. adults perceive Zika to be a major threat (13%), and only about 15% have taken protective actions. Our findings have implications for improving technology-supported risk communication.

Keywords:

Zika Virus Infection, Risk, Social Media

Introduction

One of the most important public health (PH) responses to the threat of Zika virus (ZKV) is communicating the risk to the public and empowering people with the knowledge they need to protect themselves. PH authorities continue to disseminate such information through traditional media; however, increasingly, they also leverage information and communication technologies (ICTs), such as social media. This shift is a response to the rapid growth in the public's use of the internet to get news[1] and to seek health information [2].

Recent qualitative studies have found that people do indeed seek information about ZKV from PH authority websites and, when faced with scientific uncertainty, some turn to social media to collect additional information to inform their personal decision-making, and to collaboratively assess the risk of ZKV at multiple levels [3], [4]. Unfortunately, this research also suggests that PH authorities' ZKV communication through ICTs, especially social media, has been insufficient [3], [5].

These studies have provided important insights into social media users' risk perceptions – their judgements about the risk associated with ZKV – as well as into their information needs and the extent to which PH authorities have met those needs. To improve their technology-supported risk communication, though, PH authorities need to understand the pervasiveness of different ZKV risk perceptions and the factors that predict these perceptions and taking protective actions. We aimed to address this gap by quantitatively analyzing publicly available data on different ZKV risk perceptions and their effects on self-reported mitigation behaviors.

Methods

The results of this cross-sectional study are based on data from the Pew Research Center (PRC) Wave 17 American Trends Panel Survey, which was conducted between May 10 and June 6, 2016. The dataset, as well as the detailed survey methodology, are available on PRC's website [6]. Briefly, 4,563 adult members of the national, probability-based American Trends Panel completed the self-administered online (90%) or mailed (10%) survey in English or Spanish. PRC provided survey weights to account for differential probabilities of selection into the panel and nonresponse. We utilized the data from Form 1 (of three survey forms), which was completed by 1,549 individuals. Although PRC has released a report based on this survey, [7] it did not include results related to ZKV.

We examined factors associated with two main outcomes:

1. Risk Perception: ZKV risk perceptions were measured by the survey question, "How much of a threat, if any, is the Zika virus for…" (a) "The health of the U.S. population as whole" (N=1,510); (b) "Your personal health" (N=1,525); and (c) "For women who are pregnant or trying to get pregnant today" (N=1,514). For (a)-(c) the following options were provided, "A major threat," "A minor threat," and "Not a threat." While we compared perceptions for different target groups, we focused our bivariate and multivariate analyses on (b), using weighted chi-squared tests of independence for the former and a multinomial logistic regression model for the latter.

Based on the literature, [8] we hypothesized that the following would be associated with perceiving ZKV to be a threat to personal health: higher Zika knowledge; discomfort with scientific uncertainty; social **factors (e.g., recently discussed health**); personal risk factors, including demographic variables associated with higher risk of getting ZKV (i.e., U.S. region, age). We included internet use, sex, race/ethnicity, income, and education as control variables [8], [9].

2. Risk Mitigation: We were also interested in the factors associated with U.S. adults taking protective actions. This was measured with the survey question: "Have you, personally, taken any actions to reduce your risk of getting the Zika virus, or haven't you done this?" (N=1,523). While "action" was left to respondents' interpretation, based on PH recommendations, one action may be avoiding travel to high-risk areas – areas with cases of locally transmitted ZKV. In addition, if one decides to travel to or lives in a high-risk area, it may include actions such as preventing mosquito bites.

We developed risk mitigation hypotheses and selected factors based on the widely used Health Belief Model [10]. Weighted chi-squared tests were used for bivariate analyses, and a binomial logistic regression model for multivariate analysis. Specifically, we hypothesized that taking action would be more likely among those: perceiving ZKV to be a risk to their personal health, and a more severe threat (i.e., major vs. minor); with indicators of being "activated" – higher general health self-efficacy and higher knowledge of ZKV; and with personal risk factors *(see above)*. We also hypothesized that the following may be barriers to action: comfort with scientific uncertainty, limited access and interest, lower education, and lower income. Finally, based on previous studies, we included sex and race/ethnicity as control variables [9].

Results

Risk Perceptions: A majority of U.S. adults judged ZKV to be a major threat to pregnant women (60%) and a minor threat to the U.S. population (60%). In both cases, fewer than 10% believe that ZKV poses no threat. On the other hand, only 13% of adults believe that ZKV is a major threat to their personal health, and 38% no threat at all.

***Personal Health*:** The descriptive analysis revealed that all hypothesized factors were significantly associated with ZKV risk perception (α=0.05 level), with the exception of age (p=0.10). All statistically signficant factors were included in the multinomial logistic regression model. In addition, we controlled for differences in technology use by age, and an interaction between income and education. Table 1 summarizes the significant results from the multivariate analysis.

Table 1– Results of Multinomial Logistic Regression Predicting Perceiving ZKV to be a Threat (Major and Minor) to Personal Health Compared to No Threat.

	Major Threat OR (95% CI)	**Minor Threat OR (95% CI)**
ZKV Knowledge		
Nothing at all	ref	NS
A lot	2.77 (1.50-5.10)	
Discomf. w/ Uncer.	1.57 (1.04-2.36)	NS
Social Factors		
Disc. Health & Med.	NS	1.40 (1.09-1.79)
Social Media User	NS	0.52 (0.33-0.83)
U.S. Region		
Midwest	ref	ref
Northeast	NS	2.04 (1.40-2.98)
South	1.72 (1.05-2.83)	2.49 (1.81-3.44)
Race/Ethnicity		
White non-Hispanic	ref	ref
Black non-Hispanic	3.02 (1.73-5.28)	NS
Hispanic	2.67 (1.59-4.49)	NS
Other	NS	0.63 (0.40-0.98)

Risk Mitigation: About 15% of U.S. adults have taken action to reduce their risk of ZKV infection. The descriptive analysis showed that all hypothesized predictors were significantly associated with the outcome variable. Thus, all factors were included in the binomial logistic regression model, with the exception of potential barriers with descriptive evidence of not actually being barriers – internet use, education, and income – which were excluded on a conceptual basis. Table 2 summarizes the significant results from this analysis.

Table 2– Statistically Significant Results of Logistic Regression Predicting "Taking Action to Mitigate Risk of Getting Zika."

	OR (95% CI)
Susceptibility & Severity	
Not a Threat	ref
Major Threat	1.77 (1.07-2.92)
Activation	
Self-efficacy	
Low	ref
Medium	1.78 (1.06-3.13)
High	2.71 (1.50-5.08)
Zika Virus Knowledge	
Nothing at all	ref
A little	4.89 (1.94-16.33)
A lot	8.41 (3.35-27.98)
Potential Barriers to Action	
Low Interest in ID Threat News	0.20 (0.09-0.37)
U.S. Region	
Northeast	ref
Midwest	2.39 (1.41-4.10)
South	2.31 (1.47-3.72)
Age Category	
50-64	ref
65+	2.39 (1.41-4.10)
30-49	2.31 (1.47-3.72)

Conclusions

The results of this study suggest that a relatively small portion of the U.S. population perceives ZKV to be a major risk to their personal health. This perception may be a result of intuitive judgments, research, or some combination. Since perceiving ZKV to be a major threat to personal health is a key predictor of taking action to mitigate risk of infection, our results may indicate a need for PH authorities to adjust their ZKV risk strategies and communications to reach and engage new audiences and address misperceptions. Towards this end, social media presents key opportunities, including to better engage the public in mitigating the risk of ZKV in their community.

References

[1] E. Shearer and J. Gottfried, "News Use Across Social Media Platforms 2017," *Pew Research Center's Journalism Project*, 07-Sep-2017. .

[2] "Health Fact Sheet," *Pew Research Center's Internet & American Life Project*. [Online]. Available: http://www.pewinternet.org/fact-sheets/health-fact-sheet/. [Accessed: 14-Apr-2015].

[3] X. Gui, Y. Kou, K. H. Pine, and Y. Chen, "Managing Uncertainty: Using Social Media for Risk Assessment during a Public Health Crisis," presented at the CHI 2017, 2017, pp. 4520–4533.

[4] X. Gui *et al.*, "Multidimensional Risk Communication: Public Discourse on Risks during an Emerging Epidemic (in press)," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal, QC, Canada, 2018.

[5] X. Gui *et al.*, "Understanding the Patterns of Health Information Dissemination on Social Media during the Zika Outbreak," in *2017 Annual Symposium*, Washington, D.C., 2017.

[6] Pew Research Center. American Trends Panel Wave 17. http://www.pewinternet.org/dataset/american-trends-panel-wave-17/. [Accessed: 22-Feb-2018].

[7] C. Funk, B. Kennedy, and M. Hefferon, "Vast Majority of Americans Say Benefits of Childhood Vaccines Outweigh Risks," *Pew Research Center: Internet, Science & Tech*, 02-Feb-2017. .

[8] D.-H. Choi, W. Yoo, G.-Y. Noh, and K. Park, "The impact of social media on risk perceptions during the MERS outbreak in South Korea," *Comput. Hum. Behav.*, vol. 72, pp. 422–431, Jul. 2017.

[9] A. Bish and S. Michie, "Demographic and attitudinal determinants of protective behaviours during a pandemic: A review," *Br. J. Health Psychol.*, vol. 15, no. 4, pp. 797–824, Nov. 2010.

[10] M. Sharma and J. A. Romas, *Theoretical foundations of health education and health promotion*, 2nd ed. Sudbury, MA: Jones & Bartlett Learning, 2012.

Address for correspondence

Contact Tera L. Reynolds at tlreynol@uci.edu.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190692

Status Analysis of Nursing Assessment Terminology of Neurological Conditions and Its Cross-Mapping with the International Classification of Functioning, Disability, and Health (ICF)

Ding Shan-ni[a], Pan Hong-ying[a], Yu Xue-qin[a], Zhang Jian-guo[b]

[a] Sir Run Run Shaw Hospital School of Medicine Zhejiang University, Hang Zhou, Zhejiang province, China
[b] Huzhou University, Hu Zhou, Zhejiang province, China

Abstract

Patients with nervous system disorders with an accurate nursing assessment can experience an improved prognosis and promotion of health.The lack of uniform terminology limits the accuracy of nursing in China. ICF constitutes a unified and standard language can help standardize nursing assessment terms. This study show that ICF is suitable for Chinese nursing practice by using ICF Clinical Checklist and ICF-linking-rules to map the nursing assessment terminology of neurological conditions with ICF.

Keywords:

nursing informatics; International Classification of Functioning; Disability and Health; neurological;

Introduction

According to Chinese epidemiological investigation, stroke is the leading cause of death and confers a huge burden and effort on patients and health professionals. The American Heart Association points out that about 90% of patients who survive a stroke have varying degrees of dysfunction, which seriously affects the patient's psychological status, social function, and quality of life. Neurological conditions have become a global concern. Nursing assessment reflects the functional status and changes of patients accurately and provides an objective basis for diagnosis and treatment. Patients with nervous system disorders with an accurate nursing assessment can experience an improved prognosis and promotion of health. Standardized nursing terminology is the basis for improving work communication and promoting sharing of patient information. The electronic nursing information system(e-NIS) is the product of standardized terminology. The e-NIS is widely used in clinical practice, which is helpful to improve work efficiency. The lack of uniform terminology, different hospital has different e-NIS in China, which limits the accuracy of nursing and the sharing of information. For nurses, sharing knowledge with other health care workers can contribute to a broader understanding of patients' situation. Comparability of information is essential to ensure that the widest range of information is available for any decision-maker at all levels of the health system. The ICF constitutes a unified and standard language suitable as a reference for comparability of health information. ICF has a complete set of terminology systems related to function, disability, and health. The characteristics of the terminology system are accurate definitions, neutral words, the separation of structure and function, alternatives of disability and participation disabled (humanities), category definition is consistent in meaning and logic, the concept of all kinds of categories are specific and identifiable, can accurately expression the basic properties of the concept [1]. The application of the ICF in clinical nursing practice is mainly in the field of Rehabilitation Nursing about nursing diagnoses and interventions in Europe. However, there is no study of ICF cross-mapping with nursing terms in China. It is not known whether ICF is suitable for Chinese nursing assessment terminology. **The purpose** of this study is to analyze the application status of neurological nursing assessment terms and determine whether the ICF covers nursing assessment.

Methods

Three hospitals from the different cities of Zhejiang province in China were invited to participate in this study, all of them agreed. All participants were recruited, and case documents were collected between January 2016 and December 2016. Three research staff travelled to each hospital to evaluate the patients who met the standards and collected their care documents. A total of 377 patients participated in this study.

Research Tool: ICF Clinical Checklist(ICC) [2].

At first, we set up a research team consisted of a chief nurse, a postgraduate nurse, and two nurses who have been clinically active in neurological for more than ten years. The research team leader conducted a one-week theoretical and practical training for team members. The training content includes the concept of standardized nursing terminology, concept, and intention of ICF, how to use ICC, etc. Then the three members used the ICC to conduct a one-year on-site assessment of patients in neurology and neurosurgeon department at three tertiary general hospitals, and collected electronic nursing evaluation records, medical records, self-reports, family member's reports, medical examinations, clinical records, etc., from which, extracted nursing assessment terms related to neurological conditions. Forming a term pool. Next, we mapped the nursing assessment terminology with ICF based on well-established ICF-Linking-Rules [3]. Two researchers familiar with the ICF and ICF Linking Rules were appointed who link the identified source of information independently to the ICF and then compare and contrast their linking. In the case of ambiguity, a third researcher would be consulted for final decision on the most appropriate linking. If nursing assessment terms comprised more than one concept, every single concept was linked to the ICF, meaning that one nursing assessment terms could be linked to more than one ICF category. Finally, content analysis was used to analyze the current status of electronic nursing assessment terminology of neurological disorders.

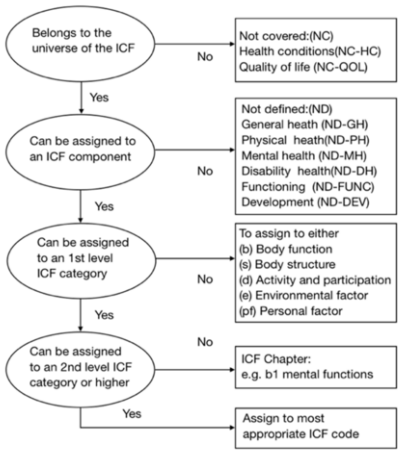


Figure 1 Linking Decision Tree

Results

One hospital has an independent daily nursing assessment form which is convenient for nurses to assess the patient's physical function and health every day. The electronic nursing assessment records of hospitals vary from one language to another. Evaluation rate of 4 (75%) neurological nursing assessment concepts were 100% (377), evaluation rate of "Dizziness" was 95.5% (360), evaluation rate of 10 (17.9%) concepts were range from 50% to 90%. There were 37.5%(21) nursing assessment concepts are closely related to neurological conditions, but did not appear in the existing electronic nursing assessment records, unrecorded rate of 9 (16.1%) categories were range from 40% to 50%, unrecorded rate of 8 (14.3%) concepts were more than 80%, of these concepts, the unrecorded rate of "orientation", "muscle tension", "fine motor skill" and "blepharoptosis" are all more than 90%. The evaluation rate of "quality of sleep" was 100%, but the evaluation rate of the sub-concepts including "on set of sleep", "maintenance of sleep" and "functions involving the sleep of cycle" were only 53.3%. Nobody evaluated and recorded the "memory"

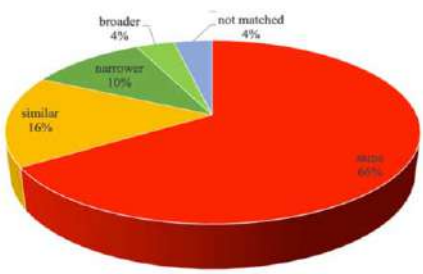


Figure 2 Mapping result of neurological diseases nursing assessment languages with ICF

Conclusions

The nursing assessment related to functional status is incomplete and electronic nursing assessment terms still lack of unified standards in China, resulting in failure to fully reflect the content of nursing work and thus fail to reflect the value of nursing.

ICF is a multidimensional "biological-psychological-social" ecological model, which can comprehensively assess the function and health status of patients. ICF can help standardize nursing assessment terms.

A total of 54 (96.4%) concepts matched with a corresponding ICF category, in terms of 37 "same" concepts (66.1%) and 9 (16.1%) concepts of "similar" granularity, which means ICF is applicable to describe the nursing assessment terms of neurological conditions. There were 6 (10.7%) of the nursing assessment terms more specific ("narrower"), such as "b134 Sleep function" in ICF, including amount of sleep, onset of sleep, maintenance of sleep, quality of sleep and functions involving the sleep of cycle, which means the original ICF categories cover the electronic nursing assessment language. ICF is applicable to describe the nursing assessment terms of neurological conditions.

Limitations

Research may be limited by the region. Zhejiang Province is one of the fastest growing provinces in e-NIS in China. It is probably that the status of nursing assessment terminology in other provinces is worse than in Zhejiang. So it is an essential issue to carry out standardized terminology as soon as possible.

Research prospects

In the next phase, the research team intends to construct a terminology system for neurological conditions nursing assessment that meets China's national conditions based on ICF and collaborates with software engineering developers to embed it into the e-NIS and apply it to clinical practice to verify its scientificity and practicality.

Acknowledgements

The authors have no conflict of interest to declare.

Funding sources: this research was supported by the Key points of Zhejiang medical and health platform.

References

[1] Wiegand N M, Belting J, Fekete C, et al. All talk, no action: The global diffusion and clinical implementation of the international classification of functioning, disability, and health[J]. American Journal of Physical Medicine & Rehabilitation, 2012, 91(7):550-560.

[2] World Health Organization. Classifications of Functioning, Disability and Health(ICF)[EB/OL]. http://www.who.int/classifications/icf/icfapptraining/en/

[3] Cieza A, Fayed N, Bickenbach J, Prodinger B. Refinements of the ICF Linking Rules to strengthen their potential for establishing comparability of health information. [J]Disability and Rehabilitation. 2016:1-10.

Address for correspondence

Ding Shanni: Sir Run Run Shaw Hospital School of Medicine Zhejiang University, Hang Zhou, Zhejiang, China;
email:linlangying@foxmail.com;Phone:15088670203

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190693

Organizing Health Data Standards Based on Knowledge Map

Liu Shen, Panpan Deng, Haixia Sun

Institute of Medical Information/Medical Library, Chinese Academy of Medical Sciences & Peking Union Medical College

Abstract

To facilitate easy use of health data standards, we collected health data standards and parsed them into more fine-grained knowledge units. Reference and inclusion relations among these standards were constructed into knowledge map. Until now there are 156 standards collected and 4796 reference relations linked within those standards. Besides, an interface was built to enable users to easily get one standard's main information without referring to numerous PDF documents.

Keywords:
ICD Codes; Data Linkage; Knowledge Bases

Introduction

The increasing volume of data in healthcare makes it a challenge to use these health data correctly and effectively. The standardization of the health data is the precondition to perform analysis methods [1]. National Health Commission of the People's Republic of China released many health data standards covering many fields in medical management activities [2]. Besides, some commercial literature databases also include some health data standards. But data in health data standards are difficult to capture and are not systematically collected and linked in those databases, usually as they are "trapped" in PDF documents that can be difficult to parse into structured fields computationally. For example, to know what data item and their allowable values in a laboratory test report would be a time-consuming task, for several databases and documents should be searched and checked. Health data standards will be collected, parsed and organized in a formalized, editable and searchable way. Also based on the reference relations among standards, linkage among those health information standards will be built into a knowledge map thus researchers, as well as clinicians, can easily get well-organized knowledge from the national health data standards.

Methods

Data Overview and Preprocessing

The health data standards can be divided into 3 main collections[2]. One part is health data elements which consist of 6 properties and focused on one subject. Take the standard 'Health data element dictionary-Part 12: Medical plan and intervention' as an example, abdominal pain degree code is one of the data elements in it. As listed in table 1, the six properties are: data element identifier; data element name; definition; data type of data element value; presentation format; and the allowable values. In that data element, the data type is S3, meaning they referred to other code systems and the presentation format is N..3, which means the code is a number with at most 3 characters, which were all defined in the primary health data standards. Sometimes the allowable values may be an enumerated type, or sometimes they may be referring to other resources like other code systems or other code value domain standards.

Table 1–Data Element for Administration Route Code

Items	Values
Identifier	DE06.00.134.00
Name	Administration route code
Definition	Indicate the route of medication in a particular coding system
Data Type	S3
Presentation Format	N..3
Allowable Values	Refer to the standard "WS 364.12-2011 Classification and coding for value domain of health data element-part 12: Medical plan and intervention", one of the common value sets CV06.00.102 Administration route code

Other important parts are standards for medical activity records, like electronic medical records and residents' health records. These standards have specifications for the document of different medical activity scenarios, like outpatient and emergency medical record and consultation record. They defined what items should be included in those standards, how these data should be organized and what standards should be referred to or adopted. Besides, standards for coding value domains are collected which consist of allowable values for some of the data element. Most of the health data collected standards documents are in a non-editable format. To parse the knowledge units in them and digitize and create linkage among health data standards, we conducted Optical Character Recognition (OCR) process to all the collected health data standards and put them in a structured database.

Knowledge Linkage Design

To construct the knowledge map, we digitized each standards' data elements, code values and document content compositions and specifications. We defined three main relations in the knowledge map: has part; refers data model and refers data value. Has part means the inclusion relationship. Refers data value indicates the value range will be based on the other standards'. Refers data model means one standard may adopt the other standards' data elements which provide not only data values but also data types and formats. Thus linkage can be constructed based on the inclusion relationships and reference information. The health data standards knowledge map linked among standards are listed below in figure 1. Besides referred

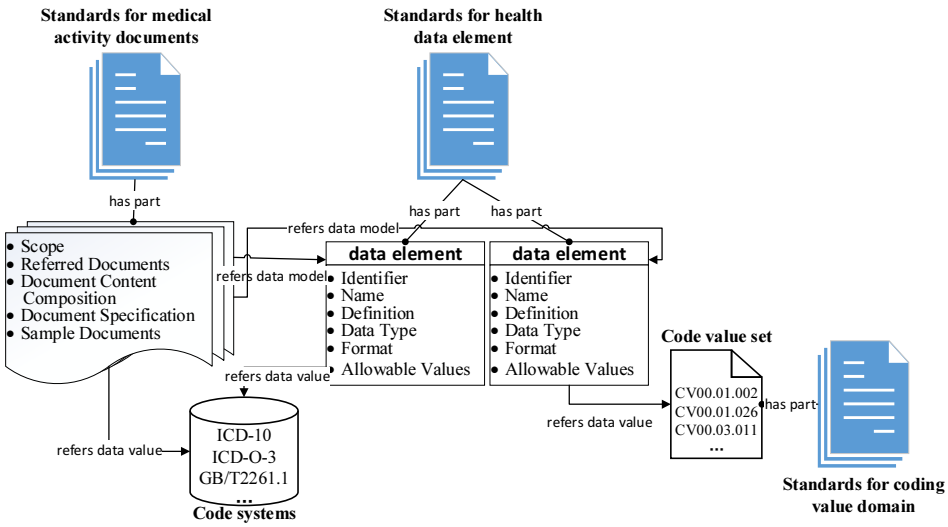


Figure 1–Health Data Standards Knowledge Map

to other standards' code values, some data elements' allowable value range may also adopt some common international code system's standard, like International Classification of Diseases (ICD) and national recommended standards. In this way, we linked all those three collections of health data standards and they can be easily reached and explored.

Results

This work provided a more fine-grained and searchable health data standards database, which currently consists of 156 health data standards, of which 37 are mainly standards for medical activity documents and 119 are mainly standards for health data elements and code values. There are 622 code value sets and 1748 data elements who will provide references for other standards. Up to the submission, there are 4796 reference relations. Besides, we also provide users an interface to search and check standards, more finely grained data elements and the reference map. The health data standards were linked by the data elements, code value sets and other code systems, and both the standards and data elements can be browsed or searched in a user-friendly interface. All of the data have been curated and standardized. Here is an example of data linkage based on the knowledge map.

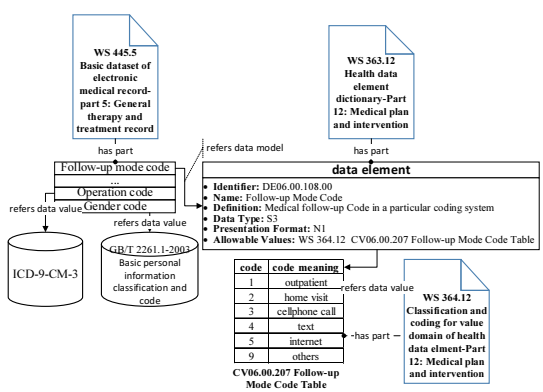


Figure 2– Data Linkage examples

Conclusions

By digitizing and connecting health data standards in a knowledge map, the detailed items and referrention relations can be more directly obtained thus in the future could improve the efficiency for researchers and medical staffs. Compared in searching abilities with WANFANG DATA and CNKI, two of the most commonly used commercial literature databases in China, and Chinese Health Information Standard Portal, our system can provide data element retrieval function within hundreds of health data standards. Properties of data elements will be easily accessed, especially the allowable value which may be the most commonly used ones. Besides, reference relations like which standards have referred to one same data element are showed clearly. However, there are still some places needed to be further explored since we haven't opened this interface to researchers and medical staffs to facilitate their work and get their feedbacks and evaluations.

Acknowledgements

This work was supported by basic scientific research operating expenses of the central public welfare scientific research (Grant No. 2018PT33024) from Chinese Academy of Medical Sciences.

References

[1] D. Liu et al., Harmonization of health data at national level: A pilot study in China, *Int J Med Inform* **79** (2010):450-458.

[2] National Health Commission of the People's Republic of China. Health Data Standard [EB/OL] http://www.nhc.gov.cn/wjw/s9497/wsbz.shtml [accessed March 23, 2019]

Address for Correspondence

NAME: Haixia Sun, EMAIL: sun.haixia@imicams.ac.cn

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190694

Quality Improvement of Blood Drawing Through Targeted Training Using an Operation Support System

Ryoko Shimono[a], Rie Akinaga[b], Norikazu Inaba[c]

[a] *Presidential Endowed Chair for "Platinum Society," The University of Tokyo, Tokyo, Japan,*
[b] *Department of Clinical Laboratory, Iizuka Hospital, Fukuoka, Japan*
[c] *Kobayashi Create Co., Ltd., Nagoya, Japan*

Abstract

In order to reduce the rate of needle re-insertion (hereafter, retaking) during a blood drawing operation, we proposed a method for training blood drawing staff, based on data analysis. The method is composed of three steps: (1) analysis of collected data for selecting staff to be trained intensively, (2) training, and (3) assessment of training results. We obtained the result that the retaking rate was reduced for the trained staff.

Keywords:
Phlebotomy, Data Analysis, Training Programs

Introduction

In order to assure quality in a blood test, phlebotomy, the act of drawing blood, is important as an upstream process step. Reducing the rate of needle re-insertion (hereafter, *retaking*) reduces adverse events and ensures smooth operation.

Akinaga et al [1] have developed a system to match the level of difficulty of a blood draw from a particular patient, and the level of skill required for the staff drawing blood. The matching system has been effective in reducing the rate of retaking and the time required to draw blood. This matching system on its own, however, does not contribute to raising the skill level of phlebotomy staff.

In order to increase the level of blood drawing skill, suitable training for staff is required. General contents of training have been developed by the World Health Organization [2] and by the Japanese Committee for Clinical Laboratory in Japan [3]. These standards are however not sufficiently precise for targeted training.

In this study, we aimed to assess the effectiveness of a targeted training, which makes use of an operation support system, in reducing the rate of blood retaking.

Methods

This study was implemented as follows:

Analysis of collected data about retaking

The authors collected data for all outpatients at Iizuka Hospital who had blood drawn in daily work. The data included detailed information on both patients who had blood drawn and phlebotomy staff.

When blood retaking occurred, the condition of the patient's blood vessel was also collected. The condition of the patient's blood vessel was described using thirteen elements that describe characteristics of the blood vessel, such as "thin," "difficult to anchor," and so on.

Then, we analyzed the collected data to select staff who required intensive training. These staff were selected based on the number and the situation of blood retaking.

Training of selected staff

The selected staff received training that included:

- lecture following the prepared procedures,
- practical phlebotomy exercise, and
- on-site training by an educational committee for one or two days.

During the on-site training, selected staff were observed peforming the phlebotomy procedure, and the correct procedure was explained in accordance with each staff member.

Assessment of training based on data analysis

In order to verify the effectiveness of the training, we conducted data analysis of the rate of retaking defined as below:

$$\text{rate of retaking (p)} = \frac{\text{number of patients with retaking (r)}}{\text{number of patients (n)}}$$

Then, we analyzed the rate of retaking by comparing the rates before and after training. We used chi-squared tests to compare the pre-training rate of retaking and the post-training rate of retaking for the eight trained staff members.

This study was approved by the institutional review boards of Iizuka Hospital.

Results

The results showed the following:

Analysis of collected data of retaking

Based on analysis of the collected data for all staff members, eight staff members who had the greatest number of retaking due to blood vessel being "thin," were selected for intensive training.

Training of selected staff

The selected eight staff members were trained through the designed training program. In the on-site training by the education committee member, the explanation for each staff was different. For example, one of the eight staff members was instructed about the position for phlebotomy, and another staff member was instructed about how to anchor a blood vessel to perform the procedure.

Assessment of training based on data analysis

We observed the rate of retaking by the selected eight staff members. *Figure 1* shows the monthly rate for six months, including the training period.

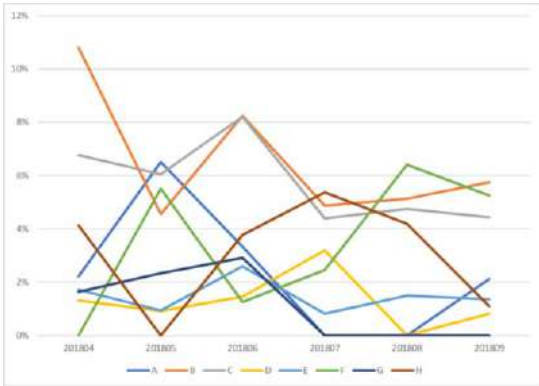


Figure 1–Monthly rate of retaking for eight trained staffs

Then, we analyzed the rate of retaking by comparing the rates before and after training. *Table 1* shows a significant reduction of the rate of retaking.

Table 1– Comparison of the retaken patients before and after training

	Before Training (n=2130)	After Training (n=4365)	P
Retaken Patients, n (%)	98 (4.6%)	133 (3.0%)	0.001

Conclusions

First, we analyzed the data collected for all outpatients who had blood drawn in daily work in order to select the staff members who received intensive phlebotomy training. Then, the assigned members trained the selected staff members. Our analysis found that the rate of retaking of trained staffs was reduced significantly when comparing post-training to pre-training retaking rates. These findings could provide a method to train assigned staff members based on data analysis in healthcare operation.

Acknowledgements

The authors would like to thank the joint research members in Iizuka Hospital for their significant contributions to this study. This study was supported by JSPS KAKENHI (Grant Number JP17K01245). And activities of the Presidential Endowed Chair for “Platinum Society” in the University of Tokyo are supported by the KAITEKI Institute, Inc., Mitsui Fudosan Co., Ltd., Shin-Etsu Chemical Co., Ltd., ORIX Corporation, Sekisui House Ltd., and East Japan Railway Company.

References

[1] Akinaga R et al., Matching of Patient’s Difficulty and Staff’s Ability to Take Blood for Improving Blood Collection Operation for Outpatients, *Japanese Journal of Clinical Laboratory Automation*, **42**(5), 599-606, 2017.

[2] World Health Organizations, WHO guidelines on drawing blood: best practices in phlebotomy, 2010.

[3] Japanese Committee for Clinical Laboratory Standards, JCCLS Standard Phlebotomy Guideline, 2006.

Address for correspondence

Ryoko Shimono, shimono@platinum.u-tokyo.ac.jp

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190695

Development of a Graphical Interface to Visualize and Analyze the Pathways of Patients During Their Hospital Stay for Thoracic Surgery

Lucile Trutt[a], Nicolas Mauduit[b], Brice Leclère[a]

[a] *Department of Medical Evaluation and Epidemiology, Nantes University Hospital, Nantes, France*
[b] *Department of Health Information Management, Nantes University Hospital, Nantes, France*

Abstract

The multidisciplinary aspect of patient care is not yet sufficiently taken into account in hospital ward management. By gathering data from the administrative database of hospital stays and the operating theaters management software, we developed a graphical user interface to visualize and analyze the pathways of patients who underwent thoracic surgery in the Nantes University Hospital of France in 2016 and 2017.

Keywords:

Data Display, Critical Pathways, Hospital Records

Introduction

In the last decades, medicine has evolved from organ-by-organ treatments to multidisciplinary patient care. But the hospital organisation and the administrative management of hospital stays are still compartmentalised, and the complexity of patient pathways is little-known. We are not currently able to determine the typical pathways taken by patients during their stay, their relative frequencies, variations, time trends, etc.

Some attempts have been conducted to visualize medical data in a dynamic way, but they focused on the evolution of care and pathology, rather than hopitalization pathways [1,2]. Medicine can benefit from the achievements of other fields, such as social science [3], energy management [4], or multi-parametric visualisation design [5].

Our objective was to develop a management support tool by means of a graphical user interface (GUI) enabling the visualization and the analysis of the patient pathways. We focused on patients who underwent thoracic surgery at Nantes University Hospital of France in 2016 and 2017.

Methods

Data

We based our work on two software databases used in the hospital: the administrative patients management software and the management software for the operating theaters. The first provided us with admission, transfer and discharge data, and the second allowed us to select patients who underwent thoracic surgery in 2016 and 2017.

These two datasets were merged into a single dataset of hopitalization sequences representing the patient pathway among medical units and operating suites. These sequences were the series of what we called "positions": combinations of the code of the medical unit where the patient went through and an index number in relation to the first surgical intervention. The data management was carried out using Python 3.6.

Interface

Our GUI was composed of two parts. The first was used for displaying a flowchart of all patient pathways within a year by a Sankey diagram. The user was able to adjust the chart through parameters like its length or precision level.

The second part was dedicated to a closer analysis of the datasets. We used a sequential pattern mining (SPM) algorithm to retrieve the most frequent patterns among the patient pathways. SPM allowed us to keep repetitions visible within the patterns and were more adapted to our linear data than algorithms relying on cycles or networks, like in process mining [6]. Then, every pattern was associated with indicators like its support (frequency) or the confidence of sequential rules between its items. The user was able to adjust parameters like the length of the pattern, requirement of specific medical unit codes, or posting of selected indicators.

The GUI was built as an interactive webpage in order to make it portable and easy to use. The SPM was carried out using the SPMF Java library [7] and the Hirate-Yamana algorithm to forbid leaps over consecutive medical units [8].

Evaluation

We evaluated the user experience of the GUI. We devised an experimental protocol with a sample group of managers and a control group of external users (with knowledge only in either medicine, statistics or computer science), and conducted user observations followed by a debriefing. We also designed five exercises to do on either part of the interface, and observed how the users used the GUI to solve them. We focused on usability and recorded all difficulties encountered by users.

Results

We retrieved 3331 total hospital stays in 2016 and 2017.

Examples of the GUI can be seen in the Figures 1 and 2. Figure 1 depicts the flow chart part for 2016, and Figure 2 shows the results of a SPM.

We observed three manager users and three control users. Overall, the user experience evaluation was satisfactory. The GUI was appealing, but some exercises showed that it can be hard to use, especially when users have a low statistical background. Participation in the construction of the tool and knowledge in the operation of hospital records seem to be the most important facilitating factors for its use and

understanding. Some areas for improvement were highlighted, especially in ergonomics. Users were overall satisfied and reported being "interested" or even "very enthusiastic" about this project.

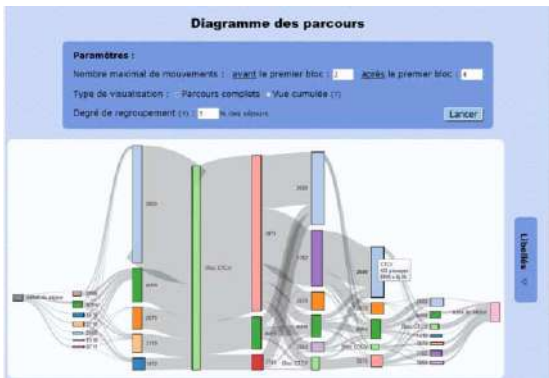


Fig.1 – Screenshot of the first part of the GUI: flow chart of patient pathways in 2017.

Fig.2 – Screenshot of the second part of the GUI: SPM beginning and ending with specific medical units in 2017.

Discussion

Overall, the user experience of the application was satisfactory; the professionals enjoyed it, and it helped them understand the patient pathways, correct the way they pictured them, and reorganize the patient coordination management properly. Some professionals from other medical specialities (e.g. surgical intensive care) who took part in meetings about this GUI even asked for a version with their own patient pathways.

However, despite its novelty and its perceived usefulness, this tool remains difficult to use, as it requires a minimum of knowledge in both statistics and hospital record operating. As requested after the evaluation, a guide has been written to help users understand how the GUI works, aside from ergonomic improvements and a context-sensitive help. Integrating users in the development of this tool was a great chance to make it more convient. Although a small amount of people were targeted, a larger evaluation of the new version should be conducted.

Another limitation of this work is that it was based on data extraction; therefore, we cannot apply it directly elsewhere with other parameters and a new database, such as in years to come or other reasons for hospitalization. A committee shall be formed to integrate this tool into the hospital information system (IS). But one of the main difficulties of this will be merging two databases from two unconnected software applications. Thus, this work points out the necessity of interoperability within IS.

Conclusions

We developed a GUI to improve the understanding of patient pathways during hospital stays. Based on two management software databases, our tool displays a Sankey diagram of all patient pathways in a year and the results of a SPM procedure. The user-experience evaluation highlighted great interest but also low user-friendliness due to a high level of specialization. This tool allowed managers to better depict the patients flows and to discuss their reorganization in a more informed way.

Despite the need for ergonomic improvement, we managed to create a helpful tool to better understand the patient pathways. Efforts should be made both to pursue this project to adapt it to suit other reasons for hospitalization and to widen its use, and to assess its usefulness on a larger scale.

Acknowledgements

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Source code: https://github.com/LucileTrutt/parcoursPatients

References

[1] A. Perer, F. Wang, and J. Hu, Mining and exploring care pathways from electronic medical records with visual analytics, *J. Biomed. Inform.* **56** (2015) 369–378.
[2] S. Cossin, D. Heve, and O. Jamet, Datavisualisation des parcours de soins, *Rev. DÉpidémiologie Santé Publique*. **66** (2018) S38–S39.
[3] G. Ritschard, R. Bürgin, and M. Studer, Exploratory mining of life event histories, in: Contemp. Issues Explor. Data Min. Behav. Sci., 1st ed., Routledge, New York, 2013: pp. 221–253.
[4] Eurostat, Sankey diagrams for energy balance, *Stat. Explain.* (2018).
[5] S. Bachhofner, I. Kis, C. Di Ciccio, and J. Mendling, Towards a Multi-parametric Visualisation Approach for Business Process Analytics, in: A. Metzger, and A. Persson (Eds.), Adv. Inf. Syst. Eng. Workshop, Springer International Publishing, Cham, 2017: pp. 85–91.
[6] M. Dumas, M. La Rosa, J. Mendling, and H.A. Reijers, Fundamentals of Business Process Management, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
[7] P. Fournier-Viger, J.C.-W. Lin, A. Gomariz, T. Gueniche, A. Soltani, Z. Deng, and H.T. Lam, The SPMF Open-Source Data Mining Library Version 2, in: Mach. Learn. Knowl. Discov. Databases, Springer International Publishing, Cham, 2016: pp. 36–40.
[8] Y. Hirate, and H. Yamana, Generalized Sequential Pattern Mining with Item Intervals, *J. Comput.* **1** (2006).

Address for correspondence

E-mail: lucile.trutt@outlook.fr

Mailing address : Dr Brice Leclère, SEME, PHU 11, Hôpital Saint Jacques, 85 rue Saint-Jacques, 44093 Nantes Cedex 1, France

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190696

An Evaluation of the Belgian Community Pharmacist's Satisfaction with the Implementation of the Electronic Prescription Within a Pharmacist's Software

Sven Van Laere[a], Pieter Cornu[b], Ronald Buyl[a]

[a] Department of Public Health (GEWE), Vrije Universiteit Brussel, Brussels, Belgium,
[b] Centre for Pharmaceutical Research (CePhar), Vrije Universiteit Brussel, Brussels, Belgium

Abstract

In a cross-sectional study, we evaluated the perception of the community pharmacist in Belgium about his satisfaction with the implementation of the electronic prescription in his software. 246 community pharmacists scored the implementation in their software with an average of 6.46±2.16(SD) on a score of 10. This satisfaction is associated with the software used ($p<0.001$), the frequency of slow responses ($p<0.001$), the perception of unavailability of systems ($p=0.003$) and the knowledge of the pharmacist ($p=0.036$).

Keywords:

Electronic Prescribing; Pharmacies; Belgium

Introduction

Electronic prescribing, or "e-prescribing" is the computer-based electronic generation, transmission, and filling of a prescription, taking the place of paper and faxed prescriptions [1]. Often the terms electronic prescribing, e-prescribing, ePrescribing and eRx are used interchangeably.

ePrescribing was introduced in healthcare primarily for increasing patient safety and reducing prescription errors. Another reason for introducing ePrescribing was the potential administrative simplification for healthcare practitioners, healthcare insurance institutions and other governmental institutions [2].

In Belgium, the government started the e-MED project in April 2007 with the intention of developing a coherent action plan for making the ePrescription of medicines in ambulatory care possible [2]. During a pilot phase in 2009-2012, infrastructure was tested. In 2013, software vendors of both physicians and pharmacists were invited for mini-labs to test all use cases related to ePrescribing and in 2014, the project was introduced to the Belgian public. As of 2017, a unique barcode (i.e., the Recip-e ID (RID)) was added on the paper proof of ePrescription. Independently, the organization Recip-e is responsible for the temporary storage of encoded ePrescriptions on a national server.

In the flow of ambulatory ePrescribing, there are three main actors: the prescriber, the patient and the pharmacist; they are sometimes referred to as the 3 Ps within ePrescribing [3, 4]. Since pharmacists are at the end of the chain and thus are most likely the party that observes the most problems and hindrance in processing ePrescriptions, we are interested in the perception of the implementation of the ePrescription within his software package and what other factors that influence this.

Methods

A cross-sectional study was conducted among pharmacists working in community pharmacies in Belgium, between March and May 2018. The survey included self-administered questions about demographic characteristics, pharmacy software characteristics, satisfaction with the implementation of the ePrescription in the software, knowledge about the ePrescription workflow (tested with 7 questions about the ePrescribing workflow and the handling with ePrescriptions inside the pharmacy), and frequency and hindrance of problems encountered in practice. The questions of problems in practice were based on an evaluation of pharmacist fora [5], where pharmacists reported the problems they encountered in practice during a one-year period, ranging from January 2017 to December 2017. In total six problems were identified: (1) unavailability of the eHealth system; (2) slow response of the software; (3) differences between paper proof and digitally stored prescription; (4) unclear error messages; (5) incorrect use of codes linked to medication; and (6) not allowed manual additions of the prescriber on the paper proof of prescription.

The survey was sent to all pharmacists that were members of the national pharmacy organization, i.e. Algemene Pharmaceutische Bond (APB), via a newsletter in their language of preference.

Multivariable linear regression was conducted to investigate what variables are associated with the pharmacist's satisfaction with the implementation of the ePrescription in their software, verifying the underlying assumptions of linearity and homoscedasticity. All p-values were 2-sided and $p < 0.05$ were considered statistically significant. For model building, both forward and backward model selection using AIC was used. Afterward, a significance check was performed for each covariate that was included in the model; covariates with a P-value of less than 0.10 were included in the final model.

Ethical clearance and approval were obtained from the Ethical Review Committee of the university hospital UZ Brussel, Brussels Health Campus (reference number B.U.N. 143201835300).

Results

The survey was distributed to 7,487 pharmacists [6] of respectively 4,943 community pharmacies. In total 4,200 newsletters were sent in Dutch (56.1%) and 3,287 newsletters sent in French (43.9%) (Table 1). A total number of 246 respondents completed the survey (response rate of 3.3%).

Respondents were asked to score how satisfied they are with the implementation of the ePrescription in their software on a scale of 1 to 10, where a score of 1 indicates very poor satisfaction and 10 indicates excellent satisfaction.

Belgian pharmacists rated the implementation of the ePrescription in their software with an average score of 6.46±2.16 (SD) out of 10 (Table 2). The minimum score obtained was 1 and the maximum score obtained was 10.

This satisfaction score was significantly associated with the software used in the pharmacy ($p < 0.001$, Table 3). For privacy reasons, no information about software vendors (n=6) was given. The perceived frequency of occurrence of a slow responsive system was associated with satisfaction, adjusted for other covariates ($p < 0.001$). A better knowledge of the workflow was associated with higher satisfaction with the implementation in their software package, adjusted for all other covariates ($p = 0.036$). A trend was observed for the perceived frequency of unavailability ($p = 0.086$). When both of these problems were perceived to occur less frequently, the community pharmacist's satisfaction was higher. If a pharmacist indicated to have problems with the unavailability of the system, the community pharmacist was asked to estimate the percentage of time that the services were down. This covariate was also significantly negatively associated with satisfaction ($p = 0.003$).

Table 1– Characteristics of mailing and respondents

	Belgium	Respondents
Pharmacists		
- Mail messages sent	7,487	-
- Language used		
Dutch	4,200 (56.1%)	143 (58.1%)
French	3,287 (43.9%)	103 (41.9%)
- Community pharmacies	4,943	246 (3.3%)

Table 2– Satisfaction of the implementation

	Score (n = 246)
Satisfaction	
Mean (± SD)	6.46 (±2.16)
Median	7
Min – Max	1 – 10
Q1 – Q3 (IQR)	5 – 8 (3)

Conclusions

In this study, the satisfaction of the Belgian community pharmacists about the implementation of the ePrescription in their software along with their knowledge and problems with it was questioned. The relation between these factors and the community pharmacist's satisfaction was observed. In general, Belgian community pharmacists are moderately satisfied with the implementation of the ePrescription in their software package. Satisfaction with the implementation of ePrescribing is mostly associated with the software package itself, the frequency of a slow responsive system, the knowledge a pharmacist has about the ePrescribing process, and perception about the percentage of unavailability of the system. A limitation of the study is that 3.3% response rate might limit the generalizability of the study. Moreover, causal associations cannot be inferred due to the cross-sectional study design.

Table 3– Multivariable analysis of satisfaction (squared transformation was applied)

	Beta ± SD	p-value
Intercept	30.950 (±3.850)	< 0.001
Software [a]	-	**< 0.001**
Frequency of slow responses		**< 0.001**
Daily (reference)	**-**	-
Weekly	13.644 (±3.392)	< 0.001
Monthly	16.710 (±4.005)	< 0.001
Less than monthly	13.823 (±4.868)	0.005
Never	14.155 (±6.091)	0.021
Knowledge	**2.563 (± 1.222)**	**0.036**
Frequency of unavailability		**0.086**
Daily (reference)	-	-
Weekly	8.787 (±3.610)	0.016
Monthly	6.053 (±4.100)	0.141
Less than monthly	9.742 (±5.422)	0.074
Never [b]	NA	NA
Perception of the percentage of unavailability	**-0.225 (±0.076)**	**0.003**

[a]: For privacy reasons no detail about the software is given;

[b]: One observation was dropped out of the analysis, because when a pharmacist indicated he never faced problems with unavailability of the system, he never obtained the question about the perception of the percentage of unavailability. (n = 245)

Acknowledgements

We are thankful to the senders of the survey to the community pharmacists on a national basis.

References

[1] eHealth Initiative (eHI), and the Center for Improving Medication Management (CIMM), A clinician's guide to electronic prescribing (2008).

[2] J. Van Calster and M.-L. Bouffioux, E-MED of het project rond het elektronisch voorschrijven van geneesmiddelen en de "authentieke bron" geneesmiddelen van het FAGG (2009),
Available: https://www.fagg-afmps.be/sites/default/files/downloads/newsletter_actua_01012009_nl.pdf [Accessed on March 19, 2019].

[3] M. Samadbeik, M. Ahmadi, F. Sadoughi, and A. Garavand, A Comparative Review of Electronic Prescription Systems: Lessons Learned from Developed Countries, J Res Pharm Pract 6(1) (2017), 3-11.

[4] P. Esmaeil Zadeh and M. C. Tremblay, A review of the literature and proposed classification on e-prescribing: Functions, assimilation stages, benefits, concerns, and risks, Res Social Adm Pharm 12(1) (2016), 1-19.

[5] L. Maus, FarmaWiki (2013). Available: http://www.farmawiki.be [Accessed on March 19, 2019]

[6] Sirius Insight, Aantal apotheken in België (2017). Available: http://siriusinsight.be/nl/whitepaper-nl/spreiding-apotheken/ [Accessed on March 19, 2019]

Address for correspondence

Sven Van Laere, Department of Public Health (GEWE), Vrije Universiteit Brussel, Laarbeeklaan 103, 1090 Brussel, Belgium.
Email: Sven.Van.Laere@vub.be – Phone: (0032) 2 477 44 24

This page intentionally left blank

IV. The Human Element in Medical Informatics

This page intentionally left blank

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190698

Implementation and Support of an Electronic Health Record in Youth Olympic Games in Buenos Aires 2018

Romina Frangella, Nicolás Vilnitzky, Manuel Farias Palma, Lautaro Olivera, Analía Baum, Santiago Esteban, María Victoria Giussi Bordoni, Leandro Alassia

Health Information Systems Office, Ministry of Health, Buenos Aires City, Argentina

Abstract

On October 2018, Buenos Aires hosted the Youth Olympic Games. The organization asked the Health Information System Office to provide the health stations and clinics with an Electronic Health Record (EHR) and bring support during the duration of the games. Four hundred and fifty healthcare professionals from different disciplines were trained in the use of the EHR.

Keywords:

Electronic Health Records, Health Plan Implementation, Massive Event

Introduction

The 2018 Youth Olympic Games (YOGs) was the most important multisport event in the history of Buenos Aires city. The event took place on October, and 4012 athletes from all over the world were housed in the Olympic Village. [1] As with other large-scale events, YOGs presented many challenges to the teams involved in their organization. The Ministry of Health had two central goals, on one hand to organize the healthcare delivery around the needs of the event and, on the other hand to implement an Electronic Health Record (EHR) in all healthcare delivery points.

Previous editions of the YOGs focused on the clinical record during the event in order to get statistical data on the most common lesions and illnesses of the athletes. This information could be valuable for decision making in the participants prevention and care. [2]

While the previous Games registered around 1000 medical encounters per Olympic game [3] [4], Buenos Aires presented a record of 2161 consultations between those registered in the clinic of the Olympic Village, the medical offices of the venues and referrals to Hospitals.

This mass event was held in four theme parks located in different parts of the city and in four independent venues around the suburbs of the Buenos Aires City.[1] The venues were scattered across all around the city district (203 km2). An Olympic Village was built to house the athletes and delegations over 100 hectares of Goverment real estate. The Olympic Village had a clinic built specifically for athletes' medical care and had an own pharmacy. 34 medical posts were built within the venues that provided healthcare for both the athletes as well general audiences. In addition, medical care delivery operated in coordination with the public emergency medical system and with 14 acute care hospitals and 2 pediatric hospitals.

Methods

Descriptive analysis of the processes carried out for the implementation of the EHR during the YOGs Buenos Aires 2018. For this description, we took the eight sociotechnical dimensions of Sittig et al. [5] as a guiding point

Results

Hardware and Software.

Each of the 34 Health posts of the venues and clinic was equipped with a desktop PC. The software used is an in-house Electronic Health Record (SIGEHOS) developed by the Ministry of Health, which allows the recording of clinical notes and the entry of structured data, such as anthropometric values and clinical problems. Connectivity was provided by the Government's Information Systems Agency.

Clinical Content

Following the International Olympic Committee (IOC) recommendations, a digital form was developed and included in the EHR, which allowed to fill specific information called "athlete injuries and diseases form" (Figure 1). In previous editions, this form was was filled out on paper. It recorded the athlete's discipline, the venue and the time of the episode to attend and days of absence from training or competition. The Injuries section included anatomical fields, type and cause of the same. The section Diseases, registered diagnosis, symptoms and causes. In addition, it was decided, in conjunction with the medical area of BAYOGOC (Buenos Aires Youth Olympic Games Committee), to add fields to record additional information such as: Patient's state of consciousness (conscious or unconscious), the treatment administered and if it needed a referral to a hospital, thus achieving a more complete record. Therefore, it was possible to comply with the requirements of the IOC information register.

Human Computer Interaction

SIGEHOS presents several modules, each of them represents a functionality. There are user profiles that organize access to each module according to the role that the user has. It was decided that both health professionals, administrative employees and site coordinators had to have access to the EHR. Within the Olympic Village also, professionals


Daily Report on
Injuries and Illnesses

Figure 1. Electronic health form

belonging to the pharmaceutical area could access another specific module called "drug administration". To speed up the process of registration and registration of the "Olympic family" (athletes, delegations, workforce, etc.) [6] it was decided to carry out a massive information transfer from the IOC, in order to have EHR available for every member and also avoiding mistakes in the records' identification.

Peopleware

The medical staff consisted of more than 450 professionals that included: doctors, pharmacists, dentists, kinesiologists, psychologists, nurses and administrative staff. The areas that took part in the coordination and development of the work protocols and the computer tools that were used were "BAYOGOC"; formed by the Argentine Olympic Committee and the Government of Buenos Aires City, the Special Olympic Games Project Unit (BAYOGOC-MED), the Health Information Systems Office, the informatic helpdesk (MDA) and the Information System Office (ASI) that It was in charge of the governmental policies that regulate the use and management of electronic media, also guaranteeing interoperability and accessibility of the electronic services of the Government of the City [7].

Communication and Workflow

The Health Information Systems Office was in charge of training and support for all the professionals who participated in the field, within the clinical devices installed in the different Olympic parks. The initial task was to train in the use of applications to professionals and administrators who were going to use the system both in the clinics and in the field of play. Training was carried out in two instances: a virtual one where the professional had a first approach with the application. The second instance was carried out in person at the offices of the Health Information Systems Office. It consisted of an explanatory class about the contents of the virtual course.

To complete the registration to the trainings, an email was sent to each professional with a link to a questionnaire. This had to be completed with the personal information that would allow him later, to generate the permissions to access the application from his work stations.

The user's request for registration would be received by the informatic helpdesk, which would then communicate to each successful discharge professional and the procedure to establish a personal password.

The health professionals transmitted the different problems of use of the system to the medical coordinator of the headquarters where they were working, and this was in charge of communicating it to the supporters of the Health Information Systems Office. In this way, they would evaluate the necessary measures for their prompt resolution in the field. This would be communicated through the same channels. The means of communication used were telephone call, instant messaging and email.

As a contingency plan, paper forms on were distributed to all offices and of hospitals' Emergency Rooms which gave assistance in order to ensure no data loss in case of unavailability of service.

Regarding the referral of patients to hospitals, it was performed by the Emergency Medical Attention System (SAME).

Conclusions

The advent of an event of the magnitude of a Youth Olympic Game involves a great planning. Regarding a Health Information System, this planning is of utmost importance for ensuring infrastructure and connectivity, an adequate software, and implementation efforts. A fourth item to have in mind regards troubleshooting and assistance for endpoint users.

The use of the Socio-Technical model described by Sittig et al [5] served as a checklist for these means. Regarding troubleshooting, it became a fortitude to have available a dedicated team in all the venues to assist the clinical caregivers.

References

[1] Guia del espectador Juegos Olímpicos available at: https://www.buenosaires2018.com/?lng=es

[2] Mountjoy, Margo, et al. "A Novel Antidoping and Medical Care Delivery Model at the 2nd Summer Youth Olympic Games (2014), Nanjing China." *British Journal of Sports Medicine*, vol. 49, no. 13, Jan. 2015, pp. 887–892., doi:10.1136/bjsports-2014-094424.

[3] Chia, J. K. K., et al. "Medical Care Delivery at the Inaugural Youth Olympic Games Singapore 2010." *British Journal of Sports Medicine*, vol. 45, no. 16, July 2011, pp. 1283–1288.,doi:10.1136/bjsports-2011-090234.

[4] Nabhan, Dustin, et al. "Sports Injury and Illness Epidemiology during the 2014 Youth Olympic Games: United States Olympic Team Surveillance." *British Journal of Sports Medicine*, vol. 50, no. 11, 2016, pp. 688–693., doi:10.1136/bjsports-2015-095835.

[5] Sittig, Dean F., and Hardeep Singh. "A New Socio-Technical Model for Studying Health Information Technology in Complex Adaptive Healthcare Systems." *Health Informatics Cognitive Informatics for Biomedicine*, 2015, pp. 59–80., doi:10.1007/978-3-319-17272-9_4.

[6] *MEDINFO: Precision Healthcare through Informatics*. IOS Press, 2017.Giussi Bordoni MV & cols. Elaboración e implementación de una Agenda Digital en Atención Primaria en la Ciudad Autónoma de Buenos Aires. CBIS 2016

[7] Greenhalgh, Trisha. "The Olympics Are over. What Have We Learned about Organizing and Delivering Healthcare?" *Journal of the Royal Society of Medicine*, vol. 105, no. 9, 2012, pp. 373–376., doi:10.1258/jrsm.2012.12k065

Address for correspondence

Lic. Romina Frangella, rfrangella@buenosaires.gob.ar
Ministry of Health, Buenos Aires City, Argentina.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190699

Design and Pilot Testing of an English and Spanish Behavioral Health Patient Survey on Data Privacy

Marcela P. Aliste[a], Adela Grando[a], Anita Murcko[a], Hiral Soni[a], Michael Todd[b], Madhumita Mukundan[a], Michael Saks[c], Caroline Horrow[d], Richard Sharp[d], Christy Dye[e], Darwyn Chern[e], Mary Jo Whitfield[f], Mark Callesen[f]

[a] *Department of Biomedical Informatics, Arizona State University, Scottsdale, Arizona, USA,*
[b] *College of Nursing and Health Innovation, Arizona State University, Tempe, Arizona, USA,*
[c] *Sandra Day O'Connor College of Law, Arizona State University, Tempe, Arizona, USA,*
[d] *Biomedical Ethics Research Program, Mayo Clinic, Rochester, Minnesota, USA,*
[e] *Partners in Recovery, Phoenix, Arizona, USA,*
[f] *Jewish Family and Children's Services, Phoenix, Arizona, USA,*

Abstract

We piloted a Spanish and English survey on data privacy. Thirty-one Latino behavioral health patients completed the survey in person with a preference for paper (78%) over electronic questionnaire. Dialect variations across Latino countries and the lack of tools to assess reading level in Spanish affected comprehension. Our experience will help others address similar tasks more effectively and encourage inclusion of Latino populations in future research.

Keywords:

Health Surveys, Privacy, Behavior, Latinos

Introduction

In the US, behavioral health conditions affect 44 million adults, a quarter of whom suffer from a serious mental illness (SMI).[1] About 17.8% of the US population is Latino or Hispanic, and of those, 15.6% had a diagnosable mental illness in the past year compared to an overall past year prevalence of 18.3%.[2] The rate of illicit drug use for Hispanic individuals ages 12 and up was 8.9%, while the national average was 10.2%.[3]

Latino patients often do not seek mental health treatment due to stigma. [4] We are moving towards more integrated care models where electronic data exchange is being facilitated by the electronic health records and health information exchanges. This means non-behavioral care providers might gain access to sensitive health information. Understanding the perceptions of Latino patients with behavioral health conditions relating to privacy and data sharing is vital to better inform and guide successful transitions to integrated care models.

The purpose of this study was to design, pilot test and refine Spanish and English surveys to elicit behavioral health patients' perspectives on data privacy.

Methods

Sample

Inclusion criteria were: 1) speak either English or Spanish; 2) no legal guardian; 3) receive care at the partnering outpatient facilities; 4) have a diagnosis of a behavioral health condition; and 5) be 21 years or older. This study was approved by the Arizona State University Institutional Review Board (IRB).

Procedures

Initial questionnaire design and organization

A 17-item English-language survey was designed to collect patient demographics (Q1-5) and behavioral health diagnoses (Q6), experience and willingness to share medical information for care (Q7-10, Q12-15), sensitivity perceptions of the information in those records (Q11) and willingness and motivation to share data for research (Q16-17).

Translation of questionnaire from English to Spanish

Two researchers, both native Spanish speakers with graduate degrees and academic knowledge of the subject, translated the survey. Back-translations were performed by a third native Spanish speaker to ensure that literacy levels commensurate with the patient's educational background and reading ability. An accuracy certificate was presented to the IRB.

Questionnaire administration

After consenting, participants completed the University of California, San Diego Brief Assessment of Capacity to Consent (UBACC) test.[5] Patients scoring lower than 15 were excluded. Eligible participants completed the survey in English or Spanish, using a paper survey or electronic tablet. The participants completed the survey twice (test-retest), 14-21 days apart. The time taken to complete the survey was recorded.

Survey revision

Questionnaire items were revised in response to reliability testing and recruiter feedback. Review of the Spanish versions of the revised items was performed by two native Spanish speakers born in Mexico. During the revision, we had access to behavioral health-specific vocabulary used by a Spanish-speaking behavioral health academy in Arizona.

Data analyses

Test-retest reliability was assessed using Cohen's Kappa and intra-class correlations (ICCs). Frequencies and percentages of item responses were tabulated.

Results

Sample characteristics

Threre were 31 (16 Spanish and 15 English) participants; 11 SMI; most were women (74%), aged 21-50 years (65%), completed high school (74%) with an annual income of ≤$20,000 (74%). Most participants presented with depression (75%) and/or anxiety or panic disorder (71%). The average UBACC score was 17.3. The paper format was preferred (78%) over electronic.

Reliability testing and revision of questionnaire

Questions yielding low ICCs < .50 and/or kappas < 0.40 were revised. Highlights below.

Q9 "*Which providers do you see outside this Clinic?"* yielded a poor kappa (-0.07 to 0.51). Response "*Yes, primary care providers. Such as a family doctor, nurse practitioner (NP), physician assistant (PA) or PCP*" was revised to: "*Primary care providers. Such as a family doctor.*"

In Q10 with kappa (-0.03 to 0.52), we replaced *"Has anyone at the facility explained to you the benefits and risks of sharing your data outside of the facility?*" with the less complex question: "*Have you been asked if you want to share your data with providers outside this Clinic?"*

Q12, ICC satisfactory, (0.53–0.86), "*Extremely willing to share*" was replaced with "*Always share.*" A new response ("*It does not apply to me*") was added for when participants did not receive care from certain types of providers. In Q13 (ICC: 0.41–0.75) and Q14 (ICC: 0.56–0.79; kappa: 0.19–0.54), difficulties in answering questions related to sharing drugs or alcohol use data, sexual transmitted diseases, etc., were noted. Recruiters reported that participants had difficulties when they did not feel a diagnosis or problem applied to them. To avoid hypothetical questions, a new option *"It does not apply to me"* was included.

Q15 and Q16 yielded poor kappa (0.13 and 0.12, respectively) and were simplified to more direct questions, e.g. Q15, *"Suppose that you don't share data with your provider. You have an emergency. Your provider wants to see all your data. Do you want your provider to see the missing data?"* to *"You have an emergency. What do you want your emergency provider to see when he looks at your data?".*

Q17 (ICC: 0.24–0.93) and Q18 (ICC: 0.37–0.95) were simplified. In Q19 when participants were asked with whom they would share data for research, the option *"State, county, or federal agencies. Such as the Arizona Department of Health Services"* was not well understood. We changed it to *"government agencies."*

Discussion

Valuable lessons learned in designing and pilot testing a bilingual data privacy survey included the following:

Tools to assess the reading level for Spanish documents are not easily accessible: Though limited methods are available to measure readability in Spanish, there is insufficient evidence of reliability or correlation to English reading levels.

Differences between Mexican and South American Spanish affected survey comprehension: Translations did not target the ancestry of the Latino population of Arizona. In response, the survey was reiwritten in Mexican Spanish and reviewed by three Mexican Spanish speakers.

Information to better characterize Latino subjects was not captured by the survey: The survey did not ask for country of origin or descent, years in the US, birth and immigration status, etc. which could healp acount for differences in culture between immigrants and the children of minorities born in the US.

Technology adoption barriers encountered when offered the electronic survey: Most of the participants preferred paper surveys and were more familiar with the term "tablet" compared to "iPad".

On-demand educational material could enhance explanation of data types: Participants often asked for clarifications on topics such as genetic data. On-demand education material could have improved comprehension.

Conclusions

The design and revision of the Spanish and English survey was complex and resource-intensive, involving a multidisciplinary team as well as native Spanish and English speakers. The result was a Spanish survey at a 6th grade reading level, high understandability, culutrally-correct behavioral health-specific vocabulary, accurate translations, and adaptation to a more Mexican-friendly Spanish. We have shared the lessons learned to help others address similar tasks more effectively and facilitate the inclusion of Latino behavioral health patients in future research.

Acknowledgements

This research was supported by the National Institute of Mental Health, through My Data Choices, Evaluation of Effective Consent Strategies for Patients with Behavioral Health Conditions (R01MH108992) grant.

References

[1] National Institute of Mental Health, Mental Illness, (2017). https://www.nimh.nih.gov/health/statistics/mental-illness.shtml (accessed October 19, 2018).

[2] SAMHSA, Racial and Ethnic Minority Populations, (2014). https://www.samhsa.gov/specific-populations/racial-ethnic-minority (accessed June 14, 2018).

[3] U.S. Census Bureau QuickFacts: UNITED STATES, (n.d.). https://www.census.gov/quickfacts/fact/table/US/RHI725216#viewtop (accessed June 14, 2018).

[4] A. Interian, I.E. Martinez, P.J. Guarnaccia, W.A. Vega, and J.I. Escobar, A Qualitative Analysis of the Perception of Stigma Among Latinos Receiving Antidepressants, *Psychiatr Serv*. **58** (2007) 1591–1594. doi:10.1176/appi.ps.58.12.1591.

[5] D.V. Jeste, B.W. Palmer, P.S. Appelbaum, S. Golshan, D. Glorioso, L.B. Dunn, K. Kim, T. Meeks, and H.C. Kraemer, A new brief instrument for assessing decisional capacity for clinical research, *Arch. Gen. Psychiatry*. **64** (2007) 966–974. doi:10.1001/archpsyc.64.8.966.

Address for correspondence

Adela Grando, PhD, agrando@asu.edu. Arizona State University, Department of Biomedical Informatics, 13212 E Shea Boulevard, Scottsdale, AZ 85259, US.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190700

Online Support Groups as a Source of Empowerment for People with Type 2 Diabetes

Abdulaziz Almanea[a], Peter A. Bath[a], Laura Sbaffi[a]

[a] *Information School, University of Sheffield, Sheffield, United Kingdom*

Abstract

People with Type 2 Diabetes (T2D) control much of their illness by making daily decisions regarding their health behaviours. They require certain skills, information, and support, which might not be obtainable from healthcare providers, and they may seek support through other media, such as online support groups (OSGs). This study seeks to understand the role of OSGs in empowering people with T2D by thematically analysing threads and posts from two UK OSGs.

Keywords:

Patient Empowerment; Self-Help Groups; Diabetes Mellitus, Type 2.

Introduction

Diabetes is a serious chronic condition that has serious comorbidities, such as increased risk of stroke, vision loss, and kidney failure. It is a growing global threat with approximately 425 million adults diagnosed worldwide in 2017, a figure that is estimated to reach 629 million by 2045 [1]. In 2017, approximately 850 billion United States Dollars (USD) were spent on the treatment of diabetes and its complications worldwide, which is around 10% of global healthcare expenditures [1]. In the UK, approximately 4.5 million people have diabetes according to estimates, including 1 million undiagnosed cases [2].

Type 2 diabetes (T2D) accounts for 90% of all diabetes cases, and people are usually diagnosed with T2D in late adulthood [1], which makes them struggle with managing the illness. Because of its ongoing nature, people with T2D manage 95% of their condition and healthcare professionals (HCPs) only have limited control [3]. Therefore, it is especially important to empower people with T2D, since the proposed behavioural changes relate strongly to their lifestyle [4]. However, with the limited information provided during and after diagnosis, they often seek other sources of information, e.g., the Web and online support groups (OSGs), which have been shown to be effective media to provide information and emotional support for patients [5].This is particularly important when these types of support are not readily available from HCPs.

Empowered patients are knowledgeable, skilled, and responsive to set and attain their own health goals [6]. Patient empowerment (PE) can lead to more effective communications between patients and HCPs [7], increased quality of decision making [8], improved control and management of the illness [8], and lower costs for healthcare providers [9]. Healthcare organisations around the world, including the UK [10], the EU [11] and the US [12], have set PE as a main component and have developed many initiatives to achieve this.

The Health Care Empowerment model (HCEM) identifies empowered patients as those who are engaged in the treatment of their illness, committed to it, collaborative with HCPs, informed about treatment options as well as their rights, and are able to cope with uncertainty, which can occur during any phase of the treatment [13]. The model proposes that a dynamic interplay of contextual factors, personal resources, and intrapersonal processes can influence PE for a patient who undertakes ongoing treatment. The aim of this study was to investigate how these elements, as well as empowerment elements identified by the model (i.e., patient engagement, being informed, collaborative, committed, and able to cope with uncertainty), are influenced by the use of OSGs for people with T2D. In particular, it seeks to answer the following research questions: 1) what are the information needs and behaviours of people with T2D using OSGs? 2) how do various contextual and intrapersonal factors and personal resources influence PE for people with T2D? 3) how do patients with T2D utilise the information they receive on OSGs, and how does this influence their health behaviours?

Methods

To answer the research questions, an inductive approach, utilising qualitative methods was adopted for this study. Two UK OSGs, *Diabetes.co.uk* (DCU) and *Diabetes-support.org.uk (DSOU)*, were selected, and threads from June 1st to July 31st 2017 in the general discussion sections in the groups were purposively sampled and analysed. In order to investigate the importance of OSGs in the context of empowering patients in the UK's National Health Service (NHS), members who indicated a non-UK location in their profiles were excluded from the analyses. A total sample containing 57 threads, which included 555 posts, were purposively selected (with an inclusion rate of 25% for Diabetes.co.uk and 17% for Diabetes-support.org.uk of the total sample). These threads clearly showed how the use of the OSGs helped people with T2D to manage their illness and overcome social, environmental, and intrapersonal barriers. In a number of threads, members reported how using the OSGs helped them to overcome the barriers. Members with T2D themselves posted the sampled threads and posts, not family members or friends. The first author analysed the collected threads thematically [14]. Tracy's eight criteria were followed to enhance the quality of the study [15]. Further research (i.e., interviews with OSG users) will be used to check the validity of the findings.

Prior to the data collection, the managers of the OSGs within the above organisations provided consent for undertaking the study. The study received ethics approval in accordance with the University of Sheffield Research Ethics Policy.

Results

The sample included posts from a total of 189 unique members with a mean number of 2.9 posts each. Preliminarily analyses uncovered several themes, as shown in Figure 1 below.

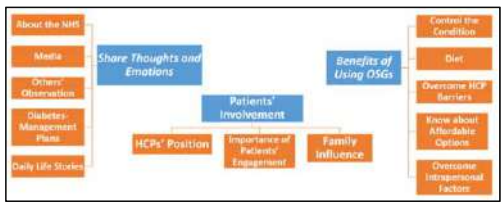


Figure 1– Map of the Themes and Sub-Themes

Members mainly benefited from the use of the OSGs to control and manage the illness, learn about dietary options, overcome barriers with healthcare professionals (HCPs), seek information about affordable options for tools to manage diabetes, and cope with intrapersonal factors (e.g., fear, anxiety, etc.). Information obtained from other members in the OSGs helped them to understand test results and set realistic plans to achieve better outcomes. Diet plans, recipes, and eating tips were widely discussed, and this clearly informed many members about eating options and how to make healthier choices. As a result, a number of members reported that the information and support they received from OSGs contributed to different elements of empowerment elements mentioned by the HCEM.

Emotions and thoughts about how members felt about the UK National Health Service (NHS), how the media was criticising the amount of spending on diabetes, and how they have been observed by others (especially employers and insurance companies) were expressed. They also discussed thoughts and plans about the management of the illness. Members also used the OSGs to share daily stories that were not necessarily related to diabetes. This was observed more in the Diabetes-support.org.uk OSG, since it is smaller in size and the members appeared to be closer to and confide more in each other.

Members had debates about whether HCPs expect them to be involved in the management of the illness, and if they are knowledgeable and capable enough to get involved in health decisions. On different occasions, they assured the importance of patients' engagement in diabetes management, especially when not receiving an effective healthcare service.

Conclusions

This study has provided a preliminary overview of ongoing research that seeks to understand the role of OSGs in empowering people with T2D in the UK. The results indicate that OSG members use the platforms to control various aspects of diabetes management. While the results answered some of the research questions, others remain unanswered; therefore, in the future, threads from the same period will be collected and analysed from a further OSG, to investigate whether these findings are present across other communities. Additionally, semi-structured interviews with current or former OSG users will be undertaken to understand in greater depth how the use of OSGs has changed their level of empowerment and to check the validity of these findings. The results of this study might help OSG moderators, HCPs, healthcare organisations, such as the NHS, friends and family members of people with T2D to understand their needs and challenges in managing the illness.

Acknowledgements

We thank the managers of the OSGs for giving permission to include these groups in the study.

References

[1] International Diabetes Federation, IDF Diabetes Atlas, 8th edn, Brussels, Belgium, (2017). Retrieved from http://www.diabetesatlas.org/.

[2] DiabetesUK, DiabetesUK Facts and Stats, (2016). Retrieved from https://www.diabetes.org.uk/Documents/Position statements/DiabetesUK_Facts_Stats_Oct16.pdf.

[3] M.M. Funnell, and R.M. Anderson, The problem with compliance in diabetes, *JAMA*. **284** (2000) 1709.

[4] L.J. Beeney, A.A. Bakry, and S.M. Dunn, Patient psychological and information needs when the diagnosis is diabetes, *Patient Educ. Couns.* **29** (1996) 109–116, doi:10.1016/0738-3991(96)00939-1

[5] J. Zhang, Supporting Information Needs of Transitional Phases in Diabetes Management Through Online Health Communities, *Companion 2017 ACM Conf. Comput. Support. Coop. Work Soc. Comput.* (2017) 107–111, doi:10.1145/3022198.3024942

[6] H.S. Wentzer, and A. Bygholm, Narratives of empowerment and compliance: Studies of communication in online patient support groups, *Int. J. Med. Inform.* **82** (2013) e386–e394, doi:10.1016/j.ijmedinf.2013.01.008.

[7] V. Zoffmann, and M. Kirkevold, Realizing empowerment in difficult diabetes care: A guided self-determination intervention, *Qual. Health Res.* **22** (2012) 103–118.

[8] E. Smailhodzic, W. Hooijsma, A. Boonstra, and D.J. Langley, Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals, *BMC Health Serv. Res.* **16** (2016) 442, doi:10.1186/s12913-016-1691-0

[9] J. James, Health Policy Brief: Patient Engagement, *Health Aff.* (2013). Retrieved from http://healthaffairs.org/healthpolicybriefs/brief_pdfs/healthpolicybrief_86.pdf.

[10] NHS, NHS England launches initiative to empower patients and communities to take control of their health, (2014). Retrieved from https://www.england.nhs.uk/2014/11/realising-value/ [accessed November 30, 2017].

[11] European Patients Forum, Patient Empowerment Campaign, (n.d.). Retrieved from http://www.eu-patient.eu/campaign/PatientsprescribE [accessed August 14, 2018].

[12] E. Chiauzzi, P. DasMahapatra, E. Cochin, M. Bunce, R. Khoury, and P. Dave, Factors in Patient Empowerment: A Survey of an Online Patient Research Network, *Patient*. **9** (2016) 511–523, doi:10.1007/s40271-016-0171-2

[13] M.O. Johnson, The shifting landscape of health care: Toward a model of health care empowerment, *Am. J. Public Health*. **101** (2011) 265–270, doi:10.2105/AJPH.2009.189829

[14] V. Braun, and V. Clarke, Using thematic analysis in psychology, *Qual. Res. Psychol.* **3** (2006) 77–101, doi:10.1191/1478088706qp063oa.

[15] S.J. Tracy, Qualitative Quality : Eight " Big-Tent " Criteria for Excellent Qualitative Research, (2010), doi:10.1177/1077800410383121

Address for correspondence

Abdulaziz Almanea
The University of Sheffield, Information School, Sheffield S1 4DP, UK. Amalmanea1@sheffield.ac.uk

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190701

Digital Literacy Program for the Use of Social Media, Aimed at Health Professionals

P Alonso Galbán[a], N Vialart Vidal[b]

[a] *Department of Special Information Services, Cuban National Center of Medical Sciences Information. Havana, Cuba.*
[b] *National School of Public Health. Havana, Cuba.*

Abstract

The WHO Strategy and Plan of Action on eHealth 2012-2017 highlights knowledge management and digital literacy as key elements for quality of care, health promotion and disease prevention, insofar as they guarantee training and better access to information in an equitable manner.

This work proposes a digital literacy program for the development of skills in social media use for information management, communication and eLearning, aimed at Cuban health professionals.

Keywords:

Health Informatics, eHealth, Information Management.

Introduction

More than 3.196 billion people use social media in the world (42%), which represents a 13% growth compared to 2017 [1]. In Cuba, the figure reaches 38% of the population [2].

The training of human resources in the use of social media in the field of health is a key element for the use of technology on the road to universal health coverage and to achieve the Sustainable Development Goals of the 2030 Agenda [3].

In the context of public health, social media favors information sharing, autonomous and networked learning, teamwork, communication, feedback, access to other related networks and contact with different experts, everything which contributes to management and collective construction of knowledge. In addition, it allows us to detect patterns and behaviors associated with the search and supply of health information.

The potential of these applications for surveillance, planning of services, development of intervention programs and health promotion are evident.

In the information society, a health professional capable of using this valuable resource in the management of information and knowledge, communication and eLearning will undoubtedly be better prepared to develop health promotion, disease prevention and medical care, based on the intensive and timely use of information and communication technology. At the same time, it will facilitate the exchange of information among professionals, academics, researchers, students and managers.

In the Cuban context, characterized by the particular importance of health issues in government public policies and, at the same time, one of the lowest rates of Internet penetration in the region (40 users per 100 inhabitants in 2017), the use of social media for health information management is a topic of particular relevance.

The WHO Strategy and Plan of Action on eHealth 2012-2017 highlights knowledge management and digital literacy as key elements for quality of care, health promotion and disease prevention, insofar as they guarantee training and better access to information in an equitable manner.

The strategy points out social media potential for the management and acquisition of new knowledge and the development of digital competencies as one of the topics to be considered in educational programs aimed at health professionals.

This work proposes a digital literacy program for the development of skills in social media use for information management, communication and eLearning, aimed at Cuban health care professionals.

Methods

A digital literacy program was designed considering the digital competencies established as necessary for the efficient use of social media in the field of health. We also considered the goals proposed in the WHO Strategy and Plan of Action on Knowledge Management and Communications [4], the WHO Strategy and Plan of Action on eHealth 2012-2017 and the PAHO IS4H project and Framework [5].

The program was designed following a human-centered design process and it was developed by four digital literacy experts from different Cuban health and academic institutions.

Results

The program consists of three postgraduate courses, one Massive Online Open Course, one blended learning course and a third in face-to-face mode, so that it can be developed according to the needs and conditions of health professionals in different contexts.

The program targets all kinds of health care professionals interested in the use of social media for health information management, communication in healthcare and eLearning, and other professionals from related sectors.

Its main objective is to develop digital skills for the efficient use of social media for information management, communication and eLearning in health, through recognition and experimentation in different social networking platforms.

Each course is organized into four thematic units covering the following content (Table 1)

The program was accredited by the University of Medical Sciences of Havana and is integrated to three master's degree programs and one diploma program. By October 2018 three blended, five face-to-face and two virtual editions had been developed, reaching 232 Cuban health professionals. Its incorporation into the self-learning resources of the Virtual Campus of Public Health of PAHO is foreseen.

Table 1

Thematic units	Contents
Internet, information and communication. From web 1.0 to web 3.0. Social media. Web 2.0 tools	History and background of networks. Internet, information and communication. Transition from web 1.0 to web 3.0. Social media. Web 2.0 tools: email lists, wikis, blogs, shared favorites, forums, websites, virtual communities of practice
Social media. Horizontal networks.	Social media: conceptual elements, emergence, characteristics and classification. Social networks in information management, communication and eLearning. Horizontal social networks: Facebook and Twitter.
Academic social networks.	Role of social media in science, scientific research and health communication. Science 2.0. Academic social networks: ResearchGate. Academia.edu, Mendeley, LinkedIn.
Infomed´s Social Networks	Infomed 2.0. Tools: discussion email lists, collaboration groups, blogs, Info links, wiki, image gallery, comments, instant messaging. Specialty and health topic websites.

Conclusions

The WHO Strategy and Plan of Action on eHealth 2012-2017 states that one of the components of eHealth is continuous education in information and communication technologies, which will facilitate its application in health. The digital literacy program that we designed can be reused for other countries that want to deploy the same kind of program. Further research about the impact of the program is needed.

References

[1] We are Social & Hootsuite, 2017 Global Digital, (2018). https://www.slideshare.net/wearesocial/2018-digital-yearbook-86862930 [accessed November 21, 2018].

[2] Social Media Stats Worldwide | StatCounter Global Stats. (2018, November 21). Social Media Stats Cuba 2017. Retrieved November 21, 2018, from http://gs.statcounter.com/social-media-stats

[3] Union Nations (2015). *Transforming our World: The 2030 Agenda for Sustainable Development.*

[4] Pan American Health Organization. (2012). *Strategy and Plan of Action on Knowledge Management and Communications.*

[5] Pan American Health Organization. *PAHO IS4H project and Framework.*

Address for correspondence

Patricia Alonso Galbán. Mailing address: Acosta Avenue No. 458a. Havana. Cuba. CP: 10500. Email: palonsogalban15121979@gmail.com. Phone: (53) 78335421. Mobile: (53) 53306645.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190702

Use of Institutional Social Media for Information Management and Communication in Healthcare in a National Health System

P Alonso Galbán[a], O Hernández Vidal[b], IR Alfonso Sánchez[c]

[a] *Department of Special Information Services, Cuban National Center of Medical Sciences Information. Havana, Cuba.*
[b] *Department of Special Information Services, Cuban National Center of Medical Sciences Information. Havana, Cuba.*
[c] *Cuban National Center of Medical Sciences Information. Havana, Cuba.*

Abstract

Social media has a growing presence in the eHealth research agenda. This research aims to characterize the use of social media in the Cuban National Health System for health information management and communication in healthcare.We specifically examine the strategy developed by Infomed and its main results.

Keywords:

Social Media, Health, Information Management

Introduction

According to the 2018 Global Digital report, published by We Are Social and Hootsuite, more than 3196 billion people use social media in the world (42%), representing a 13% growth compared to 2017 [1]. In Cuba, the number reaches 38% of the population [2].

One of the main expected trends regarding the use of Internet is the increased use of social media in the field of information search, Google search engine direct competition.

Social media has a growing presence in the eHealth research agenda. To date, their use in healthcare have been identified as a resource efficient approach to information and knowledge management and specialized communication, that could be used in many settings, and particularly where care providers and patients do not have the opportunity to have personal contact.

As part of the analysis of this new digital environment, some studies have examined social media impact on health knowledge management. In this sense, it has been pointed out the need to adapt the communication between health professionals and patients to the new scenario of the so-called electronic health, in a process that leads to the expansion of the possibilities of social interaction and the consequent empowerment of citizens for healthcare, as well as the best use of the potential of ICT [3].

In the context of public health, social media favor information sharing, autonomous and networked learning, teamwork, communication, feedback, access to other related networks and contact with different experts, everything which contributes to management and collective construction of knowledge. In addition, they allow to detect patterns and behaviors associated with the search and supply of health information. The potential of these applications for surveillance, planning of services, development of intervention programs and health promotion are evident.

In the Cuban context, characterized by the particular importance of the health issue in government public policies and, at the same time, one of the lowest rates of Internet penetration in the region -40 users per 100 inhabitants in 2017- the use of the social media for health information management is a topic of particular relevance.

Infomed is the Cuban National Health System network that coordinates all the health institutions of the country in terms of communication and information management [4]. One of its goal is that social media became channels for health information management and communication in healthcare for the entire country. Each health institution in the country is expected to develop a social media strategy that relates to its particular purpose, as well as to the goals of the Cuban national health system.

This research aims to characterize the use of social media in the Cuban National Health System for health information management and communication in healthcare.We specifically examine the strategy developed by Infomed and its main results.

Methods

A descriptive study was conducted, including all the social media profiles of Cuban health institutions, from January 2017 to October 2018. The variables used were: presence on Facebook and/or Twitter, update frequency, type of resources, topics, Facebook ranking according to LikeAlyzer [5], Twitter impact according to Twitonomy [6].

Infomed's strategy for social media was described and its main results were examined with specialized analytics tools.

Results

Ninety-seven institutional profiles were studied (including Facebook fan pages and Twitter accounts). There has been an annual increment of more than 1000 Facebook followers and about 600 in Twitter since January 2017. However, most of Cuban healthcare institutions have presence only on Facebook (64%), only 38% have also a Twitter account and 64% of them maintain a systematic update (at least once daily).

The most used types of resources were articles and pictures, mainly about news, information resources (scientific articles and books), courses and events promotion, all of this related to health. The main topics covered include hypertension, cancer, obesity, diabetes, Zika, dementia, aging, pregnancy and HIV/AIDS. 53% of Facebook profiles obtained a ranking lower than the average value for any type of page (53/100) according to LikeAlyzer. On average, 28 out of 100 tweets are redistributed, according to Twitonomy.

Infomed profiles shows a growing trend, with more than 11000 Facebook followers and 2291 in Twitter, until November 11, 2018. To date, the National Center for Medical Sciences

Information has five fan pages on Facebook and five users in Twitter**.**

The average daily range on Facebook is 86528 people. Infomed fan page rank in Facebook is 75/100, higher than the average value for any type of page (53/100), for pages of the same type -health / medicine / pharmaceutical products- (51/100) and even with respect to similar brands. (70/100).

Also, 48 out of 100 tweets are redistributed, indicating a high level of impact, committed audience and favorable impact on web statistics. According to Alexa [7], Infomed users spend about five minutes per day navigating and review four pages on average. In 2017, Infomed main website (http://www.sld.cu/) was located as the Cuban web portal with the best positioning in Alexa´s ranking. Throughout the year, it remained among the 6000 most visited portals among all Internet sites and among the top 10 most visited websites in Cuba.

The main components of Infomed social media strategy included: goals; strengths, weaknesses, threats and opportunities; audience; actions for each channel; guidelines for the design of the visual identity, formal elements and content management; interaction framework with the community; material and human resources; technological means and evaluation method.

Discussion

Social media is changing the nature of interactions in the field of health. Simple qualitative methods were used that showed increased social media use, but suboptimal use of Facebook profiles and insufficient use of Twitter within the Cuban National Health Service Network.

The analysis of the social media profiles of Cuban health institutions showed that the main difficulties were the low frequency of content updating, the scarce use of hypermedia resources, the inadequate selection of images, low interaction with users and the absence of a web content linked to the publications.

Among the studied profiles highlight the Infomed ones. Our results show that the strategy implemented by the Cuban National Center of Medical Sciences Information, which integrates the social participation spaces of Infomed with Facebook and Twitter, is a good practice. It strengthens cooperation through the structuring of information services and resources around knowledge networks and develops spaces for interaction, training, advice and discussion that favor permanent exchange among the people involved in the processes, motivating new knowledge generation and facilitating access to it.

Interaction between peers in social media is shaping a new model of knowledge management and learning, a different way of obtaining training and information, updating and developing necessary skills for professionals in healthcare sector.

Additional research is needed to further characterize the social media strategies of our health institutions and evaluate their results. Such information can be used to design interventions that improve use within a health system.

Conclusions

In the context of the Cuban National Health System, the consolidation of social media use for information management and communication in helathcare would increase access to information as a pillar of knowledge production. It would also contributed to the development of digital skills in human resources, which support the integral approach to health.

Digital social networks are considered today as the main standard of the new digital communication environment. However, this study suggests that their use in the Cuban National Health System is still incipient. The definition of institutional policies and strategies constitute the greatest challenge. The results of the implementation of Infomed´s strategy is a good example.

References

[1] We are Social & Hootsuite, 2017 Global Digital, (2018). https://www.slideshare.net/wearesocial/2018-digital-year-book-86862930 (accessed November 21, 2018).

[2] Social Media Stats Worldwide | StatCounter Global Stats. (2018, November 21). Social Media Stats Cuba 2017. Retrieved November 21, 2018, from http://gs.statcounter.com/social-media-stats

[3] R. Torres Valdés, A. Santa Soriano, Necesidad de adaptar los procesos de comunicación en salud a la Sociedad Red. Nuevas formas relacionales entre profesional sanitario y paciente, *Historia Y Comunicación Social*. **18** (2014) 651-661. doi:http://doi.org/10.5209/rev_HICS.2013.v18.44277.

[4] CNICM, Acerca De, (2018). http://www.sld.cu/acerca-de (accessed November 21, 2018).

[5] Likealyzer, (2018). https://likealyzer.com/ (accessed November 21, 2018).

[6] Twitter #analytics and much more, (2018). https://www.twitonomy.com/ (accessed November 21, 2018).

[7] A., Sld.cu Traffic, Demographics and Competitors - Alexa, (n.d.). http://www.alexa.com/siteinfo/sld.cu (accessed November 21, 2018).

Address for correspondence

Patricia Alonso Galbán.

Acosta Avenue No. 458a

Havana. Cuba. CP: 10500

Email: palonsogalban15121979@gmail.com

Phone: (53) 78335421

Mobile: (53) 53306645

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190703

Acceptance of Tele-Dental Health Education Among Head and Neck Cancer Patients in Saudi Arabia

Sarah Alradhi[a], Ahmed Albarrak[a], Meshael Alghamdi[b]

[a] Medical informatics and E-learning Unit, Medical Education Department, College of Medicine, King Saud University, Riyadh, Saudi Arabia
[b] King Fahad Medical City, Dentistry Administration, Riyadh, Saudi Arabia

Abstract

Based on the latest published cancer report by the Saudi Health Council, Saudi Cancer Registry (SCR) in 2014, the percentage of cancer incidence got higher than the previous report in 2012 by 10.2% [1]. Increasing the incidences of head and neck cancer (HNC) remains a local public health burden that provoke us to consider technology innovations to manage its serious health consequences. We aimed to evaluate patients' preception toward technology usage in reciving health care.

Keywords:

Technology Acceptance Model, Head and Neck Cancer, Saudi Arabia

Introduction

Researchers indicate that at the time of diagnosis, cancer patients retain less than 20% of the information they collect [2]. For that, dental health education needs to be warranted in the patient care routine. Effective dental health education is encouraging patients to minimize their risk factors in order to reduce and prevent severe oral complication following cancer therapy. Prevention will aid to better nutrition, enhance comfort, increase adherence to cancer therapeutic process. Moreover, prevention care will lead to better health outcomes as well as improve the quality of life [3]. The purpose of this study was to evaluate patients' preceptions of using mobile dental health education among head and neck cancer patients in Saudi Arabia.

Methods

We adopted the technology acceptance model (TAM). Since the original questionnaire was in English, we used the recommended translation process by World Health Orginzation(WHO) to translate it into Arabic. The quistionare consisted of four main sections: demogeaphic, clincial data, mobile technology using data, and TAM containing 28 items scored of 5-point likert scale measuring both PU and PEOU (14 questions for each variable). This cross-sectional study was conducted in the dental department of King Fahad Medical City tertiary care hospital in Riyadh of Saudi Arabia from March till April 2019. Targeted population were 253 head and neck cancer patients who was treated there. Data were collected by phone based intrviews.

Results

Table 1 – Demographic and Clinical Data of Patients Participating in the Study (n=253)

Varibles	No. (%)
Sex	
• Male	151(59.7)
• Female	102(40.3)
Age	
• Range	19-73
• Mean±SD	49.9±11.5
Nationality	
• Saudi	230(90.9)
• Non-Saudi	23(9.1)
Education	
• Primary	4(1.6)
• Middle	39(15.4)
• Secondary	87(34.4)
• Bachelor	115(45.5)
• Master	7(2.8)
• Doctorate	1(0.4)
Duration of disease in years	
• Range	1-18
• Mean±SD	6.4±3
Tumor type	
• Benign	1(0.4)
• Malignant	252(99.6)
Treatment type	
• Surgery	108(42.7)
• Radiotherapy	245(96.8)
• Chemotherapy	166(65.6)
Duration of radiotherapy in months	
• Min-max	0-9
• Mean±SD	4.27±1.5
Duration of chemotherapy in months	
• Min-max	0-9
• Mean±SD	4.08±2.8
Time taken in minutes between home and medical service	
• Min-max	1-900
• Mean±SD	208.9±247.4
Type of electronic device you use	
• Tablet	66(26.1)
• PC	53(20.9)
• Laptop	131(51.8)
• Mobile phone	3(1.2)
• Smartphone	248(98)
Have you ever communicated with a health care provider using telemedicine?	
• Yes	45(17.9)
• No	206(82.1)
Do you want to communicate with your healthcare provider using telemedicine tools?	
• Yes	249(99.2)
• No	2(0.8)

Results (Table 1) showed an excellent usage of technology devices among participants, 98% were using smartphones and 51.8% were laptop users. 82.1% were reported that they never communicated with healthcare providers by mobile technologies. Almost 18% were contacted with healthcare providers to reschedule their appointments or to have referrals to other departments such as speech therapy. Results also showed that 99.2% of the participants were willing to communicate with healthcare proviers using mobile dental health from home to get dental health education. The correlation between those who accepted using telemedicine and different variables such as sex, nationality and tumor type (Table 2) had no significant association with the acceptance of using telemedicine. Whereas tumor type was the only factor significantly associated with receiving health education (P-value=0.03).

Table 2 – The Correlation Between Receiving Health Education and Different Variables

Variables	Receiving dental health education n (%)/ mean±SD		P-value
	No	Yes	
Sex			
• Female	86(86)	14(14)	0.51
• Male	120(79.5)	31(20.5)	
Education			
• Secondary and less	107(83.6)	21(16.4)	0.52
• Bachelor and above	99(80.5)	24(19.5)	
Nationality			
• Saudi	188(82.1)	41(17.9)	0.92
• Non-Saudi	18(81.8)	4(18.2)	
Tumor type			
• Benign	0(1)	1(100)	0.03*
• Malignant	206(82.4)	44(17.6)	

By investigating the attitude of participants toward the importance and using dental health education program (Table 3), the range of attitude score was 52-70 with a mean ±SD= 62.3±2.96. There was a significant negative correlation between duration of disease and attitude (r=-0.148, P=0.02). Nationality and receiving education program had no significant impact on the attitude score, whereas education was significantly associated with attitude score, the higher education was significantly associated with a higher score of attitude (P-value=0.01).

Discussion

The strength of this study is that this is the first study to assess the acceptance of head and neck cancer patients toward using mobile health education in Saudi Arabia, whereas the limitations include the lack of similar studies so we couldn't compare our findings with others.

Table 3 – Correlation Between Attitude and Different Variables

Variables		Mean attitute score	Std. Deviation	P-value
Nationality	Saudi	123.48	3.08013	0.989
	Non-Saudi	123.5	2.83816	
Education	Lower Education	**122.6**	**6.38967**	**0.017**
	Higher Education	**124.4**	**5.21762**	
Receiving Education	Yes	123.6	5.86741	0.490
	No	122.9	6.07678	

Conclusions

There was acceptance among patients to communicate with health care providers by using smartphones, and they had a positive attitude toward mobile dental health education.

References

[1] S. Bazarbashi, H. Al Eid, and J. Minguet, Cancer Incidence in Saudi Arabia: 2012 Data from the Saudi Cancer Registry, *Asian Pac J Cancer Prev* **18** (2017), 2437-2444.

[2] N.M. McLeod, N.R. Saeed, and E.A. Ali, Oral cancer: delays in referral and diagnosis persist, *Br Dent J* **198** (2005), 681-684.

[3] J.S. Chang, H.I. Lo, T.Y. Wong, C.C. Huang, W.T. Lee, S.T. Tsai, K.C. Chen, C.J. Yen, Y.H. Wu, W.T. Hsueh, M.W. Yang, S.Y. Wu, K.Y. Chang, J.Y. Chang, C.Y. Ou, Y.H. Wang, Y.L. Weng, H.C. Yang, F.T. Wang, C.L. Lin, J.S. Huang, and J.R. Hsiao, Investigating the association between oral hygiene and head and neck cancer, *Oral Oncol* **49** (2013), 1010-1017.

Address for Correspondence

Corresponding Sarah Alradhi at either email or mobile number noted below:

Sarah.alradhi@gmail.com

Mobile #: +966544449898

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190704

Exploring the User Engagement Scale Short Form as a Determinant of Adherence in Digital Health Interventions

Andre Q Andrade[a], Alline MR Beleigoli[b,c,d], Tiago MS Silva[e], Maria de Fátima H Diniz[c,d], Antonio LP Ribeiro[c,d]

[a] *Quality Use of Medicine Pharmacy Research Centre, School of Pharmacy and Medical Sciences, University of South Australia, Adelaide, South Australia, Australia,*
[b] *Flinders Digital Health Centre, Flinders University, Adelaide, South Australia, Australia*
[c] *Internal Medicine Department, Faculty of Medicine, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil*
[d] *Telehealth Center of the Hospital das Clínicas of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil*
[e] *The Octalysis Group, Central, Hong Kong SAR*

Abstract

Adherence determines the impact of digital health interventions. Standard tools provide a measure for user experience and predict adherence. We evaluated the User Engagement Scale Short Form (UES-SF) during the POEmaS project, a randomized clinical trial of an online weight loss platform. We received answers from 178 participants (13.7% of the cohort) and correlated the UES-SF scores with the number of sessions attended. Our findings suggest the UES-SF is an accurate evaluation of user experience, but only one domain (reward) was associated with long-term use.

Keywords:
Patient participation, telemedicine, clinical trial

Introduction

Lack of user engagement and adherence with digital health tools is a common issue in large-scale online interventions. The User Engagement Scale (UES) is a tool that helps evaluate the main determinants of adherence. Initially developed for e-commerce, social media, and search situations, the UES is based on general psychological constructs and can be used in other domains [1]. The UES has a short form (UES-SF) [2] which is composed of 12 questions divided between 4 domains: perceived usability (PU), aesthetic appeal (AE), focused attention (FA), and reward (RW). The RW domain is a summary of three domains from the original UES: endurability, a measure of how successful the interaction was and the likelihood of recommending the application to others; novelty, a measure of curiosity and interest; and felt involvement, a measure of the feeling of being "drawn in" and having fun [2].

We evaluated the applicability of the UES-SF to the digital health domain. We explored the use of the UES-SF to evaluate how users perceive a digital health intervention for weight loss and to understand which facets of the UES are most predictive of long-term adherence.

Methods

The Online Platform for Healthy Weight Loss (POEmaS) project has been extensively described elsewhere [3]. It was a randomized controlled trial that investigated the effectiveness of an online platform which delivered a 24-week behavior change program on weight loss and lifestyle habits. The yearlong trial (September 2017 to October 2018) enrolled 1298 eligible participants (≥18 years of age, BMI ≥ 25kg/m^2, not pregnant or undergoing other treatments for weight loss) who were students or staff of the Universidade Federal de Minas Gerais (UFMG) in Brazil. The project was approved by the UFMG Ethics Committee (CAAE: 73545717.5.0000.5149) and is registered as NCT03435445.

Participants were randomly allocated at a 1:1:1 ratio into three parallel arms: 1) those given access to the web-based platform; 2) those given access to the platform plus online coaching with a registered dietitian; and 3) the waiting list group, which received access to the platform after six months. The behavior change intervention was similar for all groups [4].

Prompts for answering the UES-SF were submitted to all participants by email. Users in groups 1 and 2 (platform access) were prompted at the end of the program (six months). Users assigned to the waiting list were prompted one week after gaining access to the full platform in order to evaluate whether the scale's predictive capacity changes over time. All questionnaires were answered electronically at a webpage built specifically for the purpose and were not part of the normal usage session.

Answers were analyzed after the end of the trial. We determined Pearson's correlation coefficients between long-term engagement and the score for each domain separately, as well as the overall score. We also used Student's t-tests to compare the mean UES-SF score in each domain between users who only used the platform for one day and users who returned at least once; we did the same to compare users who attended more than seven sessions with those who did fewer than seven sessions. Statistical analysis was performed using the SciPy package v1.1.0.

Results

The questionnaire was offered to all 1298 participants of the POEmaS project. We received 178 (13.7%) responses. Responders were older than non-responders but had similar baseline body mass indices (BMI). The responder group had significantly more female users. The clinical trial arm was not determinant for whether the user answered the UES-SF questionnaire. The comparisons between the responder group and the non-responder group can be seen in Table 1.

Table 1 – Baseline Characteristics of UES-SF Responders and Non-Responders

	Non-responders	Responders	p-value
Female (%)	847 (75.6%)	149 (83.7%)	0.023
Mean age	33.35	35.27	0.020
Mean baseline BMI	29.96	29.46	0.15
Waiting list (%)	349 (31.1%)	59 (33.1%)	0.79
Total	1120	178	

The platform was highly rated by the users in the four domains of the UES-SF. The response averages can be seen in Figure 1. Platform evaluation was not significantly different between participants in different trial arms.

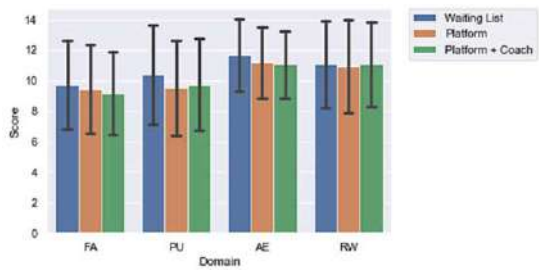


Figure 1 – Average UES-SF scores, stratified by domain and trial arm (mean and standard deviation shown). FA=focused attention, PU=perceived usability, AE=aesthetic appeal, RW=reward

The overall UES-SF score was positively associated with engagement measures. When considering the different domains, only the RW domain retained statistical significance (Table 2).

Table 2 – Association Between UES-SF Scores and the Total Number of Sessions, Stratified by Domain

	Pearson correlation coefficient	p-value
UES (all domains)	0.1864	0.013
Focused attention	0.1202	0.110
Perceived usability	0.0493	0.514
Aesthetic appeal	-0.0111	0.883
Reward	0.3011	<0.001

UES-SF scores stratified by domain between users who used the platform for only one day and users who returned at least once were significantly different for the FA, AE and RW domains (Table 3). When comparing between users who attended more than seven sessions and those who did fewer than seven sessions, only the differences in the RW domain retained statistical significance.

Table 3 – Association Between UES-SF Scores and Short-Term Engagement, Stratified by Domain

	1 session	> 1 session	p-value
Number of users	41	137	
FA (mean)	8.24	9.8	0.002
PU (mean)	9.81	10.17	0.524
AE (mean)	10.61	11.51	0.029
RW (mean)	9.76	11.39	0.001

Conclusions

The UES-SF was useful as an engagement evaluation for digital health interventions. Among the four dimensions explored in the questionnaire (reward, perceived usability, aesthetic appeal and focused attention), all but perceived usability are associated with short term engagement. However, only the reward dimension is associated with long term engagement.

Acknowledgements

This work was partially supported by CNPq and Fapemig. Alline M Beleigoli was funded by Programa de Pós Doutorado from CAPES. Prof. Antonio Ribeiro and Alline M Beleigoli are members of the National Institute of Science and Technology for Health Technology Assessment (IATS).

References

[1] H.L. O'Brien and E.G. Toms, Examining the generalizability of the User Engagement Scale (UES) in exploratory search, *Information Processing & Management* **49** (2013), 1092-1107.

[2] H.L. O'Brien, P. Cairns, and M. Hall, A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form, *International Journal of Human-Computer Studies* **112** (2018), 28-39.

[3] A.M. Beleigoli, A.Q. Andrade, M.d.F. Haueisen Diniz, R.S. Alvares, and A.L. Ribeiro, Online platform for healthy weight loss in adults with overweight and obesity - the "POEmaS" project: a randomized controlled trial, *BMC Public Health* **18** (2018), 945.

[4] A.M. Beleigoli, A.Q. Andrade, M.F.H. Diniz, R.S. Alvares, M.H. Ferreira, L.A. Silva, M.C. Rodrigues, L. Jacomassi, A.G. Cerqueira, and A.L. Ribeiro, Using the Behaviour Change Wheel for Designing an Online Platform for Healthy Weight Loss - "POEmaS", *Stud Health Technol Inform* **254** (2018), 1-10.

Address for correspondence

Andre Q Andrade, Quality Use of Medicines and Pharmacy Research Centre, School of Pharmacy and Medical Sciences, University of South Australia - GPO Box 2471, Adelaide, SA 5001

Phone: +61 8 8302 2314 | Email: andre.andrade@unisa.edu.au

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190705

Postgraduate Studies in Digital Health (eHealth): Developing a Blended-Learning Model and Real-Life Spaces

Raul Antônio Morais Melo[a], Patrícia Smith Cavalcante[b]

[a] *Medical Sciences, University of Pernambuco, Recife, PE, Brazil*
[b] *Center of Education, Federal University of Pernambuco, Recife, PE, Brazil*

Abstract

Digital Health (e-Health) is the use of the technologies of information and communication in the health area. We report the design and implementation of a course in e-Health for multi-professional postgraduate students. It was based on two key ideas: a blended-learning model and real-life spaces. The methodological triangulation approach included the educational planning of a course. In this work, we present and demonstrate the feasibility to use a blended-learning model to teach e-Health to postgraduate students based on interactive distributed learning spaces.

Keywords: *health education, learning, informatics.*

Introduction

Digital Health (e-Health) is the cost-effective and secure use of information and communications technologies in support of health and health-related fields, including health-care services, health surveillance, health literature, and health education, knowledge and research [1].

Most health professionals use information technology daily in their work, but few know how to adapt their roles and work processes to incorporate it for the most significant benefit [2]. In this respect, blended learning (b-learning) might be more suitable for health care training because of the need to combine hands-on, skills-based training at a practical level, as well as self–directed learning [3,4]. We report the development of a course in digital health, which combined both a blended-learning model and real-life spaces.

Methods

The course is an elective, open discipline, offered to post-graduate students attending the master's or doctorate on Mathematics and Technological Education (Edumatec) at the Center of Education of the Federal University of Pernambuco (UFPE) in Recife, Brazil. We, the teachers, used a design-thinking approach for the educational course planning with brainstorming and canvas map techniques, and analyses of future learning spaces and scenarios. The basic structure of the course considered the content (academic and real-life), the learning spaces (physical and virtual), and strategies (scenarios and activities).

The educational contents were digital health history and resources; and concepts, characteristics, and landmarks in education. The real-life themes included e-Health applied to the electronic health record, accreditation in the health area, innovation in health, telehealth, and development of personal competencies for health workers. All real-life contents were studied physically through visits plus chats with health or educational workers.

The physical learning spaces were the UFPE physical post-graduate classroom, a primary public health unit, a tertiary private health unit, a school of innovation, a telehealth unit, and a symposium attendance. The online learning spaces were an online classroom and a social media network. We used a freeware Learning Management System - LMS (MoodleCloud) on the Internet, office productivity programs (Microsoft Office, H5P) and interactions in social media software (Telegram), because of their availability and usability [5-8].

Three interdependent and simultaneous tracks of learning scenarios were envisaged: physical and real life (regular classroom and guided visits), online and interactive (virtual classroom and social media network) and innovative production (hackathon). As evaluation criteria for the course, we used pre- and post-tests, weekly task accomplishment, the Hackathon solution, and the degree of interaction of the participants.

Results

Thirteen students and two teachers were enrolled in the discipline for eight weeks. As a result, a thirty-hour course was developed in a b-learning model using an LMS platform, software, and social media. A myriad of activities were planned to better consolidate learning, both in real and virtual scenarios (**Table 1**). A hackathon with a real health problem was also created so that the students could follow the various phases of problem presentation, ideation, construction, validation, implementation, and solution presentation.

Table 1 - Structure and Timeline of the Course on Digital Health.

Structure / Week	0	1	2	3	4	5	6	7
Planning	Individual work Meetings	Classroom work Agreement	Guided visit Hackaton Interaction	Guided visit Hackaton Interaction	Guided visit Hackaton Interaction	Guided visit Hackaton Interaction	Guided visit Hackaton Interaction	Classroom work Products
Contents	Syllabus LMS Social Media Hackaton	Education Lanmarks Digital Health	Problem & ideation Electronic health registry	Ideation Accrediation in the health area	Construction Health innovation	Validation Telehealth	Implementation Compentencies	Presentations
Scenarios	Univerity	University	Primary public health unit	Tertiary private health unit	Technology park	Telehealth nucleus	Telehealth nucleus	University
Tools	Software	Software LMS Social media	Software LMS Social media	Software LMS Social media	Software LMS Social media	Software LMS Social media	Software LMS Social media	Software LMS Social media
Activites	Brainstorming Canvas Test Tool construction E-mailing	Conversatin circle Presentation	Visual documentation Brainstorming Bibliographic search	Guided visit Interview Wiki Canvas Biliographic search	Guided visit Interview Wiki Canvas Biliographic search	Symposium Wiki Bibliographic search	Guided visit Video confernece Wiki Bibliographic search	Concept maps Wiki Pitch
Resources	Personal experiences Scientific literature	Visual libraries Links Publications	Visual libraries Links Publications	Visual libraries Links Publications	Links to videos Visual libraries Publications	Presentations Virtual libraries Links Publications	Virtual libraries Links Publications	Virtual libraries Links Publications
Assessment	Academic approval Announcements	Pre-test Assignments	Assignments	Assignments	Assignments	Assignments	Assignments	Post-test

LMS= Learning Management System. Wiki= collaborative editing tool of the LMS.

Conclusion

We demonstrate the feasibility to design and implement a course in digital health for postgraduate studies, based on real and virtual complimentary interactive and innovative learning scenarios.

References

[1] W. W. Stead, J. R. Searle, H. E. Fessler, J. W. Smith, and E. H. Shortliffe, Biomedical informatics: changing what physicians need to know and how they learn, *Academic Medicine*, **86** (2011), 429-434.

[2] K. Rasmussen, J. M. Belisario, P.A. Wark, J.A. Molina, S. L. Loong, Z. Cotic, N. Papachristou, E. Riboli–Sasco, L. T. Car, E. M. Musulanov, H. Kunz, Y. Zhang, P. P. George, B. H. Heng, E. L. Wheeler, N. A. Shorbaji, I. Svab, R. Atun, A. Majeed, and J. Car, Offline eLearning for undergraduates in health professions: a systematic review of the impact on knowledge, skills, attitudes and satisfaction, *Journal of Global Health,* **4** (2014), 1-18.

[3] World Health Organization – WHO [homepage on internet]. *World Health Assembly Resolution WHA58.28.2005.* Retrieved from: http://apps.who.int/gb/ebwha/pdf_files/WHA58-REC1/english/A58_2005_REC1-en.pdf. [accessed March 18, 2019].

[4] J. McCarthy, Blended learning environments: using social networking sites to enhance the first-year experience, *Australasian Journal of Educational Technology*, **26** (2010), 729-740.

[5] MoodleCloud – [homepage on internet]. Retrieved from: https://moodlecloud.com. [accessed March 18, 2019].

[6] Microsoft – [homepage on internet]. Retrieved from: https://www.microsoft.com/pt-br. [accessed March 18, 2019].

[7] H5P – [homepage on internet]. Retrieved from: https://h5p.org. [accessed March 18, 2019].

[8] Telegram – [homepage on internet]. Retrieved from: https://telegram.org. [accessed March 18, 2019].

Address for correspondence

Prof. Raul Antônio Morais Melo – Faculty of Medical Sciences - University of Pernambuco – Rua Arnóbio Marques, 310 – Santo Amaro – Recife – PE – 50100.130 - Brazil – Telephone: (+558131841261) – E-mail: raul.melo@upe.br

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190706

Workforce Development Strategy for Health Information System Implementation at the Public Health System of Buenos Aires

Jacquelina Celis[a], Analía Baum[a], María Victoria Giussi Bordoni[a], Leandro Alassia[a], Adriana Stieben[a], Mariano Franco[a], Fernán Gónzalez Bernaldo de Quirós[b]

[a] *Health Information Systems Office, Ministry of Health, Buenos Aires City, Argentina,*
[b] *Government Health Advisor, Buenos Aires City, Argentina*

Abstract

One of the challenges of implementing health information and communication technology is the need for a skilled workforce that understands health care and ITC. There are also people and organizational challenges involved. This work describes the strategies to create such a workforce for the public network of healthcare facilities in the City of Buenos Aires, which include promoting the adoption of technologies and providing lines of contention for continuous training.

Keywords:

Organizational Innovation, Capacity Building, Health Information System

Introduction

Developing and managing workforces to carry out Health Information System (HIS) implementations is key and challenging. Having trained professionals in the discipline that do not meet the required task force demands when it comes to regional implementations [5] increases the challenge.

The Ministry of Health of the Autonomous City of Buenos Aires started the implementation of the Health Information System two years ago that involved 114 effectors [2]. This led to the creation of different working teams to carry out the implementation, provide support in the transformation processes, and ensure the adoption of new technologies during the management period 2015-2019 [4].

This paper describes the strategies designed and used for the conformation, growth, and consolidation of the working teams implementing HIS in the public health system of the Autonomous City of Buenos Aires.

Methods

Setting

The Buenos Aires city healthcare network is conformed by a Ministry of Health, 114 centers, structured into 12 geographical areas to organize health care delivery. The healthcare system has a total of 41,000 employees. Since June 2016, an Electronic Health Record (EHR) has been gradually implemented in the outpatient setting [2; 4]. When the Ministry of Health decided to implement the EHR, a Special Project Unit (SPU) was created and the necessary manpower was assigned to carry it out.

Study Design

This is a descriptive study about the designed strategies that were used for the conformation, growth, and consolidation of the working teams that are in charge of implementing EHR within the City of Buenos Aires. The description has been approached from four axes:

1. Talent Management: The challenge of digital transformation becomes the challenge of selecting professionals with adequate digital competences and digital culture [6]. In this sense, and considering that health systems are understood as complex systems that consist of dynamic and complicated networks of interconnected actors [1], it becomes relevant to form multidisciplinary working teams.
2. Communication and Internal Training: The working team in charge of carrying out the implementation becomes the prolocutor for organization that works on agile methodologies. The team must be trained in change planning processes, have the ability to establish a sense of urgency, form the coalition of a local team, empower people, provide short-term results and collaborate to consolidate the introduced changes [3; 7].
3. Contention lines: Training and communication programs are contention mechanisms for the working teams that can also foresee the demands that may be received by end users. Digital knowledge, information management, communication, networking, continuous learning, the strategic vision, network leadership and the user's orientation, are the thematic axes on which the training of the work team should focus.
4. Geographical distribution: The demographic distribution of the effectors has been laid out within an area of 204 Km². This distribution is a variable to be taken into account when it comes to organizing and distributing HIS implementation teams, especially if the implementation happens to be simultaneous or exponential.

Results

Talent Management

The SPU EHR is a team that combines disciplines, organized in 5 different roles and functions (Table 1).

Table 1– SPU EHR Team Profile

Role	Function	Discipline (N)
Director	Establish Strategic guidelines; coordinate with Project Leaders	Healthcare Informatics (5)
Project Leader (PL)	Establish tactical guidelines; develop the gantt and track the project	Physicians (2), Sociologist (1), Political Scientist (1), Communicator (1), Education (1), Inter-disciplinary Residency Program in Health Information Systems (*IRHIS*) (15)
Coordinator	Operate the strategic lines; manage field work	Healthcare professionals (16)
Implementer	Perform survey and implmentation tasks	Advanced Health Science students (65)
Auditor	Ensure the pattern's quality	Professionals of Administrative Careers (8)

Due to the shortage of this professional profile, the IRHIS was developed to create a training program based on in-service training. Currently, the IRHIS has 15 active professionals (physicians, psychologists, sociologists dentist, anthropologists,) who have been assigned the PL role.

The approach of all actors requires the involvement of other disciplines, ensuring the convergence of different perspectives and proving suitable solutions for end users, understanding their interests, needs and idiosyncrasies.

Communication and Internal Training

The SPU EHR designed a three-stage training process for the implementers. This training allows collaborators to pass down HIS implementation values, as well as teaching the use of their applications and finally provide certain soft tools to manage change.

Project Coordinators and Project Leaders receive their training through similar post-graduate courses such as project management programs, AMIA's 10x10 program and a master's degree in health informatics.

On the other hand, Coordinators and PL meet every two weeks to share both advances and unexpected consequences in order to generate consensus and find alternatives to continue with the implementation plan.

Contention Lines

In order to support the training of collaborators, a series of instruments as face-to-face and virtual trainings, an internal campus and meetings, were established to facilitate access to knowledge.

Likewise, prior to the launch of a new functionality in the EHR, every member of the working team must undergo virtual training designed for end users and attend every required specific training meeting.

Finally, given the complexity implied in introducing changes in health organizations and considering that implementers are those who usually receive the most unmet demands, the Project Management promoted dialogue and contention meetings with a team of experts in coaching. In these spaces, we sought to capitalize on the collaborators' experiences and provide useful soft tools to manage change within the field.

Geographical distribution

Geographical distribution of health effectors as well as the committed goals influence the way grouping and collaboration work among coordinators when simultaneously implementing several effectors. There are two main implementation scopes (primary healthcare facilities and hospitals) which are crossed by several transversal projects (contingency, printing Health records, Digital Signature, etc.)

The PL team works with the coordinator to integrate the implementation of transversal projects to the deployment of EHR positions assumed as government commitment.

Conclusions

Digital transformation does not have to be thought as the transformation of the organization. Instead, it has to be thought as the transformation of the people that are part of it. Thus, the talent management of the working team that carries out the implementation of HIS is key when adopting change. Even though the SPU EHR has put a lot of emphasis on different strategies for the development of high-performance teams, there are still evaluation instances of internal processes to be developed.

References

[1] A.a. Baum and G. García, Resistiendo: Gestión del cambio en los Sistemas de Información en Salud, in, Amazon Digital Services LLC, 2018.

[2] M.V.G. Bordoni, F. Plazzotta, A. Baum, C.F. Ilc, and F.G.B.d. Quirós, Elaboración e Implementación de una Agenda Digital en Atención Primaria en la Ciudad Autonoma de Buenos Aires, in: *XV Congresso Brasileiro de Informática em Saúde*, Goiâna, Brazil, 2016.

[3] M. Evaluation, Guide to a Team Approach to Building Capacity for Health Information Management, in, MEASURE Evaluation, 2017.

[4] M.V. Giussi, A. Baum, F. Plazzotta, P. Muguerza, and F.G.B.d. Quirós, Change Management Strategies: Transforming a Difficult Implementation into a Successful One, *Studies in health technology and informatics* **245** (2017), 813-817.

[5] W. Hersh, A. Margolis, F.n. Quirós, and P. Otero, Building a Health Informatics Workforce in Developing Countries, *Health Affairs* **29** (2010), 274-277.

[6] C. Magro, J. Salvatella, M. Álvarez, O. Herrero, A. Paredes, and G. Vélez, Cultura digital y transformación de las organizaciones: 8 Competencias Digitales para el Éxito Profesional, in, RocaSalvatella, Barcelona, 2014.

[7] J. O'Connor and J. Seymour, *Cómo descubrir y emplear la excelencia para obtener óptimos resultados personales y profesionales*, Urano, 1992.

Address for correspondence

Jacquelina Celis
Ministry of Health, Buenos Aires City, Argentina.
E-mail: jcelis@buenosaires.gob.ar
Phone Number: 01149585748 Int: 1502

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190707

The EFMI Working Group "Healthcare Informatics for Interregional Cooperation": An Evolving Strategy for Building Cooperation Bridges

Arriel Benis [a, c], Mihaela Crisan-Vida [b, c], Lăcrămioara Stoicu-Tivadar [b, c]

[a] *Faculty of Technology Management, Holon Institute of Technology, Israel*
[b] *Department of Automation and Applied Informatics, Politehnica University Timisoara, Timisoara, Romania*
[c] *European Federation for Medical Informatics, Working Group "Healthcare Informatics for Interregional Cooperation*

Abstract

The Working Group "Health Informatics for Interregional Cooperation" (WG HIIC) of the European Federation for Medical Informatics (EFMI) is dedicated to develop, to implement and to disseminate a strategy for promoting exchange of information, knowledge and experiences all around the world and more particularly, between Health Informatics arena players in the different European continent regions.

Keywords:

Europe, Social Participation, International Cooperation

Introduction

The European Federation for Medical Informatics Association (EFMI) is in Europe, since 1976 [1], is the leading organization in the field with more than 30 partner-countries affiliated societies. Among the main objectives of EFMI, a central one is promoting co-operation inside and outside Europe by supporting dissemination of academic and professional education, knowledge, standards and experience.

For matching with these goals, in a continuously changing scientifically, technologically and regulatory contexts, EFMI has workgroups (WGs) focusing on how to investigate, to disseminate and to promote a large range of sub-specialties of Medical Informatics by taking into account cultural differences. EFMI WG "Health Informatics for Interregional Cooperation" (HIIC) has been created continuing the activity of the EFMI WG Medical Informatics for Countries in Transition (MICIT) chaired until 2003 by Prof. George Mihals and from 2003 by Prof. Lacramioara Stoicu-Tivadar who proposed to EFMI Council the new WG in 2007. Since 2018 a new chair – Dr. Arriel Benis - and co-chair – Dr. Mihaela Crisan-Vida took the lead of the WG and will continue the activities in a dynamic manner for the next 4 years.

Objectives

- The EFMI WG HIIC activities consist of promoting exchange of information, knowledge and experiences all around the world and more particularly, between Health Informatics arena players in the different European continent regions. For addressing these, the EFMI WG HIIC is over the years developing and improving a dedicated continually evolving strategy [2].
- The main objectives of the EFMI WG HIIC are [3]:
  - to promote exchange of information and experiences between actors in different regions in Europe;
  - to investigate the needs, opportunities and obstacles for e-health and to review and select from different education options for developing regions;
  - to disseminate European and world-wide results and experiences across regions and between professionals;
  - to facilitate access to European groups and their facilities and outcomes by students and health professionals from developing regions;
  - to disseminate European and world-wide results and experiences across developing regions and professionals.

Implementation

During the last decade, the EFMI WG HIIC has been involved in different projects and missions, for example,

- to support the development and the implementation of the Electronic Health Record in Europe;
- to motivate national societies to join EFMI as members and also its WGs;
- to support the organization and the community involvement in the Medical Informatics Europe (MIE) conferences;
- to organize EFMI Special Topic Conferences (STCs) and workshops allowing researchers and practitioners sharing their work in a specific sub-domain [4].

Several of these EFMI WG HIIC actions during 2008 - 2017 were:

- 2008: in April the WG chair conducted a mission of evaluation and support the EHR Serbia, with participation in The Conference «Electronic Health Record, options for the Future», Sava Center, Belgrade, April 17 (from EFMI Prof. Rolf Engelbrecht, EAR, MoH Serbia); participation in MIE 2008 Goteborg and in July Organizing the Workshop "Integration of Electronic Health Records" National / International and Technical Aspects, 24-25 July; in September participation in STC 2008, London.
- 2009: during MIE 2009 in Sarajevo HIIC organized the the Workshop "Is There a Common Background to Support Better Healthcare in Central and South East Europe?" - to assess healthcare informatics stage for 6

countries in Central and East European area, and the real need for cooperation and its driving forces [5];

- 2011: HIIC chaired EFMI-STC in Laško, Slovenia, e-Health Across Borders Without Boundaries with prof. Lacramioara Stoicu-Tivadar as SPC chair and prof. Bernd Blobel and Tomaz Marcun co-chairs and editors [6].
- 2014: EFMI WG HIIC organized together with EFMI Primary Care WG the STC 2014 in Budapest, Hungary, dealing with cross-border challenges in informatics with a focus on disease surveillance and utilization of big-data, editing the volume Cross-Border Challenges in Informatics with a Focus on Disease Surveillance and Utilising Big Data [7]
- 2015: the WG supported the program and participation in EFMI MIE2015, Madrid, Spain; mediating EFMI-IMIA WGs meeting during MIE 2015; in April participation in EU-USA project Trillium bridge workshop; preparing projects for consolidating cross border cooperation with neighbours, Hungary (Szeged) and Serbia (Vrsac) in e-health projects under INTERREG frame.
- 2016: participation at STC 2016, Paris, France with scientific contributions; participation in EFMI MIE 2016, Munich, Germany and organizing the workshop on regional projects and low cost solutions for financially challenged regions: Sustainability of Telemedicine Projects for Resource Challenged Regions having as invited speaker the regional projects expert Michael Weinhara.
- to develop active cooperation at the national level between EFMI country (e.g. Romania / Hungary / Serbia since 2009), academic institutions in different regions of Europe (e.g. "Politehnica" University of Timisoara in Romania and the Holon Institute of Technology in Israel since 2017) in medical informatics education and e-health projects [8].

The results of the previous years show that interregional cooperation in the European health informatics arena is active and gradually moving from focused interested to broad one which combine academic and industrial researches and higher eduction training. The EFMI WG HIIC focused, focus and will focus on common problems that extend beyond national borders such as health, legal, political, economic and cultural factors.

Lessons and future outcomings

In the nowadays, connected world the EFMI WG HIIC is proud of the previously accomplishments.

From these experiences, the group has pointed-out new needs and challenges for taking to the next level, the interregional cooperations based on medical and health informatics projects.

An important part of the EFMI WG HIIC strategy consists for the next years to increase the numbers of its members, their academic and professional background and country of origin. Accordingly, it will allow the WG to be involved in more projects, such as taskforces, and so will be able to support them by sharing experience and expertise in running multinational and interregional projects.

The virtualization of the workspace is a huge opportunity for establishing scientific, technical, educational and mostly human active and productive partnerships in the European landscape and further.

As some examples of future outcomings, we can mention: collateral involvements in health informatics courses between academic institutions in EFMI partner-countries and the development and implementation of a collaborative and multilingual dictionary focusing on Health and Medical Informatics.

References

[1] Mihalas G., Zvarova J., Kulikowski C., Ball M., van Bemmel J., Hasman A., Masic I., Whitehouse D., Barber B., History of medical informatics in Europe - a short review by different approach. Acta Informatica Medica, 22(1) (2014), 6-10.

[2] Mihalas G., Evolution of trends in European medical informatics. Acta Informatica Medica, 22(1) (2014), 37-43.

[3] European Federation of Medical Informatics, HIIC - Health Informatics for Interregional Cooperation, https://www.efmi.org/workinggroups/hiic-health-informatics-for-interregional-cooperation

[4] Cornet, R., Stoicu-Tivadar, L., Hörbst, A., Parra Calderón, C.L., Andersen, S.K., Hercigonja-Szekeres, M., Digital Healthcare Empowering Europeans, Studies in Health Technology and Informatics, 210 (2015)

[5] Stoicu-Tivadar L., Blobel B., Kern J., Masic I, Mihalas G., Orel A., Surjan G., Urosevic V., Medical Informatics in a United and Healthy Europe, Series Studies in Health Technology and Informatics, 150 (2009), Open Access funded by EFMI and IMIA, pp. 1002-1006

[6] Stoicu-Tivadar L., Blobel B., Marčun T., Orel A., e-Health Across Borders Without Boundaries, Studies in Health Technology and Informatics,165 (2011)

[7] Stoicu-Tivadar L., de Lusignan S., Orel A., Engelbrecht R., Surján G., Cross-Border Challenges in Informatics with a Focus on Disease Surveillance and Utilising Big Data, Studies in Health Technology and Informatics, 197 (2014)

[8] Memorandum of Understanding signed between Politehnica University Timisoara, Romania and HIT Retrieved November 8, 2018, from https://www.hit.ac.il/en/news/news-and-stories/Romania_Agreement

Address for correspondence

Dr. Arriel Benis, Faculty of Technology Management, Holon Institute of Technology – HIT, Golomb St 52, Holon, 5810201, Israel

e-mail: arrielb@hit.ac.il

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190708

Development of a Research-Based Teaching Course as Blended-Learning Format in a Medical Informatics Program

Nils-Hendrik Benning, Petra Knaup

Institute of Medical Biometry and Informatics, Heidelberg University, Germany

Abstract

We consider Medical Informatics programs at universities as one of the main education resources for young scientists in our field and thus present a new design for a course teaching scientific skill at the University of Heidelberg as blended-learning format. We utilize common E-learning methods and created the whole course with respect to the concept of research-based teaching. Finally, we present our lessons learned from the current activities of the course.

Keywords:

Education, Distance, teaching, medical informatics.

Introduction

Medical informatics is a scientific discipline with variety and evolving research topics [1]. Therefore, an important aim of academic medical informatics programs is to allow students to participate in scientific work. In the medical informatics bachelor program at Heidelberg/Heilbronn (Germany), the students have a compulsory practical course where all participating students work as a research team together on an up-to-date scientific question of medical informatics.

Academic teaching in Germany is mostly organized by conducting weekly classroom instruction. This format is not so adequate for a large group of students who are working together on a project. There are times in the course of the project where intensive instruction and feedback in short intervals is necessary and there are phases in which the learners can work independently for longer periods of time. Therefore, we developed a blended-learning concept for practical training that implements research-based learning.

Blended-learning formats are used rarely (18 % of German university teachers used such formats regularly in 2017 [2]). Therefore, a process model on how to establish such a course was not available.

Methods

We used existing work from Salmon [3,4] to implement an active and social Virtual **L**earning Environment (VLE). The overall design of the VLE was created following the concepts of research-based learning by Healey and Jenkins [5] and the guidelines for research-based courses from Sonntag et al. [6]. For the implementation of the VLE, an existing instance of the open-source learning management system *Ilias* (version 5.3.7) was used. To measure the students learning success and the general satisfaction with the course, an evaluation was conducted at the end of the winter term 2018 / 2019. For this evaluation a German questionnaire was created, which utilizes the F-Komp-model from Böttcher and Thiel [7], a questionnaire that specifically addresses the evaluation of research-oriented courses. The items from the F-Komp-model were combined with the standard evaluation form of the university's quality management department.

Results

The blended-learning concept covers a 15-week term and is suited for a course which grants 6 ECTS credit points. The course is organised in weeks which include classroom teaching and exclusive E-learning weeks. The sequence of these types of weeks has to be flexible to a certain extent to allow adjustments throughout the term (see Table 1).

Structure of the VLE

During all weeks of the course the VLE is used to offer different resources, like literature, tutorials or formative tests, to the course members. The course is structured at two levels: on the first level the phases of research-based learning from Huber (see *Figure 1*) are represented as blocks in the VLE, which clearly distinguish different phases of the course from each other.

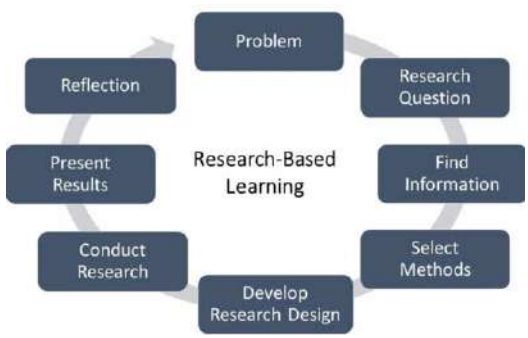


Figure 1— Research Cycle (cf. [7])

On the second level, each block uses expandable container objects for each course week. Each week consists of an overview and the actual content. The overview summarizes if classroom teaching takes place, which tasks should be completed, and which resources are needed to complete them. The content consists on the one hand of the aforementioned resources and on the other hand of interactive online-forums. The forums are used for multiple purposes: for each classroom teaching there is a forum, which shall be used to discuss organizational and content-related matters like the absence of participants or summaries of classroom discussions. Other forums represent the so-called E-tivities. These forums are an activating way of assigning tasks to learners, which not just motivates them to work on the task, but also fosters collaboration in a VLE [4]. This collaboration is supported by a teacher all the time. According to Salmon's principles of moderating, the teacher takes the role of an E-moderator, which is an important and different from the known role of university teachers [3].

The described structuring results in a course layout that is shown in Table 1.

Table 1 – General course layout

Week	Teaching Type	Content
1	E-learning	Problem: overall subject
2	Classroom	Problem: specific problems
3	Classroom	Research Question
4	Classroom	Find Information; Select Methods
5	Classroom	Develop Research Design
6-12	*Flexible*	Conduct Research
13	*Flexible*	Present Results: preparation
14	Classroom	Present Results: trial presentation
15	Classroom	Present Results; Reflection

Course Layout

The first week is used to assign the learners with reading fundamental literature to get to know to the subject matter. To make sure everyone arrived at the VLE and worked on the assignment, each learner has to write a short message introducing himself and take a short formative test about the literature that has to be passed. This introduction to the VLE results in an online brainstorming for problems in the defined research field. In week two, the preparations from the first week are used in the classroom to define the problems the learners want to work on. Week three is dedicated to the definition of the research question. To enable the learners to select the aims of their research, methodical skills are conveyed in the classroom, before the learners are divided into groups of 4 to 6 people. These groups are supported individually afterwards. The next two weeks (4 and 5) cover a deeper look in the field and the selection of research methods, which are used to create the research design for the coming weeks. These last preparing parts of the research cycle take place in the classroom. Now the learners should be ready to conduct their own research project. In this extensive part of the research cycle most teaching can be done in the E-learning setting, as the learners are expected to work independently. Classroom teaching is recommended in cases where the E-moderator finds a need for face to face motivation or discussion. The last three weeks take care of the preparation and presentation of the results. First, the preparation consists of defining requirements for the presentation, which can be done either in classroom or in the VLE. The trial presentation and the final presentation should be done in the classroom to create a setting, which is comparable to presentations at scientific events like conferences. Finally, the learners should be supported by talking about their collaboration within the course.

Implementation

We implemented the course according to the introduced concept and 18 advanced students who participate currently in our medical informatics bachelor program. The students work in the field of patient participation and consider especially applying components supporting this paradigm. Selected research questions are: “which requirements exist for application components that support patient participation?” and “which software solutions and frameworks are currently available, to support patient participation?”.

Conclusions

Generally, we found a higher workload for teachers of a blended-learning course, because the contact time with learners is not limited to fixed appointments in a work week but open for contact whenever it is needed. Furthermore, the configuration and preparation of the E-learning environment is more demanding than the preparation of materials for classroom teaching. In return, it is possible to achieve a much closer mentoring of the learners while they go through the research cycle. As this is the first time the course takes place, we expect a reduced workload for future terms, because it was implemented as a general concept, which can be instantiated by each year’s individual medical informatics research question.

With a focus on optimized results of the research projects, the time constraint of a university term limited to 15 weeks is challenging especially for the first phases of defining a research question and creating a research design that has a varying demand of time. The VLE helped us to closely accompany the students in this crucial phase of the project.

When term’s lecture period ended in January we evaluated the course with the tools mentioned. The evaluation results will be presented on the poster and we will use the results to revise the concept for the next group of students.

Acknowledgements

The course was developed in the context of the HiGHmed working package teaching and training, which is funded by the German Federal Ministry of Education and Research (grant id: 01ZZ1802A) as part of the Medical Informatics Initiative. [9]

References

[1] R. Haux, C.A. Kulikowski, S. Bakken, S. de Lusignan, M. Kimura, S. Koch, J. Mantas, V. Maojo, M. Marschollek, F. Martin-Sanchez, A. Moen, H.A. Park, I.N. Sarkar, T.Y. Leong and A.T. McCray, Research Strategies for Biomedical and Health Informatics. Some Thought-provoking and Critical Proposals to Encourage Scientific Debate on the Nature of Good Research in Medical Informatics, *Methods Inf Med* **56** (2017), e1-e10.

[2] MMB-Institut für Medien- und Kompetenzforschung, In welchem Rahmen setzen Sie digitale Medien für Ihre Veranstaltungen ein?, (2016). https://de.statista.com/statistik/daten/studie/733647/umfrage/einsatzarten-digitaler-medien-an-hochschulen-in-deutschland/ (accessed November 13, 2018)

[3] G. Salmon, *E-moderating: the key to online teaching and learning*, Routledge, New York, 2011.

[4] G. Salmon, *E-tivities: the key to active online learning*, Routledge, New York, 2013.

[5] M. Healey and A. Jenkins, *Developing undergraduate research and inquiry*, The Higher Education Academy, Heslington, 2009.

[6] M. Sonntag, J. Reuß, C. Ebert, K. Friederici and W. Deicke, *Forschendes Lernen im Seminar. Ein Leitfaden für Lehrende.*, Humboldt-Universität zu Berlin, Berlin, 2016.

[7] F. Böttcher and F. Thiel, Evaluating research-oriented teaching: a new instrument to assess university students’ research competences, *Higher Education* **75** (2018), 91-110.

[8] L. Huber, *Forschendes Lernen im Studium: aktuelle Konzepte und Erfahrungen*, UVW Univ.-Verl. Webler, Bielefeld, 2009.

[9] B. Haarbrandt, B. Schreiweis, S. Rey, U. Sax, S. Scheithauer, O. Rienhoff, P. Knaup-Gregori, U. Bavendiek, C. Dieterich, B. Brors, I. Kraus, C.M. Thoms, D. Jager, V. Ellenrieder, B. Bergh, R. Yahyapour, R. Eils, HiGHmed Consortium and M. Marschollek, HiGHmed - An Open Platform Approach to Enhance Care and Research across Institutional Boundaries, *Methods Inf Med* **57** (2018), e66-e81.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190709

A Taxonomy for Survey Measures Used to Evaluate Complex Digital Health Innovations

Tim Benson

R-Outcomes Ltd; UCL Institute of Health Informatics

Abstract

Evaluators need to measure whether innovations help patients and staff, but have lacked the tools needed to do this as part of routine care. We provide a taxonomy for the classification of survey measures, which can be used together on a pick and mix basis. These are described in the context of the evaluation of digital health innovations.

Keywords:

Healthcare evaluation mechanisms; Outcome and process assessment (health care); Healthcare quality, access and evaluation; Complexity; Patient-centred care.

Introduction

Many digital health innovations have been abandoned, failed to be adopted, scale-up, spread or be sustained [1]. We need the tools to understand not just what happens but also why.

Most patient-reported outcome and experience measures (PROMs and PREMs) address a single dimension and were developed in isolation.

This paper presents a taxonomy covering twenty-one short generic measures, which form a coherent family covering the patient needs, treatment outcomes, patient experience, other factors and key aspects of digital innovation.

These measures are designed for digital data collection and results reporting using mobile devices. This taxonomy can help people understand the role of each measure and any gaps or overlaps.

Methods

These measures were all developed in a similar way.[2]

- Recognition of the need for a new measure
- Review of existing literature and brainstorming
- Initial proposals for discussion with potential stakeholders (patients, clinicians, commissioners)
- Iteration, fine-tuning and revision
- Testing and revision
- Validation studies

Results

We created a taxonomy of measures, which may be completed by patients, staff and carers, covering a range of perceptions about their health needs, treatments, experience, social factors and innovation (see Figure 1).

Needs

The **howRu** health status measure is a short generic patient-reported outcome measure (PROM) to track and compare patients' perceptions of how they feel physically and mentally and what they can and cannot do (disability and dependence) [2].

Figure 1 Taxonomy of measures

The **Personal Wellbeing Score** (PWS) covers life satisfaction, worthwhileness (eudomonia) and positive (happiness) and negative (anxiety) affect. It is based on the Office of National Statistics ONS4 [3].

The **Health Confidence Score** (HCS) captures people's confidence in their knowledge of their health and treatment, ability ot self-manage their health, access the right help if they mneed it and shared decision-making [4].

Social isolation and **loneliness** are important determinants of health and wellbeing. People need companionship, someone to confide in, people to help them and do things with.

Sleep hygiene is another important determinant of health and wellbeing.

Treatments

Self-care, including diet, physical activity, weight and medication management.

Acceptance of loss, how people have learnt to live with events, including recognition of capabilities and change, how to do things differently and to move on with life.

Medication adherence including remembering to take medications, following instructions given side-effects or recovery and satisfaction.

Assessment of need (**howRthey**) by staff of people with cognitive loss about how much support people need

Experience

HowRwe is a short generic patient-reported experience measure (PREM), which measures their perception of the the care and service provided by one service [5].

The **Service integration** measure captures patients perception of how well different services collaborate. Perceived quality across services is often lower than that within each service.

The **Shared Decision-Making** (SDM) measure covers several aspects of how involved patients have been in making health care choices together.

Social Factors

Social determinants of health, such as education, self-esteem, housing and poverty play an large role in determining peoples' health [6]. The Social Factors measure covers these.

Neighbours, community cohesion and social capital are impacted by how well people know, trust and help each other.

Innovations

The following measures were designed for use in digital health innovation evaluation [6].

The **Digital Confidence Score** assesses people's confidence in using digital apps and similar devices. Many patients are old and frail and are on the wrong side of the digital divide.

The **Innovation Readiness** measure where people and their organisations fall on the innovativeness spectrum (innovator, early adopter through to laggard), covering whether they think new ideas are needed, their up-to-dateness, organisational support and capability.

The **Innovation Process** measure is based on Normailisation Process Theory (NPT) in terms of how innovations are implemented.

The **User Satisfaction** measure focuses on peoples perception of whether an innovation is useful, is easy to use, they can get help and their overall satisfaction.

Behaviour change, based on the COM-B model covering capability, opportunity and motivation to change behaviour [7].

Discussion

We have developed a taxonomy with five main branches, covering patient needs, treatment outcomes, experience of care, other factors and innovation-specific aspects. This taxonomy covers 21 separate measures, many of which have variants for patients, staff and carers. This structure is more complex and sophisticated than the traditional division between patient-rported outcome measures (PROMs) and patient-reported experience measures (PREMs).

This family of measures shares a common look and feel with four items and four responses. They are research-based and validated, quick and easy to use with simple unambiguous wording and low reading age, readily understood by thoise whose first language is not English. They are generic, suitable for almost all situations, irrespective of case mix across health and social care. They are channel-neutral and can be used on paper, smart-phone, tablet, PC or via text message or voice.

Any practical implementation needs to give detailed consideration to (a) selection of measures, (b) the data collection process, (c) data visualisation of results, and (d) how the results are used to improve services.

The results show trends, changes and comparisons. These are quick and easy to interpret, giving immediate feedback to all stakeholders tracking changes and differences between units.

For reporting and visualising results, a common approach is used. High is always good. Different items may be aggregated. Mean scores for groups are always reported on a 0-100 scale, with 100 indicating that all respondent chose the best option and 0 that all chose the least desirable option.

These measures were developed to meet the need for a coherent family of short generic surveys for use in evaluation and as key performance indicators of services. They are characterised by brevity and low reading age. Most other measures have been developed as academic research projects resulting in relatively longer measures, whith higher reading age and little compatibility between measures.

These measures have been used with success in evaluations of new care models, social prescribing, care homes, digital health applications, such as for diabetes self-care and AF detection.

Conclusions

This paper provides a taxonomy with five branches covering patient needs, treatment outcomes, experience of care, other factors and innovation-specific aspects. This taxonomy covers 21 separate measures, which have been used together in a pick and mix way in a wide variety of evaluation studies.

References

[1] Greenhalgh T et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Non-adoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. *J Med Internet Res* 2017; 19 (11): e367

[2] Benson T, Whatling J, Arikan S, et al. Evaluation of a new short generic measure of HRQoL: howRu. *Informatics in Primary Care* 2010; 18: 89-101.

[3] Benson T, Sladen J, Liles A, Potts HWW. Personal Wellbeing Score (PWS)—a short version of ONS4: development and validation in social prescribing. *BMJ Open Quality* 2019; 8: e000394.

[4] Benson T, Potts HWW, Bowman C. Development and validation of a short health confidence score. *Value in Health* 2016; 19 (3): A94.

[5] Benson T, Potts HWW. A short generic patient experience questionnaire: howRwe development and validation. *BMC Health Services Research* 2014; 14:499.

[6] Marmot M. *The health gap: the challenge of an unequal world*. London: Bloomsbury 2015.

[7] Benson T. Digital innovation evaluation: user perceptions of innovation readiness, digital confidence, innovation adoption, user experience and behaviour change. *BMJ Health & Care Informatics* 2019;**26:**doi: 10.1136/bmjhci-2019-000018.

Address for correspondence

Email: tim.benson@r-outcomes.com

Web: https://r-outcomes.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190710

Developing an Intuitive Intelligent Inpatient Medical Record System

Chia-Jung Chen[a], Hsien-Yi Wang[b], June-Dong Lin[a], Chung-Feng Liu[*cd]

[a] Department of Information Systems, Chi Mei Medical Center, Tainan, Taiwan
[b] Department of Medical Information Management, Chi Mei Medical Center, Tainan, Taiwan
*[*c] Department of Information Management, Chia Nan University of Pharmacy and Science, Tainan, Taiwan*
[d] Medical Research Department, Chi Mei Medical Center, Tainan, Taiwan

Abstract

Chimei Medical Center developed an intuitive intelligent inpatient medical record system which integrates structured, unstructured, and textual medical records and provides health insurance payment suggestions from the collaborative medical record writer's point of view. Combined with the use of intelligent technology (Python language with the smart Content Difference Recognition software component), the system can ensure medical record quality by focusing on the differencse in recorded progress notes.

Keywords:
Medical records, intelligence, insurance

Introduction

The medical record is an important basis for the diagnosis and treatment of patients and also reflects the quality of medical care in a hospital. Medical records document the patient's condition and the medical behaviors of the medical team members. For example, progress notes in a medical record are where healthcare professionals record details to document a patient's clinical status or achievements during the course of a hospitalization [1]. However, physicians may just copy and paste past records (known as *DITTO*) with no change or very few changes while writing notes. Therefore, using the hospital's existing data, including structured, unstructured, and textual data, combined with the use of intelligent technology such as Natural Language Processing (NLP), could free physicians from cumbersome medical records by considering a collaborative medical record writer's point of view and focusing on the differences in medical recordings to achieve the ultimate performance of collaboration within teams.

NLP is one kind of artificial intelligence (AI) technology that concerns text and speech recognition and differentiation [2]. NLP techniques are suitable for processing health record identification (e.g.[3]). In this study we introduced a system using NLP concepts to improve medical record writing in Chimei Medical Center, Taiwan. The top programming language today for data scientific exploration and development is Python [4], and Python and associated smart data analysis components are suited for processing large numbers of medical records. In this paper, we describe how an inpatient intelligent medical system was developed with Python and the AI functions of text recognition and differentiation.

Methods

JCAHO points out that human factors analysis is an essential step to designing equipment, procedures, tasks, and work environment in order to support human strengths and mitigate human weaknesses while studying patient safety and medical quality (5). Thus, the concept of human factors analysis should be kept in mind while developing the intuitive intelligent inpatient medical record system.

Chimei used the Python language and AI-based content difference recognition technology to develop a prototype of an intelligent medical record system that integrates physician orders, medical records, disease coding (ICD-10), and cost estimation (Figure 1). It provides a more convenient and smart way for physicians to write medical records. In addition, the system also displays medical record coding recommendations and medical expenses to the physician for immediate reference. This system uses the Web Flask architecture so that the user can clearly obtain the required information without installing any software components.

For smooth development, we will complete a three-stage test before going online: the first stage tested the availability of information systems with representative users (seven physician members of the Medical Record Management Committee); the second stage tested the system with seed personnel from each unit to eliminate imperfections (ten physicians from five departments); the third stage will be a full user-side introduction and system on-line installation for optimizing and adjustment to ensure the availability, integrity, and ease-of-use of the system (still under voluntary use).

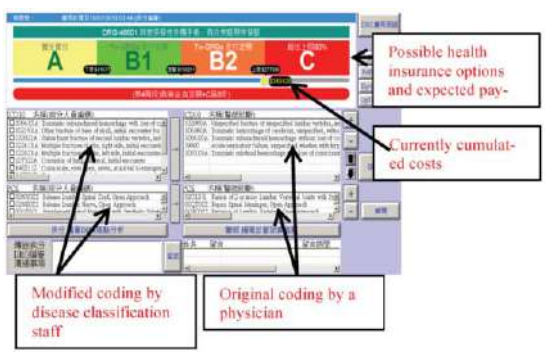


Figure 1- Integrated medical record writing with physician's orders, medical records, and disease coding as well as cost pre-calculation.

Results

For ease of operation, the system provides an intuitive visual graphical interface displaying instant interpretations of Taiwan Diagnosis Related Groups (TW-DRG) to achieve diagnostic accuracy. The design feature of the Problem-Oriented Medical Record (POMR) [6] in the system allows physicians to instantly integrate disease diagnosis with medical records to ensure consistency in disease diagnosis and medical record writing. When a physician completes a progress note and presses the "SAVE" button, the system triggers a difference comparison between the past and present records and then shows differences in colored text (Figure 2). If the difference from the past is less than 30% the record will not be saved, forcing physicians to modify the content. In addition, the system also provides cost pre-calculation for possible health insurance alternatives for physicians while considering a final decision.

Based on this rule, we computed that within nearly 40,000 progress notes recorded from May to July of 2007, the system reminded the medical record makers to correct or optimize medical records in about 1 in 10 records.

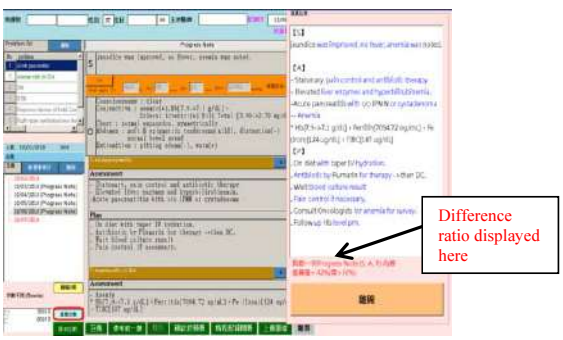


Figure 2- Medical record difference ratio displayed in red.

Evaluation

The system was derived from the viewpoint of the medical record management department with the aim of improving the quality of medical recording, and then improving medical care quality. Seven physician members of the Medical Record Management Committee and ten seed personnel of each unit were invited to pilot test this system. In general, they were satisfied with the system regarding its usefulness and ease-of-use (means=4.25, 3.88 of two five-scale questions). The core opinions are listed below:

1. The difference-rate of 30% is high for some specific departments such as pediatrics because there is little difference in patients' daily conditions.
2. The system did not work properly on old computers with Win XP OS. This problem was solved by the IS Department with Web Service technology.
3. In the initial stage, the system uses a "warning" message to remind the physicians rather than forcing the physicians to obey. More experience should be collected and discussed to decide the official launch time.

Discussion and Conclusions

Combining intelligent technology in medical record writing can improve the quality of medical records and writing efficiency. After a one-month pilot trail, this intelligent medical record system has been fully launched except for a few departments (e.g. the psychiatry department). At present, Chimei Medical Center is developing the expansion for Chinese medical records, and will try to introduce machine learning technology in the future to develop the function of optimizing the coding of medical records.

Undoubtedly, AI applications are revolutionizing how the health sector works to reduce spending and improve care quality [7]. The intuitive intelligent inpatient medical record system can be regarded as the touchstone of Chimei Hospital's AI developmemt. At present, Chimei Hospital is actively developing other AI applications. We believe that Taiwan's development of AI will be very much in line with Taiwan's excellent medical service and information communication technology.

Acknowledgements

The authors would like to thank Chimei Medical Center for funding this study (CMHCR10728). The authors also declare no conflicts of interest.

References

[1] Progress Notes, Wikipedia, Visited (11/11/2018)
[2] Natural language processing (NLP), Wikipedia, Visited (3/31/2019)
[3] N. Afzal, V.P. Mallipeddi, S. Sohn, H. Liu, R. Chaudhry, C.G. Scott, I.J. Kullo, and A.M. Arruda-Olson , Natural language processing of clinical notes for identification of critical limb ischemia, Int J Med Inform 111 (2008), 83-89. doi: 10.1016/j.ijmedinf.2017.12.024.
[4] R. Gall, Here's Why Python Is The Top Programming Language For Big Data, SmartData Collective, 2018.
[5] The Joint Commission on Accreditation of Healthcare Organizations (JCAHO), Human Factors Analysis in Patient Safety Systems, The Source 13 (2015), 7-10.
[6] L.L. Weed , The Problem Orientated Record as a basic tool in medical education, patient care and clinical research, Ann Clin Res 3 (1971), 131-4.
[7] Forbes Insights, AI and Healthcare: A Giant Opportunity, Feb 11, 2019.

Address for correspondence

Chung-Feng Liu (fredliu@mail.cnu.edu.tw)

Department of Information Management, Chia Nan University of Pharmacy and Science

No.60, Sec.1, Erren Rd., Rende Dist., Tainan City 71710, Taiwan (R.O.C.)

TEL: +886-6-2664911 ext 5301 FAX:+886-6-3660607

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190711

"D/C the CC (Carbon Copy)" - Improving the EHR Signal-to-Noise Ratio for Clinicians by Selective Feature De-Implementation

Ling Chu[a,b], Seth M Toomay[c], Cameron S Ginnings[b], Kyle G Gabriel[b], Ethan A Halm[a], Duwayne L Willett[a,b]

[a] *Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas, USA*
[b] *Health System Information Resources Department, University of Texas Southwestern Medical Center, Dallas, Texas, USA*
[c] *Department of Radiology, University of Texas Southwestern Medical Center, Dallas, Texas, USA,*

Abstract

De-implementation of a 10-year EHR configuration resulted in over 50% decrease in the volume of the most-common InBasket message type received by PCPs. Pro-actively seeking out ways to not only (a) implement helpful new EHR features but (b) de-implement detrimental ones offers an opportunity to accelerate improvement in the S/N ratio and reduce clinician frustration and dissatisfaction with the EHR. Balancing governance decision agendas with de-implementation opportunities can enhance the clinician experience.

Keywords:

Electronic Health Records; Burnout, Professional; Cognitive Science

Introduction

Anticipated clinician benefits of migrating from paper medical records to an electronic health record (EHR) stubbornly continue to lag behind expectations. Initial high hopes for EHRs making clinical work more efficient, clinical decision-making more streamlined, and seamless access to information across systems gave way to unintended consequences of decreased productivity, increased documentation burden, and information overload. While physician burnout is certainly multifactorial, the concept of "cognitive load" is now seen to play an important role [1].

The term "signal-to-noise" (S/N) ratio originated in the field of electrical engineering and can be defined as the ratio of the power of a desired electrical signal to the level of background interference or noise. While this phrase's origins in the deterministic world of electrical engineering may seem far afield from the highly variable environment of modern clinical care and the psychology of physician cognition and burnout, important similarities emerge. Both electrical signal-to-noise and the psychological cognitive load concept consider two information types: relevant to the situation at hand ("signal") vs. irrelevant ("noise"). Accordingly, improvement strategies can involve increasing relevant information (boosting signal), decreasing irrelevant information (dampening noise), or both. In our local EHR, we sought an opportunity to measurably reduce the S/N ratio's denominator and improve clinician experience.

Methods

Software:

The authors' institution has used Epic (Verona, WI, USA) for its electronic health record (EHR) continuously since 2001 in its clinics and since 2009 across its hospitals. Clinician patterns of use of the EHR are available for review via an online portal (signal.epic.com), with appropriate security rights. EHR data is extracted via Epic's Clarity reporting database into a Data Warehouse (SQL Server, Microsoft, Redmond, WA, USA) for analysis. For this report, data analysis and visualization of EHR data were performed using PowerBI (Office 365 suite, Microsoft).

Procedures:

Project initiation:

One author (LC) noted that available data on patterns of EHR use by clinicians at our organization revealed that General Internal Medicine (GIM) physicians were receiving three times as many "CC Chart" (Carbon Copy of Chart to other than the primary recipient) messages in their InBasket as specialists in the same organization. By inspection, it appeared that GIM was receiving CC Charts from all specialist visits, not just ones with information meant for the GIM physician. In discussions with fellow GIM clinicians, consensus emerged that many CC Chart InBasket messages were extraneous and irrelevant, and wading through them had proven burdensome.

Generating support:

The opportunity for reduction in InBasket overload was brought to the GIM Division Chief, who strongly supported addressing the root cause of the problem, then to the Ambulatory Chief Medical Officer, who further encouraged looking at the issue Health System wide.

Investigation:

The lead physician (LC) worked with the EHR team (CG and colleagues) to assess the sources of the high volume of CC Chart messages.

Results

Results of Investigation:

In a review of 3 years of InBasket messages received by PCPs (in GIM and Family Medicine), CC Chart messages were found

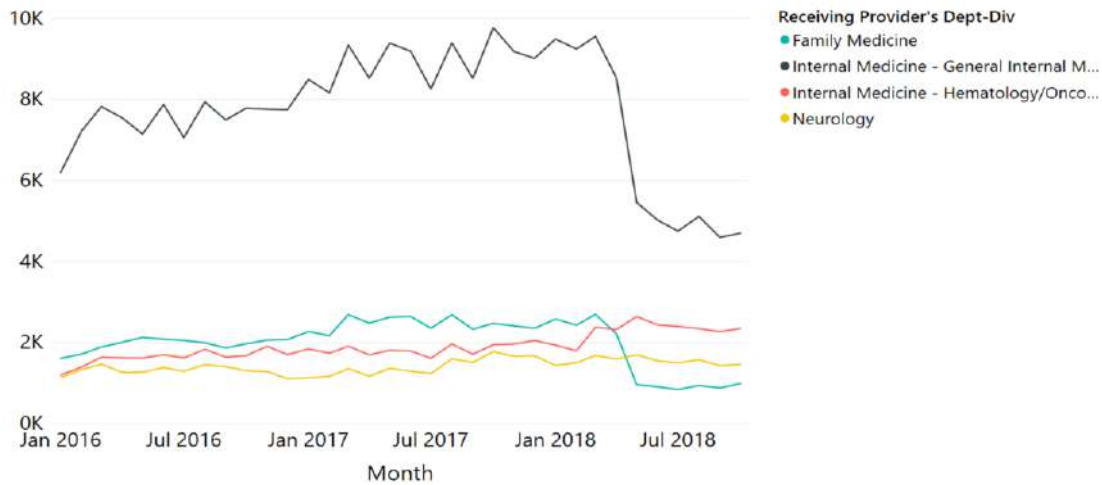


Figure 1: CC Chart InBasket message volume by Department-Division, thru Sept 2018

to be the largest volume message by a considerable margin: 332K total CC Chart messages received by these primary care physicians, followed by 256K for Results messages, and 234K for Patient Call messages, with other message types lower still.

Approximately 50% of CC Chart messages were found to be due to a programming point that automatically sends a copy of each specialist's clinic note to the Primary Care Physician on file when the encounter is closed. Further investigation revealed this practice had started early in the ambulatory roll-out of the EHR at our institution (>10 years prior), then propagated to over 50 clinic-specific settings over the years, covering the vast majority of specialist clinic physicians. Consensus was reached among physician leadership and the practicing clinicians—GIM and specialists alike—to abandon this practice and use only intentional CC Chart routing. The EHR team located where each automated programming point was set, and on a designated date removed them all. To mitigate concern for any unintended consequences from PCPs being unaware of key specialist visits, specialists remained able to "CC" the PCP on any individual note deemed important for PCP awareness.

Results of Intervention:

Prior to the EHR intervention, the frequency of CC Chart messages received by GIM physicians was growing steadily, exceeding 9,500 CC Chart messages per month (Figure 1). After the intervention, this frequency was cut in half, to under 5,000 messages per month. Though a smaller practice, Family Medicine physicians saw an even larger relative decrease (60%) in CC Chart messages, from approximately 2,500 per month to less than 1,000 per month. Non-primary care Departments-Divisions served as controls and saw no contemporaneous decrease in their CC Chart messages.

Discussion

De-implementation of a 10-year EHR configuration resulted in over 50% decrease in the volume of the most-common InBasket message type received by PCPs. Pro-actively seeking out ways to not only (a) implement helpful new EHR features but (b) de-implement detrimental ones offers an opportunity to accelerate improvement in the S/N ratio and reduce clinician frustration and dissatisfaction with the EHR [2]. This perspective may hold implications for local EHR governance committees, where agendas often overwhelmingly contain only proposed new features. Balancing governance decision agendas with de-implementation opportunities can enhance the clinician experience.

Conclusions

An initiative to reduce the denominator of the EHR "signal-to-noise ratio" for primary care clinicians succeeded in reducing unwanted instances of their most common type of InBasket message by 50-60%. De-implementing unhelpful old EHR features can measurably lessen clinicians' cognitive load, and complements adding helpful new EHR features. Studies of varied additional de-implementation initiatives are underway.

Acknowledgements

Funding provided by the Health System Information Resources Department, UT Southwestern Medical Center.

References

[1] E. Harry, R.G. Pierce, P. Kneeland, G. Huang, J. Stein, and J. Sweller, Cognitive Load and Its Implications for Health Care, in: NEJM Catalyst, https://catalyst.nejm.org/cognitive-load-theory-implications-health-care/; NEJM Group, Massachusetts Medical Society, 2018.

[2] M. Ashton, Getting Rid of Stupid Stuff, N Engl J Med 379 (2018), 1789-1791.

Address for correspondence

Ling Chu, MD, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390, USA
E-mail: Ling.Chu@UTSouthwestern.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190712

Improving Education of Medical Students Through Telehealth

Magdala de Araújo Novaes[a,b], Amadeu Sá de Campos Filho[a,b], Paula Rejane Beserra Diniz[a,b]

[a] *Telehealth Center, Clinics Hospital, Federal University of Pernambuco, Recife, Pernambuco, Brazil*
[b] *Department of Internal Medicine, Federal University of Pernambuco, Recife, Pernambuco, Brazil*

Abstract

Telehealth can be used in medical education to promote teaching-service integration as a low-cost solution to broaden the field of medical practice, especially in rural areas of low and middle-income countries. This article presents our experience in the use of telehealth to broaden the practice field of medical students. In 2017, two itinerants' actions were carried out in municipalities in the state of Pernambuco. In both actions, 396 patients were assisted by a team of professionals, teachers and students in various specialties, with emphasis on mental health and dermatology. 9 students experimented with the use of an electronic patient record system, telehealth platform and a mobile application for tracking of disorders and diseases. The students reported high satisfaction with digital distance practices, and the enrichment of their learning. This digital approach to medical education has fostered greater collaboration among students, faculty and staff, teaching students the skills necessary for their future digital practices.

Keywords:

Medical education; Telehealth; Digital Health

Introduction

Health has changed faster than other areas of knowledge and in this scenario, preparing health professionals to succeed in their careers is the greatest challenge for universities. One strategy to improve students' learning is to broaden their field of practice by learning from other health realities, but for this to happen it is necessary to ensure supervision of these students by university teachers. We present our experience with collaborative interprofessional practice at a distance, in which students participate in an itinerant health service provided by the Telehealth Centers Network of Pernambuco [1] connected to the university through the telehealth service.

Methods

Students of the Medicine Course were invited to participate in field actions carried out by the university Telehealth Center. These actions aim to assist patients who were waiting for medical attention in vulnerable municipalities. In 2017, 9 students participated in two itinerant actions in small towns in the rural area of the state of Pernambuco. The students and staff were previously trained and had access to notebooks, tablets, diagnostic equipment, an electronic health record system (EHR) and a telehealth platform.

Results

In the first itinerant action, 3 medical students participated, where 335 patients were attended in the areas of medical clinic, ophthalmology, cardiology (consultation and ECG test) and mental health. The students used the EHR systems and performed mobile device screening in 26 elderly patients (Figure 1).

Figure 1 - Medical student screening mental disorders in elderly patients during the itinerant health service event.

Figure 2 - Mobile data collection interface for telehealth platform tracking.

In the second one, 6 medical students participated in an itinerant action in Dermatology, where 61 consultations were carried out and educational activities in skin cancer prevention were performed for patients who were waiting. Skin exams were performed and those altered were launched on the telehealth platform for evaluation and discussion by experts and students.

In both action students and staffs analyzed the data. Exams results for ECG and Dermatoscopy were returned to patients through the primary care units delivered by the internet using the telehealth platform (Figure 3).

Figure 3 - Telehealth Platform interface used by students to connect with teams and access test results.

Conclusion

The use of telehealth as a strategy to strengthen the teaching-learning process is being implemented at our university. The students who participated in the telepathic-supported activities were reported to have high satisfaction, had access to a greater number of patients in different health situations, including the digital practices that they would not have if they were only in the hospital unit, and this enriched their training. Since 2003, the Telehealth Center, in partnership with the Health Informatics Discipline of the Medicine Course, has made efforts to insert telehealth in the pedagogical project of the Medicine Course. So far, these insertions have been carried out in a non-systematic way, but from 2017, with the good results by itinerant actions, the high level of satisfaction of the students who participated, followed by evidence in the literature of the benefits of distance health practices [2; 3], we believe we are on the right path for the future.

Acknowledgments

To Brazilian Health Ministry by financial support. Rede de Núcleos de Telessaúde de Pernambuco (RedeNUTES) and the Brazilian Telehealth Program.

References

[1] Diniz PRB, Sales FJR, Novaes MA. Providing telehealth services to a public primary care network: the experience of Rede NUTES in Pernambuco, Brazil. Telemedicine and e-Health. 2016;22(8):694-8.

[2] L. S. van Galen, C. J. Wang, P. W. B. Nanayakkara, K. Paranjape, M. H. H. Kramer & J. Car (2018) Telehealth requires expansion of physicians' communication competencies training, Medical Teacher, DOI: 10.1080/0142159X.2018.1481284

[3] Edirippulige, Sisira, et al. "It's important, but not important enough: eHealth as a curriculum priority in medical education in Australia." Journal of telemedicine and telecare 24.10 (2018): 697-702.

Address for correspondence

Prof. Magdala de Araújo Novaes

Núcleo de Telessaúde, Hospital das Clínicas, 2º andar, UFPE, Av. Prof. Moraes Rego CEP: 50670-420 CDU - Recife, PE, Brazil. +55 81 2126-3903, magdala.novaes@nutes.ufpe.br

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190713

ApiAppS: A Project to Study and Help Practitioners in Recommending mHealth Apps and Devices to Their Patients

Jean-Charles Dufour[a], Julien Grosjean[b,c], Stefan Darmoni[b,c], Mobin Yasini[d], Guillaume Marchand[d], Christian Simon[e], Aline Sarradon-Eck[f], Marie Préau[g], David Darmon[h,i], Matthieu Schuers[j], Parina Hassanaly[i], Roch Giorgi[a]

[a] Aix Marseille Univ, APHM, INSERM, IRD, SESSTIM, Hop Timone, BioSTIC, Marseille, France
[b] Department of Biomedical Informatics, Rouen University Hospital, Rouen, France
[c] Sorbonne Université, UMR_S 1142, LIMICS, Paris, France
[d] Department of Research and Development, DMD Santé, Paris, France
[e] SILK Informatique, Angers, France
[f] Aix Marseille Univ, Institut Paoli-Calmettes, INSERM, IRD, SESSTIM, CanBios, Marseille, France
[g] Social Psychology Research Group (GRePS), Lyon 2 University, Lyon, France
[h] Université Côte d'Azur, Département d'enseignement et de recherche en médecine générale, Nice, France
[i] Aix Marseille Univ, INSERM, IRD, SESSTIM, Marseille, France
[j] Department of General Practice, Department of Biomedical Informatics, Rouen University, Rouen University Hospital, LITIS EA 4108, Normandie University, Rouen, France

Abstract

The ApiAppS ongoing project aims to provide physicians with a decision support system for the prescription / recommendation of mHealth technologies. We describe the context and the components of the project which includes: 1) a technical part on modelling and implementing the decision support system, and 2) a psychosocial investigation part designed to have a better knowledge of general practitioners (GPs) and patients' expectations, beliefs and practices.

Keywords:

Mobile health, Computer Systems, Health comunication.

Introduction

Mobile Health (mHealth) can play an effective role in various domains affecting individual's health conditions (well-being, prevention, management, monitoring or follow-up of known or suspected pathologies). The use of mHealth (apps or devices) can reduce the number of medical visits, increase convenience for the patient while minimizing healthcare system cost [1]. However, several factors hinder this potential: diversity and plethora of frequently confusing products; relative variability and/or relevance of a solution depending on the context of its use; difficulties in assessing its qualities and added values; multitude of covered areas (health and / or wellbeing); and diversity of users/patients characteristics (clinical and paraclinic data, lifestyle, individual preferences). Furthermore, there is a lack of implementing knowledge of the psycho-social aspects/effects related to the arrival of this new technological third-party in the patient-doctor relationship [2].

As well as medical prescriptions, general practitioners (GPs) have a crucial role to play in ensuring that the use of mHealth can really be at the service of the patient and be properly integrated into his/her follow-up and health management [3]. We believe that it is necessary to move towards a mechanism of technical integration (in practice and patient management software) and human integration (in the physician-patient relationship) of mHealth comparable to the mechanism of a usual medical prescription. In order to achieve this, as it is the case with drugs prescription tools, GPs must have the adequate tools to appreciate (as best as possible): the quality, the indications and contraindications, the possible effects of mHealth apps/devices and their evolution/revision (new market entrant, removal, modification of features). In addition, GPs need to know the possibilities of integrating and analysing the data produced by these mHealth apps/devices and how they can be efficiently used to enrich the relationship with the patient.

Cataloguing mHealth apps and devices is already underway. There are indeed specialized health platforms, which offer such catalogues (for examples: myhealthapps.net, appscript.net, mhealth-quality.eu, medappcare.com). Some technical or conceptual problems still have to be resolved: advanced categorization of apps, rich and standardised medical terms indexing, indications concerning tools complementarity.

Beyond the intrinsic improvements that can be made to these platforms, the key step to guide the choice of an mHealth app/device that meets the patient's needs, is to couple these platforms with one or more information sources which give a detailed level of the medical context. The electronic medical record (EMR) is a valuable source because it contains information rigorously collected, numerous, diversified, semantically rich and medically validated with an individual detail level (concerning the patient). Medical knowledge repositories/databases (clinical practice guidelines, medical drugs characteristics) are other sources to take into account for at least two reasons. First, consulting such repositories reflects the user's particular concern (a medical context) for the accessed information (for example: consulting web pages about asthma or a beta2-mimetic drug expresses a very specific concern). Second, these repositories contain validated knowledge (evidence-based practice recommendations, drugs indications or effects to monitor) that must be taken into account when suggestion of an mHealth app/device consistent with this knowledge is desired (for example: suggest an app for daily steps tracking is consistent with guidelines indicating that regular physical activity is recommended for diabetic persons).

To this date, EMR, medical knowledge databases and platforms cataloguing mHealth apps/devices (mHealth app web store), function in silos without interoperating. mHealth app web

stores operate as standalone applications and are based on rudimentary data to guide the choice of an mHealth app/device, limited to what a user can (and agrees to) fill through a search form. Using EMR and/or medical knowledge databases, GPs do not have access to mHealth apps/devices selected for their relevance in the management of a given patient.

The objective of the ApiAppS project is to identify, explore and, when possible, remove technical and psycho-social obstacles so that GPs can use a system to be of help in the prescription of mHealth apps / devices adapted to their patients.

Methods

The ApiApps project has two fundamental components:

1. **ApiAppS SHS investigations (Social and Human Sciences)**: psycho-social, anthropological, usability studies.
2. **ApiAppS Techno**: implementation of a decision support system and interoperability of the 3 sources of information.

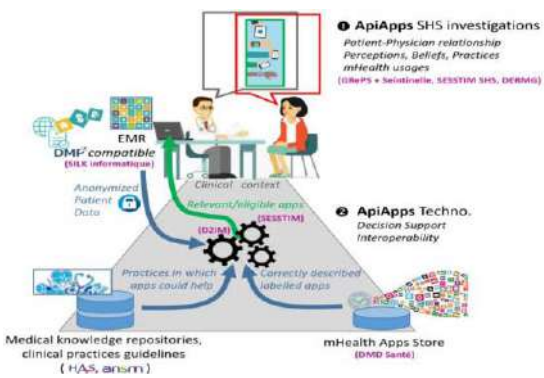


Figure 1 – ApiAppS big picture

ApiAppS implementation will integrate the psychosocial aspects related to the mHealth usage and to their prescription. These aspects are identified and analysed in a dual approach (anthropological and psychosocial), with qualitative and quantitative studies, in order to better document:

- Representations, expectations, beliefs and practices of GPs and patients, related to mHealth apps and devices. Are mHealth apps/devices perceived as tools to enhance personal responsibility (patient empowerment) or is mHealth perceived more as control and standardization tools by health stakeholders and authorities?
- The place of mHealth apps in the doctor-patient relationship [4]: Are they perceived as intrusive, disruptive to the relationship, or conversely, as an element that can reinforce it (or even on a symbolic dimension, the extension of the relationship as for medical drugs)? It will also be questioned whether the prescription itself induces new ethical problems and fears in terms of data security.

ApiAppS project aims to provide a web-service for decision support so that the selection/prescription of a mHealth app/device is adapted to the patient's profile and is relevant to the doctor-patient relationship and the care process. We try to resolve the following technology and information challenges:

1. What kind of information should be available in the electronic medical record, in the knowledge databases, and in mHealth app web stores to allow a basic coupling between these systems? Which data will allow an optimal coupling?
2. What are the coding format(s) and the terminology(ies) adapted to such data?
3. Are the existing interoperability frameworks adaptable to the problem of communication with a store dedicated to mHealth apps/devices?

Results

ApiAppS is a work in progress (from 2018 to 2021). A first qualitative study [5] (semi-directive interviews carried out with 20 GPs) shows there is a gap between mHealth innovation and GPs' everyday practice. GPs perceive mHealth as a potential tool in the relationship with patients but are not familiar with mHealth tools and have doubts and questioning about: How to integrate mHealth into a medical consultation? What place should be given to mHealth in medical follow-up? What to do with the produced data? Is there a risk of isolating patients and/or giving them illusions of safety? Others qualitative and quantitative studies focused on patients' groups are ongoing.

ApiAppS Technological components will lead to the definition and implementation of an interoperability framework in compliance with EMR programs, knowledge databases and mHealth app web stores. First results of the project will be communicated during the poster session.

Conclusions

mHealth is a means to have supplementary options for the patient follow-up or for other aspects in the patient-doctor relationship. This project is in line with: the current evolution of the healthcare mobility; the integration of mHealth apps and devices in care prescription; and the need to better inform practitioners and patients of their relevant/irrelevant use in certain situations. ApiAppS project is a contribution to remove the bottlenecks that constrain rational use of mHealth.

Acknowledgements

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR), grant ANR-17-CE19-0027.

References

[1] H.E. Payne, C. Lister, J.H. West, and J.M. Bernhardt, Behavioral functionality of mobile apps in health interventions: a systematic review of the literature, *JMIR MHealth UHealth* 3 (2015).

[2] C. Pearce, Computers, Patients, and Doctors—Theoretical and Practical Perspectives. In: Shachak A, Borycki EM, Reis SP, editors, Health Professionals' Education in the Age of Clinical Information Systems, Mobile Computing and Social Networks, Academic Press, 2017, p. 5-22.

[3] F. Lopez Segui, C. Pratdepadua Bufill, N. Abdon Gimenez, J. Martinez Roldan, and F. Garcia Cuyas, The Prescription of Mobile Apps by Primary Care Teams: A Pilot Project in Catalonia, *JMIR MHealth UHealth* 6 (2018), e10701.

[4] A.E. Tozzi, E. Carloni, F. Gesualdo, L. Russo, and M. Raponi, Attitude of families of patients with genetic diseases to use m-health technologies, *Telemed. J. E-Health Off. J. Am. Telemed. Assoc* 21 (2015), 86–89.

[5] L. Auroy, and A. Sarradon-Eck, "The digitally engaged doctor": the discourse and the reality, *SGA-SSE-SSA Annual Congr. Zurich* (2018).

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190714

Biomedical Informatics Workforce in Croatia: Qualitative Analysis of Teachers' Opinions on Needs and Employment Opportunities

Kristina Fišter[a], Hrvoje Belani[b], Danko Relić[a], Marijan Erceg[c,d]

[a] *University of Zagreb School of Medicine, Andrija Štampar School of Public Health, Zagreb, Croatia*
[b] *Ministry of Health, Zagreb, Croatia*
[c] *Croatian Public Health Institute, Zagreb, Croatia*
[d] *University of Applied Health Sciences, Zagreb, Croatia*

Abstract

We report a qualitative analysis of higher education biomedical informatics (BMI) teachers' opinions on the needs for BMI workforce and employment opportunities in the Croatian health care system. Needs were perceived as considerable; however this was not coupled with adequate employment opportunities. Barriers were identified to correcting this imbalance, including the failure of the health care system planners to recognise BMI experts as an important part of the workforce and lack of adequate educational opportunities.

Keywords:

Medical Informatics/education; Medical Informatics/organization and administration; Qualitative Research

Introduction

Although medicine has been slow to adopt the rapidly developing information and communication technologies, digital medicine has become a thing of the present as well as future [1,2]. How to prepare the health care workforce and how to organize services in the digital age has been a topic of much debate. It seems clear that what is needed is educating and employing biomedical informatics (BMI) specialists, but also acquiring important new competencies related to the field by all health professionals [3].

A course in BMI at the University of Zagreb School of Medicine was first introduced in 1967 [4]. Postgraduate BMI studies as well as preparations for starting BMI subspecialisation programs for medical doctors of all specialties, which were under way in the late 1980s, were interrupted in the last decade of the 20th century, not to be renewed.

Describing the current educational landscape of BMI in Croatia, we previously reported teachers' characteristics and their attitudes about educational opportunities [5]. Here, we report the teachers' opinions on the needs and employment opportunities for BMI workforce in the health care system.

Methods

We conducted an online questionnaire survey of BMI teachers in Croatia. We acquired the initial list of participants from the Croatian Society of Medical Informatics (CroSMI). This list was extended by online searches for higher education courses and programs in BMI in Croatia and affiliated teachers.

The final sample consisted of all teachers of BMI in Croatia, 44 women and men who were invited by e-mail to participate and received up to three reminders during the eight days while the survey was active in the spring of 2018. For data collection we used LimeSurvey. Participation was anonymous and voluntary.

The questionnaire comprised 24 items, spanning participants' demographic characteristics, educational and academic status, as well as experience in teaching BMI and details of courses they teach. Attitudes about BMI education in Croatia were assessed by 13 items with Likert scale answers, which has been reported previously [5]. In addition, participants were asked to express their opinions in free text about the need for and work opportunities for medical informaticians in the Croatian health care system, which we are reporting here. Opinions were assessed through three open ended items:

1. Please describe the needs for BMI cadre in the Croatian health care system.
2. Please describe employment opportunities for BMI cadre in the Croatian health care system.
3. Final thoughts, comments, and suggestions.

Before the analysis was undertaken, all participants were assigned pseudonyms. Two researchers (KF and DR) independently analysed the answers using descriptive thematic approach [6], which included coding and refinement of themes and subthemes. Disagreements were settled consensually. We used QSR International's NVivo 12 software for qualitative data analysis.

Results

Twenty-two teachers responded to the survey (response rate 50%) with equal numbers of women and men. Median age was 47.5 years (range 30-69), median time spent working in BMI education 11 years (0-40), and median length of overall working experience 23.5 years (4-42). Fourteen participants volunteered their answers to the three open ended items, seven women and seven men.

The word frequency query showed that the most common word mentioned by the participating teachers was '*the system*', in several declension forms of Croatian language, with a total number of 35 occurrences. The next most common words were

'*medical*' with 18 occurrences as well as '*informatics*' and '*needs*' with 13 occurrences each.

The health care system was depicted as rigid, sluggish, poorly organised, unable to govern data and processes, and lacking awareness of the importance of evidence-based decision making, with quotes such as:

- "The system is crying for quality [BMI] cadre, but due to its rigidity in opening new positions as well as lack of recognition of the importance of good organisation of informatics services in health care, employment opportunities are limited. … BMI is often learnt informally when needed in the workplace. This leads to … inability to govern data and processes." (Marta, female, 47)

Teachers unequivocally recognised a great need for BMI experts at all levels of the health care system, but they also pointed out a lack of recognition of these needs by the health care planners as well as lack of adequate employment opportunities.

- "The needs for BMI cadre in the Croatian health care system are increasing, yet what it currently missing, in my opinion, is exactly a systematic, multidisciplinary approach. … Although the lack of health care cadre is becoming increasingly conspicuous in Croatia, I'm unconvinced planners are aware of the need for the BMI cadre." (Đuro, male, 30)

Other participants used words such as *'huge'*, *'urgent'*, and *'inestimable'* to describe the needs for BMI cadre in the health care system. To describe employment opportunities, words such as *'limited'* and *'mediocre'* were used. Further subthemes identified regarding the health care system were the inability to ensure adequate pay for BMI experts, as well as lack of regulation of employment opportunities but only in the public health care sector.

Not only does the lack of recognition of needs for BMI experts exist in the health care system; similar is apparent in the academic community. No dedicated BMI studies exist, and individual courses that are taught to future medical doctors, nurses, and other health professionals are often positioned too early in the curriculum or are considered less important courses for which assigning a numerical grade is not needed. Consequently, competencies are often inadequate, and this reflects poorly on the current workforce.

- "Besides, the scientific circles do not recognise medical (biomedical, health) informatics, considering that medical informatics is like any other informatics, i.e. it is neither science nor profession so it should not be viewed as such." (Koraljka, female, 69)

Offering thoughts on solutions to these problems, several of the participants identified the need for congruence between employment and educational opportunities as key.

- "Solutions should be sought through cooperation of health care and educational institutions in clearly articulating the needs for BMI cadres and defining educational programmes for all health care workers at all levels of education. Rapid development of infomation technology also imposes the need for life-long learning, which is conceivable in cooperation of educational and health care institutions." (Gašpar, male, 56)
- "Education in BMI should be expanded/viewed from the aspect of the application of information/digital technologies in the daily work of staff employed in different institutions in health care - which primarily relates to the use of various health care - clinical - business information systems." (Maja, female, 52)

Participating teachers expressed hope that two bodies in the making, a central national e-health body being set up at the Ministry of Health Care and a CroSMI education working group, would help with recognition of the BMI as a discipline, improvement of educational opportunities, as well as correcting the imbalance between the needs for BMI experts and employment opportunities in the health care system.

Conclusions

Congruence between educational and health care systems is needed to correct the imbalance that currently exists between the needs and employment opportunities for BMI experts in the health care system. The many challenges we face in setting up the BMI workforce can be best tackled by concerted multi-sectoral, multi-institutional efforts, led at the national level by a dedicated e-health body supported by all stakeholders including CroSMI.

Acknowledgements

This work received Research Support by the University of Zagreb, Croatia.

References

[1] R. Wachter. *The Digital Doctor: Hope, Hype, and Harm at the Dawn of Medicine's Computer Age*. McGraw-Hill Education, New York City, NY, 2015.
[2] Health Education England. The Topol review: preparing the healthcare workforce to deliver the digital future. Interim report 2018. Available from *https://www.hee.nhs.uk/our-work/topol-review*. Accessed: 18 November 2018.
[3] D.B. Fridsma, Health Informatics: A Required Skill for 21st Century Clinicians, *BMJ* (2018) k3043. doi:10.1136/bmj.k3043.
[4] D. Šimić, Editor, *Medical Informatics - Qualifications and Professions* [in Croatian], Zagreb University, Faculty of Organisation and Informatics, Varaždin, 2015.
[5] K. Fišter, P. Hrabač, D. Relić, B. Išgum, and M. Erceg, Educational Landscape of Biomedical Informatics in Croatia: Who Are the Teachers and What Are Their Attitudes, *Studies in Health Technology and Informatics* (2018) 217–221. doi:10.3233/978-1-61499-921-8-217.
[6] V. Braun, and V. Clarke, Using thematic analysis in psychology, *Qualitative Research in Psychology* **3** (2006) 77–101. doi:10.1191/1478088706qp063oa.

Address for correspondence

Asst Prof Kristina Fišter, MD, MSc, DSc; E-mail: kfister@snz.hr.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190715

Analysis of Financial Factors Which Driving the Operating Margins of 880 Public Hospitals in Japan Using a Business Intelligence System

Yuta Fushimi[a], Takahito Takagi[a], Yutarou Kishizuchi[a], Takanori Uchiyama[a], Tokiharu Miyahara[b]

[a] Department of Medical science, Faculty of Medicine, Kawasaki Medical School, Kawasaki Medical school, Kurashiki, Okayama, Japan,
[b] Department of Healthcare Information Resource Management, Faculty of Medicine, Kawasaki Medical School, Kawasaki Medical School, Kurashiki, Okayama, Japan

Abstract

Management improvement was strongly required at public hospital and Local Incorporated Administrative Agency hospitals in Japan. We analysed financial statements of 880 public hospitals using Business Intelligence System.

In order to raise the efficiency of management, it is important to have patients with a high profitability (seriously ill patients) at municipal hospitals; whereas, in Local Incorporated Administrative Agency hospitals, avoiding producing ordinary losses and not incurring excessive expenses will be important for management improvement.

Keywords:

Financial statements, Business Intelligence System, public hospital

Introduction

With the aging of patients, medical costs in Japan have steadily increased year by year, and now account for 10.9% (2016) of GDP [1]. Even under ordinary circumstances, management improvement was strongly required even at municipal hospitals and Local Incorporated Administrative Agency hospitals, which undertook the management of unprofitable sectors of the national public insurance. The purpose of this research is to analyze the financial situation of public enterprise hospitals nationwide and to extract factors for efficient management of public hospitals.

Methods

We used the financial statements of public hospitals publicly announced by Japanese Ministry of Internal Affairs and Communications.

We collected financial data from the balance sheets and profit-and-loss statements of about 880 municipal hospitals and 88 Local Incorporated Administrative Agency hospitals. In addition, non-financial hospital data such as the number of hospital beds, the number of inpatients, the number of outpatients, etc. were collected [2]. Approximately 30 financial indicators were calculated from the balance sheets and profit-and-loss statements. Among them, we analyzed which factors affect the representative indicators that represent revenue and expenses. For the analysis, a business intelligence system (IBM, Cognos Analytics 11.1) was used, and correlation analysis by scatter diagram and driver analysis by spiral visualization were performed.

Results

Forecasting a single driver affecting the profit margin of medical service, the most influential factor on the profit margin of 880 municipal hospitals was “outpatients unit profit per person per day (56%) and inpatients unit profit per person per day (52%)”.

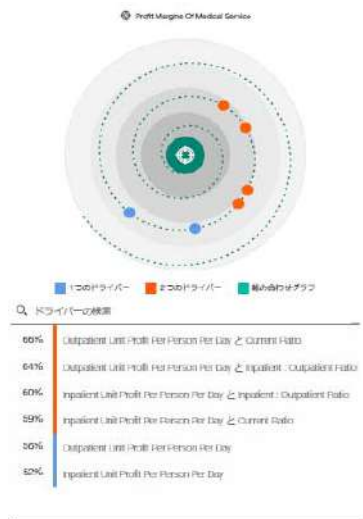


Figure 1– target analysis for profit margin of municipal hospitals

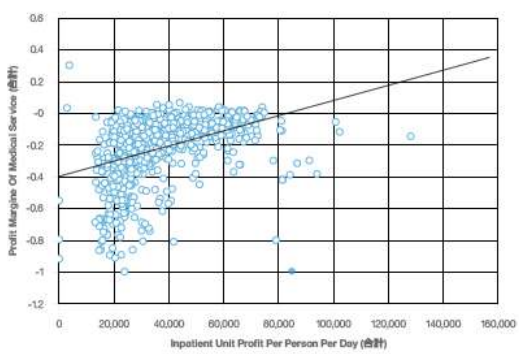


Figure 2– correlation between profit margin and inpatient unit profit per person per day of municipal hospitals

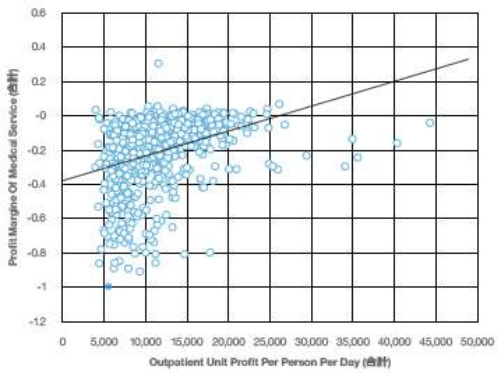


Figure 3– correlation between profit margin and outpatient unit profit per person per day of municipal hospitals

Meanwhile, the driver which influenced the profit margin of medical service of 88 Local Incorporated Administrative Agency hospitals was "ordinary loss per one bed (54%)".

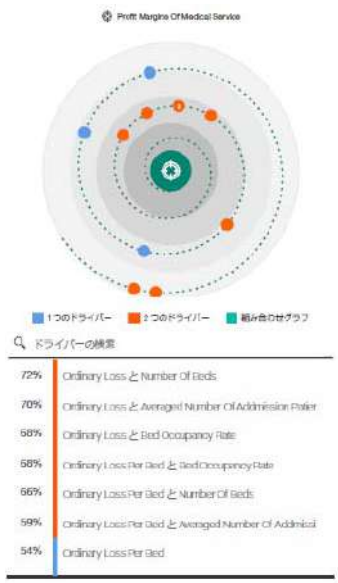


Figure 4– target analysis for profit margin of Local Incorporated Administrative Agency hospitals hospitals

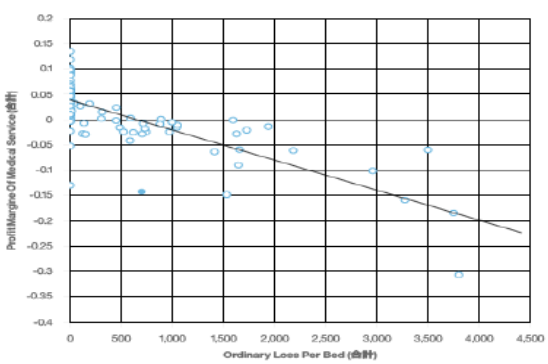


Figure 5– correlation between profit margin and ordinary loss of Local Incorporated Administrative Agency hospitals

Conclusions

In order to raise the efficiency of management, it is important to have patients with a high profitability (that is, seriously ill patients) in both the outpatient and inpatient units of municipal hospitals; whereas, in Local Incorporated Administrative Agency hospitals, avoiding producing ordinary losses and not incurring excessive expenses will be important for management improvement.

Acknowledgements

There is no conflict of interest. We appreciate Ms.Yoshida and Dr. Akiyama's technical support and spiritual support.

References

[1] Ministry of Internal Affairs and Communications, *About social security (2018.4),* https://www.mof.go.jp/about_mof/councils/fiscal_system_council/sub-of_fiscal_system/proceedings/material/zaiseia300411/01.pdf

[2] Ministry of Internal Affairs and Communications, *Financial situation of public hospitals and comparative chart of financial analysis. (2016*, http://www.soumu.go.jp/main_sosiki/c-zaisei/hospital/kessan-bunseki/h28.html

Address for correspondence

Tokiharu Miyahra

577 Matsushima, Kurashiki city, 701-0192

t.miyahara@med.kawasaki-m.ac.jp

TEL. ++81-864-62-1111

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190716

Improving Patient Participation in Cancer Clinical Trials: A Qualitative Analysis of HSRProj & RePORTER

Lynette Hammond Gerido[a] and Zhe He[a]

[a] *School of Information, Florida State University, Tallahassee, Florida, 32306 USA*

Abstract

Enrollment and representativeness, in cancer trials, have been a problem resulting in studies with poor generalizability to the real-world population. This study uses qualitative analysis of two publicly available datasets (HSRProj and RePORTER) to explore funded research projects, which employ strategies at the macro, meso, and micro levels. Our research shows that although the number of projects designed to increase participation in cancer clinical trials peaked in the 2000s, most employ a single level strategy.

Keywords:

Clinical Trial, Patient Selection, Qualitative Research

Introduction

Cancer is the second leading cause of death in the United States, exceeded only by heart disease [1]. According to the statistics of the Centers for Disease Control and Prevention, over half a million people in the U.S. died due to cancer in 2015 [1]. Since early in the 20th century, cancer research has grown rapidly [2]. Appropriately designed clinical studies, especially Randomized Controlled Trials (RCTs), provide gold standard evidence for determining the efficacy and safety of medical interventions [3]. Evidence in RCTs is used by healthcare providers to guide their clinical decisions, and by the regulatory agency to support their approval and adoption of new therapies into clinical practice. However, the enrollment and subsequent population representativeness in cancer trials have been a problem in clinical research [4], resulting in studies with poor generalizability to the real-world population

The passing of the National Institutes of Health (NIH) Revitalization Act in 1993 sparked increased efforts to raise awareness about cancer clinical trials and to reduce barriers to participation. In 2014, Heller et al. conducted a systematic review of the literature to assess the use of complex and interwoven strategies to increase participation in clinical trials at multiple levels; research system level, community level, and individual level [5]. This study qualitatively analyzes the abstracts and metadata of funded projects in two publicly available databases to classify goals, identify target populations being served, and to identify frequency patterns and trends in multilevel strategies to improve participation in cancer clinical trials over the past three decades.

Methods

The National Institutes of Health (NIH) funds the clear majority of medical research to advance cancer research and improve cancer care. One tool offered by the National Library of Medicine (NLM) is the Health Services Research in Progress (HSRProj) database, which contains descriptions of health services research in progress that has been funded by federal and private grants and contracts before results are available in a published form. The researchers queried the HSRProj database to identify projects related to increasing participation in Cancer Clinical Trials (CCTs). After iterative discussions between the researchers, the following search string was formulated and used: "*cancer" AND "clinical trial" AND "participation" NOT "conference".* The query included all available content regardless of fiscal year or active status.

The NIH Research Portfolio Online Reporting Tool's Expenditures and Results module (RePORTER) is a repository of NIH-funded research projects and resulting publications and patents. The same query used in HSRProj was executed using the advanced text search query. The search was limited within the project title and abstracts. Then researchers exported the titles, abstracts, project start dates, and the funding amounts for all available projects initiated prior to 2018. The HSRProj database resulted in 196 abstracts for review, whereas the initial RePORTER results were 1062 studies but when restricted to unique project numbers, there were 218 projects.

We imported the data into Nvivo 11 for coding and analysis. Each abstract was reviewed to determine if it was in fact intended to increase cancer clinical trials participation. Then we identified abstracts as primary when it directly indicated increasing participation as a major goal or outcome of the project. The final number of projects with a primary goal was 94 (52 HSRProj and 42 RePORTER).

Abstract Coding

Descriptive information about each abstract was recorded, including: (1) project title, (2) start and end dates of research, (3) funded entity, and (4) unique project database ID. A directed content analysis was conducted to identify funded entities, under-represented populations, and strategies to increase participation in cancer clinical trials. To develop a coding schema, the researchers independently reviewed a sample of cases to make observations about themes among the cases. Together they reconciled the observations and then devised a scheme based on the reconciled observations. One researcher carried out the coding and another reviewed to confirm agreement.

Strategies

Health-related decision-making and health care delivery is generally distinguished using three levels; micro, meso, and macro [6]. The micro level targets the individual patient. It involves interactions between patients and providers at the

bedside, clinic, or health center. The meso level represents families, groups, or communities. It involves interactions between institutions, regional networks, and local agencies. The macro level represents research intended to benefit large populations. Each strategy was categorized as belonging to one or more of the four categories associated with the three levels: Micro (1a) *consumer*: outreach and education projects to improve consumer awareness and knowledge about clinical trials. Some projects include translation of educational materials; (1b) *provider*: workforce development and recruitment initiatives to increase accessibility to clinical trials; Meso (2) *system*: enterprise improvements, technological innovations, and strategic alliances to improve regional cancer care delivery, and Macro (3) *research*: studies to better understand barriers, develop interventions, and generate best practices for increasing participation.

Results

Charateristics of Included Projects

The studies in the sample have start dates between 1983 and 2015. The number of studies initiated in each decade, appears to increase linearly each decade. In both databases, a majority of the research was conducted by a university. The HSRProj sample shows a slightly greater proportion of research conducted by universities (N =26, 50%) than the RePORTER sample (N=19, 44%). National Cancer Institute's funded entities, NCORP and CCOP, represent a little over one third. In each sample, nearly half of the studies targeted a specific race or ethnic group as the underrepresented population (HSRProj: 49% (N = 22) vs. RePORTER: 51% (N=18)). In HSRProj, each of the other underrepresented populations (women, elderly, or rural) combined were identified in less than 23% (N=12) of the studies. A similar trend was observed in the RePORTER sample with less than 40% (N=17) of the studies identifying women, elderly, or rural underrepresented populations.

Strategies for Improving Participation

The researchers aligned the three levels and with respective strategies that projects employed to improve participation in cancer clinical trials. There appears to be consistent frequencies between the level at which the projects acted.

Single-Level Approaches

Table 1 shows both HSRProj and RePORTER, have projects acting at the macro level by using research as the sole strategy to increase participation in cancer clinical trials (HSRProj N = 12; 23% vs. RePORTER N=10; 24%).

Table 1 – Comparison of Strategy-Level Combinations.

Level(s)	HSRProj n, %	Level(s) n, %
Macro Only	12, 23.1%	10, 23.8%
Meso Only	7, 13.5%	5, 11.9%
Micro Only	3, 5.8%	7, 16.7%
Macro/Meso	5, 9.6%	4, 9.5%
Macro/Micro	5, 9.6%	4, 9.5%
Meso/Micro	13, 25.0%	8, 19.0%
Macro/Meso/Micro	7, 13.5%	4, 9.5%

Multilevel Strategy Combinations

Some projects employed more than one strategy in the attempt to improve participation in cancer clinical trials. It should be noted, when a program employed more than one strategy of the same type (e.g., two consumer outreach projects within a single project), the researchers counted the project only once in association to its relevant strategy classification. Table 1 provides a breakdown of frequencies of studies based on approaches targeting multiple levels and using two or more strategies. The most used multilevel strategy combination in both HSRProj or RePORTER was the meso-level paired with the micro level (HSRProj N = 13; 25.0% vs. RePORTER N=8; 19.0%).

Discussion

Clinical trials are the gold standard for cancer research. The cancer patient population is diverse [7]. Therefore, it would be expected that the development and testing of new cancer therapies would include a representative sample of individuals from diverse racial and ethnic groups, of different genders, including the elderly, and rural populations. This has not been the case, despite several decades of efforts to address these disparities. Barriers to clinical research participation are complex and interwoven between multiple levels and stakeholders [8-9]. Successful projects need more than one strategy to address multiple barriers in study design, at provider enrollment sites, for the individual, and within the community [5]. Our research shows that there have been increases in projects designed to increase participation in cancer clinical trials. Despite these efforts, the accrual strategies employed tend to be employed alone and not combined with others.

References

[1] FastStats [Internet]. Center for Disease Control and Prevention (CDC). 2018 [cited 2018 Apr 26]. Available from: https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm

[2] I.B. Weinstein, K. Case. The History of Cancer Research: Introducing an AACR Centennial Series. Cancer Res. 2008 Sep 1;68(17):6861–2.

[3] T. Kennedy-Martin, S. Curtis, D. Faries, S. Robinson, J. Johnston. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials. 2015 Nov 3;16:495.

[4] S. Fu, L. McQuinn, et al. Barriers to Study Enrollment in Patients With Advanced Cancer Referred to a Phase I Clinical Trials Unit. The Oncologist. 2013 Dec 1;18(12):1315–20.

[5] C. Heller, J.E. Balls-Berry, et al. Strategies addressing barriers to clinical trial enrollment of underrepresented populations: a systematic review. Contemp Clin Trials. 2014;39(2):169–182.

[6] H.J. Sutherland, J.E. Till. Quality of life assessments and levels of decision making: differentiating objectives. Qual Life Res. 1993 Aug 1;2(4):297–303.

[7] USCS Data Visualizations [Internet]. [cited 2018 May 2]. Available from: https://gis.cdc.gov/grasp/USCS/DataViz.html

[8] J.M. Robinson, W.M.K. Trochim. An Examination of Community Members', Researchers' and Health Professionals' Perceptions of Barriers to Minority Participation in Medical Research: An Application of Concept Mapping. Ethn Health. 2007 Nov 1;12(5):521–39.

[9] C.R. Baquet, K. Henderson, P. Commiskey, J.N. Morrow. Clinical Trials: The Art of Enrollment. Semin Oncol Nurs. 2008 Nov 1;24(4):262–9.

Address for correspondence

Zhe He, PhD. School of Information, 142 Collegiate Loop, Tallahassee, Florida, USA 32306-2100

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190717

Disruptive Analysis of Closed Questions Assessments at Medical School, Interest of Massive Multi-Choice Tests

Marice Fourcot[a], Lionel Di Marco[a], Vanda Leungo[b], Pierre Gillois[a]

[a] UGA, CNRS, Public Health Department, CHUGA, TIMC Laboratory UMR 5525, Themas, Grenoble, F-38041,
[b] Sorbonne Université, CNRS, Laboratoire d'Informatique de Paris 6, F-75005

Abstract

Descriptive statistics, Classical Test Theory (CTT), and Item Response Theory (IRT) are used to analyze learning. These results are a powerful synthesis for students to evaluate their progression into their personal pathway of learning, for teachers to improve the quality of different multiple-choice questions (MCQ) series, and for administration to assist in making pedagogical decisions. Our study explores tools to evaluate MCQ tests, and MCQs. Answers to tutorials and final examination of students of the first year of health contest were collected and analyzed.

Keywords:

Psychometrics, Educational measurement, Learning

Introduction

Ongoing improvements in the quality and efficiency of education is a key effort. This is a huge opportunity for researchers because learning management systems can track various details of the interaction between the student and the learning platform. Pertinent interaction details include: time spent on a lecture, responses to tests, questions submitted online. These details represent an enormous amount of data, and are thus substantial work for researchers to analyse. In an effort to improve it and to optimize education, dashboards for students or teachers have been developed. Additionally, the fields of learning, teaching, and academic analytics let us analyse closed question series, such as multiple-choice questions (MCQ), with different tools. Such tools include: descriptive statistics, Classical Test Theory (CTT) and Item Response Theory (IRT) [2].

Since 2006 at the Grenoble faculty of medicine, lecture courses have been substituted with blended learning, where lectures have been deleted in favour of multimedia resources and skills interactions [1]. The learning process is divided in 4 sequences: self-learning using multimedia resources, questions submitted online, meetings for interactive question-answer sessions with teachers, and tutorials in preparation for the rank exams using MCQs [3]. Analysis of the answers provide a student dashboard, which serves as a tool for monitoring and motivation. The objectives of the dashboard is to assess the student's level more accurately than with a single mean, and to provide ranking information among other students. The objectives of the teacher's dashboard are to provide an evaluation of students' overall level and evolution, and to highlight students having difficulties. The teacher's dashboard also improves testing by assisting instructors to select MCQs based on quality. Thus, we propose how to perform a qualitative and quantitative statistical analysis that will provide second step recommendations in order to build a dashboard that serves the learner, teacher, and institution.

Methods

Classical descriptive statistics are used to characterize each test: the number of students who passed it, thenumber of MCQs, success rates, mean score of 20, median score and Cronbach's alpha. Classical Test Theory (CTT), or True Score model, is based on the assumption that for a particular test with a sample, the observed score on a test is the sum of the "true score" and of the random "error". CTT is a collection of many statistics, including the average score, item difficulty defined as the proportion of examinees who did not respond in the keyed direction, discrimination index that indicates how well different levels of knowledge can be differentiated and the test's reliability. Finally, CTT is used to analyse tests, rather than to check MCQs.

More recently, the development of the Item Response Theory (IRT) permitted the detection of problematic MCQs, such as questions that are too difficult or those that are misunderstood. In addition, the IRT allows the calculation of the 'true score' of a learner at a test, which can be useful for students to assess their real knowledge independently of the difficulty of the test. The main assumption of the IRT is that a test is built to measure an ability, called "latent trait". There are two types of IRT model depending on the assumption made on the latent trait; use a unidimensional model if the trait is unique, or use a multidimensional model if it can be assumed that multiple abilities are involved in the test.

For unidimensional latent and dichotomous correction, logistic functions are mainly used. Unidimensional models are characterized by the number of parameters they include. In a three-parameter logistic model (3PL) the parameters are difficulty, discrimination and pseudo-guessing. The most well-known model is the Rasch's model, which is considered as a one-parameter model, and calculates difficulty and discrimination too, but assumes that the discrimination level is equal for all items.

In this study, we built each type of model, and then chose the best one for a given test using the Akaike information criterion (AIC). Graphical approaches, such as the Item Characteristic Curve (ICC), indicate the difficulty parameter. This parameter is defined as the level of ability for a probability of a keyed response of 0.5. The pseudo-guessing parameter is readable as the intercept. The Test Information Curve (TIC) as the sum of

the Item Information Curve (IIC) of each item in the test, and determine ability levels that can be differentiated by the test.

Results

During the 8-year study period, 12,478 students were enrolled. 472 MCQ tests took place in which 16,731 MCQs asked for 15,915,916 answers.

Global overview with CTT per item

We reviewed MCQs on one test in order to characterize questions. On this test, no MCQs appear difficult, the max was found to be 0.75. However, this index is linked to students' results, so in our case of a rank examination, where students work a lot, the difficulty is probably underestimated.

Non-sigmoid curve detection for students abilities adequation with IRT by Item

In Figure 1, ICCs are shown as sigmoid spread along ability levels with a significant slope so that the test may distinguish differences in students' ability. The ICCs for MCQ 4 (in blue), 7 (the turquoise one on the left of the group), 21 (turquoise), 23 (the red one on the right), are abrupt across space sigmoid from each other. A flat ICC indicates that the item measured something else other than the test's objective. An MCQ adapted to all student levels and should not be constant or decreasing.

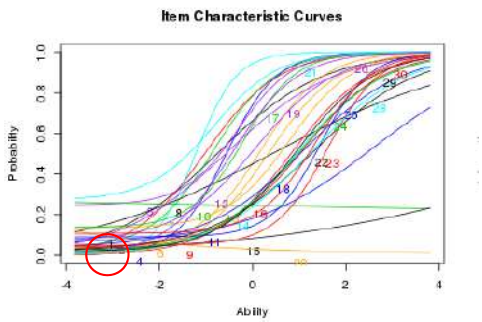


Figure 1- ICC graphs based on BioPhysics 3 2015-16

Selection of questionable MCQs

Figure 2 highlights only questionable MCQs: those with an ICC's intercept higher than 0.2, too high pseudo-guessing parameters (box a), a probability of keyed answer for an ability at 4 lower than 0.8, too difficult (box b), a maximum level of information lower than 3 times the maximum level of information in the test, and not informative.

Discussion

Our study explored several tools to evaluate MCQ tests, and MCQs themselves. The diversity in available tools gives flexibility in analyzing an MCQ, but there is also a complexity factor. None of the tools can be considered as better than the others. Thus we choose to show CTT index to the student and

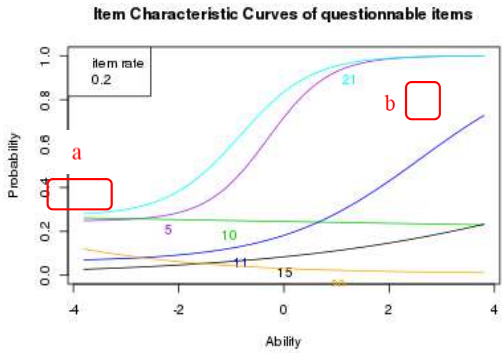


Figure 2- ICC graphs of questionable items

IRT parameters to the teacher, because IRT needs explanation to be understood, while students might want instant results. Our use of IRT presents an issue based on the type of model chosen. Logistic models are the best ones for IRT, but they need a dichotomous explanatory variable. While it is not the type of correction used during tests, and polytomous correction gives fine grain understanding of the score. To remedy this situation polytomous IRT has been developed but it is not possible to provide quickly understandable conclusions. The use of these Rasch-type methods, in addition to the traditional approach, is now essential in the face of the emergence of educational big data, particularly in health. These methods require a pedagogical step of explanation with teachers who are not specialists in these fields, in order to guide them, but the gain in using the IRT can no longer be called into question. The era of "learning and teaching analytics" has already begun, it should not be missed.

Acknowledgements

This work has been supported by the HUBBLE project (ANR-14-CE24-0015).

References

[1] P. Gillois, J.L. Bosson, C. Genty, J.-P. Vuillez, and J.-P. Romane, The impacts of blended learning design in first year medical studies, *Studies in Health Technology* **210** (2015), 607-611.

[2] C. Gu and R. Gutman, Combining item response theory with multiple imputation to equate health assessment questionnaires, *Biometrics* **73** (2016), 990-998.

[3] S. Houssein, L.D. Marco, C. Schwebel, V. Luengo, P. Morand, and P. Gillois, Consequences of Switching to Blended Learning: The Grenoble Medical School Key Elements, *Studies in health technology and informatics* **247** (2018), 356-360.

Address for correspondence

Pierre Gillois
pierre.gillois@univ-grenoble-alpes.fr.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190718

Evaluation of a Mobile Application to Enhance Medication Management Following Hospital Discharge: Study Protocol for a Pilot Randomized Controlled Trial

Bettina Habib[a], Santiago Marquez Fosser[a], David L Buckeridge[a,b], Daniala L Weir[a,b], Melissa Bustillo[a], Manish Thakur[a], Thai Tran[a], Aude Motulsky[c,d,e], André Bonnici[f], Emily G McDonald[c,f], Todd C Lee[c,f], Robyn Tamblyn[a,b,c]

[a] Clinical and Health Informatics Research Group, McGill University, Montréal, Québec, Canada
[b]Department of Epidemiology and Biostatistics, McGill University, Montréal, Québec, Canada
[c]Department of Medicine, McGill University, Montréal, Québec, Canada
[d]Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Montréal, Québec, Canada
[e]School of Public Health, Université de Montréal, Montréal, Québec, Canada
[f]McGill University Health Centre, Montréal, Québec, Canada

Abstract

Over 3 million hospitalizations and 17 million ER visits occur in Canada each year. A substantial proportion of these encounters are preventable and attributable to medication non-adherence. Non-adherence to medication changes during discharge increases the risk of adverse events post-discharge. A mobile application was developed to improve medication management of post-discharge patients. A pilot randomized controlled trial was conducted to assess the application's usability and efficacy in decreasing non-adherence to medication changes made at discharge.

Keywords:

Medication therapy management, Medication adherence, Mobile applications

Introduction

Over 3 million hospitalizations and 17 million emergency department visits occur in Canada each year [1]. Seventy percent of hospitalized patients experience an adverse event involving medication(s) within 30 days of discharge [6]., At least 58% of medication-related adverse events are preventable and are the result of prescribing or dispensing errors, incomplete drug information, and the underuse or overuse of medications [3; 8]..Patients are often discharged on a substantially different medication regimen than the medication regimen prior to their admission[2]. . A previous study of 2,655 patients discharged from one academic health centre determined 55% were non-adherent to at least one medication change made at discharge. Patients non-adherent to all changes had a 35% increased risk of adverse events compared to those who were adherent to all changes [11].

Patient interview data suggest that unclear communication with patients about medication changes may be driving non-adherence, as might difficulties in managing complex drug regimen information and dosing schedules. We must consider innovative approaches to implementing and evaluating patient support mechanisms to reduce medication non-adherence post-discharge. Mobile technologies are a potentially powerful tool to provide individual-level health support. Studies suggest that mobile health(mHealth) technologies are effective in improving health behaviours [7]. With 77% of adults in the United States reporting smartphone ownership in 2017, an opportunity has been created to assess the efficacy of mHealth technologies in improving health behaviours and outcomes [9]. Smart About Meds (SAM) is a mobile application designed to improve medication management for patients following discharge. The study protocol aims to evaluate SAM's usability and efficacy in reducing non-adherence to medication changes made at hospital discharge.

Methods

A pilot randomized controlled trial of 100 discharged patients will be conducted at an academic health Patients in the intervention arm will receive access to the app, while patients in the control arm will receive usual care. A usability assessment will compare, between the two arms the rates of non-adherence to medication changes in the 30 days post-discharge. The pilot begins in mid-2019 and will be completed by March 2020.

Study Population

The study population will comprise 100 patients discharged from the internal medicine unit of one academic health centre. Patients must be covered by the Quebec public insurance drug plan, be ≥18 years old, own a tablet or smartphone with internet, and must speak and read English or French.

Randomization and Blinding

Permuted block randomization with varying block sizes of two and four will be used to randomize 100 patients 1:1 to the control or intervention group. Data analysts are blinded to group allocation while study participants and personnel are unblinded.

Intervention

Control arm (Usual care)

At discharge, patients will be given a prescription, and may or may not receive written or verbal instructions about changes made during medication therapy management.

Intervention arm (Mobile application)

At discharge, patients will receive access to the SAM application. SAM was developed using a user-centred design and agile development processes that incorporated user feedback in iterative cycles of design, testing, and evaluation. The application retrieves medications prescribed to a patient from the discharge prescription and dispensed medications via real-time linkage with the provincial pharmacy claims database.

The app creates a patient-friendly list of prescribed/ dispensed medications, grouped by therapeutic class, and offers tools targeting barriers to adherence:*Integrated adherence monitoring and feedback:* Alerts patients, caregivers, and hospital pharmacists to adherence problems. Discharge prescriptions are linked to dispensed medications and decision algorithms that generate alerts: 1) prescribed medications not filled, 2) medications filled as the wrong dose, and 3) discontinued medications that are refilled.*Pharmacy connect:* Patients connect with hospital pharmacists about discharge prescription(s), therapy changes, and alerts through a secured messaging service.

Symptom checker: Patients determine which of their medications has side-effects similar to experienced symptoms.

Social connect: Patients share experiences of their medications (e.g. effectiveness and side-effects). Patients can view comments from other patients using the same medications.

Caregiver connect: Caregivers access to the app

Health professional dashboard: Pharmacists manage a group of patients, receive and respond to questions, manage alerts or potential medication-related problems, transmit and manage requests for consultation from treating physicians and pharmacists, and document services.

Other features: User access to directives, pill images, drug information, dosing schedules, and daily pill reminders.

Beta-testing the app

A beta-test was conducted using the think aloud protocol [5] during the final stages of design and development. Eight users completed six standardized tasks in the application. The users were audio-recorded as they performed the tasks, and completed the System Usability Scale (SUS) survey [4]. The mean SUS score of 79.1 suggested that the app is usable and likely to be recommended by users to a friend. Audio recordings were analyzed and were integrated into the final version of the application.

Outcome Measures

Usability assessment

One week post-discharge, the intervention group completes a questionnaire that assesses satisfaction with the app and collects feedback on its usability. The questionnaire has been designed using extended versions of the technology acceptance model [10]. The think aloud protocol [5] will be used to assess the app's (1) ease of use, (2) user-friendliness, (3) efficiency, and (4) features that may cause confusion, frustration, or user errors.

Non-adherence to medication changes

Pharmacy claims data from the provincial health insurer will be used to identify medications dispensed within 30 days post - discharge. These data will be compared with the discharge prescription to identify non-adherence to medication changes, defined as (1) failure to fill newly prescribed medications, (2) refilling a prescription discontinued at discharge, or (3) filling a previous prescription where the dosage was modified at discharge, but was dispensed at the incorrect daily dose.

Data sources

Pharmacy claims data from the Régie de l'assurance maladie du Québec will be used to identify prescriptions filled post-discharge. Medications prescribed and changed at discharge will be determined using patients' discharge prescriptions.

Data Analysis

Descriptive statistics will be used to summarize patient demographic and clinical characteristics, technology assessment questionnaire results, and non-adherence rates for each group and overall. The two arms will be compared to measure, the average number of medication changes not adhered to in the 30-day follow-up period using the Student's t-test. Quantitative analyses will be conducted using SAS version 9.4 and the think aloud protocol.

Conclusion

The medication management app can be an effective tool to increase adherence to medication changes made at discharge. The use of the app can potentially reduce adverse events post-discharge. Results of the pilot will be used to support a larger clinical trial to assess impact on adverse events. This data will provide supporting evidence for the use of health informatics tools to promote patient outcomes.

Acknowledgements

This project was funded by the Quebec Ministry of Economy, Science and Innovation. Authors have no conflicts of interest.

References

[1] Access Data and Reports, in, Canadian Institute for Health Information (CIHI), Ottawa, ON.

[2] M.H. Beers, J. Dang, J. Hasegawa, et al., Influence of hospitalization on drug therapy in the elderly, *J Am Geriatr Soc* **37** (1989), 679-683.

[3] C.M. Bell, J. Bajcar, A.S. Bierman, et al., Potentially unintended discontinuation of long-term medication use after elective surgical procedures, *Arch Intern Med* **166** (2006).

[4] J. Brooke, SUS: a 'quick and dirty' usability scale, in: *Usability Evaluation in Industry,* P.W. Jordan, B. Thomas, B.A. Weerdmeester, and I.L. McClelland, eds., Taylor & Francis, London, UK, 1996.

[5] E. Charters, The Use of Think-aloud Methods in Qualitative Research: An Introduction to Think-aloud Methods, *Brock Education Journal* **12** (2003).

[6] A.J. Forster, H.D. Clark, A. Menard, et al., Adverse events among medical patients after discharge from hospital, *Cmaj* **170** (2004), 345-349.

[7] C. Free, G. Phillips, L. Galli, et al., The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review, *PLoS medicine* **10** (2013).

[8] L.L. Leape, D.W. Bates, D.J. Cullen, et al., Systems analysis of adverse drug events. ADE Prevention Study Group, *Jama* **274** (1995), 35-43.

[9] J. Poushter, C. Bishop, and H. Chwe, Social Media Use Continues to Rise in Developing Countries but Plateaus Across Developed Ones: Digital divides remain, both within and across countries, in, 2018.

[10] V. Venkatesh and H. Bala, Technology Acceptance Model 3 and a Research Agenda on Interventions, *Decision Sciences* **39** (2008), 273-315.

[11] D.L. Weir, A. Motulsky, M. Abrahamowicz, et al., Failure to follow medication changes made at hospital discharge associated with adverse events in 30-days: A prospective cohort study of hospitalized medical and surgical patients (submitted), (2018)

Address for correspondence

Robyn Tamblyn, BScN, MSc, PhD, FCAHS, CM
1140 Pine Avenue West, Montreal, QC H3A 1A3
Tel. 514-934-1934 ext 32999; robyn.tamblyn@mcgill.ca

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190719

Assessment by Patients of a Connected System for Telerehabilitation: Lessons Learned from a Randomized Qualitative Study

Maurice Hayot[a], Roxana Ologeanu-Taddei[b], Wafa Bouaynaya[c], Bronia Ayoub[a], François Bughin[a]

[a] *Department of Clinical Physiology and PhyMedExp, University of Montpellier, CNRS, INSERM, CHU Montpellier, Montpellier, France*
[b] *University of Montpellier, Montpellier, France*
[c] *LEMNA, University of Nantes, France*

Abstract

This poster presents the design of a connected system for telerehabilitation for patients with obesity and the assessment of the system through a randomized qualitative study on a sample of 15 patients. The patients expressed positive motivation but negatively assessed (as a deficiency) the system operation. All patients found that the system was neither intuitive nor easy to use.

Keywords:

Telerehabilitation, telemedicine, rehabilitation.

Introduction

Rehabilitation is a multidisciplinary intervention process combining retraining with effort, therapeutic education, nutritional programs, and psychosocial and behavioral interventions. This process can improve physical and psychological conditions and decrease multimorbidity mortality. Rehabilitation takes place in specialized centers, but accessibility is limited to a small number of patients. Nevertheless, the benefits of rehabilitation decrease over time due to loss of patients' motivation.

The number of mobile health applications (mHealth), especially those dedicated to the management of chronic diseases, has increased to improve access and efficiency and reduce costs in some areas of health. In the context of obesity, applications provide an additional tool that can help in behavioral management programs by promoting the modification of regular lifestyle habits. Therefore, the recent development of telerehabilitation [1] is a promising approach that has been addressed mainly in pilot studies.

Given technology advances, French healthcare professionals have designed a telerehabilitation system (Telerehab) to empower a population of obese patients to modify their behaviors regarding physical activity and nutrition. Telerehab was designed in 2017 by a French hospital team composed of two physicians specialized in rehabilitation, a nutritionist, and a physiotherapist in collaboration with a health app publisher and a rehabilitation center. The goal is to help patients with obesity self-monitor their nutritional and physical activities to maintain their motivation and behaviors related to physical activity and nutritional programs. The informational and educational content of the system is based on the French guidelines for rehabilitation for patients with obesity. The final objective of the designers of Telerehab was to evaluate in a prospective randomized controlled study on whether a 3-month treatment with Telerehab makes it possible to reduce the amount of fat in obese patients with respect to a control group.

In the meantime, a qualitative study aimed to assess the usability of Telerehab in order to improve it and to evaluate the final objective. This paper provides the methods and results of this assessment. Telerehab is composed of an application and a website that provide a program of adapted physical activity (video capsules) and nutritional information, complemented by medical teleconsultations by videoconference and by connected devices (pedometer, heart rate monitor, cycle ergometer, bike). Patients can track their nutrition and physical activity using the declarative data entered in a form. Technical support was provided by email and phone to help patients use the system.

Methods

A mock-up of the system was tested in February 2018. A randomized controlled study was conducted for 50 patients to assess the clinical objectives. Patients were recruited from the Departments of Clinical Physiology and Internal Medicine in Montpellier University Hospital (France) and from the rehabilitation centers of Groupe 5 Santé (France). The inclusion criteria were patients aged from 25 to 65 who were considered obese with a body mass index (BMI) greater than 30 kg / m². The study was conducted on two groups, an experimental group and a control group, each composed of 25 patients. The experimental group used Telerehab for 3 months. The ethics committee approved of the study (CPP 17/39 - F. Vasseur, clinical trial agreement UF 9799).

The patients from the experimental group received a short training session (6 h) by a research assistant who was recruited for the study. A technical support person would come to the patient's home after 48 h and 4 and 8 weeks to help the patient use the application. Patients had access to teleconsultation with a physician. This physician would schedule a teleconsultation for a patient according to the medical parameters recorded. The qualitative study, consisting of 15 semistructured interviews, was conducted among voluntary patients of the experimental group, 3 months after Telerehab implementation. The interview guide focused on the ease of use and usefulness of the application as well as any positive aspects and aspects that should be improved. The demographic characteristics of the patients interviewed are summarized in Table 1.

Table 1– Demographic characteristics of the patients

Sex	**Number**
Men	8
Women	9
Age	**Number**
25-29 years	2
30-39 years	1
40-49 years	3
50-59 years	9
60-65 years	2

The analysis was performed using open coding, which is an "*analytical process through which concepts are identified and their properties and dimensions are discovered in the data*" ([2] 101). Data were coded independently by two authors of this paper. Then, they compared the coded data and, in case of disagreement, they discussed their results in order to reach a consensus. Moreover, these authors also looked at the negative or positive assessment of the themes that we identified.

Results

Patients in the experimental group expressed that information recording was time consuming and lacked benefits, e.g., personalization of advice for meals. They also emphasized that the application suffered from functionality overload and a cumbersome system. However, a majority of patients emphasized the potential usefulness of the system for motivation.

The issue of personalization often came up in the various reflections by the patients. Personalization is one of the things that has been of most interest to patients. They consider that their use of the application should result in more personalization:

"*With the video capsules you should be able to propose personalized tricks adapted to the person, because as we enter many data I would like to receive, for example, a personalized menu with respect to my weight and my size*." (Patient 26)

Personalization of the system is also associated with other aspects of perceived utility, such as the desired added value of the information offered. Nevertheless, patients were satisfied by the teleconsultation aspect of the application.

Another issue is related to the usefulness of connected objects, especially the pedometer, mainly for technical reasons. Indeed, the majority of patients were concerned about the synchronization between the cycle ergometer and the application and believed that the data collected from the cycle ergometer was biased. The main themes that were coded are summarized in Table 2. We have to mention that these themes were assessed negatively, that is, the table shows instances where those themes were found lacking (with the exceptions of software slowness, human support, usefulness of teleconsultation and overall motivation). Most patients considered that information input was very time consuming, while at the same time they did not receive any information output, such as relevant alerts and notifications.

Table 2–Themes coded

Themes	Number of patents	Number of quotes across all patients
Intuitiveness	15	39
Utility of the information recorded	13	36
Technical issues, i.e., failures	12	24
Ease of use	11	22
Overall motivation	11	24
Slowness	10	20
Data preservation	9	13
Motivation for physical activity	8	19
Human support	8	14
Customization	6	9
Usefulness of teleconsultation	6	8
Behavior change	5	8
Added value	5	7
Perceived clinical effects	4	10

Discussion

These results show that Telerehab was not considered easy to use or useful by patients. While patients emphasized that Telerehab should increase their motivation for healthy activities, e.g., physical activity and adapted meal plans, they considered that the system could not yet achieve this objective because of an unsatisfactory design.Previous studies have shown that functionality comprising ease of use and quality of information is an important criterion for health apps [3]. Clinical outcomes, clinical processes, healthcare utilization and costs have all been evaluated in assessing telerehabilitation [4]. Our study adds the importance of technical issues and usefulness, i.e., customization, usefulness of teleconsultation, behavioral changes, added value and perceived clinical effects, to the list of issues that should be considered when studying methods of telerehabilitation. Moreover, our results highlight the importance of technical problems related to connected objects in the context of the lack of interoperability of these objects [5].

Conclusions

This paper presents the design of a medical connected system for telerehabilitation and a randomized qualitative study to assess this system. The results highlight several issues: technical problems, lack of ease of use and little usefulness of the information recorded.

References

[1] T. Deborah, T. Russell, and R. Latifi, Telerehabilitation: current perspectives, Studies in health technology and informatics 131 (2008), 191-210.

[2] A. Strauss and J. M. Corbin, *Basics of qualitative research: Grounded theory procedures and techniques*, Sage Publications, Inc, 1990.

[3] S. R. Stoyanov, L. Hides, D. J. Kavanagh, O. Zelenko, D. Tjondronegoro, and M. Mani, Mobile app rating scale: a new tool for assessing the quality of health mobile apps, *JMIR mHealth and uHealth*, 3(1), 2015.

[4] D. Kairy, P. Lehoux, C. Vincent, and M. Visintin (2009). A systematic review of clinical outcomes, clinical process, healthcare utilization and costs associated with telerehabilitation. Disability and rehabilitation, 31(6), 427-447.

[5] R. V. Tuckson, M. Edmunds, and M. L. Hodgkins, Telehealth, New England Journal of Medicine, 377(16) (2017), 1585-1592.

Address for correspondence

roxana.ologeanu-taddei@umontpellier.fr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190720

Cost-Effectiveness of Digital Wound Care Education in a Healthcare Organization

Anna-Maria Hiltunen[a], Riikka Laurila[a,b], Katariina Silander[a], Timo Kuosmanen[b]

[a] HEMA Institute, Department of Industrial Engineering and Management, Aalto University
[b] School of Business, Department of Information and Service Management, Aalto University

Abstract

Advantages of digitalization are understood, but implementation to healthcare is slow. Cost savings and quality improvements are needed in healthcare. Continuous education of healthcare professionals is essential for quality, and digital education (DE) enables that cost-efficiently.

The aim was to evaluate the cost-effectiveness of a DE for wound care by comparing it to lecture education (LE).

DE enabled a slightly better learning outcome than LE. However, combination resulted in superior outcome. DE provided best cost-effectiveness.

Keywords:

Education, Continuing, Cost-Benefit Analysis, Information Science

Introduction

Continuous development of professional skills is essential to ensure good-quality care. Wound care is the key competence of nurses as it is estimated that approximately one percent of the population in developed countries suffer from a hard-to-heal wound during lifetime [1]. Wound care skills will be emphasized in the future as the population gets older.

Even though it is important that professionals are knowledgeable, lecture education (LE) is expensive and takes time. With digital tools costs can be reduced and education offering scaled up [2], especially as digital tools are as effective as traditional learning method in professionals' behavior [3]. Cost-effectiveness (CE) analysis is considered to be a good tool to help the decision making, and can be used to evaluate digital education (DE). The aim is to study the CE of a DE compared to traditional LE.

The three main research questions are:

1. What is the relationship between the wound care education method and health care professionals' knowledge level?
2. Do the costs of the DE differ from LE?
3. What is the CE of the DE in wound care compared to LE?

The study was done in a health and social care organization, which had two ways for educating professionals for wounds: 1) traditional classroom LE, and 2) a DE, which included three modules (ABC of wounds, Local treatment of open wounds and Prevention of pressure ulcers). The DE had written material and visuals to support the learning. The DE included final exam. Neither visits to the DE platform nor exam sits were limited, but the employee was allowed to visit the DE platform and sit the final exam as many times as they wished. The DE platform could be accessed from the organization as well as the employees' homes. The DE was provided by Duodecim Medical Publications Ltd (later Duodecim), a Finnish Company publishing information content for medical and healthcare professionals. Since 2017, the organization had shortened the LE from a day to half a day and decided the DE to be a prerequisite for the half day LE. Before 2017, the one-day LE alone covered the education in wound care.

Methods

Data on knowledge level and professionals' background factors were collected with a web-questionnaire, which was distributed to all healthcare professionals in departments possibly facing wounds regularly or randomly (approx. 2k professionals).

The knowledge was tested with a knowledge test, which was identical to the original final exam of the DE. The background questions related to demographics, experience in the industry, and the participants education. All the participants were categorized into groups based on completed wound education. First, the participant data was divided into three: 1) completed one-day LE, 2) completed half-day LE, 3) no LE. Second, the groups were split in two: 1) completed DE, 2) no DE completed.

The knowledge tests were graded and the data analysis was carried out by IBM SPSS Statistics. The average scores between different groups were tested by Kruskal-Wallis H-Test and Mann-Whitney U-test. Finally, a multiple linear regression model to control the background variables was conducted.

Information regarding education costs was estimated by the organization and Duodecim. The data included the costs of DE, one-day and half-day LE, and organizational cost of one nurse per hour. The estimated cost consists of the (1) direct costs from acquiring LE or DE (fees and teaching), and (2) indirect costs (working time). The estimated direct cost of one DE per participant was allocated from the total acquirement cost of the DE license. CE of the DE was evaluated with a CE ratio: cost of additional (incremental) intervention and the additional outcome (effect caused by the outcome, i.e., knowledge level).

Results

Description of the participants

In total, 94 professionals (=N) responded to the questionnaire, corresponding approximately 5% response rate.

The age range was from 23 to 64 years with a mean of 42. There were no great age differences between groups. The overall average of experience was 17 years, but the group with half-a-day LE and DE had lower average experience in healthcare industry than other groups. The participants had worked in the organization for 12 years on average, shortest in the group of DE and half day LE, longest in the group with only one day LE. On average, the respondents had worked in their current department for 10 years. Most of the participants were females (98%) and registered nurses (66%). The frequency of practical wound care work was divided more evenly between participants, and 52% announced to work with wounds at least once a week, 23% at least once a month and the rest less frequently. Total 20 respondents (21%) were wound care nurses. In DE, 60% had completed at least one of the module, and 33% had accomplished the organization's one-day LE about basics of wounds. Total 14% of the respondents had accomplished the half-day LE, which was available from 2017. The rest of the respondents (53%) had not taken part in the wound care LE in their organization.

Knowledge test outcomes

The distribution of the knowledge test scores is close to normal. The variation was -2 to 20 (out of 27). The average score was 10.05 points. The most frequently earned score was eight points. The more education the participant had, the higher the knowledge test score was. All the score averages in DE groups were higher than in the groups without any DE. Also, the longer the LE had been (one-day > half-day > no LE), the higher the scores were. The one-day LE together with the DE resulted in more points than the half-day LE and the DE combined (11.90 vs. 11.10). If the participant had only carried out the DE alone, the average (9.4) was still better than if the participant had carried out only the one-day LE alone (average 9.2). However, this difference was not statistically significant.

Statistical significance of the effect on knowledge was tested for six variables (research group, LE, work experience, wound care frequency, DE, and wound care nurse). The group with an accomplished DE had higher score average than without DE (10.7 vs. 8.9, p=0.035), and wound care nurses' average was higher than other respondents (12.1 vs. 9.5, p=0.024).

Cost estimation

The direct cost for the DE calculation was 0.55€ per person per one module. The referential time to execute the DE is three hours in total. Because mostly nurses are participating the DE and LEs, the cost of nurses is used to identify the indirect cost: it is counted from the average cost of a nurse per hour. The average cost (incl. side costs, excl. holiday compensations) is 17.85€ per hour.

The direct cost of LE per nurse consists mostly of the compensation for lecturer. The direct cost estimation per person for the one-day LE is 50€ and for the half-day 30€. The length of one working day is approximately seven, five hours and for half-day course it is four hours.

Cost-effectiveness analysis

DE is inexpensive and has good learning results, thus the DE cost (55.19€) per outcome (9.4) ratio was 5.87€ per knowledge unit, and the relative CE ratio was 0.29. The one-day LE together with the DE achieved the best absolute outcome (11.9). However, this combination was the most expensive (239.02€). Thus the cost per outcome ratio was 20.09€ and and relative CE ratio was 1.

The half-day education is not typically executed alone without the DE, so it is irrelevant to be analyzed individually. The cost (156.57€) per outcome (11.10) ratio of a combination of the DE and a half-day LE was 14.10€ per one knowledge unit, and the relative CE ratio was 0.71. The cost for one-day education alone was 183.84€ and the outcome was 9.2. Thus the cost per outcome ratio was 19.98€ and relative CE ratio was 1.

The incremental cost-effectiveness (ICE) ratio was defined as the ratio between the additional cost and additional outcome. The additional outcome refers to the knowledge level of participants with education minus with participants with no wound education (8.9). The score from the DE is 9.4, so the additional effect on learning is 0.5 units. The additional learning was highest in DE and one day LE (3). For one day LE additional learning was 0.3, and in combination of DE and half-a-day LE 2.2. ICE ratio for DE was 110.37€, for DE combined with half-a-day LE 71.17€, for one day LE 612.79€ and for combination of DE combined with one day LE 79.67€.

Conclusions

The relationship of wound education and wound knowledge is proved to be positive based on the results of this study. The more education, the higher knowledge test scores. DE can enable cost savings, increase the amount of professionals trained or enable better learning outcomes. Even though DE performed well in terms of CE, the additional LE improved the outcomes. With the comparison of ICE ratios, the optimal choices for organizations would be the combinations of DE and either half-day or one-day LE.

However, the decision between the learning methods does not need to be fixed. A relevant response to the results is segmentation. Professionals could be segmented based on the different levels of requirements concerning wound care. The priorities of an organization and value judgments between cost and learning must be taken into consideration along with CE analysis.

Acknowledgements

This work was supported by a grant from Business Finland. We also acknowledge the support Kainuu Social and Health Care Joint Authority and Duodecim Medical Publications Ltd. provided us.

References

[1] F. Gottrup, J. Apelqvist, and P. Price, Outcomes in controlled and comparative studies on non-healing wounds: recommendations to improve the quality of evidence in wound management, *J Wound Care* **19** (2010), 239–268.

[2] D. Zhang, and J. F. Nunamaker, Powering e-Learning in the new millennium: An overview of e-Learning and enabling technology, *Inform Syst Front* **5** (2003), 207–218.

[3] P.M Sinclair, A. Kable, T. Levett-Jones, and D. Booth, The effectiveness of Internet-based e-learning on clinician behaviour and patient outcomes: A systematic review, *Int J Nurs Stud* **57** (2016), 70–81.

Address for correspondence

Anna-Maria Hiltunen, anna-maria.hiltunen@aalto.fi, +358404109205

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190721

Virtual Student Collaboration: Connecting Student Health Professionals

Michelle Honey[a], Kim Young[b], Hazel Cowls[b]

[a] *School of Nursing, University of Auckland, Auckland, New Zealand*
[b] *School of Nursing & Midwifery, University of Plymouth, Plymouth, United Kingdom*

Abstract

This pilot study aimed to provide students with a learning activity to develop a global perspective of health care. Senior nursing students from New Zealand and the United Kingdom (n=15) were allocated to eNetwork small-groups. Students met weekly for one month via audio/video-conferencing for discussions. Students reported that this collaborative activity enabled them to engage purposefully with other students. While cultural differences were noted, similarities were also found and multiplestudents expressed interest in working overseas.

Keywords:

Students, Nursing; Pilot Projects; Videoconferencing

Introduction

Information and communication technologies (ICT), such as Skype, can provide opportunities for engagement across the globe [1]. The ways in which individuals communicate within their personal and professional lives is rapidly changing [2]. Health professional students, including nurses, are considered 'digital natives' as they commonly utilise a range of internet-sourced communication media to support their knowledge development and interaction with others [3]. Extending the use of ICT provides an opportunity for health professional students to colaborate demonstrating international eNetworks in action. This poster describes a pilot which aimed to provide students with a learning activity to give a more global perspective of nursing and health care.

Advances in social media can facilitate opportunities to develop understanding of different countries and cultures. Nursing has a global presence, yet students can have little knowledge of the health and social care needs and provision outside their local environment and country [4]. Exchange activities are well-established in higher education and facilitate excellent opportunities for intercultural learning. Such scholarly collaborations are usually abroad. The outcomes for student exchange are often related to global appreciation, cross-cultural knowledge and the skills required in a diverse nursing workplace. These attributes are important for all health professionals given the cultural dexterity required in the delivery of health and social care [5].

The faculties at both schools of nursing who participated in this pilot study already had experience in using social media activities with students [6,7]. This project sought to increase student participation and collaboration using technological media to facilitate opportunities to collaborate and learn from each other. It was designed to help prepare nurses for their professional career in a global environment where eNetworking will be valuable to address diverse population health issues.

Methods

The aim of this pilot study was to determine student nurses' perceptions of an international collaborative learning activity mediated by ICT. Participants were volunteers drawn from students enrolled in the third year of their undergraduate nursing programmes at the University of Plymouth in the United Kingdom (UK) and the University of Auckland in New Zealand (NZ). The nursing students were placed in one of three groups, with each group having three UK and two NZ students. Ethical approval was obtained from each university.

Students were asked to contribute during four, 30-minute audio/video meetings (using Zoom) with their small group during February 2018. Discussion guidelines were provided for each week's meeting.

As this was the first time this learning opportunity was offered, the researchers were interested in the students' experience of the small group meetings and what they may have learnt from virtually meeting students from a different country. An anonymous evaluation survey comprised of open- and closed-ended questions was emailed to each student as a link to Microsoft Forms at the conclusion of the month-long learning activity. The evaluation form included ten Likert-scale statements. In addition, there were three open-ended questions: *What aspect of the collaborative learning activity was most helpful for your learning? What improvements would you like to see?* and *What, if anything, do you intend to do as a result of participating in this collaborative learning activity?* Data analysis comprised collating the Likert items and thematic analysis for the open ended responses.

Results

Nine students volunteered from the University of Plymouth, and six volunteered from the University of Auckland. Of the 15 participants, 13 responded to the survey (87% response rate). All students agreed (31%) or strongly agreed (69%) that this collaborative learning activity enabled them to engage purposefully with other nursing students. Similarily, all students agreed or strongly agreed with the statement: 'I would recommend other student nurses participate in future collaborative learning activities with international nursing students.

Thematic analysis of open-ended responses identified four themes:

1. Cultural differences

2. International connections & future employment
3. Benefits of e-engagement
4. Limitations of e-engagement

Overall students decsribed this pilot as “an interesting experience” and that it was “positive to engage with students from another country”. However, eNetworking did not negate the idea that a “lovely travel grant to enable us to meet face to face” would be ideal.

Discussion

This study describes an international collaborative virtual engagement activity, reports, the students’ experiences, and identifies practical issues. The findings from this study confirm that ICT provides students with opportunities to gain a global perspective of nursing and healthcare, concuring with earlier studies which also identified that students can gain experience and intercultural skills without having to leave their homes [8-10]. Internationalisation is high on the agenda for tertiary education institutions as they produce graduates that are global citizens [11]. The dialogue between students brought about an interest in future employment and working overseas which supports some intriguing results from web-based interaction with international peers who were ready for international work [8].

Students appeared to enjoy the virtual engagement activity making international connections and seeing friendships beginning to form. This was despite the inconvenience of time differences between the two countries, although time differences are a known issue when two hemispheres, or multiple countries are involved [6,12]. It was evident that the students developed an awareness of similarities, as well as differences, within healthcare provision, thus enhancing their understanding of health with a global perspective, which has been noted in other uses of virtual student engagement [4,6].

The limitations of this study were the small number of participants involved. Areas for further research include maintaining contact with the participants to see whether their eNetworks developed over time. Also, as it has been recognised that newly qualified nurses may experience mixed emotions upon qualifying from a sense of achievement to apprehension [13], it would be interesting to establish whether the challenges facing newly qualified nurses are the same globally. Furthermore, students who participated in this study may be able to support each other as newly qualified nurses.

Conclusions

Health professionals are part of global community, so encouraging and supporting virtual student interaction can help to connect professionals develop a broad perspective on health care. eNetworking amongst the students enabled identification of cultural differences and similarities in their programmes of study, health care systems, and issues being faced. Barriers included using synchronous audio/video-conferencing and the ensuing problems of managing time differences.

References

[1] J. Oates, Use of Skype in Interviews: The Impact of the Medium in a Study of Mental Health Nurses, *Nurse Researcher* **22,** (2015), 13-17.

[2] S. Weller, *The Potentials and Pitfalls of Using Skype for Qualitative (Longitudinal) Interviews: NCRM Working Paper,* National Centre for Research Methods, Southhampton, UK, 2015.

[3] S. Bennett, K.A. Maton, L. Kervin, The Digital Natives Debate: A Critical Review of the Evidence, *British Journal of Educational Technology* **39,** (2008), 775-786.

[4] P.M. Procter, Brixey, Juliana J., F. Todhunter, M.L.L. Honey, Social Media Providing an International Virtual Elective Experience for Student Nurses, *Informatics* **4,** (2017), 1-8.

[5] F. Todhunter, B. Hallawell, D. Pittaway, Implementing a Virtual Exchange Project for Student Nurses in Queensland and Nottingham, *Nurse Education in Practice* **13,** (2013), 371-376.

[6] D. Greenwood, M. Honey, A. Clancy. Preliminary Findings of an International Public Health Nursing Student Collaboration. Paper presented at: *3rd International Public Health Nursing Conference: Making the Difference*; August 25 -27, 2013; Galway, Ireland.

[7] R. Jones, J. Kelsey, P. Nelmes, N. Chinn, T. Chinn, T. Proctor-Childs, Introducing Twitter as an Assessed Component of the Undergraduate Nursing Curriculum: Case Study, *Journal of Advanced Nursing* **72,** (2016), 1638-1653.

[8] E. Carlson, M. Stenberg, B. Chan, et al., Nursing as Universal and Recognisable: Nursing Students’ Perceptions of Learning Outcomes from Intercultural Peer Learning Webinars: A Qualitative Study, *Nurse Education Today* **57,** (2017), 54-59.

[9] B. Wachter, An Introduction: Internationalisation at Home in Context, *Journal of Studies in International Education* **7,** (2003), 5-11.

[10] M. Wihlborg, E.E. Friberg, K.M. Rose, L. Eastham, Facilitating Learning Through an International Virtual Collaborative Practice: A Case Study, *Nurse Education Today* **61,** (2018), 3-8.

[11] C. Shiel, Developing the Global Citizen, *The Higher Education Academy: Academy Exchange* **5,** (2006), 18-20.

[12] M.L.L. Honey, I. Doherty, A.M. McKillop, Factors for a Successful International Collaboration to Develop an Evidence Based Practice E-Learning Course, In: Luca J, Weippl ER, eds. *ED-MEDIA 2008: World Conference on Educational Multimedia, Hypermedia & Telecommunications*Vienna, Austria, June 30 - July 4, 2008, 2597-2602.

[13] S. Kumaran, M. Carney, Role Transition from Student Nurse to Staff Nurse: Facilitating the Transition Period, *Nurse Education in Practice* **14,** (2014), 605-611.

Address for correspondence

Michelle Honey
m.honey@auckland.ac.nz

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190722

Factors Affecting Smartphone Usage Self-Report Levels

In Hye Yook[a,b], So Jin Park[c], Mun Joo Choi[a,b], Dai-Jin Kim[c], In Young Choi[a,b]

[a] Department of Biomedicine & Health Sciences, The Catholic University of Korea, Seoul, Republic of Korea
[b] Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea
[c] Department of Psychiatry, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

Abstract

For the treatment of smartphone addiction, it is important to understand users' smartphone usage patterns. Most of the studies are based on self-report surveys. However, there are differences between self-reported usage and real usage. For a better understanding of usage patterns, this study identified demographic and social factors that affect smartphone usage self-report levels. Also, it was confirmed that the influencing factors differ depending on the smartphone usage content by application category.

Keywords:
Smartphone, Self Report, Data Accuracy

Introduction

As smartphone addiction becomes a social problem, it is necessary to understand how people use smartphones. Most of the previous studies were based on self-report surveys, and the diagnosis and treatment of smartphone addiction also relied on self-reports [1,2].

Some studies using both self-report surveys and real log data showed there was a difference between them [3,4,5]. However, there was a lack of confirmation as to which factors made that difference. Since self-report usage is a major indicator of the diagnosis of smartphone addiction, it is necessary to identify the cause of the difference between actual and self-reported usage.

In this study, we used self-reported usage, real usage, demographic, and sociological information to find factors influencing the under or over estimation of smartphone usage. We also identified whether the factors are different for the under or over estimation of smartphone usage depending on the application (app) category.

Methods

We developed a mobile application system to collect the mobile use log. The system was implemented for four weeks to 382 participants who lived in a metropolitan area. At the same time, we obtained all participants' self-reported usage, demographic, and social information through online surveys.

Then we categorized participants to the under or over estimation group using self-reported and real smartphone usage. We used a Chi-square test to find factors influencing the under or over estimation of smartphone usage. In addition, to figure out whether a factor was different depending on the content, the top five categories of smartphone usage were analyzed.

Results

There was no perfect match between self-reported usage and real usage. For this reason, people were divided into two groups: under or over estimation of smartphone usage. Table 1

Table 1 – Self-report levels by factor (N=382)

Factor		Self-Report (%)		χ^2
		Under	Over	
Gender	Male	62 (29.2)	150 (70.8)	8.55***
	Female	28 (16.5)	142 (83.5)	
Age	20's	20 (18.0)	91 (82.0)	2.73
	30's	34 (25.2)	101 (74.8)	
	40's	36 (26.5)	100 (73.5)	
Marriage	Single	35 (21.1)	131 (78.9)	1.26
	Married	54 (25.7)	156 (74.3)	
	Divorce/ Separation	1 (16.7)	5 (83.3)	
Education	High school	22 (27.5)	58 (72.5)	0.87
	Graduate school	68 (22.5)	234 (77.5)	
Family emotional support	Never	3 (10.7)	25 (89.3)	5.01*
	Sometimes	47 (21.9)	168 (78.1)	
	Usually	40 (28.8)	99 (71.2)	
Bullied	Have	11 (14.5)	65 (85.5)	4.35**
	None	79 (25.8)	227 (74.2)	
Suicidal impulse within a year	Have	20 (17.2)	96 (82.8)	3.69*
	None	70 (26.3)	196 (73.7)	
Family history of mental disorders	Have	4 (12.5)	28 (87.5)	2.37
	None	86 (24.6)	264 (75.4)	

***: <.1, **: <.05, ***: <.01**

shows the number of people who are in the under or over estimation group by factor. We found four factors, Gender, Bullied, Family Emotional Support, and Suicidal impulse within a year, were related to an over estimation of smartphone usage. Females, people without Family emotional support, people who had experiences of being Bullied, and people who had a Suicidal impulse within a year had a greater over estimation of their smartphone usage.

Table 2 – Influencing factor by app category (N=382)

Factor		SNS (%)		χ^2
		Under	Over	
Age	20's	9 (8.1)	102 (91.9)	610**
	30's	19 (14.1)	116 (85.9)	
	40's	26 (19.1)	110 (80.9)	
Bullied	Have	4 (5.3)	72 (94.7)	5.28**
	None	50 (16.3)	256 (83.7)	
Factor		**Game (%)**		χ^2
		Under	**Over**	
Suicidal impulse within a year	Have	24 (20.7)	92 (79.3)	9.77**
	None	100 (37.6)	166 (62.4)	
Family history of mental disorders	Have	24 (20.7)	92 (79.3)	9.77**
	None	100 (37.6)	166 (62.4)	
Factor		**Entertainment (%)**		χ^2
		Under	**Over**	
Family emotional support	Never	6 (21.4)	22 (78.6)	5.39*
	Sometimes	51 (23.7)	164 (76.3)	
	Usually	19 (13.7)	120 (86.3)	
Factor		**Web (%)**		χ^2
		Under	**Over**	
Gender	Male	34 (16.0)	178 (84.0)	4.54**
	Female	14 (8.2)	156 (91.8)	
Marriage	Single	23 (13.9)	143 (86.1)	8.74**
	Married	22 (10.5)	188 (89.5)	
	Divorce/ Separation	3 (50.0)	3 (50.0)	
Factor		**Finance (%)**		χ^2
		Under	**Over**	
Age	20's	43 (38.7)	68 (61.3)	4.69*
	30's	40 (29.6)	95 (70.4)	
	40's	57 (41.9)	79 (58.1)	
Bullied	Have	17 (22.4)	59 (77.6)	7.58**
	None	123 (40.2)	183 (59.8)	
Family history of mental disorders	Have	6 (18.8)	26 (81.2)	4.01**
	None	134 (38.3)	216 (61.7)	

***: <.1, **: <.05, ***: <.01**

Table 2 shows the factors affecting the self-report level in the top five app categories of Social Network Service (SNS), Game and Entertainment, Web, and Finance, of smartphone usage. The self-reported level of SNS was influenced by Age and experience of being Bullied. Game and Entertainment was influenced by the social factors of Suicidal impulse within a year and Family history of mental disorders. The Web was influenced by Gender and Marriage while Finance was influenced by Age, experience of being Bullied, and Family history of mental disorders.

Conclusions

As a result of the research, we found there are different influencing factors between self-reported levels of total smartphone usage and self-reported levels of the content. For future work, it will be helpful to consider this difference. In addition, it is presumed social factors are closely related to the self-reported levels of smartphone usage. In future research, it will be necessary to study this correlation.

Acknowledgements

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014M3C7A1062893).

References

[1] M. Kwon, D.J. Kim, H. Cho, and S. Yang, The smartphone addiction scale: Development and validation of a short version for adolescents, *Public Library of Science One* **8** (2013) e83558.

[2] H. Kim, Exercise rehabilitation for smartphone addiction, *Journal of Exercise Rehabilitation* **9** (2013) 500-505.

[3] J. Boase and R. Ling, Measuring mobile phone use: Self-report versus log data, *Journal of Computer-Mediated Communication* **18** (2013) 508-519.

[4] C. Montag, K. Blaszkiewicz, B. Lachmann, R. Sariyska, I. Andone, B. Trendafilov, and A. Markowetz, Recorded behavior as a vlauable resource for diagnostics in mobile phone addiction: Evidence from psychoinformatics, *Behavioral Sciences (Basel, Switzerland)* **5** (2015) 434-442.

[5] H. Lee, H. Ahn, T.G. Nguyen, S.W. Choi, and D.J. Kim, Comparing the self-report and measured smartphone usage of college students: A pilot study, *Psychiatry Investigation* **14** (2017) 198-204.

Address for correspondence

In Young Choi

222 Banpo-daero, Seocho-gu, Seoul 06591, Korea

Tel: +82-2-2258-7648, E-mail: iychoi06@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190723

An Efficient Simulation-Based Optimization Approach for Improving Emergency Department Performance

Ibtissem Chouba[a,b], Farouk Yalaoui[a], Lionel Amodeo[a], Taha Arbaoui[a], Philippe Blua[b], David Laplanche[b], Stéphane Sanchez[b]

[a] *Institut Charles Delaunay, LOSI, Université de Technologie de Troyes, FRE2019, CNRS, 12 Rue Marie Curie, CS42060, Cedex, 10004, Troyes, France*
[b] *Département d'Information Médicale, Centre Hospitalier de Troyes, 101 Avenue Anatole, 10000, Troyes, France*

Abstract

In recent years, health care organizations, in particular emergency department (ED), have come under increasing pressure to provide quality care. In this context, human resources are a central aspect: a good utilization of health worker could improve quality of care. In this paper, a simulation model is proposed. The model represents an ED coupled with an optimization method to optimize the allocation of medical and para-medical human resources in the hospital center of Troyes. We aim to improve the quality of services offered to patients through the minimization of Average Waiting Time (AWT) and Average Inpatient Stay (AS). The proposed approach has proved to be effective to reduce AWT and AS by 12 minutes and 21 minutes respectively.

Keywords:

computer simulation, resource allocation, emergency departments

Introduction

In recent years, health care organizations, in particular emergency departments, have come under increasing pressure to provide quality care while overcoming increasing volume pressure. In the other side, simulation has become an effective decision-making tool to optimally allocate the often-limited resources in health care to improve patient flow while minimizing the costs of providing care and increasing patient satisfaction. In addition, combined optimization and simulation tools allow decision makers to efficiently find optimal emergency department configurations, even for complex integrated installations.

Over the last decade, there have been several studies in the literature dealing with the optimization of human and material resources in a health care through the technique of coupling simulation with optimization [1, 2, 3, 4].

In this paper, in order to improve the quality of services offered to patients in the ED of the hospital center of Troyes (CHT) (France), a simulation-based optimization approach is developed. This method uses a discrete-event simulation model, combined with a genetic approach. This tool allows the users to evaluate the Average Waiting Time (AWT; the time that the patient spends from their arrival to their first medical exam) and the Average Inpatient Stay (AS; the time the patient spends from the first exam to their exit) of patients for different organizations of the service. The expected value of patient arrivals is obtained through forecast models developed in a previous work, and that reached a very good performance (up to 91.24 % for the annual total flow forecast) and robustness to epidemic periods [5].

Methods

Simulation model

The developed simulation model requires three kinds of inputs: 1) patient arrival rates, 2) the frequency of each pathway, and 3) the processing times of services. These were obtained from data collected from the emergency database for the whole 2017 year. We modeled the considered probability distribution using the "fitdistrplus" package of R® software [6]. These data were used to determine the best probability distribution to represent each stage of the patient flow. The time of inter-arrival process (T_i) follows an exponential distribution with average of $Y_j * T_h$, i.e $T_i \sim e^{Yj * Th}$, where T_h is the average arrival rate observed for the hour h of the day, and Y_j the number of daily arrivals for the day j.

Optimization

The optimization problem considered in this work aims to minimize the AWT and the AS of patients. In order to reach the optimization goal, the problem was solved with a genetic algorithm coupled with the simulation model (as an objective function evaluator) in order to optimize emergency staff planning. Thus, the optimization problem can be represented as follow:

Minimize $F(x)$

Subject to

$$\sum_{j=1} X_{ij} = Ci \qquad \forall i = 1 \dots 4 \quad (1)$$

$$X_{ij} \in N \qquad \forall i = 1 \dots 4 \;\; \forall j = 1 \dots 4 \; (2)$$

Where X_{ij} is the decision variable that represents the number of available staff per slot time of 6h. $i(= 1 \dots 4)$ is the index of staff type: 1 for emergency physicians, 2 for resident physicians, 3 for nurses and 4 for caregivers, and $j = 1 \dots 4$ is the index of the slot time of 6h per day. The constraint (1) ensures that the same number of hours worked by each type of staff is fixed because the ED of CHT, as a first step, does not seek to increase or decrease its workforce.

Table 1 shows the availability of interchangeable staff per 6 hour time slot.

To define the best parameters of the genetic approach, the experimental design methodology proposed by [7] is used. Then, the parameters used for our GA are: population size = 100, probability of mutation = 0.9, number of maximal iterations = 500 and probability of crossover = 0.9. The time limit was set to 1h30

Table 1: Availability of staff per slots of 6 hours

	00h -6h	**6h - 12h**	**12h - 18h**	**18h - 00h**
Emergency physician	2	2	2	2
Resident physician	3	3	3	3
Nurse	3	3	3	3
Caregiver	3	3	3	3

Results

To assess the efficiency of the proposed simulation model, the performance indicators obtained from our simulation model were given to CHT's decision makers who compared them to daily indicators of the hospital. The assessed gap between real and simulated indicators is less than 9%. Based on this assessment, the model was considered reliable and able to support experiments.

The best solution of the emergency staff planning returned by the genetic approach is presented in table 2. In terms of performances, the solution reduces the AWT by 12 minutes and the AS by 21 minutes.

Table 2:Optimized solution of the annual planning

	00h - 6h	**6h - 12h**	**12h - 18h**	**18h - 00h**
Emergency physician	1	1	3	3
Resident physician	2	2	4	4
Nurse	2	2	5	3
Caregiver	1	2	4	3

Discussion and Conclusion

After improving the current staff planning by keeping the same number of staff, the effect of increasing the number of staff by 1 for each type is studied.

Results show that the nurse resource has the greatest impact on the system. Adding a nurse allows the hospital to reduce the AWT by 25 minutes and the AS by 32 minutes.

To conclude, a simulation-based optimization approach that aims to minimize the AWT and the AS for patients in the hospital center of Troyes, France is presented. The proposed solution provides a gain of 12 minutes in AWT of patient and a gain of 21 minutes in AS which represents a noticeable reduction in waiting times for patients.

This work can be extended to include the planning of human resources, in which the assignment of resources can further reduce waiting time by effectively assigning resources on overloaded shifts.

References

[1] K. Ghanes, O. Jouini, Z. Jemai, M. Wargon, R. Hellmann, V. Thomas, and G. Koole, A comprehensive simulation modeling of an emergency department: A case study for simulation optimization of staffing levels. *Proceedings of the 2014 Winter Simulation Conference* (2014) (pp. 1421-1432).

[2] M. A. Ahmed, and T. M. Alkhamis, Simulation optimization for an emergency department healthcare unit in Kuwait. *European journal of operational research*, *198* (2009), 936-942.

[3] D. Steward, T.F. Glass, and Y.B. Ferrand, Simulation-based design of ed operations with care streams to optimize care delivery and reduce length of stay in the emergency department. *Journal of medical systems*, *41* (2017), 162.

[4] B. Majid, H. Pasha Abgarmi, S. Yousefi, A. Alizadeh, and H. Mahmoudzadeh, Simulation-based optimization for improving hospital performance. *International Journal of Hospital Research*, 6, (2017), 72-84.

[5] M. Afilal, F. Yalaoui, F. Dugardin, L. Amodeo, D. Laplanche, and P. Blua, Forecasting the emergency department patients flow. *Journal of medical systems*, 40, (2016), 175.

[6] M. L. Delignette-Muller, and C. Dutang, fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, 64, 2015, 1-34.

[7] M. Pillet, *Les plans d'expériences par la méthode Taguchi* (2012), p. 250.

[8] Davis L. Handbook of genetic algorithms, (1991).

Address for correspondence

Chouba Ibtissem

ibtissem.chouba@utt.fr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190724

The SNIK Graph: Visualization of a Medical Informatics Ontology

Franziska Jahn[a], Konrad Höffner[a], Birgit Schneider[a], Anna Lörke[a], Thomas Pause[a], Elske Ammenwerth[b], Alfred Winter[a]

[a] Institute of Medical Informatics, Statistics and Epidemiology, Leipzig University, Germany
[b] Institute of Medical Informatics, UMIT – University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria

Abstract

SNIK, a medical informatics ontology, combines knowledge from different literature sources dealing with the management of hospital information systems (HIS). Concepts and relations were extracted from literature, modeled as an ontology and visualized as a graph on a website. We demonstrate the potential of the graph visualization for tuitional scenarios. SNIK complements teaching and learning with conventional literature by concentrating knowledge that is scattered over different pieces of text around one node of a graph.

Keywords:
Semantic Web, Information Management, Hospital Information Systems

Introduction

Although computer scientists have provided theory and applications for ontologies, they rarely apply ontologies to their own fields of research. There are well-known biomedical ontologies [1], but there are almost no ontologies structuring medical informatics knowledge. With SNIK, the semantic network of information management in hospitals (German: Krankenhäuser) we developed an ontology that describes HIS management from a functional point of view: Who performs which function and which information is needed or updated by this function? The role "CIO", for example, is responsible for the function "strategic information management" which updates the information artifact "strategic information management plan". Statements like these are coded as RDF (Resource Description Framework) triples in SNIK.

A typical use case of SNIK is teaching HIS management at universities [2]. Due to its use in teaching, we needed a way to visualize the ontology, so that students and lecturers can browse through the knowledge and can easily discover how different concepts are linked to each other. In previous work [3], we tested the visualization of SNIK with the help of graph visualization tools. Although offering useful functionalities, these desktop tools are not easy to handle by occasional users. Moreover, the ontology could only be used after having downloaded the latest SNIK file and having installed the tool. Now, with SNIK being available as Linked Open Data together with a web-based graph visualization based on cytoscape.js [4], users can easily analyze the linked concepts of HIS management. It is the aim of this work to demonstrate;

- the visualization capabilities of web-based SNIK ("SNIK graph") and
- their use for specific teaching and learning scenarios.

Methods

SNIK is based on knowledge from text books, frameworks, and expert interviews dealing with HIS and their management. As a first step to develop SNIK, literature was selected in accordance with the learning objectives of lectures on HIS management at Leipzig University. In these lectures, students shall learn a clear terminology for roles, functions, and information artifacts of HIS management. For its use in teaching, SNIK had to provide answers to the questions "Who performs which HIS management function and which information is needed or updated by these functions?" Thus, the SNIK metamodel was specified, which, determines the rules for extracting knowledge from texts. Relevant concepts found in texts have to be assigned to one of the metamodel classes "role", "function" or "entity type", the latter representing types of information artifacts. For linking concepts to each other, 16 different relationships were defined. Extraction from the texts was then done pairwise by reading through a chapter and extracting knowledge to spreadsheets with predefined columns. Each pair of extractors prepaired a consolidated spreadsheet for a chapter. All spreadsheets were checked again by another group in order to have one consolidated .csv-file containing the textbook's subontology. In addition, interlinks between different subontologies were specified within a further spreadsheet. The .csv-files were then converted to RDF, published on the SNIK website, [5] and the extracted knowledge became a part of the SNIK graph.

Results

The SNIK graph, which represents the current status of the SNIK SPARQL endpoint, consists of 2,845 entity types, 1,090 functions, and 234 roles (16.11.2018). It comprises knowledge from five different sources which are linked to each other by 704 interlinks (Figure 1).

The following teaching and learning scenarios can be visualized within the SNIK graph (http://www.snik.eu/graph):

(1) A teacher prepares a lecture on strategic HIS directing. He or she uses the "circle star" visualization for the function "Strategic HIS Directing" and gets a mind map to structure the lecture (Figure 2).

(2) A student has decided to be a project manager in a students' hospital project and wants to know his or her responsibilites. The "circle star" around the role "project manager" shows 24 functions and 8 entity types from three different sources a project manager is responsible for.

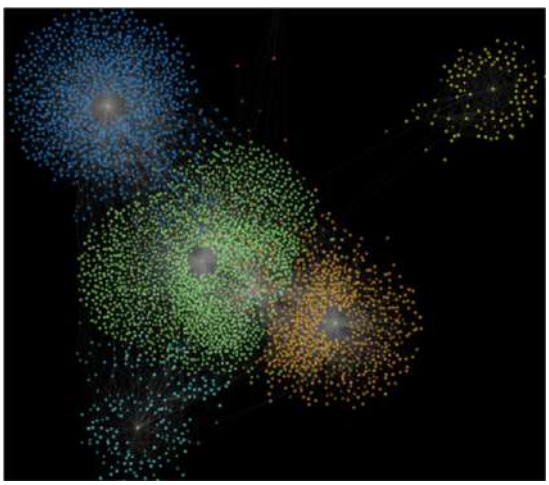


Figure 1 – Five colored clouds representing subontologies of SNIK. The clouds contain knowledge from three books (dark blue, green and orange), a CIO interview (light blue) and a framework (yellow).

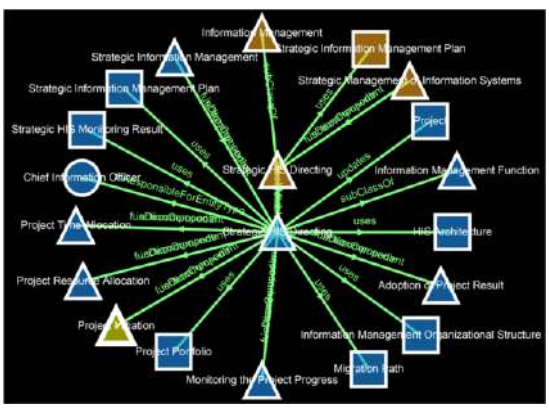


Figure 2 - Circle star for the function "Strategic HIS Directing" (rectangles represent entity types, triangles represent functions and circles represent roles).

(3) Students learn new concepts about HIS quality by linking them to concepts already learned. A teacher asks a student to find out how the new concept "Quality of Data" is linked to the "Patient Identification Number". The student connects the two concepts by using the spiderworm visualization and learns that a patient identification number is associated with object identity. Object identity is a subclass of integrity of data. Besides integrity of data, there are also 13 other criteria for quality of data (Figure 3).

Conclusions

The SNIK graph enables students and teachers to browse through the knowledge of HIS management on a website. With the help of the visualization capabilities like "circle star" and "spiderworm" specific questions arising during learning, the preparation and the execution of lectures can be answered. The SNIK graph complements texts on HIS management by a highly interconnected map. Knowledge about a concept, which is often scattered over different chapters or even books, can be found around one node of the graph visualization. Other possibilities of teaching support by SNIK, like the automatic generation of multiple choice tests (e.g [6]), are subject of further research. Nevertheless, due to its size, SNIK is not free of errors. Hence, every user can participate in the continuing quality assurance process and report bugs directly within the graph visualization.

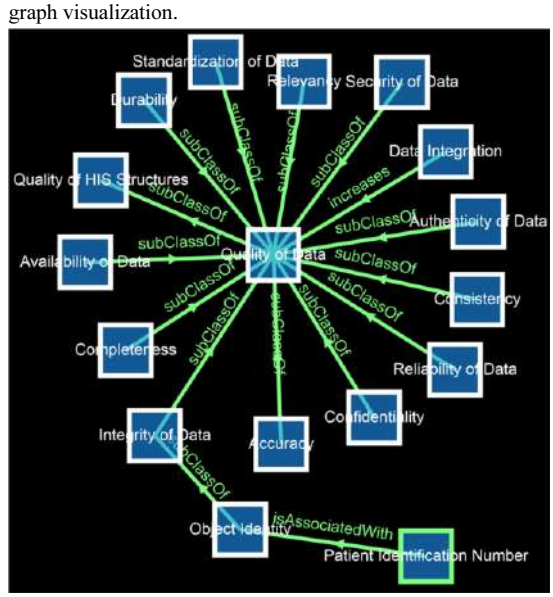


Figure 3 - A spiderworm connecting "Patient Identification Number" and "Quality of Data"(surrounded by its neighbors).

Acknowledgements

This work was supported by the DFG (German Research Foundation) under the Project SNIK, grant no. 1605/7-1.

References

[1] J.J. Cimino and X. Zhu, The practical impact of ontologies on biomedical informatics. *Yearb Med Inform* (2006), 124–135.

[2] F. Jahn, M. Schaaf, C. Kahmann, K. Tahar, C. Kücherer, B. Paech, and A. Winter, An ontology-based scenario for teaching the management of health information systems. In *Exploring complexity in health: Proceedings of MIE2016 at HEC2016,* A. Hoerbst, ed. IOS Press, Amsterdam, Netherlands, 2016, 359–363.

[3] M. Schaaf, F. Jahn, K. Tahar, C. Kücherer, A. Winter, and B. Paech, The visualization of large ontologies from a tool point of view. In *Exploring complexity in health: Proceedings of MIE2016 at HEC2016,* A. Hoerbst, ed. IOS Press, Amsterdam, Netherlands, 2016, 349–353.

[4] M. Franz, C.T. Lopes, G. Huck, Y. Dong, O. Sumer, and G.D. Bader, Cytoscape.js: a graph theory library for visualisation and analysis. Bioinformatics **32** (2016), 309–311.

[5] The SNIK Team, SNIK Website, http://www.snik.eu [cited 2018 November 12].

[6] M. Al-Yahya, Ontology-based multiple choice question generation. The Scientific World Journal (2014), 1–9.

Address for correspondence

Franziska Jahn, Leipzig University, Faculty of Medicine, IMISE, Härtelstr. 16-18, 04107 Leipzig; franziska.jahn@imise.uni-leipzig.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190725

Implementation of an Inpatient Portal Integrated to an EHR: First Stage Evaluation

Liliana Giraldo[a], Janine Sommer[a], Eunice Correa[a], Mariana Simón[a], María Grande[b], Santiago Márquez Fosser[a], Fernando Plazzotta[a], Daniel Luna[a]

[a]*Department of Health Informatics, Hospital Italiano de Buenos Aires, Ciudad de Buenos Aires, Argentina*
[b]*Department of Internal Medicine, Hospital Italiano de Buenos Aires, Ciudad de Buenos Aires, Argentina*

Abstract

Technology has led to the communication of the hospitalized patient and their family with clinical knowledge. The objective of this study is to describe the rate of use of the Inpatient Personal Health Record of Hospital Italiano de Buenos Aires. A cross-sectional study was conducted between May and September of 2018. The rate of use was 2.10% (95% CI 1.89-2.32). The most used functionalities were analyzed. New strategies are needed to improve the rate of use.

Keywords:

Personal Health Records, Inpatients, Electronic Health Records

Introduction

Health organizations are moving towards greater health information technology, allowing more comprehensive patient care, and improving communication within the health team. Often the implementation of these technologies in hospitals has been directed mainly to health personnel, more specifically, to the health care professional instead of the patients and their families [1,2].

The scope of 21st century medical care has generated strategies for Effective Health Care, and one of the established requirements is the "Empowerment of patients and their families in the effective management of health care decisions, and their implementation, including personal health records (PHRs), education about the individuals conditions and options, and support of timely and focussed communication with professional health care professionals" [3]. This scope is not only necessary in the home and ambulatory care environments, when the patient is hospitalized, it should also be considered.

A PHR is one of those tools, which provides convenient access to personal health information from any location with Internet access [4]. The development of online PHRs involved the patient in the process of medical care, generally in the outpatient environment [5]. The initiatives of inclusion of health information technologies in the hospitalization field, encourage the institutions to offer patients electronic access to some clinical data [6,7]. Interventions that provide patients with clinical information have been effective in promoting patient participation in health-related decision-making, increasing the patient's adherence to their own care plans [8,9].

The Department of Health Informatics developed and implemented an Inpatient Portal (IP), which is being used since May 2018. This work aims to describe the rate of use of the IP and the characteristics of the users in our Institution from the implementation to the present.

Methods

Design

A cross-sectional study was carried out that included all admissions of adult patients, hospitalized during the period between May 20 and September 30, 2018, at Hospital Italiano de Buenos Aires.

Statistic Analysis

We defined to report the patient's rate of use with the amount of patient that login to the IP during the hospitalization. It is constructed with a denominator, defined as the totality of patients admitted to the hospital within the study period; and numerator, defined as users who log in to the IP during their hospital stay. It will be reported as a percentage with its respective 95% confidence interval (95% CI), calculated through normal approximation.

Quantitative variables will be described as mean and standard deviation or median and interquartile range, as appropriate for distribution. The categorical ones will be described as relative frequency and percentages, with their respective 95% CI.

Results

The characteristics of the patients who were admitted are detailed in Table 1.

Table 1: Patient Characteristics

Variable	Porcentage	Users
Age 18-20	4.0%	15
21-30	12.1%	45
31-40	18.9%	70
41-50	10.5%	39
51-60	11.0%	41
61-70	15.1%	56
71-80	13.2%	49
81-90	11.8%	44
91-99	2.9%	11
Female	58.9%	218
Healthcare Insurance Plan	53.2%	384
Total	100%	370

The rate of inpatients who made login to the IP during the study period was: 2.10% (95% CI 1.89-2.32). Out of a total of 17,612 patients only 370 login. Figure 1 shows the income by month.

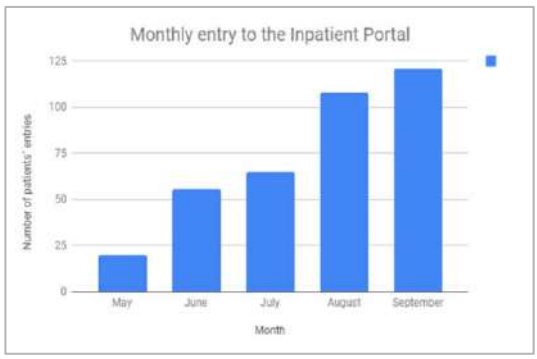


Figure 1: Monthly Entries to the Inpatient Portal

The 370 users who login the IP, made 9249 clicks. The most frequently used sections Vital Signs, Hospitalization data and location transfers. (Table 2)

Table 2: Used Sections

Sections	Use %	Total use
Vital Signs	24.6%	2278
Hospitalization data	15.6%	1444
Location transfers	14.5%	1350
Health care team	14.4%	1341
Food service	13.1%	1217
News	7.0%	656
HIBA TV	6.2 %	579
Phone numbers	4.1%	384

Conclusions

The rate of use between May and September 2018 was 2.10% (95% CI 1.89-2.32). Of the 370 users who logged the IP, the majority were female, affiliated with Healthcare Insurance Plan and youth. The most used functionalities were: vital signs, hospitalization data, and location transfers. New planning and implementation strategies are needed to improve the rate of use.

Acknowledgements

This study was carried out under the Residency Program in Health Informatics and the Area of Information Technology Oriented to the Community of the HIBA. We want to thank the entire interdisciplinary team (doctors, nurses, designers, developers) for the work you have done.

References

[1] M. Berg, C. Langenberg, I. vd Berg, and J. Kwakkernaat, Considerations for sociotechnical design: experiences with an electronic patient record in a clinical context, *International Journal of Medical Informatics* **52** (1998), 243-251.

[2] W.J. Winkelman and K.J. Leonard, Overcoming structural constraints to patient utilization of electronic medical records: a critical review and proposal for an evaluation framework, *Journal of the American Medical Informatics Association* **11** (2004), 151-161.

[3] National Research Council, *Computational technology for effective health care: immediate steps and strategic directions*, National Academies Press, 2009.

[4] M. Hagland, Smart rooms, smart care delivery: UPMC clinician leaders leverage technology for greater effectiveness in patient care, *Healthcare Informatics: the Business Magazine for Information and Communication Systems* **28** (2011), 36, 38-39, 42.

[5] D.L. Bronson, A.S. Rubin, and H. Tufo, Patient education through record sharing, *Quality Review Bulletin* **4** (1978), 2.

[6] P.C. Tang, J.S. Ash, D.W. Bates, J.M. Overhage, and D.Z. Sands, Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption, *Journal of the American Medical Informatics Association* **13** (2006), 121-126.

[7] J.D. Halamka, K.D. Mandl, and P.C. Tang, Early experiences with personal health records, *Journal of the American Medical Informatics Association* **15** (2008), 1-7.

[8] S. Greenfield, S. Kaplan, J. Ware Jr, E. Yano, and H. Frank, Patients' participation in medical care: Effects on blood sugar control and quality of life in diabetes, *Journal of General Internal Medicine* **3** (1988), 448-457.

[9] S.H. Kaplan, B. Gandek, S. Greenfield, W. Rogers, and J.E. Ware, Patient and visit characteristics related to physicians' participatory decision-making style: results from the medical outcomes study, *Medical Care* (1995), 1176-1187.

Address for correspondence

Liliana Giraldo BSN, Department of Health Informatics, Hospital Italiano de Buenos Aires. Juan Domingo Perón 4190, C1181ACH. Ciudad de Buenos Aires, Argentina. email: liliana.giraldo@hospitalitaliano.org.ar

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190726

Chronic Kidney Disease and the Use of Social Media as Strategy for Health Education in Brazil

Juliana Gomes Ramalho de Oliveira[a], Marjan Askari[b], Marília Girão Nobre Fahd[c], Gabriel Araújo Pereira[c], Aglalberto Lourenço da Silva[d], José Eurico Vasconcelos Filho[a], Maria Helena de Agrela Gonçalves Jardim[e] and Geraldo Bezerra da Silva Junior[f]

[a] Public Health Graduate Program, Health Sciences Center, University of Fortaleza. Fortaleza, Ceará, Brazil
[b] Erasmus School of Health Policy & Management, Health Services Management & Organization, Erasmus University. Rotterdam, The Netherlands
[c] School of Medicine, University of Fortaleza. Fortaleza, Ceará, Brazil
[d] Laboratory of Innovation in Information Technology, Technology Sciences Center, University of Fortaleza. Fortaleza, Ceará, Brazil.
[e] Health School, Universidade da Madeira. Funchal, Região Autônoma da Madeira, Portugal
[f] School of Medicine, Public Health Graduate Program, Health Sciences Center, University of Fortaleza. Fortaleza, Ceará, Brazil.

Abstract

Chronic kidney disease (CKD) is a Public Health problem worldwide. Treatment in complex and depends on patient education to achieve adequate adherence. We describe in this paper novel strategies for patients' education based on internet (youtube and instagram), through videos, images and texts information directed for patients in a project developed in Brazil.

Keywords:

Chronic kidney disease; Internet; Social media; Health education.

Introduction

It is estimated that 10% of world population has chronic kidney disease (CKD) [1]. This trend is related to the increase in life expectancy and the prevalence of chronic non-communicable diseases such as hypertension [2] and Diabetes Mellitus [3], the main causes of CKD. Studies have been developed to investigate the knowledge of population about CKD and the results show that groups of all parts of the world have low awareness about the disease which can reflect in adherence to prevention [4,5]. However, there is an even greater aggravation in situations of low knowledge about CKD: studies show that many individuals that make up the groups at risk of developing the disease, as hypertensive and diabetic, are not aware of this [6].

The repercussions of low knowledge on CKD can be seen in the high percentage of patients with late diagnosis, that is, in the more advanced stages of the disease, which unfortunately is related to the reduction of life expectancy in all age groups [7], and higher mortality during treatment [8]. The study of Ricardo et al. [9] has shown that low health literacy was associated to low estimated glomerular filtration rate (GFR). The aim of this study was to describe the use of social media as strategy for health education in the context of CKD.

Methods

This study is part of the Renal Health Project which develops several strategies for prevention and treatment of CKD. The project was funded by the International Society of Nephrology (ISN) through the Clinical Research Program, and is supported by the Brazilian Society of Nephrology. We created two tools of communication with the internet users, a channel on YouTube and a profile on social media Instagram in order to share digital content about the CKD. Although the goal of both is to disseminate information about prevention and treatment of the disease, it is believed that these two channels can achieve different population groups that have low knowledge about CKD, as general population [9, 10], non-dialysis patients [11, 12] even between health team [13, 14].

Channel on YouTube

YouTube is indisputably the biggest online video platform worldwide and the second most famous social network website worldwide as of July 2018, ranked by number of active users (1 900 millions) [15]. The YouTube channel Renal Health proposal Created in December 2017, (https://www.youtube.com/channel/UC3-GHeHAndcrRmbE4I_qE_w). It was designed to describe, in a logical sequence of videos, from kidneys definition and functions through the CKD diagnosis to the guidelines about treatment of this disease. A doctor called "Kidney" was chosen to answer the questions made by a woman voice.

Profile on Instagram

The Instagram App is one the most popular social networks worldwide. While the number of monthly active Instagram users from January 2013 was 90 million in June 2018 was a 1 billion. In that period Brazil was in third position on number of Instagram users (63 millions) [15]. Initially, we choose this social media to test the acceptability of general populations and CKD patients. The profile on Instagram (https://www.instagram.com/renal_health) was developed in June 2018, and arose from the need to broaden the spectrum of health education actions. The published content is based on the most current guidelines in Nephrology and contents are often shared by great and reliable institutions such as the Health Ministry of Brazil, World Kidney Day, World Health Organization and Brazilian Society of Nephrology.

Results

Currently, the channel has four videos and a total of 157 views. The first and the longer has (3:01 minutes), it is about

notions of Nephrology and concepts of CKD, the second on diagnosis "How the doctor discovers that a person has CKD", the third is a tutorial for the use of the application Renal Health, developed by our research group, and the last is about of CKD stages. Until the end of this work the Instagram profile counted on 223 publications and 832 followers. It is noteworthy that on August 28th the profile had 300 followers, reaching 600 on October 16th, 2018, showing significant growth. Only from October 8th throw 14th, 2018, according to Instagram's management data, publications had been viewed 8,904 times, from 1,366 accounts. A post on the CKD stages was viewed 1,074 times, indicating the range of reach of the publications. As for followers, 33% live in Fortaleza, Ceará, Brazil, followed by São Paulo (6%), São Luís (3%), Rio de Janeiro (3%) and Recife (2%); 41% are in the age range of 25 and 34 years old, 30% are 35 to 44 and 73% are women.

Discussion

For CKD patients the aim of this channels is to familiarize them with the aspects related to their health status because from the acknowledge of an expanded understanding. The patients have the autonomy to discuss with the health team about the best behaviors to be implemented in their treatment, which is, what makes them the co-managers of their health care. On the profile on Instagram it is noticed that many CKD patients feel free to post comments on their personal experiences demonstrating identification with the topics covered. For Moorhead [16], social platforms enable chronic patients to share their health issues, anguish, and experiences with other patients, generating empathy for those who once accommodated regarding the disease or felt alone.

Conclusions

Based on the increasing number of followers in both channels developed there is interest by general population and CKD patients of to improve the knowledge about CKD. Although it is known that knowledge does not determine behavior, it is evident that the absence of knowledge hinders the adherence to measures aimed to prevent health problems. Therefore, the population, especially the groups at risk for the disease development, after having access to more information on strategies for the protection of renal function and early diagnosis, might be able to make conscious decisions about lifestyles and their effects on their health status.

References

[1] A. K Bello, Global Kidney Health Atlas: A report by the International Society of Nephrology on the current state of organization and structures for kidney care across the globe. *International Society of Nephrology*, Brussels, Belgium. 2017.

[2] V. Jha, A.Y.M. Wang, H.Y. Wang, The impact of CKD identification in large countries: the burden of illness, *Nephrology Dialysis Transplantation* **27** (2012).

[3] M. N. Ozieh, C. E. Dismuke, C. P. Lynch, L. E. Egede, Medical care expenditures associated with chronic kidney disease in adults with diabetes: United States 2011, *Diabetes Research and Clinical Practice* **109** (2015), 185–90.

[4] A, Khalil, M. Abdalrahim, Knowledge, attitudes, and practices towards prevention and early detection of chronic kidney disease, *Int Nurs Rev* **61** (2014), 237-45.

[5] J.W. Stanifer, Knowledge, Attitudes, and Practices Associated with Chronic Kidney Disease in Northern Tanzania: A Community- Based Study, *PLoS One* **11** (2016).

[6] N. V. Faria, C. M. S. Teixeira, S. F. L. Nunes, Conhecimento dos usuários do programa HIPERDIA sobre a doença renal crônica, *Journal of Management & Primary Health Care* **5** (2014), 4-9.

[7] United States Renal Data System, USRDS *Annual Data Report.* Volume 1 – CKD in the United States 2016.

[8] H. Diegoli, M. C. G. Silva, D. S. B. Machado, C. E. R. N. Cruz, Late nephrologist referral and mortality assotiation in dialytic patients, *Brazilian Journal of Nephrology* **37** (2015) 32-37.

[9] Chow et al. Public lacks knowledge on chronic kidney disease: telephone survey, *Hong Kong Medical Journal* **20** (2014), 139-44.

[10] P. Roomizadeh, D. Taheri, A. Abedini, M. Mortazavi, M. Larry, B. Mehdikhani, S. M. Mousavi, F. A. Hosseini, A. Parnia, M. Nakhjavani, Limited knowledge of chronic kidney disease and its main risk factors among Iranian community: an appeal for promoting national public health education programs, *International Journal of Health Policy and Management* **2** (2014), 161-166.

[11] K. L. Moraes, Letramento funcional em saúde e conhecimento de doentes renais em tratamento pré-dialítico, *Revista Brasileira de Enfermagem* **70** (2017), 155–162.

[12] P. A. Lopez-Vargas, Knowledge deficit of patients with stage 1–4 CKD: A focus group study, *Nephrology* **19** (2014), 234–243.

[13] R. Evans, P. Rudd, U. Hemmila, H. Dobbie, G. Dreyer, Deficiencies in education and experience in the management of acute kidney injury among Malawian healthcare workers, *Malawi medical journal: the journal of Medical Association of Malawi* **27** (2015), 101–3.

[14] A. Schreider, N. M. S. Fernandes, Assessment of knowledge on renal replacement therapy in health care workers of the regions: Juiz de Fora, São João Nepomuceno e Santos Dumont, *Jornal Brasileiro de Nefrologia* 37 (2015) 382–384.

[15] The Statistics Portal, *Instagram – Statistics & Facts*. Available: < https://www.statista.com/statistics/578364/countries-with-most-instagram-users/>

[16] Lien, Y. H. H. Mystery of Chronic Kidney Disease Awareness, *The American Journal of Medicine* **125** (2012).

Address for correspondence

Prof. Geraldo Bezerra da Silva Junior. University of Fortaleza. Av. Washington Soares, 1321, Bloco S, sala S-01. CEP: 60811-905. Fortaleza, Ceara, Brazil. Phone/Fax: +55 85 34773280 / +55 85 34773424. E-mail: geraldobezerrajr@unifor.br

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190727

Evidence-Based Usability Principles for Safe Computerized Provider Order Entry (CPOE) Interface Design

Gurprit K. Randhawa[a,b], Ashley Garnett[b], Sihong Huang[c], Parvin Dhot[b], and Mary-Lyn Fyfe[a]

[a] *Island Health, Victoria, BC, Canada*
[b] *University of Victoria, Victoria, BC, Canada*
[c] *Fraser Health, Surrey, BC, Canada*

Abstract

There is a dearth of evidence-based tools to design the safest Computerized Provider Order Entry (CPOE) system possible. An evidence-based list of usability principles for the design of the CPOE interface was developed following a literature review, and validated with the Chief Medical Information Officer and CPOE team at Island Health. The list includes 11 usability principles that can be used to inform ongoing CPOE interface design and evaluation efforts to improve patient safety.

Keywords:

Medical Order Entry Systems, User-Computer Interface.

Introduction

The design of a Computerized Provider Order Entry (CPOE) system is a significant decision factor in reducing medication errors and maximizing safe use. Although the first CPOE system was implemented in 1971, there are several outstanding challenges concerning the usability of CPOE [1] even 47 years later. The underlying issue "appears to be that there is no consolidated source of evidence that provides organizations with tools to design the safest CPOE system possible" [1]. Davis and Stoots' 2012 article on EHR adoption emphasizes that organizations ought to incorporate checklists and tools to minimize any unintended consequences [2]. They suggest simple tools such as checklists act as a highly-effective quality-control instrument that "can prevent potential costly mistakes and even fatal medical errors" [2]. The literature highlights that though there is no "best" method in designing a CPOE system, it requires trial and error work at an organizational level [1]. Sengstack (2010) underlines that there is a need for tools that:

1. Are based on lessons learned and documented in the literature;
2. Can be used to evaluate CPOE systems; and
3. Can be used as a guide for informatics specialists at all levels as the iterative process of improvement continues to cycle [1].

A study from Sunnybrook Health Sciences Centre in Toronto also highlights the importance of organizational commitment to CPOE design to improve efficiency, usability and safety prior to system implementation [3]. Similarly, Yuan et al. (2011) underscores the importance of using a design framework to identify requirements and functional specifications for the User Interface (UI) design of clinical decision support (CDS) [4]. To address the aforementioned recommendations from the CPOE design literature, Island Health developed evidence-based guidelines for the design and evaluation of its CPOE system, as there is currently no comprehensive usability tool to inform CPOE interface design. Three CPOE design tools were developed for CDS Alerts, the CPOE User Interface, and references within CPOE Order Sets. This paper summarizes the usability guidelines developed and adopted by Island Health for the safe and usable design of the CPOE interface. Island Health is a health care organization in British Columbia, Canada with over 23,000 staff and physicians.

Methods

A literature review of CPOE interface design was conducted from November 2017 to January 2018 using Google Scholar. Various search terms were used, including "CPOE design" AND "colour" OR "color"; "CPOE design" AND "font"; "CPOE design" AND "alerts"; "CPOE design" AND "screen"; "CPOE design" AND "icons"; "CPOE design" AND "lists." The inclusion criteria for the literature review included primary or secondary articles that (1) were in English, (2) published between January 2000 and January 2018, and (3) that discussed CPOE design principles. The references of included articles were also mined for relevant articles. Recommendations for CPOE interface design were extracted from the included articles. A thematic analysis of the recommendations was conducted to identify themes from the articles, as well as to develop general usability principles for CPOE interface design. To confirm the congruence of the CPOE interface design principles with general user interface design principles (e.g., Nielsen's ten general principles for interface design), usability heuristics were also reviewed. The resulting CPOE interface design usability principles and sub-principles were iteratively reviewed and refined to produce a final list usability design principles. Face validity was evaluated with the Chief Medical Information Officer (CMIO) and CPOE order set team at Island Health. In addition to a list of usability principles, a designer-centred checklist for CPOE interface design was developed for use by the CPOE team and clinical informatics specialists at Island Health. The checklist presented the usability principles in the form of a question.

Results

In total, thirty articles were included in the literature review, and a list of 11 CPOE interface design usability principles was developed. The list of usability principles for CPOE interface design fall into 11 categories: (1) layout, (2) terminology, (3) use of headings and sub-headings, (4), placement and proximity, (5) general use of colour, (6) use of colour combinations, (7) text style, (8) use of punctuation, (9) use of icons, (10) use of lists, and (11) alert visibility. These principles align with 5/10 of Nielsen's ten general principles for interface design. Due to space limitations, only the sub-principles for Principle 1 (Layout) are presented below. The full (detailed) CPOE interface design checklist with principles, sub-principles, and corresponding literature/evidence will be presented in the poster.

Principle 1: Layout

1. Minimize the layers of screens (to a maximum of 3 layers) to facilitate user's navigation in the system [5].
2. Use visual cues (i.e., colour, font or conspicuous spatial arrangement) consistently. For example, use visual cues within the ordering screen to indicate the difference between the latest results vs. previous results) [5,6,7]
3. Organize the screen elements into logical groups, visually separated by space, alignment and borders; their meaning should be easily recognized by users [5].

Discussion

Island Health's comprehensive and validated list of usability design principles for safe CPOE interface design addresses the current gap in the literature surrounding the need for a consolidated source of evidence for designing safe CPOE systems. The list and check-list of design principles have been adopted by Island Health for its CPOE interface design practices. However, it is interesting to note that the authors did not come across any literature published on the use of special characters in CPOE design, such as asterisks. This was an area of particular interest for Island Health, as asterisks had been previously used in CPOE interface design in the organization. As such, it is unclear if the use of special characters may have any effect(s) on the quality and safety of CPOE interfaces.

Given the iterative process of continuously improving CPOE design, it is important to note that the consolidated list of usability design principles presented in this paper is a "living and breathing" guideline that should be continually updated based on usability-related lessons learned by health care organizations. Based on ongoing CPOE usability evaluations, health care organizations across the globe should regularly document and make recommendations for CPOE interface design to ensure that the design principles are based on emerging best practices in the field. Specifically, pre- and post- evaluation design may be used to evaluate changes in the usability of CPOE before and after adopting the design principles.

A major limitation of this research is that a scoping or systematic review of the CPOE interface design literature was not conducted due to time constraints, and only one database (Google Scholar) was searched. It is recommended that a scoping or systematic review be conducted. Additionally, the list of CPOE usability design principles was only validated with the CMIO and CPOE team at Island Health. The list should be further validated with stakeholders in other health care organizations at the provincial, national, and international levels.

Conclusions

This paper contributes a list of 11 evidence-based usability design principles for CPOE interface design, including principles related to: (1) layout, (2) terminology, (3) use of headings and sub-headings, (4), placement and proximity, (5) general use of colour, (6) use of colour combinations, (7) text style, (8) use of punctuation, (9) use of icons, (10) use of lists, and (11) alert visibility. Specifically, this paper addresses the suggestions made by Sengstack [1] by developing a usability principles tool for CPOE interface design that (a) is based on the literature, (b) can be used to evaluate CPOE usability, and (c) can be used to guide health informatics specialists at all levels in CPOE design and quality improvement. As such, the list of principles has implications for usability evaluation of CPOE; the usability design principles may be applied by health care organizations as a check-list to support the evidence-based design, adoption, ongoing evaluation, and improvement of CPOE interfaces. Given that the authors did not find published literature on the use of special characters (e.g., asterisk) in CPOE interface design, it is recommended that further research be conducted to evaluate the usability and use of special characters to ensure safe CPOE interface design.

References

[1] P. Sengstack, CPOE configuration to reduce medication errors a literature review on the safety of CPOE systems and design recommendations, *Journal of Healthcare Information Management* **24**(4) (2010), 26-34.

[2] C. Davis, & M. Stoots, *A Guide to EHR Adoption: Implementation Through Organizational Transformation*. Chicago, IL: HIMSS, 2012.

[3] J. Chan, K.G. Shojania, A.C. Easty, & E.E. & Etchells, Does user-centred design affect the efficiency, usability and safety of CPOE order sets? *Journal of the American Medical Informatics Association,* **18** (2011).

[4] M.J. Yuan, G.M. Finley, J. Long, J., C. Mills, C., & R.K. Johnson, Evaluation of user interface and workflow design of a bedside nursing clinical decision support system, *Interactive Journal of Medical Research* **2** (2013), e4.

[5] R. Khajouei, M.W. Jaspers, The impact of CPOE medication systems' design aspects on usability, workflow and medication orders. Methods of information in medicine **49** (2010), 3.

[6] A.Z. Hettinger, & R.J. Fairbanks, Recognition of Wrong Patient Errors in a Simulated Computerized Provider Order Entry System, *Academic Emergency Medicine* **1** (2011), S237.

[7] J. Horsky, G. Schiff, D. Johnston, L. Mercincavage, D. Bell, & B. Middleton, Interface design principles for usable decision support: a targeted review of best practices for clinical prescribing interventions. *Journal of Biomedical Informatics* **45** (2012), 1202-1216.

Address for correspondence

Gurprit K. Randhawa, gurprit@uvic.ca

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190728

My Little Smart Personal Assistant: A Co-Designed Solution to Ensure an Optimized Ageing-Well at Home in Rural European Settings

J-B. Kern[a,b,c], S. Strola[d], J. Quintas[e], T. Moulaert[f], J-P. Jacquet[c], P-Y. Benhamou[g]

[a] Department of General Practice, Grenoble Alpes University, Grenoble, France,
[b] TIMC-IMAG Laboratory, CNRS UMR 5525, Grenoble Alpes University, Grenoble, France
[c] European Rural and Isolated Practitioners Association, EURIPA France, Neuilly-sur-Seine, France
[d] CIC 1406 (Innovative Technology), Grenoble Alpes University, INSERM, CHU Grenoble Alpes, Grenoble, France
[e] Laboratory of Automatic and Systems, Instituto Pedro Nunes, Coimbra, Portugal
[f] PACTE Social Science Laboratory, School of Political Studies, Science Po Grenoble, CNRS UMR 5194, Grenoble Alpes University, Grenoble, France
[g] LBFA Laboratory, CNRS UMR 1055, Grenoble Alpes University, Grenoble, France

Abstract

My Little Smart Personal Assistant is a co-designed remote connected device with an interactive vocal assistant that provides a panel of social/medical services for the rural European elderly population. The aim is to create a new patient-centered solution to improve quality of life, self-autonomy, and integration within local community. This should improve aging-well at home in rural settings.

Keywords:
aging, patient-centered care, remote consultation.

Introduction

Background

The aging society is one of the major challenges for the healthcare system. About 17% of the overall population is aged 65 years or older in members of the OECD (Organisation for Economic Co-operation and Development), and particularly higher in France and Portugal (18 and 20%, respectively)[1,2]. These statistics will rise in the next years. A main concern associated with aging is the forthcoming frailty, which affects around 10% of people 65-74 years and increases within older population[3].

Besides frailty, the prevalence of chronic conditions and co-morbidities also increase with age. In Europe, 8 in each 10 people older than 65 suffer from at least one chronic disease. Frailty/chronic diseases are particularly threatening for isolated aged patients, as they do have little or none informal caregivers helping them to manage properly on their basic needs. This situation is even worse for those living in rural areas where access to health ressources is usually challenging.

Proposed solution

Our consortium brings together actors from the academic, technological, research network, associative, and industrial sectors in Portugal and France.

My Little Smart Personal Assitant (MLSPA) proposes to develop a remote connected device which integrates an interactive vocal personal assistant aiming to support health and social care services delivered in rural areas. This package is co-designed starting from an existent technology - the Virtual Social Companion (ViSoCo) developed by the Instituto Pedro Nunes (IPN) in Portugal[4]. IPN leads the technological development of the solution and works closely with GLINTT (Global Intelligent Technologies), a Portuguese company managing the market and business tasks to implement this solution in the European healthcare systems. Our endeavour takes advantage of their consolidated experience in Portugal and Spain.

The consortium also benefits from the participation of EURIPA (European Rural and Isolated Practitioners Association), which is charged with the implementation in rural areas of the solution in Portugal and France. In addition, MaDoPa (Centre Expert en Technologies et Services pour le Maintien den Autonmie à Domicile des Personnes Agées) assists the co-developpement and evaluation of healthcare and autonomy innovative solutions.

Université Grenoble Alpes (UGA) enriches MLSPA ensuring the overall management of the project and takes advantage of the competences of INSERM (Institut national de la santé et de la recherche médicale) for field investigation, experimental methodology, data analysis, usability, and social evaluation.

Together, these multidisciplinary competences will cover the needs required for the first evaluation of MLSPA solution. The goal is to validate a panel of multilevel indicators obtained from a patient-centered approach consolidating a market deployment strategy.

Methods

Approach and experimental plan

MLSPA proposes an innovative approach which allows the overall maturation of the project through a first co-construction phase (i.e. sinergy with end-users and stakeholders) and a following field-test in real settings (i.e. rural communities).

The maturity process of our solution follows the methodology proposed by INSERM via the CIC-IT department of Grenoble (Centre Investigation Clinique – Innovation Technologique) for innovations in the healthcare domain. It allows to validate the prototype faster when compared to the time reported in other projects (3-4 years instead of 8-10 years). The innovation maturity cycle will decrease the costs for innovation

(technology implementation) and increase potential collaborations in healthcare domains[5]. This methodology improves the relationship amongst the consortium partners. In respect to an efficient behavior, a dynamic work emerges to ensure a beneficial agreement that provides sufficient incentives to market actors in order to constantly improve products and services in the healthcare system[6]. In addition, a technological maturity of the device is expected to escalate to TRL 6 (technology readiness level).

In particular, MLSPA solution will be tested by the design of an experimental plan deployed for free in 2 rural places in Portugal and 2 in France, with a total of 100 volunteers (25 patients in each place). The patients will be selected by their general practionners (GPs), and the cohort has been calculated in order to ensure an appropriated pilot study. The population selected for this phase are people over 65 years old living in rural settings, demonstrating frailty (assessed by the FRAIL scale – Fatigue Resistance Ambulation Illnesses and Loss of Weight)[7], with chronic condition, absence of cognitive decline (MMSE "Mini Mental State Examination" over 23), and an "Index of Independence in Activities of Daily Living" (ADL) over 3.

During a visit, GPs will collect the following clinical data using a standardized questionnaire and evaluation protocol: age, FRAIL scale, MMSE, ADL, weight, height, BMI (body mass index), IADL (Instrumental activities of daily living), Up and Go test, self-reported exhaustion by patient, number of treatments (molecules), consultations (number in 2018), hospitalisations (number in 2018), self-perceived frailty by patient (numerical scale), and perceived frailty by GP (numerical scale). Additionally, non-clinical data will be collected at the inclusion including: demographic data, existing internet connection at home (or possibility to obtain access), social environment (i.e. local community activities). Data collected will be recorded electronically with the support of a personal health data hosting provider.

MLSPA will be compliant to the ethical standard, the clinical study regulations and the patient data treatment (e.g. European General Data Protection Regulation 2016/679). All data acquired by this project will be in encrypted databases (e.g. personal healthcare authorized data host). Furthermore, we will ensure that the data flow and their communication will be handled only under secure connections (e.g. HTTPS protocols, Secure Sockets Layer, and Virtual Private Network).

Impacts and Conclusions

The short-term objective is to demonstrate our solution in terms of benefits on the follow-up of the global activities of this population.

In addition, this study will provide social-medical-economic indicators that will be used by the stakeholders for a further large-scale deployment plan (i.e. market access strategy, reimbursement policy, healthcare companies involvement…). Finally, data on the participation of GPs and social worker will be collected to optimize and correctly address the project in terms of usability (user's feedback), participation, and personalized services.

We aim to enable older adults to live with dignity in their homes and to improve their social engagement in rural communities. MLSPA will improve the follow-up of physical activity, healthy eating, monitoring of medical treatment for chronic diseases and social life involvement and connection of the participants. This paves the way to a future empowerement of the self-health management by citizens, providing them an active and participative role.

The final goal is to provide support and guidance features related to self-management and health literacy associated with frailty prevention. This may promote social prescribing and reduce the impact of frailty and chronic morbidity on the targeted population.

All together, this will facilitate positive, active, and healthy aging at home, reduce the risk of dependency, and avoid unexpected hospitalisations in decreasing the cost for the healthcare system.

Figure 1 below displays MLSPA project vision.

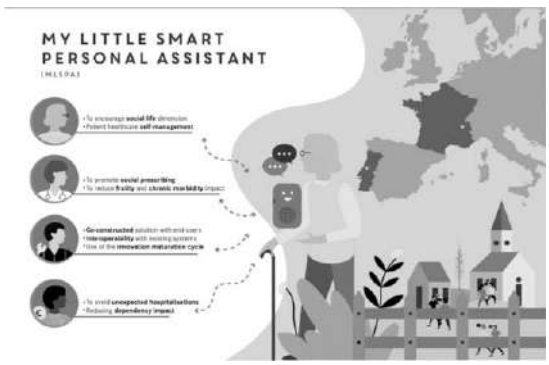


Figure 1: MLSPA vision

References

[1] T.W. Bank, World Bank Annual Report 2017, The World Bank, 2017. http://documents.worldbank.org/curated/en/143021506909711004/World-Bank-Annual-Report-2017 (accessed November 3, 2018).

[2] OECD, Health at a Glance 2017: OECD Indicators, OECD, 2017. https://www.oecd-ilibrary.org/social-issues-migration-health/health-at-a-glance-2017_health_glance-2017-en (accessed November 3, 2018).

[3] C.R. Gale, C. Cooper, and A. Aihie Sayer, Prevalence of frailty and disability: findings from the English Longitudinal Study of Ageing, *Age Ageing*. **44** (2015) 162–165.

[4] C. Tsiourti, M. Ben Moussa, J. Quintas, B. Loke, I. Jochem, J.A. Lopes, and D. Konstantas, A Virtual Assistive Companion for Older Adults: Design Implications for a Real-World Application, in: Y. Bi, S. Kapoor, and R. Bhatia (Eds.), Proc. SAI Intell. Syst. Conf. IntelliSys 2016, Springer International Publishing, 2018: pp. 1014–1033.

[5] A. Moreau-Gaudry, and L. Pazart, Development of innovative technologies in health: The CREPS cycle, Concept – Research – Experiment – Product – Safe Care, *IRBM.* **31** (2010) 12–21.

[6] L.A. Marsch, and D.H. Gustafson, The Role of Technology in Health Care Innovation: A Commentary, *J. Dual Diagn.* **9** (2013) 101–103.

[7] J.E. Morley, T.K. Malmstrom, and D.K. Miller, A simple frailty questionnaire (FRAIL) predicts outcomes in middle aged African Americans, *J. Nutr. Health Aging*. **16** (2012) 601–608.

Address for correspondence

Jean-Baptiste Kern. General Practice Department, Faculty of Medicine, Domaine de La Merci, 38706 La Tronche Cedex, France. E-mail: jean-baptiste.kern@univ-grenoble-alpes.fr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190729

Education in Biomedical and Health Informatics: A European Perspective

Aikaterini Kolokathi[a], Arie Hasman[b,g], Catherine Chronaki[c,g], Inge Madsen[d,g], Anne Moen[e,g], Rebecca Randell[f,g] and John Mantas[a,g]

[a]Health Informatics Laboratory, National and Kapodistrian University of Athens, Athens, Greece,
[b]Department of Medical Informatics, AMC – University of Amsterdam, Amsterdam, The Netherlands,
[c] Secretary General, HL7 Foundation, Brussels, Belgium,
[d]Centre of Clinical Guidelines, Department of Health Science and Technology, University of Aalborg, Aalborg, Denmark,
[e] Institute for Health and Society, University of Oslo, Oslo, Norway,
[f] School of Healthcare, Baines Wing, University of Leeds, Leeds, UK,
[g] AC2 – Accreditation and Certification Committee, European Federation for Medical Informatics Association, Lausanne, Switzerland

Abstract

In higher education, programs in specialization in Health Informatics, Medical Informatics, Health Engineering are continuously growing. In this research, almost 1800 universities and colleges were checked in order to find related educational programs at all academic levels. Approximately 1000 academic leading degree programs in those domains have already been identified. The detailed records of the related educational programs will help to understand the current educational needs and priorities. Although, the growth of the related educational programs is not the same in each country.

Keywords:
Medical Informatics, Education, Europe

Introduction

Information technologies and telecommunications improve prevention, public health, quality of health care and biomedical research. The majority of new technological tools focus on access, processing and management of medical data and information, enhancing the decision-making process. Thus, it is important for the health scientists and professionals to be well trained by instructive programs in order to develop and operate these systems successfully [1; 2].

The past few years, in higher education, programs such as biomedical informatics, health informatics and biomedical engineering are continuously growing. Recent domestic and international articles emphasize curricula with specialization in Health Informatics, Biomedical Informatics, Medical Informatics, Medical Engineering and Biomedical Engineering of the European Universities which are offered at undergraduate and postgraduate level. However, there is a remarkable absence of detailed recordings of the European educational programs in these specializations. Thus, the scope of this research is to fill in this gap. The established EFMI Accreditation and Certification Committee (AC2) tries to fill in this gap among other tasks as well. This initiative was launched at that time by the EFMI President and endorsed by the EFMI Board in consultation with the EFMI WG EDU Chair. The main task of the AC2 Committee is to search for ways to develop and implement the Accreditation and Certification initiative in Europe. Since, the Committee in order to achieve its aim, has detailed records of the educational programs in specialization in Health Informatics, Medical Informatics, Biomedical Informatics, Bioinformatics, Nursing and Dental Informatics, Health and Medical Technology, Health Engineering, Medical and Biomedical Engineering at all academic levels [1-3].

Methods

This project follows a series of steps in order to achieve the aim. First of all, online searching was conducted on Coursera, EdX, FutureLearn and Udemy databases. Secondly, a full list of the universities, colleges, institutions etc. of each European country was compiled via Google searching. Thirdly, the official website of every university was carefully checked so as to locate educational programs related to our subject. It is a fact that the duration of the search process has been conducted a couple of years before and the research is still in progress [2]!

It is worth mentioning that the research is limited only to European countries that are members of the European Federation for Medical Informatics – EFMI (Table 1) [4].

Table 1- The list of EFMI Countries-Members

Countries-Members of EFMI			
Armenia	Austria	Belgium	Bosnia-Herzegovina
Croatia	Cyprus	Czech Republic	Denmark
Finland	France	Germany	Greece
Hungary	Iceland	Ireland	Israel
Italy	Republic of Moldova	Netherland	Norway
Poland	Portugal	Romania	Russian Federation
Serbia	Slovenia	Spain	Sweden
Switzerland	Turkey	Ukraine	United Kingdom

Specifically, this study has covered almost 27 EFMI countries-members out of 32 [2]. In more detail, Figure 1 shows the countries have completed this process and the countries will have been completed in a few weeks.

Almost 1800 universities and colleges were checked in order to find related educational programs at all academic levels. About 180000 courses in total were checked.

Results

More than 1000 educational programs in these domains have been found in a wide variety of undergraduate and postgraduate degrees so far.

Figure 1-Twenty seven EFMI Countries-Members

Specific information was collected for each educational program. All these elements are included in Access Database (Figure 2). The information is being collected for each study program: university/ies, department/faculty, study program name, academic level (e.g. undergraduate / postgraduate / doctoral / postdoctoral Studies), type of education (full time – part time – combined), mode (on campus – e-learning / distance learning), specializations, director of the education program, details about contact person of the program, curriculum, time table, learning outcomes, competencies and program's language (English – local – bilingual). Additionally, Program's ECTS and academic staff's details are contained. Needless to say, the authors' interest focuses only on academic degree programs [1; 2].

Figure 2- Catalogue of Educational Programs

Furthermore, this database includes educational programs in specific areas such as Nanomedicine, Medical Electronics, Clinical Informatics, Clinical Technology, Clinical Engineering, Computational Biology, Life Science Informatics, Clinical Data Management, Big Data in Healthcare, Data Mining in Healthcare and Medicine, Digital Health Systems, E-Health, Telemedicine, Healthcare Analytics, Wireless Networks in Healthcare, Internet of Thing in Healthcare etc. Until now, the authors' have no view on specific countries such as the Russian Federation, Turkey and Ukraine due to local issues.

Finally, the data on this database is published in the appropriate established website (http://www.bmhi-edu.org). This website is under development and is available on the network. As verification of the included information is still in progress the information is not still available to the public. Only the registered users have access to the content.

Discussion

Earlier efforts in developing databases of educational programmes at an international level have been made. It is worth mentioning the effort by WG1 on health and medical informatics education of the International Medical Informatics Association (IMIA). The initiative of IMIA had aimed at the creation of an online database that would provide information about programs and courses in Health and Medical Informatics worldwide [5]. In our case, the data of the research comes from European Countries and provide information about programs in Health and Medical Informatics, Biomedical Informatics, Bioinformatics, Medical and Biomedical Engineering etc. In addition, this study has included only the academic degrees of bachelor, master or doctoral and postdoctoral level in related educational programs. While the certificates, short courses, modules, summer schools, exchanges studies, lectures and seminars were excluded. Due to the limitations of this research, there is an unclear viewpoint for specific countries such as Belarus, Estonia, Georgia, and Latvia.

AC2 Committee strongly supports the promotion and provision awareness of the educational initiative to the wider biomedical and health informatics community in Europe. Consequently, this information to explore in order to understand the current educational needs and priorities of each European country. Presently, the educational priorities in Biomedical and Health Informatics for Europe are closely interlinked with the European Union's educational framework [3]. Thus, the Educational Policy should be focused on local, state, national and international activities and addresses short- and long-term needs [6]. The Committee will work diligently to seek collaborations and invite other international organizations to cooperate [3].

Conclusions

The present educational programs are offered at undergraduate and postgraduate level including a variety of specializations such as health informatics, bioinformatics etc. However, the growth of educational programs is not the same in each country. It would be useful to see if in each country a satisfying number of educational programs is provided with well-detailed curriculum at each level depending on their specialization [1; 2].

Acknowledgements

The authors would like to thank the European Federation for Medical Informatics (EFMI) and AC2 Initiative for supporting this project. This research is an on-going work.

References

[1] E. Manifava, A. Kolokathi, and J. Mantas, Exploring the biomedical and health informatics educational programs in Europe, *Stud Health Technol Inform* **202** (2014), 67-70.

[2] A. Kolokathi, and J. Mantas, Education in Biomedical and Health Informatics: A Mapping Approach, *Stud Health Technol Inform* **251** (2018), 313-316.

[3] C. Lovis, G. Mihalas, J. Mantas, N. Peek, R. Balicer, I. M. Masic, L. Stoicu-Tivadar, and Anne Moen, EFMI, *Yearbook of medical informatics* **27** (01) (2018), 296-301.

[4] EFMI, EFMI Countries (2017), available from URL https://www.efmi.org/about/efmi-countries/.

[5] R. Haux, J. Frank, and P. Knaup, The IMIA WG1 database on health and medical informatics programs and courses: a call for participation, *Methods of information in medicine* **36** (1997), 233-234.

[6] N. Al-Shorbaji, E-health in the Eastern Mediterranean Region: a decade of challenges and achievements, *Eastern Mediterranean Health Journal, Special Issue* **14** (2008), S157-S173.

Address for Correspondence

Aikaterini Kolokathi, MSc, Health Informatics Laboratory, National and Kapodistrian University of Athens, e-mail: katiakolokathi@yahoo.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190730

Relating Factors for Acceptance of Health Care Technology: Focus on Mental Workload

Lisanne Kremer[a], Sumona Sen[b], Bernhard Breil[a]

[a] *Faculty of Health Care, Niederrhein University of Applied Sciences, Krefeld, Germany*
[b] *Faculty of Industrial Engineering, Niederrhein University of Applied Sciences, Krefeld, Germany*

Abstract

Medical information systems and care robots are two typical examples of human computer interaction in health care. Although used in a stressful environment, effects on mental workload and acceptance are hardly evaluated. We conducted an experimental design including collaborative robotics and eye tracking in a nursing situation to test the practicability and plausibility of eye tracking as a measuring method for workload. Results showed that eye tracking is feasible if context factors are adjusted. Data reduction and classification of tasks are necessary.

Keywords:

Workload, Eye Movements, Health Information Systems

Introduction

Health information system (HIS) has already become an important part of daily health care. Investigations of usability of HIS often focus on technical issues and not on the whole system (technical, human). Interacting with a computer while operating daily working tasks means to switch attention between patients, computer and other competing tasks. This can lead to a high workload. Comparable to the use of HIS in medical settings are Human Robot Interactions (HRI) in health care: care assistance robots have the potential to compensate gaps in patient-centered care caused by demographic change [1]. The topic of HRI poses the challenge of safety issues as well as usability and acceptance by users. Meanwhile, physical limits in interaction with robots are well defined, there remain open questions – as well as in the topic of HIS - regarding mental issues like stress and strain and acceptance by users [2]. Working and interacting with a robot or a HIS for a whole working shift needs a high amount of attention. As a result, the workload can be too high to manage all competing tasks.

"Workload is defined as the physical and/or mental requirements associated with a task or combination of tasks" and "refers to that portion of the operators limited capacity actually required to perform a particular task" [3]. Accordingly, mental workload (MWL) may be an imbalance of resources of the operator and requirements of the task influencing the operator's performance. O'Donnell & Eggemeier addressed the lack of valid, objective measurement methods regarding psychophysiological correlates that can be applied in real work situations [4]. Advantages of psychophysiological methods are that they are uncontrollable and unmodifiable by the proband as well as spontaneously in response to stimuli. They can be an effective approach for workload analysis in a highly demanding work situation like a health care situation [5].

One aspect of analysing HIS is the integration of physiological (MWL) and intentional measures (for example, acceptance) [6]. For analysing "acceptance", there exist several models, e.g. the *unified theory of acceptance and the use of technology* (UTAUT) identifying four key factors (performance expectancy, effort expectancy, social influence and facilitating conditions) and four moderators (age, gender, experience and voluntariness) [7]. Several studies investigated mostly one or two of the factors, but only a few studies showed extensions as including individual characteristics to the UTAUT model. Leaving out individual characteristics is one of the main criticism of the UTAUT model as well as in health care research [8]. We propose MWL as an additional factor to the UTAUT model that influences behavioral intention.

As MWL is determined by task complexity and difficulty, it is probably a good predictor to explain the high amount of stress of health care staff [9]. Work psychology states that mental stress emerges of strain resulting from task requirements [10]. Knowing about a lack of acceptance of new technologies in health care – especially HIS – leads us to further hypotheses. Our main objectives are to validate eye movements as well as pupil width as physiological correlates for MWL measurements in health care settings to be able to test possible extensions of the UTAUT model.

Methods

Embedded in a students project, we developed an experimental simulation design. We used a simulated hospital room that holds environmental conditions stable (noise, lights). It contains a nursing bed, an over-bed table, a collaborative robot, a flatscreen TV as well as decoration equipment to make the hospital situation more realistic to the probands. Sirens and other hospital sounds are played for a more realistic scenario. The patient sits in bed with the over-bed table beside his right side, the collaborative robot is placed behind the table. The executing task of the robot is to handle the patient medication, something to drink as well as something to eat. The robot puts the three different objects on the over-bed table one after another and puts it back afterwards. The patient needs to lift up the objects in the time between and before the robot takes the object back 30 seconds later. There's a planned pause in HRI when the patient tries to solve a Sudoku. The described procedure is repeated twice; at the end of the setting, the proband is asked to fill out an adapted UTAUT questionnaire. The setting (without the UTAUT questionnaire) takes about 10 minutes. As the experimental setting – as well as real nursing settings – is complex and cannot be interrupted for standard subjective measurement methods, we use the eye tracking method for measuring cognitive parameters. We propose the use of mobile *pupil labs* binocular glasses that include two infrared (IR) spectrum eye cameras for dark pupil detection and a scene camera. The mobile using solution makes the pupil labs glasses very flexible as well as the very low weight of 37gr which supports the flexible research setting.

Additionally to physiological measures of workload, we use an adapted form of the UTAUT questionnaire including some questions addressing MWL and leaving out questions not matching to our research question. We already conducted pre-tests, each execution with another proband. The planned sample size is a minimum number of 90 probands, divided into two groups considering differences between "digital natives" and people over the age of 40.

Results

The first pre-tests – conducted by a group of students – showed the practicability of the study design and the potential of the method of eye tracking to measure MWL during a high demanding task situation. The pupil labs glasses only weigh around 37 grams and as reported by probands are well suited and pleasant to wear for a long time (around 15min). Wearing the eye tracker reportedly does not disturb the visual field of the proband, nor do the wires that are connecting it to the mobile phone. Those parameters ensure that the probands in this stationary – but as well as probands in more flexible – study designs are not extra loaded by the eye tracker. We asked our pre-test probands if the simulated area the experiment is performed in, felt like a realistic scenario for them. As one main goal of our study is to test eye tracking on its plausibility and practicability in a real nursing setting, it was important to create a realistic scenario. We optimized some objects concerning critic points the probands pointed out, like adding more hospital associated decoration to the walls.

Measuring stress via eye tracking in the first step means extracting data concerning pupil width. As we wanted do define MWL in interaction with technology (in this case: human robot interaction), we focus on extracting pupil width data at timestamps that refer to the collaboration of the proband and the collaborative robot. First data analyses within the pupil labs software resulted in a great amount of data as the world camera produces additional data. The results in the test setting are promising as parameters of pupil width and fixation were recognized correctly and could be correlated to specific situations in a qualitative manner. As Pupil Capture ejects a gaze map directly after recording, we could already point out that fixations on the robot took much longer than needed actually. While the robot already paused, the probands still fixated his arm.

While it does not seem reasonable to assess results to our UTAUT questionnaire with only n = 3 probands, we asked our probands about comprehensibility and content-related relevance.; we changed the wording of some questions to improve critical aspects

Discussion

We discussed possible extensions of the UTAUT model and identified first hints for MWL as a possible predictor to influence acceptance of new supporting technologies used in health care settings (MIS and Robots). This study demonstrated that it is feasible to use eye tracking in a real setting as glasses are small and lightweight. The investigation of an experimental setting including a collaborative robot seems to be a good starting point for testing our hypotheses in a protected realistic – but laboratory setting. Our goal was to validate whether the eye tracking method can be transferred to similar settings, we see some first results.

Nevertheless, there are some limitations which are relevant to consider. In the current setting we worked with inexperienced students, there was less equipment (i.e. medical records, information systems) and fewer sources which attracted attention (i.e. no interaction with other colleagues). Although the simulation is as close as possible to a real scenario, we expect more distraction in a real clinical setting. Therefore, our results may be less transferable.

Conclusions

We conducted a first pre-test for the eye tracking method as a valid measurement method for measuring MWL. While first pre-tests showed some promising hints for further studies, we identified the need for reducing relevant data as well as the need for another method measuring MWL to correlate the matching parameters.

References

[1] E. Broadbent, R. Tamagawa, A. Patience, B. Knock, N. Kerse, K. Day, and B.A. MacDonald, Attitudes towards health-care robots in a retirement village, *Australas J Ageing* **31** (2012), 115-120.

[2] DGUV, Deutsche gesetzliche Unfallversicherung). Kollaborierende Robotersysteme. Planung von Anlagen mit der Funktion „Leistungs- und Kraftbegrenzung". *DGUV Information* (2017) FB HM-080.

[3] S. Gudipati, A. Punnathur, Workload assessment techniques for job design (2016) [Last accessed: 25. 11. 2018]. Available at: http://semac.org.mx/archivos/6-9.pdf

[4] T. Eggemeier, G.F. Wilson, Performance-based and subjective assessment of workload in multi-task environments. I*n Diane L. Damos, Eds., Multiple-Task Performance. Taylor and Francis*, 1991, Washington, DC

[5] L.E. Nacke, An Introduction to Physiological Player Metrics for Evaluating Games. *In M. S. El Nasr, A. Drachen, & A. Canossa (Eds.), Game Analytics: Maximizing the Value of Player Data*, 2013, 585 - 619, London, UK: Springer London.

[6] E. Ammenwerth, S. Graber, G. Herrmann, T. Burkle, and J. Konig, Evaluation of health information systems-problems and challenges, *Int J Med Inform* **71** (2003), 125-135.

[7] V. Venkatesh, M.G. Morris, G.B. Davis, and F.D. Davis, User Acceptance of Information Technology: Toward a Unified View, *MIS Quarterly* **27** (2003), 425-478.

[8] V.D. Dwivedi and R.M. Gibson, An ERP investigation of quantifier scope ambiguous sentences: Evidence for number in events, *Journal of Neurolinguistics* **42** (2017), 63-82.

[9] S. Ribback, Psychophysiologische Untersuchung mentaler Beanspruchung in simulierten Mensch-Maschine-Interaktionen, (2003).

[10] D. Manzey, B. Lorenz, and V. Poljakov, Mental performance in extreme environments: results from a performance monitoring study during a 438-day spaceflight, *Ergonomics* **41** (1998), 537-559.

Address for Correspondence

Lisanne Kremer, M. Sc.

Faculty of Health Care

Niederrhein University of Applied Sciences

Reinarzstraße 49, 47805 Krefeld

lisanne.kremer@hs-niederrhein.de

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190731

Nursing Informatics as a Specialization in India: Present and Future

S. K. Meher, Sanjay Gupta, Shailendra Sharma, Mohamed Ibrahim M, Kamal Ajmera

Department of Computer Facility, All India Institute of Medical sciences, Ansari Nagar, New Delhi, India

Abstract

Nurse informatics specialist is the new concept in India moreover it's a kind of challenge in the highly populated All India Institute of medical sciences, a quaternary care premium medical institute in India to digitalise all the medical / administrative paper work.

The paper emphasizes on nature of duties for nursing Informatics practicner, desirable skills, challenges and finally implementing nursing Informatics concept in India, in All india Institute of Medical sciences. For the first time the nursing informatics concept is used and implemented in India, and I am the part of this concept. I will be further sharing my experiences of Nursing informatics practice in India.

Keywords:
Nursing Informatics, NIS, NI future.

Introduction

Nursing informatics is a career that focuses on finding ways to improve information management and communications in nursing to improve efficiency, reduce costs and enhance the quality of patient care.

The Institute has comprehensive facilities for teaching, research and patient-care. As provided in the Act, AIIMS conducts teaching programs in medical and para-medical courses both at undergraduate and postgraduate levels and awards its own degrees. Teaching and research are conducted in 52 disciplines. In the field of medical research AIIMS is the lead, having more than 1500 research publications by its faculty and researchers in a year. AIIMS also runs a College of Nursing and trains students for B.Sc. (Hons.), Nursing (Post-certificate) degrees.

Primary emphasis of nursing care is Documentation so the main emphasis of nursing informatics too is documentation and patient education. It's well known fact that quality care depends on effective communication among healthcare providers, since healthcare providers communicate primarily through the notes they write in a patient's health record/ in-patient chart, nurse informaticists seek to continually improve the speed, timeliness and accuracy of patient charting. The doctor can access all the patient records from anywhere and anytime at his own ease, which increases the efficiency of doctor and decreases his anxiety about patient's condition. When health workers have access to more up-to-date, complete patient notes, they can make better decisions about patient's care.

Some of the nurse informatics specialists perform patient care duties but most focus on developing, improving, testing or training nurses to use EMR and different modules of Hospital information system. Recognizing the potential of nurse informatics to improve quality care and reduce costs, some hospitals and health systems are creating staff roles for nurse informaticists.

Their titles include clinical analyst, informatics nurse specialist, and director of clinical informatics or clinical informatics coordinator.

Methods

The fully in house developed Hospital Information system is implemented first time in AIIMS, New Delhi. the system includes all the modules of HIS including ADT, laboratory management, Inventory management system, Stores management system, radiology module, OT management system etc. once the HIS developed and implemented in the hospital there was the requirement of trained man power to manage and further train others. Then the volunteers were asked from the Nursing practitioners to be part of it, and new concept of nursing informatics was born in India. The highly motivated and dedicated 60 nurses were selected and they were trained. They were successful managing the system with the help of IT staff. I was the team leader of the Nursing Informatics practitioners

Nurses are the major users of the EHR in the hospital because they use most of the modules of HIS and are responsible for a large portion of the documentation that addresses quality measures, safety measures and the overall clinical picture of the patient.

Need of nursing informatics in hospital information system:

- Analyse clinical and financial data.
- Promote and facilitate access to resources and references.
- Provide nursing content in standardized languages.
- Enable cost saving and productivity goal.
- Nurses focused to get a device that are integrated, voice activated, handheld, use biometrics, provide translation are portable are wireless, auto settle, and are mainly smart.
- Greater nurse satisfaction leads to greater patient satisfaction.

- Nurse Informaticists promote and facilitate access to resources and references'.
- Support for their mission to deliver high quality, evidence based care.
- Support for better services by facilitating true interdisciplinary care.
- Improvement in key relationships with provider and care recipients.
- Enable cost saving, time saving and productivity goal.
- Facilitate change management
- Enhance continuity of care
- Support and improve health information system
- . In All India Institute of medical sciences we have started our manual lab system to completely computerised lab results, tracking is also a most useful and productive part in health care system.
- In All India Institute of medical sciences Nurse informatics specialist helped hospital information system to full digitalise stores and all inventory system in hospital.
- We have successfully made all paper note and documents electronic.

Researchers are also been done and available through hospital information system

Results

Information management is integrated into nursing process and to digitize Assessment, Planning, Implementations, evaluation of nursing practice and complete healthcare system.

Discussion

The role of Nursing Informatics is to Standardized Documentation by collection of information

And its Management by reconciling the patient data.

They are the key to successful implementation of any electronic health related activities.

It plays a major role in Research and Evidence Collection

Conclusions

Nursing Informatics is developing field of study that is highly interdisciplinary. It is strongly connected to education, business and computer sciences.

References

[1] Canadian Nurses Association. (2001). What is nursing informatics and why is it so important? *Nursing Now*, 11

[2] Hebert, M. (2000). A national education strategy to develop nursing informatics competencies. *Canadian Journal of Nursing Leadership* 13(2).

[3] Staggers, N., & Bagley Thompson, C. (2002). The evolution of definitions for nursing informatics: A critical analysis and revised definition. *Journal of the American Medical Informatics Association, 9* (3), 255 – 262.

Address for correspondence

Dr. Susil Kumar Meher
Head, Health Informatics
Department of Computer Facility
AIIMS, Ansari Nagar, New Delhi 110029
Ph: 91-11-26594674, 26593570, 26593589
Fax: 91-11-26588663
Mobile : 91-8383024670
Email : skmeher@aiims.ac.in

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190732

Result and Effectiveness of Malicious E-mail Response Training in a Hospital

Hye Sook Lee, Da Na Jeong, Su In Lee, Shin Hae Lee, Kyung Hwan Kim, Hae Young Lee, Hyun Jai Cho, Sae Won Choi, Taehoon Ko

Office of Hospital Information, Seoul National University Hospital

Abstract

Malicious e-mails sent intentionally to institutions have caused an increase in data breaches. Measures against these methods must be taken by healthcare institutions to prevent leakage of sensitive personal medical information. As a form of training, we conducted a phishing simulation to gauge the response of the hospital staff to similar email attacks, and to raise awareness about information security protocols.

Keywords:
Education, Electronic mail, Hospital.

Introduction

In recent years, risk of personal information leakage has increased due to malicious e-mails sent intentionally by hackers. [1] These malicious e-mails are usually sent to staff accounts [2] and can infect the employee's computer to compromise and damage important data when opened. The infected computer can then be used to spread to other computers through the intranet and cause secondary damages. Because personal health information is extremely sensitive information, data leak involving hospital records may result in serious harm to the patients involved and damage the reputation of the hospital. For these reasons, Seoul National University Hospital (SNUH) has preemptively conducted malicious e-mail simulation training. This training is designed not only to understand the level of security awareness of the hospital employees, but also to prevent information leakage by raising the awareness of the employees.

Methods

Malicious e-mail response training was organized by the Office of Hospital Information of Seoul National University Hospital and was conducted with support from an external security consulting company. Employees who have intranet accounts with access to hospital information system, general staff of the information office and human resources department, system administrators of other departments, and clinical fellows were selected as the target group.

Response training was organized in three stages: preparation, training, and analysis. In the preparation stage, the target group was selected after setting up a response training plan. Information protection and personal information protection campaigns were conducted through banners and notices on the hospital intranet to raise employee's awareness. In the training stage, a fake phishing mail was randomly sent to the subjects in three different scenarios: modification guide of personal information from a shopping mall, penalty payment from Seoul Gangnam police station, and notice about certificate leakage by malicious code. The reaction information was collected and divided into cases of just reading, clicking on a link, or downloading an attachment. All cases were considered to be infected by malicious code and subjects were notified via a pop-up to contact information security officer of the information/system security team. In the analysis stage, the results of the response training were analyzed to derive future security strategy.

Results

Among 405 employees targeted, 222 people (55%) viewed the malicious mail, 63 people (16%) of the viewers clicked on the link, and 66 (16%) people downloaded the attachment. Of the 222 people required to report malicious mail, 70 (32%) reported the incident to the information/system security team.

The details of the training, risk of malicious code infection, and compliance were shared through intranet with the entire hospital staff.

Conclusions

The number of reports of suspected malicious mail to the information security team increased after the malicious email response training. It is expected that the awareness of personal information protection will be improved by expanding the number of trainees in the future and periodically conducting malicious e-mail response training with scenarios modified to better suit the hospital's characteristics.

References

[1] Eom Jungho, *the improvement Plan of a Customized Cyber-Training Structure for enhancing the Capability of Cyber Security*, Journal of Security Engineering, 2015

[2] Moon Jeayeon, Chang Younghyun, *Ransomware Analysis and Method for Minimize the Damage*, the Journal of the Convergence on Culture Technology, 2018

Address for correspondence

Hye Sook Lee

Mailing Address: Office of Hospital Information, Seoul National University Hospital, 101, Daehak-ro, Jongno-gu, Seoul, Republic of Korea

E-mail:amaranth@snuh.org

Phone: +82 2 2072 2774, Fax number: 82 2 2072 0050

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190733

Comparison of Medical/Health Informatics Education at the Best Global Universities for Clinical Medicine in Mainland China, Japan and South Korea

Jialin Liu [a b], Siru Liu [c], Yong Li[a]

[a] *Department of Medical Informatics, West China Medical School, Sichuan University, Chengdu, China;*
[b] *Department of Otolaryngology, West China Hospital, Sichuan University, Chengdu, China*
[c] *Department of Biomedical Informatics, University of Utah, Salt Lake City, USA*

Abstract

The aim of this study is to better understand the differences between medical/health informatics education in mainland China, Japan and South Korea. We compared medical/health informatics education at the top 10 universities in these three countries. Japan and South Korea have developed modernized education systems in medical/health informatics showing core features in the education of medical/health informatics. The universities in mainland China offer very few curriculum systems for medical/health informatics.

Keywords:

Education, Health informatics, Medical informatics

Introduction

Education in medical/health informatics has been discussed from various points of view around the world. The need for education in medical/health informatics has also been well recognized [1]. Curriculum development is a core issue, and there is a medical/health informatics program designed particularly for undergraduate and graduate students in medical school [2]. Asian countries such as Japan, mainland China, and South Korea have developed medical/health informatics programs. Compared with Japan and South Korea, the development of medical/health informatics as a discipline in mainland China is relatively lagging behind [3]. This study is to provide an overview of the medical/health informatics education in these countries and compare them. The aim of this study is to better understand the differences between medical/health informatics education in mainland China, Japan, and South Korea in order to further advance the development of medical/health informatics in mainland China and other developing countries.

Methods

We chose the top 10 universities for clinical medicine in mainland China, Japan and South Korea according to the U.S. News Rankings from 2018 [4]. We gathered information directly from the websites of the selected universities. Some information was found through Google and Baidu by constructing a query composed of one or more keywords. The search terms used were combinations of "education," "medical informatics," "health informatics", "undergraduate medical education," "graduate medical education," "curriculum", "course" and "university name". All the data were collected through the websites of government agencies, universities, academic societies, associations of practitioners, and relevant organizations. These data were categorized into undergraduate medical/health education, and graduate medical/health education, and the educational system in each category was compared across the countries.

Results

According to the U.S. News Rankings 2018, the top 10 universities for clinical medicine in mainland China, Japan and South Korea rank: 1. Japan from 85 to 299 (mean rank 204.5, mean score 57.86±4.57); 2. mainland China from 125 to 336 (mean rank 227.5, mean score 56.22±5.59); 3. South Korea from 72 to 418 (mean rank 248.5, mean score 55.18±8.30) (Table 1). According to the top 10 universities for clinical medicine in Japan, medical/health informatics courses are taken by both undergraduate and graduate students. Among the top 10 universities in South Korea, nine of them provide medical/health informatics course for both undergraduate and graduate students. However, among the top 10 universities in mainland China, they have not offered medical/health informatics courses.

Table 1. The Top 10 Universities in Mainland China, Japan and South Korea (clinical medicine)

No.	Mainland China	Japan	South Korea
1	Peking Univ.	Univ. of Tokyo	Seoul Nat'l Univ.
	125/64.5	85/68.8	72/69.9
2	Fudan Univ.	Kyoto Univ.	Yonsei Univ.
	131/64.1	136/63.7	146/63.0
3	Shanghai Jiao Tong Univ.	Osaka Univ.	Sungkyunkwa Univ.
	157/62.1	187/59.6	147/62.9
4	Sun Yat-sen Univ.	Tokyo Med & Dent Univ.	Univ. of Ulsan
	184/59.8	196/58.0	167/61.0
5	Capital Med. Univ.	Tohoku Univ.	Korea Univ.
	231/55.2	198/57.9	233/55.1
6	Central South Univ.	Keio Univ.	Catholic Univ. of Korea
	272/52.4	214/56.6	257/53.4
7	Huazhong Univ. of Sci & Tech	Kyushu Univ.	Kyung Hee Univ.
	272/52.4	226/55.7	301/50.3
8	Nanjing Med Univ.	Nagoya Univ.	Pusan Natl Univ.
	279/52.1	248/54.2	363/46.7
9	4th Military Med Univ.	Hollaido Univ.	Kyungpook Natl Univ.
	288/51.3	256/53.6	381/45.7
10	Shandong Univ.	Okayama Univ.	Ewha Womans Univ.
	336/48.3	299/50.5	418/43.8

Discussion

This study selected higher education institutions in three major Asian countries to analyze, compare, and contrast their medical/health informatics education. The top 10 universities have shown strength in producing research and developing disciplines. Therefore, we examined the medical/health informatics curriculum system in the top 10 universities for clinical medicine in mainland China, Japan, and South Korea. The curriculum development is a core component in the educational process[5]. The curriculum development in medical/health informatics has an exceptionally broad scope. It is not only about the students, the teachers, and the school, but also about the development of society and the discipline. Through studying the design of medical/health informatics curriculums in these universities, we may better understand the development of medical/health informatics in the countries. There are no medical/health informatics curriculum systems in the top 10 universities in mainland China. The result may be associated with the scarcity of teaching staff and professionals with comprehensive backgrounds in medical informatics, and the unclear teaching direction of the discipline in mainland China [3]. This comparison study provided information about the further development of medical/health informatics education in mainland China and developing countries. The educational needs of health and medical informatics should be well recognized and provide various forms of educational opportunities [6,7]. The findings of this study may facilitate medical/health informatics curriculum development in mainland China and developing countries.

The data collection in this study was based on official websites. Due to funding, time, and regional limitations, there is a lack of survey data, which may have some impact on the accuracy of this research output on medical/health informatics education.

Conclusions

Medical/health informatics has been recognized as a standard curriculum in the top 10 universities in Japan and South Korea. However, there is not a standard curriculum of medical/health informatics in the top 10 universities in mainland China. The design of the medical/health educational systems in Japan and South Korea may provide useful information for the development of medical/health informatics education in mainland China and other developing countries.

Acknowledgements

This project was supported by the National Natural Science Foundation of China (Grant Number:71273182/G0308)

References

[1] C.J. Green, G.H. van Gyn, J.R. Moehr, F.Y. Lau, and P.M. Coward, Introducing a technology-enabled problem-based learning approach into a health informatics curriculum, *Int J Med Inform* **73** (2004), 173-179.

[2] M. Okada, K. Yamamoto, and T. Kawamura, Health and Medical Informatics Education in Japan, *Med Inform* **1** (2005), 193-198.

[3] J. Lei, Q. Meng, Y. Li, M. Liang, and K. Zheng, The evolution of medical informatics in China: A retrospective study and lessons learned, Int J Med Inform **92** (2016), 8-14.

[4] Best Global Universities 2018, https://www.usnews.com/education/best-global-universities, accessed on December 26, 2018.

[5] K. Kranthi, Curriculum Development, *IOSR Journal of Humanities and Social Science* **22** (2017), 1-5.

[6] R. Haux, F.J. Leven, J.R. Moehr, and D.J. Protti, Health and medical informatics education, *Methods Inf Med* **33** (1994), 246-249.

[7] G.I. Mihalas, and L. Stoicu-Tivadar, History of Romanian Medical Informatics: Learning from the Past to Reshape the Future, *Med Inform* **27** (2018), 243-251.

Address for correspondence
Dr. Jialin Liu
327 Room, 8th Teaching Building, No.37 Wai Nan Guo Xue Xiang Street. Chengdu, Sichuan, China. 610041
Telephone number: +86-28-85422306
Fax number: +86-28-85582944
Email: DLJL8@163.com;

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190734

Current Status and Trends in Health Informatics Research: A Bibliometric Analysis by Health Technology and Informatics

Siru Liu[a], Jialin Liu[b,c] , Tao Zheng[c]

[a] Department of Biomedical Informatics, University of Utah, Salt Lake City, USA
[b] Department of Medical Informatics, West China Medical School, Sichuan University, China
[c] Information Center, West China Hospital, Sichuan University, Chengdu, China

Abstract

The aim of this study was to understand status and trends in health/medical informatics research. We used the Scopus database to extract all papers published in Studies in Health Technology and Informatics from 2008 to 2017. This study presented the key bibliometric indicators such as annual publications, top 10 authors, institutions, countries, and co-occurrence of keywords. These findings can be used to enhance our understanding of health/medical informatics research.

Keywords:
Health informatics, Medical informatics, Research, Bibliometrics

Introduction

The Studies in Health Technology and Informatics is a book series published by IOS Press. "This book series was started in 1990 to promote research conducted under the auspices of the EC programmes' Advanced Informatics in Medicine (AIM) and Biomedical and Health Research (BHR) bioengineering branch" [1]. It presents the proceedings of many international health/medical informatics conferences (e.g. MedInfo). The book series "has developed into a highly visible global platform for the dissemination of original research in this field, containing more than 250 volumes of high-quality works from all over the world" [2]. Our study aimed to provide a macroscopic overview of the main characteristics of the book series based on a bibliometric analysis and visualization map. It will bring a panoramic view of health/medical informatics research to scholars and help them to identify suitable researching orientation as well.

Methods

A literature search was conducted in Scopus for publications from January 1, 2008 to December 31, 2017, and used the following publication name: "Studies in Health Technology and Informatics". All data were collected by two authors and downloaded in text format.

After that, the records including all essential information (such as paper title, abstract, keywords, authors' names, affiliations, and references) were compiled for further analysis. Data extraction and analysis were performed using Microsoft Excel 2016. VOSviewer (Leiden University, Netherlands), a software tool for constructing and visualizing bibliometric networks, was used to construct a knowledge map of all keywords relations.

Results

Of the 8129 articles retrieved, the majority were conference papers (n = 7227, 88.90%), articles (n = 682, 8.39%), editorials (n=123, 1.51%), reviews (n = 59, 0.73%) and conference reviews (n=38, 0.47%). They were all published in English. Subject areas included medicine (33.4%), engineering (33.3%) and health professions (33.3%). Figure 1 shows the annual distribution of publications from 2008-2017.

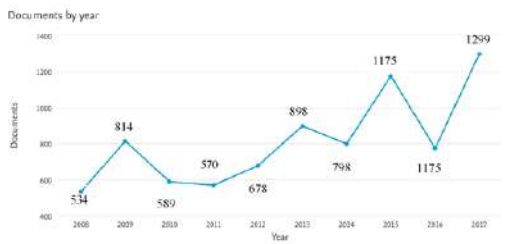


Figure 1 Annual distribution of publications

Institutions and Countries

The publications of Studies in Health Technology and Informatics originate from 111 different countries/regions and 14374 institutions/organizations. The top 10 productive institutions and countries are shown in Table 1. The greatest amount of research activity contribution in countries was from the United States (n=1592, 19.6%), followed by Germany (n = 818, 10.1%) and Canada (n = 660, 8.1%). The University of Victoria was the most productive institution (n = 284, 3.5%), followed by Inserm (n = 215, 2.6%), and Aalborg University (n= 155, 1.9%).

Contributing Authors

17476 authors contributed the 8129 papers. The top 10 most prolific authors in the book series, with their number of publications, are shown in Table 2. Borycki EM clearly obtains the first position with 95 (1.17%) articles.

Co-occurrence of author keywords analysis

The period ranging from 2008 to 2017 included 13438 keywords from 8129 articles, 1047 keywords appeared 5 or more times. The term maps show 500 keywords having the greatest total link strength (Fig 2a, 2b). It grouped 12 clusters. The most common keywords are "electronic health record/records", "virtual reality", "telemedicine", "ehealth" and "patient safety".

Table 1 Top 10 most productive institution and countries

No.	Institution	Record	Country	Record
1	Univ of Victoria (Canada)	284 3.5%	USA	1592 19.6%
2	Inserm (France)	215 2.6%	Germany	818 10.1%
3	Aalborg Univ (Denmark)	155 1.9%	Canada	660 8.1%
4	Univ of Athens (Greece)	110 1.4%	UK	560 6.9%
5	UiT The Arctic Univ of Norway (Norway)	83 1.0%	Australia	490 6.0%
6	Acad Med Cent, Univ.of Amsterdam (Netherlands)	81 1.0%	France	448 5.5%
7	Univ Cattolica del Sacro Cuore (Italy)	79 1.0%	Italy	367 4.5%
8	Univ of Amsterdam (Netherlands)	77 0.9%	Japan	325 4.0%
9	Karolinska Institutet (Sweden)	75 0.9%	Norway	295 3.6%
10	Univ of Tasmania (Australia)	74 0.9%	Austria	294 3.6%

Table 2 Top 10 of most productive authors

Authors	Record	Institution	Country
Borycki EM	95	Univ of Victoria	Canada
Kushniruk, AW	86	Univ of Victoria	Canada
Blobel B	83	Univ of Regensburg	Germany
Riva G	75	Univ Cattolica del Sacro Cuore	Italy
Househ M	56	Kin Saud Bin Abdulaziz Univ for health	Saudi Arabia
Mantas, J	52	Sci	Greece
Lovis C	49	National and Kapodistrian Univ of Athens Univ Hosp of Geneva	Switzerland
Kushniruk A	44	Univ of Victoria	Canada
Nøhr, C	43	Aalborg Univ	Denmark
Saranto, K	43	Univ of Eastern Finland	Finland

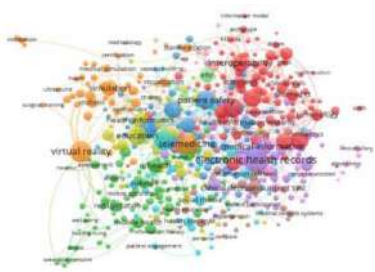


Figure 2a Network visual keywords analysis

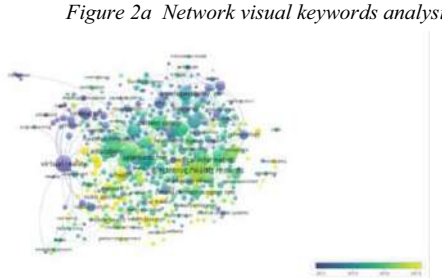


Figure 2b Overlay visualization keyword analysis

Discussion

The number of articles produced annually has obviously increased over this period (2008-2017). The number of publications has been considered a measure of scientific productivity and interest in health informatics research.

A co-occurrence analysis of authors keywords of the publications can provide insight into the main topics and research trends in health/medical informatics. In co-occurrence, the relatedness of terms is determined based on the number of publications in which two terms occur together. The result of the terms analysis is presented in Figure 2a-2b. The size of the circles represents the occurrence of a term (the bigger the circle, the higher the occurrence of a term in the author keywords). The overall distance between terms offers information on their relatedness. The shorter the distance between two terms, the stronger their relation. The relatedness of terms is determined by counting the number of times that terms occur together in the keywords [3]. The colors are used to distinguish between different clusters. The term map of overlay visualization is the timeline view of the network (Figure 2b). The color of a term indicates the term's average publication year (the terms with yellow show the times with more recent publication year). It shows chronologically the most applied keywords which are "big data", "universal design'', "mhealth", "social media", "health literacy", and "vocabulary". This transformation shows researchers are increasingly interested in these research themes.

Conclusions

These outcomes help us perceive the evolution of research in the field of health/medical informatics over a 10 year period. The study also shows Studies in Health Technology and Informatics is an important global conference proceeding in digital health and health/medical informatics.

Acknowledgements

This project was supported by the National Natural Science Foundation of China (Grant Number:71273182/G0308)

References

[1] https://www.scimagojr.com/journalsearch.php?q=19582&tip=sid [accessed on October 21, 2018]
[2] http://ebooks.iospress.nl/bookseries/studies-in-health-technology-and-informatics [accessed on August 20, 2018]
[3] S.P. Rodrigues, N.J. Van Eck, L. Waltman, and F.W. Jansen, Mapping patient safety: a large-scale literature review using bibliometric visualisation techniques, *BMJ Open 4(*2014),1-8.

Address for correspondence
Dr.Jialin Liu
327 Room, 8th Teaching Building, No.37 Wai Nan Guo Xue Xiang Street. Chengdu, Sichuan, China. 610041
Telephone number: +86-28-85422306
Fax number: +86-28-85582944
Email: DLJL8@163.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190735

Educational Game as an Aid to Good Practices in Dentistry

Maria Lúcia Bezerra Feitosa[a], Márcia Maria Pereira Rendeiro[a], Ana Emília Figueiredo de Oliveira[b]

[a] *State University of Rio de Janeiro (UERJ), Rio de Janeiro RJ, Brazil,*
[b] *Federal University of Maranhão (UFMA), São Luís, Maranhão, Brazil*

Abstract

The use of games may constitute an innovative strategy for training.

Objective: Develop an educational quiz type game, construct 60 assessment items and validate this content.

Methodology: Elaboration of the didactic-pedagogical project,

development of the game, creation of the assessment items and validation of the content.

Results: the technical evaluation obtained a Content Validity Index (CVI) over 80%.

Conclusions: The game can be used as a tool for making information available, thus contributing to the democratization of knowledge.

Keywords:
Experimental games, dentistry, infection control.

Introduction

Provision of health care is always associated with occupational hazards related to work practice. In several studies, there is a lack of compliance with current legislation concerning biosafety norms, which draws attention to the need for professional training [1,2]. Several researches are proposing the use of Serious game as an interactive, playful object, as well as a learning facilitator and supporter [3,4].This study aims to develop an educational game and validate, pedagogically and technically, the 60 assessment items to be used within the game, as an aid to educational practices on biosafety in Dentistry.

Methods

Elaboration of the Didactic-Pedagogical Project

The game, called "Biosafety in Dentistry", consists of 60 assessment items created from the main theme and arranged in a database. This number of items allows the random use of 20 questions in each game. The defined target player involves dental surgeons, undergraduate and graduate students in Dentistry. The developed assessment items underwent pedagogical and technical validation.

Development of the Serious Game / Motivation strategies

The educational game uses a 2D interface. The technologies used to develop this tool are three (03): Unity3D®, the game engine, Adobe Illustrator CS6®, arts authoring tool and Ableton Live®, for background sound editing.

To develop the game, participation and interaction of several professionals was necessary, all from several areas such as: Pedagogy, Technology, Health, Graphic and Instructional Design.

This computational system was developed and registered at the National Institute of Industrial Property (Portuguese acronym INPI), through process nº 51 2017 000649 0, by the Federal University of Maranhão Foundation - UFMA.

Five (05) developmental scenes were created, which are listed below.

Figure 1- Game opening

Figure 2- Character selection

Figure 3- Tutorial

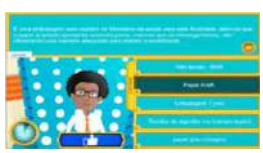


Figure 4- Gameplay

Figure 5- Final ranking

The game presents different tools and strategies: tutorials, scores, response time, opportunities to get tips and ranking, as well as the game topics, images and sounds, with characters in motion.

Creation of Assessment Items

The content is related to biosafety standards and hazard prevention.

Pedagogical Validation

It was carried out with the support of (02) evaluators of the Pedagogical Department of the Universidade Aberta do Sistema único de Saúde (UNA-SUS) and of the UFMA.

Technical Validation

The questions on the biosafety in Dentistry game were evaluated by (05) professionals of reference in biosafety, with expertise in the area.

Results

The 60 questions were submitted to the statistical analysis of data. For this analysis, the Content Validity Index (CVI) was used.

The concordance index between them was > 80%, referring to the criteria of textual clarity, practical relevance, suitability to the target player and response time.

Conclusions

At the end of this study, we conclude that the objectives of developing an educational quiz type game on biosafety and validating the 60 questions were pedagogically and technically achieved, with values over 80% for all the analysed criteria.

References

[1]Berlet LJ, Ascari RA, Silva OM, Trindade LL, Krauzer IM, Jacoby AM. Factors that influence the quality of the sterilisation process. J Nurs UFPE on line. 2014;8(7):1997-2003. https://periodicos.ufpe.br/revistas/revistaenfermagem/article/viewFile/11527/13422. Access in January 22, 2018.

[2]Mendonça ACC, Bezerra ALQ, Tipple AFV, Tobias GC. Indicadores de qualidade de processamento de produtos para a saúde em autoclaves a vapor. Rev Enferm UFPE on line. 2017;11(Supl. 2):906-14. https://periodicos.ufpe.br/revistas/revistaenfermagem/article/view/13459/16148. Access in April 2, 2018.

[3]Rung A, Warnke F, Mattheos N. Investing the use of smartphone for learning purposes by Australian dental students. JMIR. MHealth UHealth.2014;2(2): e20. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4114424/. Access in February 7, 2018.

[4]Rondon S, Sassi FC, Andrade CRF. Computer game-based and traditional learning method: A comparison regarding students' knowledge retention. BMC Medical Education, London, 2013; 13(30): 1-8. https://bmcmededuc.biomedcentral.com/track/pdf/10.1186/1472-6920-13-30 Access in January 3, 2018.

Address for Correspondence

Maria Lúcia Bezerra Feitosa
luciafeitosa1@gmail.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190736

A Framework for Enhancing and Updating Study Programs in Public Health and Medical Informatics Fields in Montenegro

Andriana Magdalinou[a], John Mantas [a], Ramo Šendelj [b], Ivana Ognjanović [b], Petra Knaup [c], Elske Ammenwerth [d], Orsolya Varga [e], Goran Nikolić [f], Anđela Jakšić Stojanović [g], Dragan Đurić [h]

[a]*Department of Nursing, National and Kapodistrian University of Athens, Athens, Attica, Greece,*
[b]*University of Donja Gorica, Podgorica, Montenegro,*
[c]*Institute of Medical Biometry and Informatics, University of Heidelberg, Heidelberg, Germany,*
[d]*University for Health Sciences, Medical Informatics and Technology, Tirol, Austria,*
[e]*Department of Preventive Medicine of Faculty of Public Health, University of Debrecen, Debrecen, Hungary,*
[f]*Department of Medicine, University of Montenegro, Podgorica, Montenegro,*
[g]*University Mediterranean, Podgorica, Montenegro,*
[h]*Institute of Modern Technology Montenegro, Podgorica, Montenegro*

Abstract

Montenegro plans to enhance and modernize the curricula and programs in public health fields in line with EU standards and hence the Erasmus-Phelim project developed a framework to develop, implement, and evaluate the education process. A stepwise approach consisting of three dimensions per step was implemented for workshop development.. For the evaluation, a train-the-trainer approach was developed and a self-regulation concept consisting of three phases was applied.. Semi-structured interviews with the workshop participants were conducted and results suggested that self-regulation is an understandable concept and can be applied as a training and knowledge transfer method.

Keywords:
Education, Public Health Professional, Self Concept

Introduction

Montenegro plans to enhance and modernize the curricula and programs in public health fields in line with EU standards. For this purpose the Erasmus-Phelim project developed a framework to plan, implement, and evaluate the educational process in an attempt to improve public health professionals' skills in creating sustainable and flexible health system that delivers good quality health care services and protects health of the citizens [1]. It is an important opportunity to explore to make educational activities, services, and programs an integral part of the community and to ensure their longevity [2]. Additionally, the sustainable educational approach will be developed at all levels from undergraduate to doctoral level that, includes evidence-based approach through innovative National Platform for Education and Research in Public Health.

During our project implementation, the educational activities required a train-the-trainer approach to workshops in which the Self-regulation concept was applied as shown in Figure 1. It consists of three phases: Forethought, Performance Control and Self-reflection [3].

Figure 1 – The Three Phases of the Self-regulation Concept

Furthermore, selected awareness programs about health prevention and public health promotion campaigns were communicated with Montenegrin citizens to engage them with public health issues and to improve the public health.

Methods

A new master's degree program was designed based on an analysis of the current academic programs in other European countries. A stepwise approach consisting of five phases was implemented during the project, with each step consisting of three dimensions by simultaneous and coherent actions. For the planned workshops, a train-the-trainer approach was developed and the Self-regulation concept was applied. The Self-regulation encompasses three phases: Forethought, Performance Control, and Self-reflection. Qualitative semi-structured interview with the workshop participants were conducted.

Results

The Stepwise Approach

A cyclical stepwise approach was implemented as illustrated in Figure 2. The first step was to provide detailed analyses of public health education in Montenegro and cross-matched with best EU standards and practices. The second step required the creation of core elements of formal educational system in public health (including all levels, from undergraduate to post-graduate studies, including master's and doctoral studies). At the third stage sustainable strategies for training of professionals in multidisciplinary public health fields were established. In the final step, evaluation of training strategy was implemented.

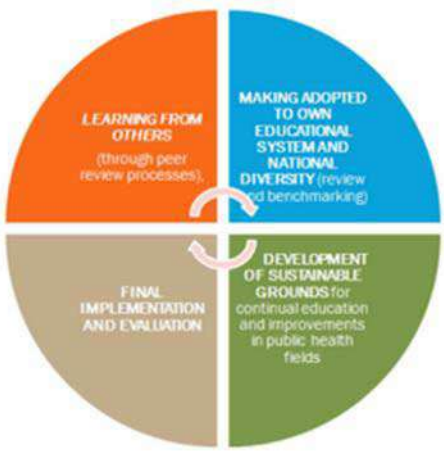


Figure 2 – The Stepwise Approach in Enhancing the StudyPprograms

The stepwise approach consists of three dimensions as shown in Figure 3 including the development of the curricula and training activities, the ongoing quality assurance of educational process. The public health promotion results are demonstrated in Figure 4.

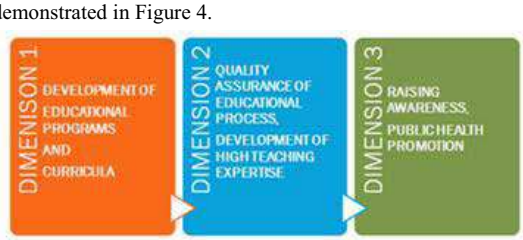


Figure 3 – The Three Dimensions of theStepwise Approach

Within this framework , the Self-regulation concept was applied in two training workshops. Qualitative Semi-structured interviews with the nineteen workshop participants were conducted and the results are shown in Figure 4.

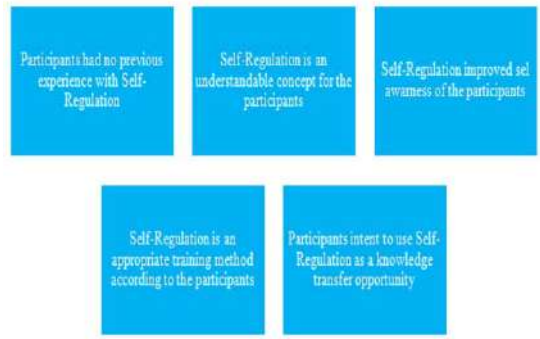


Figure 4– The Results of the Semi-structured Interviews

Conclusions

A stepwise approach for the development of the education process in Montenegro was applied, with each step consisting of three dimensions. A train-the-trainer approach was developed during the workshops and the Self-regulation concept was applied. The qualitative semi-structured interviews with the participants of the workshops suggested that Self-regulation was an understandable concept that can be used as an appropriate training method and a knowledge transfer opportunity to improve participants self-awarness.

Acknowledgements

Enhancement of study programs in Public Health Law, Health Management, Health Economics, and Health Informatics in Montenegro (PH-ELIM) is an Erasmus-Phelim project for Capacity building in higher education EAC/A04/2015 Selection Year 2016.

References

[1] I. Ognjanović , R. Šendelj, P.Knaup,J. Mantas, R. Adany, E. Ammenwerth,G. Nikolić,A J. Stojanović, and D. Đurić, The Stratified Framework for Enhancement of Study Programs in Public Health in Montenegro, *IOS press Studies in Health Technology and Informatics*, (2018), 301-304.

[2] Office of Adolescent Health, *Building Sustainable Programs: The Framework Office of Adolescent Health*, 2014.

[3] D. H. Schunk, *Learning Theories: An Educational Perspective, Sixth Edition*, Pearson, 2012.

Address for Correspondence

Andriana Magdalinou
E-mail address: alfami@nurs.uoa.gr

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190737

Construction of a Home Digital Signage System to Promote Walking as a Physical Activity

Akira Ochiai[a], Tadamasa Takemura[a]

[a] *Graduate School of Applied Informatics, University of Hyogo, Kobe City, Hyogo Prefecture, Japan*

Abstract

Physical inactivity is a social problem related to health. Based on the concept of health communication, this study aims to construct a mechanism to provide continuous information to improve physical activity. We built a "home digital signage system" which consists of a function of collecting the localized information from the websites and distributing to the tablets. The results of the introduction of this system have been suggested to promote walking by increasing media exposure.

Keywords:

Health Communication, Exercise, Information Seeking Behavior

Introduction

Physical inactivity is thought to increase the risk of non-communicable diseases [1]. World Health Organization (WHO) guidelines for physical activity are recommended for walking that is easy to carry out with walking or shopping. The target values are also shown. However, in Japan, the number of steps from 2010 to 2015 tends to be flat [2]. With regard to such social problems related to health, there is health communication as a concept to effectively transform the behavior of the person. Health communication is defined that a process of promoting behavioral transformation and social change through message exchange and improving health results [3]. There are previous studies on health communication which carried out the campaign of improvement of physical activity and eating habits in the inter-mountain areas of Japan. Shimazaki et al. investigated the characteristics of the target group in advance and distributed the messages prepared according to gender and lifestyle through various information media. As a result, paper media such as leaflets and public relations magazines distributed to all households are reported to be effective [4]. The high awareness of information from media with many opportunities to see it is explained in the media exposure theory in a study that showed the impact of the provision of health information by mass media. The media exposure theory has been suggested how much exposure to health information affects information perception and behavior [5; 6]. However, providing continuous information in a wide area through paper media is difficult.

On the other hand, mass-media such as television, newspapers and magazines can provide to a wide area, but we have expected that the information becomes uniform and the cost becomes high. Meanwhile, the network environment has been improved, and it is becoming possible to provide dynamic information using a system that can always communicate. In advertisement media, digital signage for dynamically presenting information is spreading. Digital signage is possible to constantly display information in the living space, change to information of high interest according to the situation of the area and individual, and continuously provide information for a long time. This characteristic may increase media exposure.

Therefore, in this study, we built a digital signage system for home as a mechanism to promote walking as physical activity using information terminals such as tablet-type devices to continuously provide information according to the situation. We also examined whether information provision by these mechanisms affects the promotion of walking as physical activity.

Methods

Review of system requirements

We examined the mechanism of "home digital signage" from three perspectives: home digital signage terminal, information gathering, and distribution, user interface (UI).

Home digital signage terminal

We considered that digital signage terminals installed indoors need different requirements from the conventional terminals installed outdoors. Our preliminary interview about the size and installation place suggested to use tablets as terminals because most opinions were that "it is better not to be too big" and "I want to put it in a living room with a TV". In this study, we used iPad Air and installed in easy to see places.

Information collecting and distribution

There were many possible targets for information reminding of physical activity, but this time we focused on walking and used local information; nearby events, shopping facilities, tourist guides, and restaurant guides. These were based on our hypothesis “local information leads to leaving the house.”

In order to keep updating the information, we have needed a mechanism to automatically collect by using web scraping. However, due to the websites of the data sources often uploading data in the form of non-machine-readable content, we also needed human intervention. Regarding to distribution, the information should be automatically sent to the terminals.

User interface

Previous studies on methods of presenting health information have shown that using images and simple letters contribute to understanding of contents [7; 8], so the system should use images when available. With reference to the criteria for sighted people, including the elderly [9], the height of letters in this system should be set the body to 9 mm and the title to 12 mm. The animation sliding from the right to the left is easy to notice that the screen has switched [10], so we adopted a left slide show. The display time could be considered for 30 seconds with reference to the display time of one slide in a large digital signage tends to be 11 seconds or more and 30 seconds or less.

System construction

This system consists of a function to collect contents and a function to distribute contents to digital signage terminals. MySQL for database as technology, Python program for web scraping, content delivery, and screen display using HTML, PHP, and JavaScript.

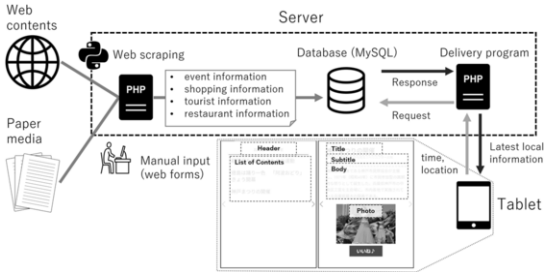


Figure 1– System outline and screenshot

As a specific operation, web scraping is periodically executed on the server, and information such as the date and time, location and contents of event information and shopping information is collected. In cases where web scraping would be difficult, we used input forms to obtain information directly. The collected information is accumulated in the database on the server as the distribution content of the digital signage. The tablets terminals periodically accesses the server to extract the latest local information, download the content to be delivered in webpage format, and display it.

Evaluation

In order to evaluate the influence on the promotion of media exposure and physical activity by this system, we conducted a hearing survey and measured the number of steps for subjects. We interviewed about the awareness of information and the willingness to practice walking, which are considered as factors related to the implementation of physical activity in the media exposure theory. The subjects recruited middle-aged and elderly non-workers in order to eliminate the influence of commuting. For ethical considerations, we explained that the purpose of this survey, protection of personal information, and right to cancel in the case such as negative condition occurred. There were three subjects: a man in his 90's, a woman in her 80's, and woman in her 50's. We investigated the frequency of access to the home digital signage terminals and the intention to go outside to walk. We also measured the number of steps through the wearable activity monitor "Fitbit Charge 2" for five days before the introduction of the system and for five days afterward. Fearing that the introduction of an activity monitor could influence the subjects' activity, we had them wear the activity monitors for a month to get used to them.

Results

With our constructed system, we were able to continuously display information to the digital signage terminal installed at home. The daily average number of steps increased after the introduction of the system (Table 1).

Table 1– Comparison of the daily average number of steps

Subject	Old	Sex	Before	After	Diff.
A	50s	Female	4,707	5,332	+625
B	80s	Female	2,482	2,603	+121
C	90s	Male	2,286	2,814	+528

The subjects answered that they would access the terminal once every one to three hours, for reasons such as "I wanted to know whether there was any new information" and "I checked it because it changes every once in a while." Regarding their will to walk after the introduction of the system, they made comments including the following: "I would see information on events and, when going out for other business, stop at the events" "In a month's time, I feel like I would go out and shop more" "I feel more stimulated to walk" "If the weather and my physical condition allow, I want to go out".

Conclusions

With this system, it became possible to continuously present the collected information. Subjects are browsing about once every 1 to 3 hours in anticipation of displaying the latest information, and the home digital signage system is thought to be effective in increasing media exposure. From the results of the step count data and the hearing survey, it has been suggested to promote walking by increasing media exposure. However, further examinations are required to assess the system's long-term effects on physical activity.

References

[1] World Health Organization, Global recommendations on physical activity for health, (2010). http://www.who.int/iris/handle/10665/44399 (accessed October 24, 2018).

[2] Ministry of Health, Labor and Welfare, National Health and Nutrition Survey Report, (2015). http://www.mhlw.go.jp/bunya/kenkou/eiyou/dl/h27-houkoku.pdf (accessed October 25, 2018).

[3] R. Schiavo, Health communication: From theory to practice, *Jossey-Bass*, San Francisco, 2007.

[4] T. Shimazaki, Health Communication: Support for making habits of health behavior, *Waseda University Press Section*, Tokyo, 2016.

[5] B.J. O'Hara, E.B. Adrian, P. Philayrath, Using mass-media communications to increase population usage of Australia's Get Healthy Information and Coaching Service, *BMC Public Health* **12** (2012), 762.

[6] D.R. Morris, M.P. Rooney, R.J. Wray, M.W. Kreuter, Measuring exposure to health messages in community-based intervention studies: A systematic review of current practice, *Health education & Behavior* **36** (2009), 979-998.

[7] P.S. Houts, C.C. Doak, L.G. Doak, M.J. Loscalzo, The role of pictures in improving health communication: A review of research on attention, comprehension, recall, and adherence, *Patient Education and Counseling* **61** (2006), 173-190.

[8] S. Kripalani, R. Robertson, M.H. Love-Ghaffari, et al., Development of an illustrated medication schedule as a low-literacy patient education tool, *Patient Education and Counseling* **63** (2007), 368-377.

[9] Foundation for Promoting Personal Mobility and Ecological Transportation, Basic survey report on character signature in public facilities etc., (2017). http://www.ecomo.or.jp/barrierfree/report/data/27_03_moji.pdf (accessed October 25, 2018).

[10] J. Endo, K. Hayami, J. Nakamura, Evaluation of the Effect of Transition in Digital Signage Screen Designs, *Journal of Graphic Science of Japan* **45** (2011), 3-10.

Address for correspondence

Akira Ochiai

Mailing-address: 7-1-28-618, Minatojima-minamimachi, Chuo Ward, Kobe City, Hyogo Prefecture, 6500047, Japan

E-mail: akira.ochiai@me.com

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190738

Health Information Sharing Among Muslim Women in a Japanese Mosque

Katsumasa Ota[a], Haruka Furusho[a], Ayana Mawaki[a], Yukari Niimi[b], Chikako Ikegami[c], Naoko Arakawa[d], Jukai Maeda[e]

[a] *Department of Nursing Sciences, Graduate School of Medicine, Nagoya University, Nagoya, Japan*
[b] *Department of Nursing Science, Seirei Christopher University, Hamamatsu, Japan*
[c] *Department of Nursing Science, Nagano College of Nursing, Komagane, Japan*
[d] *College of Life and Health Sciences, Chubu University, Kasugai, Japan*
[e] *Faculty of Nursing, Tokyo Ariake University of Medical and Health Sciences, Tokyo, Japan*

Abstract

With the increase of foreign residents in Japan, most of them are suffering from inadequate health information. In reference to an ethnography, we conducted fieldwork using a participant observation and an interview with 36 foreign Muslim women. Our study clarified they especially needed health information on childcare because of a lack of such information in their native languages, and needed the support of Japanese Muslims because of problems in communication with healthcare personnel.

Keywords:

Information Seeking Behavior, Cross-Cultural Comparison, Islam

Introduction

The number of foreign residents living in Japan has recently exceeded 2.3 million. Because they are living in a different cultural society, foreign residents tend to face language problems, difficulties accessing the healthcare service system, and preconceptions by Japanese people [1]. Moreover, it has been suggested that emotional and informational support is the most important for foreign mothers [2]. Studies of immigrant health that work with immigrant institutions are an important part of health promotion [3]. According to earlier research, faith communities can be effective partners in reducing health disparities among residents and promoting community health. An American study underscores the potential utility of sermons for health education programs and health behavior interventions in American mosques [4]. However, few studies on medical care focus on the community of foreign residents, especially regarding female Muslims.

Our study, therefore, targets Muslim women who experience many religious limitations and cultural differences in a Japanese context. In addition, we also wanted to present the actual situation of healthcare information shared among them using a social networking service. Our results may provide suggestions for health centers and medical institutions regarding potential support for those foreign residents.

The purpose of this study is to clarify the needs for health information by Muslim women and the role that mosques play to protect the health of foreign resident women and ensure they receive the proper healthcare.

Methods

In reference to an ethnography, we perfromed fieldwork to describe the basic behavior patterns of individuals and groups in this specific culture. In addition, we used the following models proposed by Leininger to understand participant groups because of differences between the researcher and participants in terms of country of origin, culture, and religion.

1. Stranger-Friend Model
2. Observation-Participation- Reflection Model (OPR)

Participants

Participants in this study were women, aged 20 and older who visited an Islamic religious institution (mosque) in "A" Prefecture.

Data collection

Participant observation

When the study began, we were not acquainted with any of the participants. Therefore, we first participated in worship services, dinners, study sessions, and meetings at the religious institution and started to observe the conversations and actions of the participants. We deepened our understanding of their circumstances and backgrounds by simply observing for a certain period of time before beginning to participate in a more direct way. We determined key informants who were regarded as reflecting the norms, values, beliefs, and general lifestyles of the culture, and who were usually interested in and willing to participate in the study. We consistently evaluated the relationship with the group, and when the relationship deepened, we strengthened our position as insiders and collected information. We recorded data in field notes instead of using a recording device.

Interview

After we examined the detailed information shared by participants during the observation process, if necessary, we individually questioned key informants about details with help from other Muslims who could speak Japanese.

Ethical considerations

All data collection began after IRB approval (17-1486). We obtained oral agreements during observations and informed consent from participants at interviews.

Data analysis

Before analyzing the data, we clarified our own personal faith, cultural backgrounds, and particular factors that would bias the data analysis. We scrutinized descriptions of the interviews based on a content analysis method. We analyzed and categorized the data from both "emic" (as an insider) and "etic" (as an outsider) viewpoints. In addition, we endeavored to analyze the data within its cultural context by confirming the content that we had obtained through interviews with key informants.

Results

Informant attributes

The 36 informants were in their 20's–50's. The participants came from countries in Asia and Africa: Afghanistan, Iraq, Indonesia, Uzbekistan, Egypt, Syria, Bangladesh, Malaysia, and Japan.

Results from participant observation

We summarized three themes from the 20 sessions of participant observation.

Emotional support

In the community room in the mosque, dinner was offered between worship times in the afternoon and evening every Saturday. The worship attendants could freely participate in the meal.

Strict times were observed for entering the room where the meal was held to keep men and women separate. A female researcher participated in the times for women. Therefore, we could observe the situation while only women and their children ate dinner. The participants formed groups of several people, talked, and sometimes shared information. Indonesian women, in particular, gathered together in groups of five or six each time and chatted pleasantly during dinner.

> *"Here [in the mosque, people are] the most reliable. Their [original] country is the same [as mine] and they are Muslim. There are some Indonesians who gave up coming to pray because [they live in] Japan. People visiting here are really reliable. [They are] friends."*

Discussions about childcare

Informants frequently conversed about childcare. These conversations included a varieth of content, such as school and after-school club activities. Sometimes, participants discussed topics related to Islam, such as

> *"Even if tomorrow is Eid[Islamic celebration], I will not let [my daughter] be absent from school [tomorrow]."*

In addition, the Malaysian participant described her experience of consulting with a Malaysian who had recently come to Japan.

> *"Vaccination. Japanese [vaccinations] are different from Malaysian [vaccinations]. Indeed, Japan has fewer [kinds]. She asked me what her children should have because [her children] may go back to Malaysia in the future".*

Requests for supports

In the Japanese mosque, Japanese Muslims are often asked by foreign Muslims to explain and translate school and hospital documents. We observed that Japanese Muslims explained referrals to a large hospital to foreign Muslims.

One Japanese Muslim who was very supportive of foreigners said,

> *"They ask me for various things such as reading [Japanese] documents. Sometimes I clearly express the Japanese in other words, then I feel [that it is] not my specialty..."* .

In addition, an Indonesian Muslim stated the following:

> *"Local anesthesia makes my heart beat fast. Though I told the dentist about that, he said just "relaaaax." He doesn't trust and understand me because I'm a foreigner"* .

Some foreigners feel that doctors do not carefully listen to what they say because they are foreigners.

Conclusions

Community within mosqus as emotional support

Even though they live in a cross-cultural society, the participants have developed emotional connections with people sharing a common attribute, such as religion or country of origin. Although they utilized a social networking service [SNS], they still needed face-to-face communication with each other. Further observation is required to determine how some groups, other than Indonesians, discuss and share information.

Concerns about childcare topics and requests for information about childcare

Conversations about childcare occurred frequently, indicating that these were topics that concerned participants a great deal. Because they were part of cultural and religious minorities, the participants faced troubles concerning their childcare and discussed those issues with each other. Future research should clarify these problems more concretely.

Support by Japanese Muslims

In addition to the support provided by participants from the same country, our results indicate that foreign Muslims often asked Japanese Muslims, whose nationalities are Japanese, for their support which could not be provided by SNS. Although Japanese Muslims did not have special knowledge about the topic of consultation, they easily understood the participants' problems because they usually talked with them. In addition, sometimes Japanese Muslims stood between foreign Muslims and other native Japanese, which made it easier to communicate even if the other Japanese person was prejudiced against foreigners. We should clarify the role of Japanese Muslims as advocates. These results could provide useful suggestions for improving the understanding of healthcare workers at health centers when they contact those foreigners.

Acknowledgements

Funding was provided solely by our university.

References

[1] M. Teraoka, Y. Muranaka, Aspects of cross-cultural experience perceived by foreigners living in Japan when using its healthcare services, *Japan Academy of Nursing Science* **37**(2017), 35–44.

[2] S. Imai, S. Kita, H. Tobe, M. Watanabe, A. Nishizono-Maher, K. Kamibeppu, Postpartum depressive symptoms and their association with social support among foreign mothers in Japan at 3 to 4 months postpartum, *International Journal of Nursing Practice* (2017), doi: 10.1111/ijn.12570.

[3] Y.Covington-Ward, K. Agbemenu, A. Matambanadzo, "We feel like it was better back home:"Stress, Coping, and Health in a U.S. Dwelling African Immigrant Community, *Journal of Health Care for the Poor and Underserved* **29** (2018), 253‑265.

[4] A.I. Padela, S. Malik, S. Ahmed, Acceptability of Friday sermons as a modality for health promotion and education, *Journal of Immigrant and Minority Health* (2017), doi:10.1007/s10903-017-0647-8.

Address for correspondence
Katsumasa Ota kota@met.nagoya-u.ac.jp

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190739

mHealth Apps for Self-Management of Chronic Conditions in France: What Is out There?

Luiza Prado[a], Camille Carpentier[a,b], Marie Préau[b], Anne-Marie Schott[a], Alexandra Dima[a]

[a] EA 7425 Health Services and Performance Research, Université Claude Bernard Lyon 1, Lyon, France
[b] EA 4163 Groupe de Recherche en Psychologie Sociale, Université Lumière Lyon 2, Lyon, France

Abstract

mHealth apps use is rapidly increasing and has high potential value for people with chronic conditions, supporting behaviour change for self-management and improving patient-provider communication. To build a well-functioning app development environment, we need to examine the range of available apps according to their variability and content. A review of Google Play apps revealed limited availability for chronic self-management in France among substantial unrelated content. Most apps' content was difficult to understand and act upon.

Keywords:

Telemedicine, Chronic Disease, Self-Management

Introduction

mHealth apps are widely downloaded worldwide; over half of mobile phone users in the US have downloaded at least one health-related app in 2015 [1]. As people tend to carry their smartphones with them at all times, apps have the potential to help people improve their lives through learning to cope with different situations [2]. For chronic conditions, behaviour change is essential for reaching and maintaining quality of life and preventing premature death [3][4]. Best benefits can be achieved by targeting mHealth to the 4 groups of chronic conditions responsible for more than 70% of deaths worldwide: cardiovascular diseases, cancer, respiratory diseases, and diabetes [5]. Additionally, to evaluate if these tools are effective in supporting behaviour change, it is useful to consider them as health-related materials. As such, the information presented needs to be accurate; easily understood by persons with different communication competencies, styles, and health literacy levels; and must offer suggestions of actions the user can take concerning health events to optimize their reach and enhance health decision making [6].

We reviewed mHealth apps available in France on Google Play store in April 2018 to examine their content and relevance for self-management of chronic conditions. Here, we aimed to better understand the current offer to guide future improvements in the development and use of these tools, using the two main ways of finding self-management apps on marketplaces: the "TOP" list and keyword search. We also assessed the selected apps' understandability and actionability levels using the Patient Education Material Assessment Tool for audiovisual materials (PEMAT-A/V) [7]. This work is part of a larger project that evaluated mHealth apps in chronic conditions.

Methods

We extracted the first 500 free and 55 paid apps labeled as "TOP" on the "Médecine" category on Google Play store. We considered 500 apps to be a representative sample, and 55 was the display limit for paid apps. We also performed a keyword search using 12 popular words related to the 4 groups of conditions mentioned before (e.g., maladie cardiaque, accident vasculaire cérébral, maladie respiratoire, cancer, diabète). We included the first 20 apps for each keyword. Name, description, number of downloads, stars, ratings, version, last update, and developer information were extracted for all the apps. Apps in French were classified in 8 categories according to their store descriptions (Chronic conditions, Reproductive health, Students and professionals, Information for the general public, Nutrition, Pranks and fake tests, Other). This was an inductive categorization, following previous studies [8]. Each category was then subdivided following specific app purpose described on the marketplace. We compared the number of chronic condition self-management apps identified with "TOP" versus "keyword" search strategies.

When possible, apps were downloaded and assessed using PEMAT-A/V, a systematic method to evaluate and compare understandability and actionability of patient education materials. It includes 13 items in five topics (Content, Word Choice & Style, Organization, Layout & Design, and Use of Visual Aids) to evaluate understandability, and 4 items for actionability. Each item was rated with 0 (If Disagree) or 1 (If Agree), while not applicable items were labeled "N/A."

Results

Figure 1 shows the screening process for selecting the apps analyzed in this review.

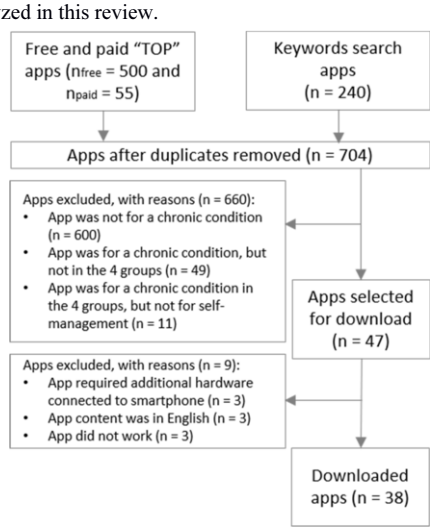


Figure 1 - Apps screening process

Of the 704 apps included after duplicate removal, 167 apps (23.72 %) were not in French (Table 1). From the 541 apps in French, 104 (14.77 %) targeted a chronic condition (such as migraine, back pain, speaking and hearing disorders, mental disorders, etc.). The "Other" category (17.47 %) comprised apps for finding the geographic location of pharmacies or defibrillators, for calling emergency services, for clinical analysis laboratories or hospitals, etc.

Table 1 - Apps categories and examples of subcategories

Categories	Examples	Freq	%
Chronic conditions	Cardiovascular diseases; Respiratory diseases; Diabetes and weight disorders; Chronic pain	104	14.77
Other	Health insurance, hospitals, laboratories; Games	123	17.47
Students and professionals	Interactive learning; Diagnostic, treatment or dosage aid	120	17.05
Information for the general public	Anatomy, biology, medical terms and tests; Diabetes and nutrition; Cancer	90	12.78
Reproductive health	Period tracking; Pregnancy, baby growth, delivery	48	6.82
Pranks and fake tests	Glucose, blood pressure, cholesterol, HIV fake tests	25	3.55
Consultations	Find professionals and book appointments; Home and online consultations	20	2.84
Nutrition	Food and weight diary	7	0.99
Other language	English; Spanish; Portuguese	167	23.72

In the "Chronic Conditions" category, 47 apps targeted self-management of conditions in the 4 groups with highest mortality risk. Of these, 22 were found only in the keyword search, 18 only in the "TOP" list, and 7 in both searches. Therefore, the keyword search resulted in finding 29 mHealth self-management apps out of 240 (12.1 %).

We were able to download and analyze 38 of these apps and understandability and actionability scores were computed (Table 2). Twenty-six apps (68.4%) presented understandability levels below 50% and 25 (65.8%) apps had null actionability scores.

Table 2 – Number of apps by PEMAT scores quartiles (range 0-100%)

	0-25%	*26-50%*	*51-75%*	*76-100%*
Understandability	*8*	*18*	*8*	*4*
Actionability	*25*	*5*	*4*	*4*

Conclusions

We conducted a descriptive analysis of mHealth apps available in France on Google Play store. We found a limited number of apps supporting self-management for people with chronic conditions in a considerably diverse market. Using keywords related to these conditions did not necessarily lead to finding more relevant self-management mHealth apps.

Apps presented low scores of understandability and actionability, which means users may not understand the information and may not be supported to take action. This is likely to limit the potential of these apps to help people with chronic illness to make positive changes on how they manage their condition in their daily lives.

It is important to mention that among the list of apps described here, there were several apps of questionable utility posing considerable risks of misinterpretation and dangerous health-related decisions, such as pranks and fake tests, especially for blood glucose and blood pressure measurement.

Potential users of mHealth apps, including both patients and healthcare providers, might require support to search and select relevant apps for themselves or their patients' needs. Research on mHealth apps selection and use patterns is required to inform the efficient adoption of such technology.

Acknowledgments

Authors declare no conflict of interest. This work has been funded by the European Commission (MSCA-IF n°706028) and the Chaire PreventHorizon.

References

[1] P. Krebsand D.T. Duncan, Health app use among US mobile phone owners: A national survey, *JMIR Mhealth Uhealth* 3 (2015).

[2] P. Klasnja and W. Pratt, Healthcare in the pocket: Mapping the space of mobile-phone health interventions. *J Biomed Inform* 45 (2012), 184-198.

[3] S. Hamine, E. Gerth-Guyette, D. Faulx, B.B. Green, A.S. Ginsburg, Impact of mHealth chronic disease management on treatment adherence and patient outcomes: A systematic review. *J. Med. Internet Res.* 17 (2015), e52.

[4] World Health Organization, *Preventing chronic diseases: A vital investment*, World Health Organization, 2005.

[5] World Health Organization, *Non-communicable diseases*, (2018). http://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases (accessed June 11, 2018).

[6] G.L. Kreps, The relevance of health literacy to mHealth, *Stud Health Technol Inform* 240 (2017), 347–355.

[7] S.J. Shoemaker, M.S. Wolf, and C. Brach. Development of the Patient Education Materials Assessment Tool (PEMAT): A new measure of understandability and actionability for print and audiovisual patient information. *Patient Educ Couns* 96(3) (2014), 395-403.

[8] A. Huguet, S. Rao, P.J. McGrath, L. Wozney, M. Wheaton, J. Conrod, and S. Rozario. A sSystematic review of cognitive behavioral therapy and behavioral activation apps for depression. PloS One 11(5) (2016).

Address for correspondence

Luiza Prado, luiza.siqueira-do-prado@univ-lyon1.fr.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190740

InfoSAGE: Usage Pattern of a Family-Centric Care Coordination Online Platform

Yuri Quintana[a,b], Juan Henao[a], Eli Kaldany[a], Max Gorenbeg[a], Yi-Pei Chen[a], May ADRA[a,b], Lewis Lipsitz[c], Charles Safran[a,b]

[a] Division of Clinical Informatics, Beth Israel Deaconess Medical Center, Boston, MA, USA
[b] Harvard Medical School, Boston, MA, USA
[c] Institute for Aging Research, Hebrew SeniorLife, Boston, MA

Abstract

Globally, there is an expanding elderly population, and families are finding it increasingly challenging to coordinate care for their older family members. This paper reports on the usage patterns of InfoSAGE, an online private social network that has tools for communication and care coordination for elders and their families. This descriptive analysis describes the types of family networks using the platform and types of functionality most used by elders and their family members.

Keywords:
frail elders, care coordination, smartphone, social networks

Introduction

The population of people over the age of 60 are the fastest growing demographic group worldwide, and are projected to triple from 901 million (12.2%) in 2015 to 3.2 billion in 2100 (28.5%) [1]. By 2050, it is expected that those aged 75 years or older will account for 11.0% of the population, up from 5.4% in 2010. This older population is projected to strain the healthcare system, with an average growth in health expenditure of 5.8% per year, and an estimated 20% of the total economy by 2025 [2]. Furthermore, when compared to previous generations at similar ages, baby-boomers experience higher rates of chronic disease). Families face increasing amounts of information and communication needs between family members involved in the care of frail elders. Those that cannot afford costly long-term care solutions will necessarily turn to family caregivers to fill the gap, with informal caregiving valued at $450 billion in 2009 alone [3]. Informal caregiving is associated with higher levels of stress, poorer quality of life, increased strain on family relationships, and substantial time commitments [4]. Often, informal caregiving falls disproportionately on a single individual, who may act as the coordinator of care in a family structure. Maintaining an accurate medication is a significant problem for elders and studies have shown that 30% - 80% of patients have a discrepancy between the medicines ordered in the hospital and those they were taking at home [5]. Caregivers and elders are increasing using the Internet for health care information and care coordination tools [6]. The use of communication applications on the Internet, for the care coordination of elders, has the potential to distribute tasks and alleviate stress among family members. Unfortunately, little has been reported on the use of the Internet or smartphones for those of 75 years and older for care coordination.

Methods

We developed InfoSAGE to understand how online and mobile technologies could families frail elders. The InfoSAGE website (https://www.infosagehealth.org) is a free-to-use Internet platform that is built around an elder that we call the Keystone. This Keystone is connected to care participants, such as family members, care providers, and neighbors and friends. When users sign up for InfoSAGE, they can choose to be an elder or a caregiver. Each family network has a minimum, a family dyad: an elder over 75 years of age (the Keystone) and an identified study partner ('Proxy' or 'Caregiver'), a family member who provides support or care. The platform allows the elder to control who is in their network. The system has a search tool that can be used to search curated resources for elder care; a message board for elder and family communcation; a shared calendar and task list; and a medication manager. Medications can be access-restricted to Keystones and proxies. InfoSAGE allows an elder or proxy to control and restrict access to certain features by type of user. Keystones and proxies can adjust the tiers at any given time. We partnered with local senior organizations, including two large retirement and continuing care communities in metro Boston. These organizations both helped to recruit potential participants and provided early feedback and helped shape the overall project design. We advertised the presence of the site through newsletters and town hall meetings. A member of the study team met with prospective users and families to help with enrollment and the online signing up process. Details of the system features have been described in detail through other works [7].

Results

From January 1, 2014, to April 6, 2018, there were 285 registered users and 18,499 page views. One-hundred-and-sixty-two users provided their year of birth, and 113 did not. Based on those who did report their year of birth, the average age of the Keystone users was 75.5 years, and the average age of the caregiver users was 56.6 years (Keystone median age 82, IQR: 23.5; caregiver median age 56.5, IQR: 11.0). There were 26 dyads (elder and caregiver pairs) that signed up for the detailed survey study. Of these users, 54% self-reported as Caucasian, 4% as African American, and the rest were unknown. Of those who did describe their relationship with a Keystone, 47% were daughters, 25% were sons, 9% were spouses, and 19% were other, including formal caregivers. We looked at what devices were used to log onto the website and, for all users, desktop/laptop devices accounted for 87.6%, while mobile/tablet accounted for 12.4%. Users in the

longitudinal survey were asked 6 months after sign-up "Do you feel comfortable using this system?" and they responded Strongly Agree=6%, Agree=19%, Neutral=56%, Disagree=19%, and Strongly Disagree 0 %. In analyzing our care networks, we observed that family members were distributed across 122 communities, so providing more geo-specific resources could be useful additions. Figure 1 shows a family network.

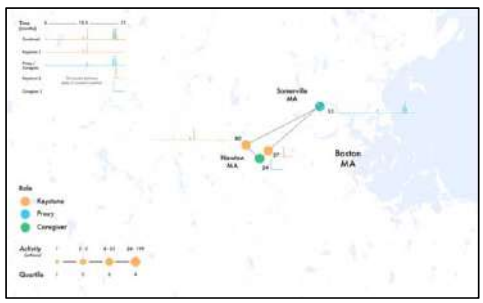


Figure 1 - Example Family Network

Figure 2 compares Keystone usage versus non-Keystone usage. The majority of InfoSAGE networks exist as dyad pairs, which may have resulted from the study design, although 24 family networks consist of three or more users. Network sizes range from one, in networks with a single keystone, to seven users, with several Keystones. The largest networks on InfoSAGE are multi-keystone clusters, encompassing multiple family networks joined by single or multiple users common to each individual network. These extended networks comprise 5% of all networks.

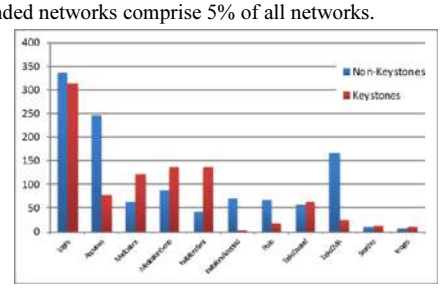


Figure 2 - Keystone usage versus non-Keystone usage

Roles within each network are managed by that network's Keystone or proxy. Of the 285 InfoSAGE users, 157 are keystones, 115 are proxies, 19 are caregivers, and 5 are participants, with 24 users having more than one role due to membership in more than one network. One example of a multirole user is in the case of married elders who are often keystones in their own network and proxy to their spouse.

Acute medical events often require enormous coordination of care, involving the distribution of information, at a sensitive and emotionally distressing time. Networks on InfoSAGE that have had sudden events occur to a family's keystone have seen a surge in task assignment and management, and communication associated therein. One such network had 18 tasks created and assigned over a span of two days, compared to two tasks in the preceding five months. During the following 30-day period, the creation and assignment of tasks returned to 'baseline' levels. This suggests that the ability to organize a list of assignable actions in one central, online source within a family network, that may be geographically dispersed, can bring a measure of order to a chaotic and overwhelming time.

The InfoSAGE platform is more useful if the user (elder or family) is already registered and familiar with the system before needing it, meaning that InfoSAGE, or other technologies like it, designed to support elders in their homes, have many functions and potential uses, such as medication lists, calendars, to-do lists, microblogs, personal stories, etc., that may play a role at different points in care. From a family's perspective, these tools are most useful during the transitions of care. For example, the system may be more valuable during a visit to the emergency room, or a discharge from hospital to home. However, the family needs to be familiar with and using the tools before these transitions occur in order to make information available at the time they are needed.

Reduction in isolation requires greater family support and communication. One barrier to the adoption of InfoSAGE in this context is the family's perception, structure, and support for using the tools. The perception of 'no-need', alternate forms of contact, or if there are no family caregivers, are reasons for non-use. One possible gap here is the difficulty of incorporating formal caregivers (home health aids, visiting nurses, social workers) and informal caregivers within one network. There needs to be more financial models for compensating formal care givers involved in remote care of older adults.

Conclusions

Our study shows that it is feasible to establish an online platform for elders over the age of 75 and their families and caregivers for information exchange and care coordination. Future research will explore barriers of adoption and sustainability in more depth. The barriers may be related to an added time burden for physicians, the lack of an informal or family caregivers, difficulty in accessibility to technology or Internet.

References

[1] United Nations, Department of Economic and Social Affairs, Population Division (2015). World Population Prospects: The 2015 Revision, Volume I: Comprehensive Tables (ST/ESA/SER.A/379)

[2] Keehan SP et al. (2016). National health expenditure projections, 2015–25: Economy, prices, and aging expected to shape spending and enrollment. Health Affairs: 2016 Aug 1;35(8):1522-31.

[3] Sloane PD et al. (2014). What will long-term care be like in 2040? North Carolina Medical Journal 75(5): 326-330.

[4] Ekwall A et al. (2004). Dimensions of informal care and quality of life among elderly family caregivers. Scandinavian Journal of Caring Sciences 18(3): 239-248.

[5] World Health Organization. (2014). High 5s Project. Interim Report. 2014. Available at URL: http://www.who.int/patientsafety/implementation/solutions/high5s/High5_InterimReport.pdf

[6] Greenwood S. (2016) Social Media Update 2016. Washington, DC: Pew Internet & American Life Project. Available at URL: http://www.pewinternet.org/2016/11/11/social-media-update-2016/

[7] Quintana Y, Crotty B, Fahy D, Lipsitz L, Davis RB, Safran C. Information sharing across generations and environments (InfoSAGE): study design and methodology protocol. BMC Med Inform Decis Mak. 2018 Nov 20;18(1):105. doi: 10.1186/s12911-018-0697-4.

Address for correspondence

Yuri Quintana, Email: yquintan@bidmc.harvard.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190741

Safe Surgery: Application for Logistic Support for Safe Surgery

Marizete Campioni[a], Grace T M Dal Sasso[b]

[a] *Bridge Laboratory, Federal University of Santa Catarina, Florianópolis,, Santa Catarina, Brazil*
[b] *Health Sciences Center, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil*

Abstract

We developed and evaluated the usability of Check-up Surgery, a surgery application from the mApp® Platform. The Usability Engineering Cycle was used as method of development. The results showed that the Check-up Surgery application had excellent usability criteria with an average of 4.63 to 4.70 by 81% of the evaluators. For all the interviewees there was statistical significance in relation to the usability criteria of the application, with pValue ≤ .05.

Keywords:

mHealth, Medical Informatics, User-Computer Interface

Introduction

Every two years, the Global Patient Safety Alliance organizes a Global Challenge to motivate practitioners' perceptions of health safety and to establish public policies for good health practices [1]. From 2007 to 2008, the Global Challenge of Surgical Safety Practices centered on the implementation from the motto: *Safe surgeries save lives.* Therefore, whatever the reported problem, it must be corrected before the surgery begins [2]. Throughout the world, increased safety in surgical procedures provides dialogue oportunites to deepen the knowledge regarding the surgical act, both for the patient and for the healthcare team [3]. Safe surgeries require communication, teamwork and recognition of the importance of patient safety by the surgical team [4].

Studies also point out that there are still difficulties in fulfilling the WHO objectives checklist and the aspects for this are still unclear. [5].

Communication and information technologies (ICT), in addition to facilitating the dissemination of knowledge in the health area, can also support the critical decision-making of professionals, contributing to the achievement of reliable diagnoses and therapeutic guidelines as well as behaviors qualified intended for patients and users of the service. [6,7]. It is also emphasized that accessing information in real time and/or remotely contributes to resolving health problems/needs in different geographical regions, promoting broad coverage in terms of specialized healthcare, undertaken in the major urban centers. In this scenario, emphasis is placed on the phenomenon of the mobile technologies in particular, the use of applications. Apps are conceptualized as a set of tools designed to undertake specific tasks and jobs [6]. As a result, this study aimed: To develop and analyze the usability of an application for Surgical Safety, from the LAPETEC/GIATE App® Platform.

Methods

The Usability Engineering Cycle was used as method of development of the application. Three people participated in the development: a content designer, an evaluator and a systems programmer. After the application development, the usability assessment was performed by the Usability Instrument based on NBR ISO 9241-11, from June to July 2018. The sample was intentional non-probabilistic and consisted of 32 health professionals from a General Hospital of Florianópolis, Santa Catarina, composed of seven nurses, four anesthesiologists, eight physicians, two administrators, one pharmacist, six information technology professionals and four administrative technicians. For data analysis, statistical and descriptive and inferential statistics were used, with the ANOVA test presented for a 95% confidence interval and a pValue≤.005 significance index. The usability tool is organized into 19 items that correspond to the usability criteria of the application on a likert scale from 1 to 5. Being 5 = Very Good (MB), 3 = Good (B), 2 = Regular (REG) and 1 = Bad (R). It was considered in the evaluation that the values of the average between 1 to 1,5 received the classification (Bad); from 1.51 to 2.5 (Regular); from 2.51 to 3.5 (Good); from 3.51 to 4.5 (Very Good) and from 4.51 to 5 (Excellent). It was also defined as the minimum target average for the application to have usability criteria ≥ 3.

Results

The developed system can be accessed at the following link: http://site.erue.giate.ufsc.br:8080/marizete/.
An example of a developed screen is shown in figure 1.

The application was built to be instructive and allow user autonomy. In the next figure 2, the interrelationship of the components is presented in a graph of dependencies between the application flows. Fuzzy logic was applied. That is, according to the characteristics of the procedure and the patient, the application indicates the necessary materials and advises all the surgical team of the necessary logistics.

It was organized so that the screens are sequential in stages, to signal the information. If you need to replace some information that has already been flagged, you can return to that screen and make the desired change. Once the corrections/changes to the screen have been completed, simply go to the next screen and the flow continues in the direction to complete the planning.

Figure 1 *– Laryngeal procedure screen*

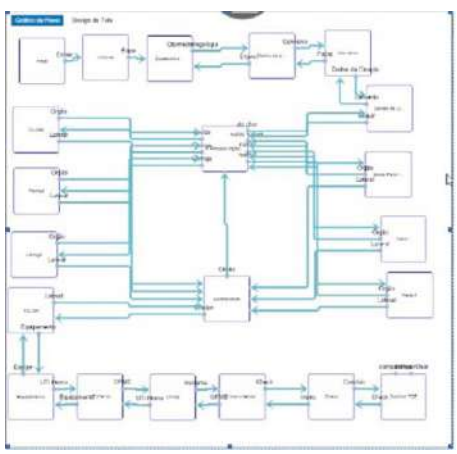


Figure 2: *Dependency Graph among Application Flows*

After application development was complete, the usability instrument was applied, yeilding the following results. The 'excellent' satisfaction rate is 69%, and the 'very good' rating was 21%. Followed by 7% who selected 'good', 3% as 'regular', and 1% 'bad'. The general average among the Administrators was 4.7 (Excellent) and pValue <.0007. The overall average among physicians was 4.75 (Excellent) and pValue<.0000. Among nurses and pharmacist, the overall mean was 4.65 (Excellent) and pValue <.00001. The overall average among IT professionals was 4.01 (very good) with pValue <.001122.

Discussion

The Usability Engineering Cycle, a method used in the development of the Check-up Surgery application, proved efficient and contributed to user-centered construction and followed the steps of analysis, design and testing. The application has usability criteria among the professional categories studied: doctors, nurses, pharmacists, anesthetists and IT professionals with pValue ≤.005. The interactive design based on fuzzy logic allows support for health professionals' decision making for the development of surgery.

Higher means of evaluation were observed by health care professionals directly involved with the patient and were not observed in the institution's managers. The content of the application was developed after extensive observations of the work performed in the institution and after the usability evaluation by the participants, these differences between averages presented by the institution's team can be related to the fact that the main usability problems are the distribution of information in multiple views, identification problems of the selected patient and the confidence of the basic daily tasks in the memory of the users. Thus, the results show that the relationship between usability, errors of use and patient safety should be understood and considered in the design of the health information system [8].

Conclusions

The use of computerized information systems cannot be considered as the solution to all problems of professional practice with regard to safety. This is due in part to concrete cultural changes in the behavior and the work process of the professionals involved, and a action in all decision-making instances. Despite the small sample that represents it, it is understood that this application Check-up Surgery, can contribute to the safe surgery especially for integrating the surgical team in the same information related to the equipment needed for the surgery. This is expected to reduce delays or lack of equipment during surgery and avoid exposing the patient to unnecessary risks.

References

[1] World Health Organization & WHO Patient Safe- ty. The second global patient safety challenge: safe surgery saves lives, 2008 Geneva: World Health Organiza- tion.
[2] World Health Organization. (WHO), Guidelines for Safe Surgery 2009. Safe Surgery Saves Lives, Geneva, Switzerland.
[3] A. Pancieri, et al. Checklist de cirurgia segura: análise da segurança e comunicação das equipes de um hospital escola, Rev. Gaúcha Enf. Porto Alegre, Rio Grande do Sul, (2013), 71-78.
[4] K.K Park, J.K. KIM, M.K. Kin, Operating room nurses' experiences of securing for patient safety. Journal of Korean Academy of Nursing, (2015) 761-772.
[5] D.J. Leaper, et al, Surgical site infection: poor compliance with guidelines and care bundles, International Wound Journal, (2015),357-362.
[6] A.A. Filipova, Electronic health records use and barriers and benefits to use in skilled nursing facilities. Computer, Informatics, Nursing (2013), 305-18.
[7] S. Roberts, W.Chaboyer, R.Gonzalez, A.Marshall, Using technology to engage hospitalised patients in their care: a realist review, BMC Health Serv Res (2017),388.
[8] E. Hautamäki, U-M Kinnunen, S. Palojoki, Health information systems´ usability-related use errors in patient safety incidents, FinJeHeW, (2017), 1-12.

Address for correspondence

Grace T M Dal Sasso.

Campus Universitário s/n - 88040-900 Fpolis – SC. Brazil. Caixa Postal: 476

email:grace.sasso@ufsc.br or gracetmds@gmail.com
Phones:+55(48)999891409 or +55(48)37216495

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190742

Use of Eye-Tracking in Studies of EHR Usability – The Current State: A Scoping Review

Yalini Senathirajah[a], Elizabeth M. Borycki[b], Andre Kushniruk[b], Kenrick Cato[c], Jinglu Wang[d]

[a] *Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA*
[b] *University of Victoria, Health Information Science, Victoria, British Columbia, Canada*
[c] *Nursing Scholarship and Research, Columbia University, New York, New York, USA*
[d] *Cornell University, New York, New York, USA*

Abstract

Eye-tracking has long been used to assess usability for the public web. Recently, it is used to assess user behavior with electronic health records (EHRs). We conducted a scoping review of studies involving eye-tracking for usability of EHRs to determine the current state. Three main themes emerged: studies of usual use of systems, development of new methods, and studies of new features. Detailed user behaviors revealed by eye-tracking can contribute valuable information to redesign efforts.

Keywords:

Eye Movements, Electronic Health Records, User-Computer Interface

Introduction

In recent years, the use of eye-tracking has spread from mainstream psychology and usability communities to its use by informaticians to study the use of health information technologies (HIT), particularly, electronic health records (EHRs). As usability of EHRs is a critical problem, we undertook a literature review to determine the current state and what kinds of studies involving eye-tracking are being done in the area of healthcare.

The tool permits detection of where on the screen a user is looking during task performance, giving us a much more precise measure than just clicks, screen transitions, think-aloud, and similar tools. It is routinely used in commercial usability studies and commercial design. The recent development of lower cost eye trackers and built-in eye trackers for certain gaming laptops adds to their availability and usefulness. Here, we review the current EHR usability studies and discuss future possible studies useful for the improvement of EHR design.

Methods

Searches were done in the literature with Pubmed, Google Scholar, CINAHL, web of science, and IEEE. Figure 1 shows the number of articles retrieved, inclusion/exclusion criteria, and final set of 15 articles reviewed here. Search terms were: Eye track* (or eye-track* or eye track*) AND "ElectRONIC MEDICAL Record (or EMR), Eye track* (or eye-track* or eye track*) AND electronic health record (or EHR), Eye track* (or eye-track* or eye track*) AND electronic patient record. English and French articles, qualitative, quantitative, and mixed methods studies including feasibility studies were included. Only completed studies were included.

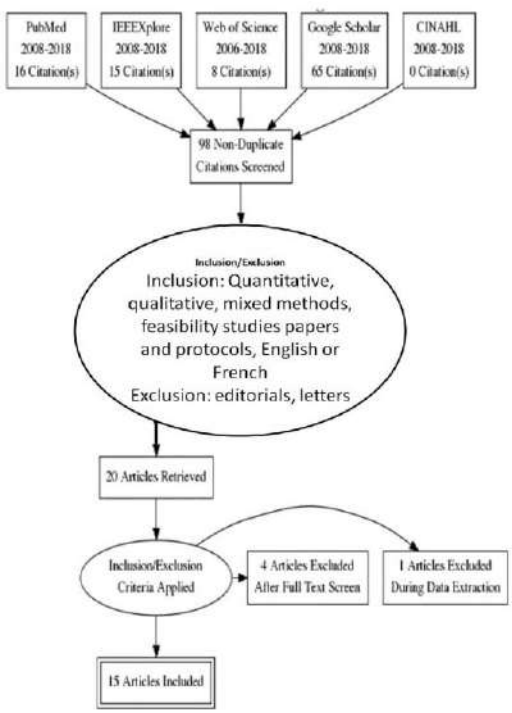


Figure 1 - Literature search and exclusion/inclusion criteria and results

The initial searches retrieved 98 non-duplicate papers after 2005; three researchers (EMB, YS, and JW) excluded 78 articles based on titles and abstracts, leaving 20; five were then excluded after full reading, for a total of 15 covered in this review. These were read and summarized by all the authors, using a summary table including bibliographic information, main research question, study design, sample, setting, outcome measures, results, discussion, limitations, and relevance to our research question. As the focus is how eye-tracking is being used to evaluate EHR usability we did not assess study quality.

Results

Overview

Within the selected studies there were three main areas of focus: use of eye-tracking to understand physician use of their usual

EHRs (five papers), development of new methods (four papers), and use of eye-tracking to evaluate new innovations in EHR design (six papers). Fourteen studies used simulations in academic health center settings. One tested only pharmacists while all others had physician subjects; one also included graduate students. All studies had generally low numbers of subjects (four to 34). Most were conducted in laboratory settings; there were also studies in an empty ICU bay, simulation centre, emergency department, and pharmacy. Both in situ studies had practical difficulties and thus incomplete information.

Findings

New methodological innovations.

King et al. compared performance of high- and low-cost eye trackers and tested the feasibility of automated detection of eye gaze [1] and found it both feasible and 88% accurate, which was adequate for most usability studies. This might become a method for mass collection of user behavior data, paving the way for predictive modelling and delivery suitable to specialty and task.

Studies of routine use of usual systems.

Gold et al. used a simulated rounding task with patient safety issues, finding that performance correlated (p=.0004) with patterns of rapid data scanning, increased numbers of screens viewed, mouse clicks, and saccades [2]. Mosaly et al. compared subjective and objective measures: eyetracking and performance measures, real and predicted [3]. They found heuristic measures are highly accurate in identifying usability issues. Additional measures were useful to identify differences in task complexity, and the unsurprising finding that note structure affects information reviewed. Amster BD et al. found in a simulated handoff task that more than 60% of the time was spent on impression and plan, with a pattern of skimming other sections. This pattern was affected by their level of trust in the author of the note [4]. Calvitti et al. conducted observations in two outpatient clinics to develop a picture of the physician, patient, EHR activity, and workload. They found different EHR functions or interfaces did not result in large differences in use patterns or objective measures [5].

Testing new interaction designs

Belden et al. found collapsible note sections and placement of impression/plan at the top of the note were associated with faster performance on their CET (author-developed Composite Eye-Tracking score) [6]. Moacdieh and Sarter found significant effects of cluttered displays in information extraction time, interaction with stress, noticing tasks, and screen time, but clutter was not significant in search accuracy or duration of specific tasks [7]. Wright et al. tested existing use with in-depth interviews. They discovered needs for dynamic display of new information, configurable displays which aid filtering and trending of the most recent data, and fast proven pathways to reduce navigation complexity [8]. Yoon et al. studied nurses' interpretation of a 3D infectious disease transmission visualization, finding a high degree of engagement and attention, and specific user preferences among the visualizations [9].

Conclusions

Eye-tracking is a useful technique for studying EHR use in three main areas to reveal new details about user cognitive behavior. Further evaluation and validation studies are needed to fully elucidate where it may be used productively in usability assessments and system redesign.

Acknowledgements

This work is funded by the U.S. Agency for Health Research and Quality, AHRQ R01HS023708.

References

[1] A.J. King, H. Hochheiser, S. Visweswaran, et al., Eye-tracking for clinical decision support: A method to capture automatically what physicians are viewing in the EMR, In *AMIA 2017 Joint Summits on Translational Science Proceedings*, San Francisco, CA, USA, 2017, 512-521.

[2] J.A. Gold, L.E. Stephenson, A. Gorsuch, et al., Feasibility of utilizing a commercial eye tracker to assess electronic health record use during patient simulation, *Health Inform J* **22** (2016), 744-757.

[3] S. Rick, A. Calvitti, Z. Agha, and N. Weibel, Eyes on the clinic: Accelerating meaningful interface analysis through unobtrusive eye-tracking, In P*roceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare* (*PervasiveHealth*), Istanbul, Turkey*,* 2015, 213-216.

[4] P.R. Mosaly, L.M. Mazur, and B. Lawrence. Usability evaluation of electronic health record system (EHRs) using subjective and objective measures, In *Proceedings of the ACM Conference on Human Information Interaction and Retrieval (CHIIR*), Chapel Hill, NC, USA, 2016, 313-316.

[5] B.D. Amster. *The process by which physicians extract information from electronic progress notes during handoffs*, Masters Thesis, University of Massachusetts Amherst, 2012.

[6] A. Calvitti, H. Hochheiser, S. Ashfaq, et al., Physician activity during outpatient visits and subjective workload, *J Biomed Inform* **69** (2017), 135-149.

[7] J.L. Belden, R.J. Coopman, S.J. Patil, et al., Dynamic electronic health record note prototype: Seeing more by showing less, *J Am Board Fam Med* **30** (2017), 691-700.

[8] B.J. Arthurs, V. Mohan, K. McGrath, et al.. Impact of passive laboratory alerts on navigating electronic health records in intensive care simulations. *Sage Open* **8** (2018), 1-10.

[9] N. Moacdieh and N. Sarter, Clutter in electronic medical records: Examining its performance and attentional costs using eye-tracking, *Human Factors* **57** (2015), 591-606.

[10] M.C. Wright, S. Dunbar, B.C., Macpherson, et al., Toward designing information display to support critical care. A qualitative contextual evaluation and visioning effort, *Appl Clin Inform* **5** (2016), 912-929.

[11] S. Yoon, B. Cohne, K.D. Cato, et al., Visualization of data regarding infections using eye-tracking techniques, *J Nurs Scholarsh* **48** (2016), 244-253.

[12] L. Stephenson, J.A. Gold, V. Mohan, and A. Gorsuch. Use of eye-tracking software to assess physician use of an electronic health record, *Am J Respir Crit Care Med* **189** (2014), A5360.

[13] T. Kocejko, K. Goforth, K. Moidu, and J. Wtorek, Visual attention distribution based assessment of user's skill in electronic medical record navigation*, J Med Imaging Health Inform* **5** (2015), 951-958.

Address for Correspondence

Yalini Senathirajah, PhD, Department of Biomedical Informatics, University of Pittsburgh, U.S.A., yalini@pitt.edu, 412-648-9214.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190743

Study of the Usability of an Automated Coding Software for Causes of Death in an African Context

Seydou Golo Barro[a,c], Grégoire Rey[b], Pascal Staccin[a]

[a] Université Nazi Boni, Bobo-Dioulasso,Burkina Faso, [b] CepiDC-Paris,French,
[c] LabSTIC-Santé, Côte d'Azur Université, Nice, French

Abstract

IRIS is an automated coding software for the causes of death. It is used in many European countries for the production of death statistics. The purpose of our work was to study the usability of this software in Africa where the quality of statistics is insufficient. For this, we have developed a device consisting of two software: "collector" and "encoder" cooperating via the same database.

Keywords:

Cause of Death, Automated Data Processing, User-Computer Interface

Introduction

WHO considers that mortality statistics in OECD countries are sufficiently reliable to plan actions to improve health and prevent disease [1]. The reliability of these statistics depends on the quality of the death certificate established by the physician and the quality of the coding of the causes of death [2]. To ensure the quality of mortality data, WHO has developed standard procedures [3]. Despite the existence of these rules, manual coding contains errors and divergent interpretations [3]. That is why developed countries have adopted an automated coding system [4], named "IRIS" [5], consistent with the International Classification of Diseases (ICD) and the American MMDS system [5], [6]. In Burkina Faso, death data is unreliable because there is not yet an effective system of coding the causes of death [7] Mortality statistics are obtained by estimation or by population survey. The objective of this work is to study the possibility of using the IRIS software in an African context.

Methods

In countries where coding software is used, vital statistics data are imported from another source (INSEE for France) and matched with medical data for the production of mortality statistics [8].

In order to personalize IRIS and appreciate its use in the context of Burkina Faso, we have developed an application called "Mimir" that works with a MySQL database interconnected with the IRIS database. Mimir integrates three modules: 1) INSD module which provides the codes of localities of Burkina Faso; 2) ID module for each deceased person; 3) the MDD module in which Medical (textual) information and Demographics Data are filled.

The device was tested on the deaths of the intensive care unit of Souro-Sanou University Hospital Bobo-Dioulasso. The intensive care unit was chosen because it is the department that

has the most varied and complex clinical records. 14 voluntary investigators completed the paper version of death certificates in accordance with WHO rules. Two anesthetists in intensive care were responsible for validating the causes of death. An investigator-coder was responsible for coding the causes of death using the IRIS software.

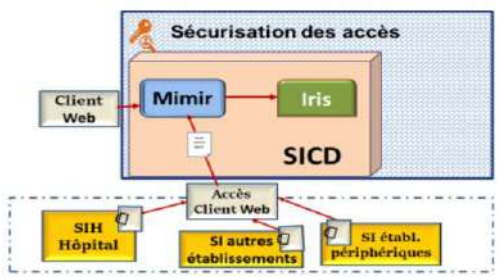


Figure 1: Model of information system on causes of death Usability

An investigator-coder was responsible for proceeding with the coding of causes of death using Mimir linked to IRIS. It was expected that the IRIS encoder could: 1) correctly recover data from Mimir; 2) anonymize death certificates; 3) assign an ICD-10 code to each textual cause of death; 4) automatically select the initial cause of death.

Results

As shown in Figure 2, the IRIS software retrieved all the registered certificates and anonymized them by assigning a unique key. For example, for patient A130027, IRIS automatically assigned a CIM code for each cause of death and selected the underlying cause of death for hematological malignancies (code C9.69) which is the origin of the death. In our study, IRIS was able to automatically encode 90% of the causes of death.

The conclusions of the study favored the deployment of the device at CHUSS. An organizational model has been set up allowing doctors to directly record the death certificate in electronic form or fill out the paper form in case of technical difficulty. All the actors have been trained [9] and the device has been functioning since 2014.

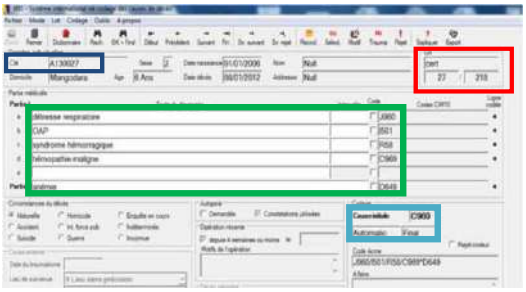


Figure2: Data Recovery and coding of causes of death by IRIS Relevant changes

After twelve months of exploitation, the statistics produced show that diseases of the circulatory system are the most common, followed by infectious diseases and parasites, but attention must also be paid to perinatal conditions.

Table 1: Distribution of causes of death by disease groups

ICD Code	Diseases	Number	%
A00-B99	Certain infectious and parasitic diseases	412	18%
C00-D48	Neoplasms	133	6%
E00-E90	Endocrine, nutritional and metabolic diseases	229	10%
I00-I99	Diseases of the circulatory system	431	19%
J00-J99	Diseases of the respiratory system	67	3%
K00-K93	Diseases of the digestive system	318	14%
L00-L99	Diseases of the skin and subcutaneous tissue	15	1%
N00-N99	Diseases of the genitourinary system	121	5%
O00-O99	Pregnancy, childbirth and the puerperium	96	4%
P00-P96	Certain conditions originating in the perinatal period	293	13%
V01-Y89	External causes of mortality	153	7%
	Total	2268	100%

Conclusions

The results of our study show that IRIS could be used to improve the quality and comparability of mortality statistics in Burkina Faso. Indeed, the coding with Iris can be regarded as reliable and complete because 100% of the certificates are recovered by the software and 90% of the causes are coded automatically [10]. Eventually, the device could be extended to the entire population of the country for the registration of community deaths through a verbal autopsy system [11,12].

Acknowledgements

This work was done as part of the process of modernizing the information system of the Bobo-Dioulasso hospital without special funding. However, the IT infrastructure and stakeholder training were paid from hospital funds.

Conflict of interest: No conflicts of interest were reported

References

[1] É. Jougla, A. Aouba, F. L. Montero, et G. Pavillon, 19 La certification des causes de décès: principe et intérêt épidémiologique, *Tmlds-Main*, janv (2010).

[2] Organisation mondiale de la santé, *Classification statistique internationale des maladies et des problèmes de santé connexes,* Genève: Organisation mondiale de la santé, 2009.

[3] G. Pavillon et F. Laurent, Certification et codification des causes médicales de décès, *Bull Epidemiol Hebd*, **30**(31) (2003), 134–138.

[4] R. A. Israel, Automation of mortality data coding and processing in the United States of America, *World Health Stat. Q. Rapp. Trimest. Stat. Sanit. Mond.*, vol. 43(4) (1990),259-262.

[5] L. A. Johansson et G. Pavillon, IRIS: A language-independent coding system based onthe NCHS system MMDS, présenté à WHO-FIC Network meeting, Tokyo, Japan, 2005.

[6] T.-H. Lu, S.-M. Tsau, et T.-C. Wu, The Automated Classification of Medical Entities (ACME) system objectively assessed the appropriateness of underlying cause-of-death certification and assignment, *J. Clin. Epidemiol.*, 58(12) (2005) , 1277-1281.

[7] L. Boussari et R. Somé, Evalutation du système national d'information sanitaire du Burkina Faso, Ministère de la santé, Burkina Faso, 2008.

[8] E. Jougla, Numéro thématique la moratilité prématurée en France, 2003.

[9] S. G. Barro, J. C. Dufour, M. Somda, S. Malo, et P. Staccini, Cause-of-death certification in Africa: analysis of a model implemented at the teaching hospital of Souro Sanou of Bobo-Dioulasso (Burkina Faso), *Stud. Health Technol. Inform.*, 205 (2014), 1110-1114.

[10] T. H. Lu, Using ACME (Automatic Classification of Medical Entry) software to monitor and improve the quality of cause of death statistics, *J. Epidemiol. Community Health*, 57(6) (2003), 470–471.

[11] C. RAO *et al.*, Evaluating national cause-of-death statistics: principles and application to the case of China, *Health Inf. Syst.*, 83(8) (2005), 618-625.

[12] Application of the verbal autopsy during a clinical trial, *Soc. Sci. Med. 1982*, 31 (1990), 585-91.

Address for correspondence

Dr. Seydou Golo Barro
Center for Training and Research in Medical Technology
School of Computer Science (ESI),
Nazi Boni University (UNB), Bobo-Dioulasso, Burkina Faso
01 BP 1091 Bobo-Dioulasso 01
Email: seydou_golo@yahoo.fr.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190744

Data Quality Governance at the Emergency Department: A Qualitative Study Influenced by Grounded Theory in Nine Swedish Emergency Departments

Niclas Skyttberg[a,b] , Joana Vicente [c], Rong Chen[ad] and Sabine Koch[a]

[a]*Health Informatics Centre, LIME, Karolinska Institutet, Sweden*
[b]*InterSystems, Sweden,* [c]*Linnaeus University, Sweden,* [d]*Cambio Healthcare Systems, Sweden*

Abstract

Vital Sign Data Quality is essential for successful implementation of clinical decision support systems in emergency care. Studies have shown that data quality is inadequate and needs improvement. This study shows that data quality is dependent on both technical and human factors and provides a conceptual model of data quality governance and improvement in the emergency department.

Keywords:

Electronic Health Records, Vital Signs, Quality Improvement

Introduction

Vital signs are central to decision making in the emergency department (ED). Vital signs data is used in calculations of risk and warning scores, where clinical decision support systems have shown promise in improving healthcare quality and productivity [1]. However, in order to reuse vital signs data in emergency care for clinical decision support systems, it needs to be correct, complete and current. Previous studies show that data quality in the ED is inadequate [2]. This study aims to explore how the quality of vital signs data may be governed and improved in EDs.

Methods

Qualitative data collection using semi-structured interviews, observations, and collection of documents in the EDs and field notes were performed. Data collection occurred in nine EDs, which were purposefully selected to represent a variety of size, regional spread and academic level. Data was analysed alongside the data collection using thematic analysis with constant comparison of new incoming data and the data already analysed [3]. Parts of the results presented here have been previously published [2].

Inclusion Criteria

This study included registered nurses and medical doctors that had worked a minimum of five years in the emergency care context. The participants were identified through the quality managers at the hospital level or through the head of staff at the included hospitals. Authors identified and invited a total of 16 participants through their e-mail address at work. All invitees accepted to participate in the study.

Data Collection

A total of nineteen interviews were conducted as evenly as possible between the nine sites. The interviews were about 45 minutes long using a semi-structured interview guide. The guide was developed and adjusted after peer-review and a pilot interview at one of the included sites. These aimed to cover participants' experience of data quality management with focus on completeness, currency and correctness [4]. The interviews were performed, either on-site or by telephone, and transcribed from a voice recording. During the data analysis three main workflows in vital sign documentation emerged and three sites were picked to represent these workflows in the observations. Observations were about 60 minutes per site and data were collected using a semi-structured observational protocol. During the observations, field notes were taken and additional documentation with relevance to vital sign data quality were collected.

Data Analysis

The transcribed recordings were analysed, coded and abstracted into categories and themes. This was done continuously using constant comparison during data collection. Coding and abstraction were performed separately by two of the researchers. However, discussions were held continuously on the emerging results within the research team, leading to an evolving model of data quality governance in EDs. Observational data and collected material from the EDs were included in the analysis. Over time fewer new codes emerged and the major concepts and themes stabilized. Upon researcher consensus that saturation had been reached, three additional interviews were performed with the participants to confirm the findings.

Ethics and Personal Integrity

The study was approved by the Stockholm Ethical Committee (Dnr 2014/1207-31/4). The researchers contacted the managers at the included sites, and they were able to grant permission to contact selected staff through work email. Each participant was given written information in advance and informed consent forms were collected for the study. All participation was voluntary, and participants had to opt-in for the data collection and they were given the right to opt-out at any stage during the analysis. Data were anonymized after collection and all participants were assured that no published information would be traceable.

Results

Three main outputs emerged from the analysis: identification of main workflows of vital sign documentation in EDs, categorization of factors influencing data quality, and creation of a conceptual model as to how these factors affect data quality governance and improvement in the ED.

Vital Sign Documentation Workflow

Only four out of nine hospitals documented vital signs data directly in the electronic health record. Five of the sites still used documentation based on paper-based templates. The main reasons behind the persisting paper-based workflows related to the interoperability and management categories described below.

Factors Influencing Data Quality

Human and technical factors were found to be the main themes affecting data quality at EDs as presented in Table 1 and Figure 1. All sub-categories had an impact on all aspects of data quality as described by Weiskopf and Weng [4], but the contribution to each one of the aspects varied between the sub-categories detailed in Table 1.

Table 1- Themes and categories affecting data quality and the order of impact on data quality aspects

Influencing Factors			Data Quality Aspects ○ Completeness √ Correctness ⏱ Currency		
Technical Factors	Interoperability and Functionality	Interoperability between systems and/or devices	○		√
		Interoperability within a system	○		√
	Standardization	Standardization of the triage process	○	√	
		Failure to comply	○	√	
		Lack of standard	○	√	
		Standardization of documentation	√	○	
		Patient Safety Standards	√	○	
Human Factors	Education and Competence	Clinical Competence	√	○	
		Method competence and knowledge	√	○	
		Sloppiness and carelessness	○	√	
		Control of staff competence	√	○	
		Experience	√	○	
	Management	Quality control	√	○	
		Change management		○	√
	User needs	Workflow support	○	√	
		Documentation support		○	√

The results show that in order to improve data completeness and currency, interoperability and user needs are the most important factors. To improve correctness, standardization, education and competence were revealed to be the main contributing factors.

Data Quality Governance Model

The way main categories and themes influence data quality development and governance are described in a circular model (Figure 1). The governance and improvement cycle was found to start with a step that aimed to standardize the workflow and structure the documentation. In this step participants mentioned that documentation should have been connected to standardized terminology and specific keywords. Using such standards was perceived to link the documentation to the emergency care context. In the next step, results showed that both education and training were necessary to make the new workflow known and implemented. Management factors were perceived as important to ensure there was compliance in the organization to the agreed workflows, as well for follow-up and feedback on outcomes in the EDs. Such feedback was perceived to improve compliance and stimulate quality improvement initiatives. If there was lack of data in the follow-ups, the workflow effectiveness had to be evaluated and here the user needs would need to be balanced with the support provided by the functionality in the information systems. When required functionality would ideally improve and increase. An example mentioned was new functionality for the automatic calculation of national early warning scores. When such functionality was added it opened opportunities for new and more structured workflows to be developed, implemented and evaluated. This governance and improvement cycle were perceived to impact all three data quality categories: correctness, completeness and currency.

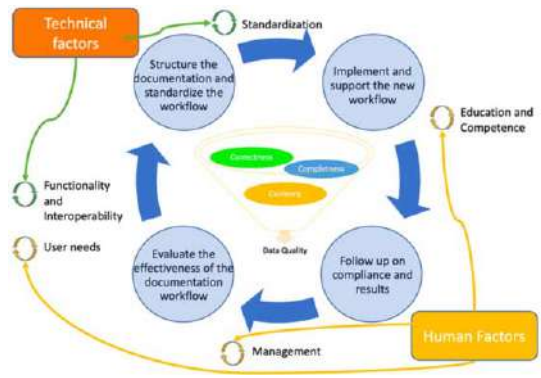


Figure 1 – The impact of human and technical factors on data quality governance in the emergency department

Conclusion

This study shows that the documentation workflow in Swedish EDs is still surprisingly manual and paper-based. Further, results show how technical and human factors impact data quality governance and improvement process in Swedish EDs. The study findings are relevant to anyone who wishes to improve data quality or reuse data from the EDs.

References

[1] T.J. Bright, A. Wong, R. Dhurjati, E. Bristow, L. Bastian, R.R. Coeytaux, G. Samsa, V. Hasselblad, J.W. Williams, M.D. Musty, L. Wing, A.S. Kendrick, G.D. Sanders, and D. Lobach Effect of clinical decision-support systems: a systematic review, *Ann Intern Med* **157** (2012), 29–43, doi:10.7326/0003-4819-157-1-201207030-00450.

[2] N. Skyttberg, J. Vicente, R. Chen, H. Blomqvist, and S. Koch, How to improve vital sign data quality for use in clinical decision support systems? A qualitative study in nine Swedish emergency departments, *BMC Med Inform Decis Mak* **16** (2016), 61,, doi:10.1186/s12911-016-0305-4.

[3] H. Starks and S. Brown Trinidad, Choose your method: a comparison of phenomenology, discourse analysis, and grounded theory, *Qual Health Res* **17** (2007), 1372-1380, doi:10.1177/1049732307307031.

[4] N.G. Weiskopf and C.H. Weng, Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research, J Am Med Inform Assoc **20** (2013), 144-151, *doi:10.1136/amiajnl-2011-000681.*

Address for correspondence

niclas.skyttberg@ki.se

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190745

The Use of Theory in Mobile Health Interventions for Patient Self-Management of Chronic Diseases

Ting Song[a,b], Siyu Qian[a,b], Tingru Cui[a], Ping Yu[a,b]

[a] *Centre for IT Enabled Transformation, School of Computing and Information Technology, Faculty of Engineering and Information Sciences, University of Wollongong, Wollongong, New South Wales, Australia,*
[b] *Illawarra Health and Medical Research Institute, University of Wollongong, Wollongong, New South Wales, Australia*

Abstract

This study aims to investigate the use of theory in mobile health interventions for patient self-management of chronic conditions, with a focus on hypertension. We identified six theories in ten studies that were reviewed. These included Health Belief Model, Social Cognitive Theory, Self-Determination Theory, Theory of Planned Behavior, Transtheoretic Model, and Technology Acceptance Model. The findings are useful for further advancement of mobile health interventions to support chronic disease management.

Keywords:

mHealth; Social Theory; Self-Management

Introduction

Mobile Health (mHealth) interventions have provided patients with a unique opportunity to engage in self-management of hypertensive conditions. For example, patients with hypertension can use a mobile phone-based application to ensure their medication intake on time, to maintain a healthy lifestyle, and to monitor their blood pressure anytime, anywhere. In comparison of non-theory based interventions, theory-based mHealth interventions have been found more effective in changing health behavior [1-3]. Therefore, this study aims to investigate the use of theory in mHealth interventions that support patient self-management of hypertension. It focuses on addressing the following two questions:

- Q1: What are the theories commonly used to guide mHealth interventions that support patient self-management of hypertension? How are these theories used?
- Q2: To what extent are these theories reported?

Methods

A literature review was conducted following the guideline of Preferred Reporting Items for Systematic Reviews and Meta-Analyses [4]. We searched seven electronic databases in June 2018: Scopus, PubMed, PMC, CINAHL Plus with full text, MEDLINE with full text, PsycARTICLE, and PsycINFO. Terms and MeSH headings, such as "mHealth", "self-management", and "hypertension", were used in combination to identify English publications from 2009 to 2018 by the first author. Also, we conducted a manual search to ensure adequate coverage. The first two authors carried out the literature screening separately as per the following criteria, with further judgement by the fourth author if any disagreement occurred. Inclusion criteria were empirical studies that reported (1) mHealth hypertension intervention, (2) at least one theory, and (3) hypertension-related outcomes. Exclusion criteria (1) were studies in which the theory has nothing to do with the intervention, and (2) were reviews, protocols, or guidelines in the topic area. The Mixed Methods Appraisal Tool assessed quality [5].

Data were extracted using Endnote and an Excel spreadsheet, and data were synthesized and analyzed using an inductive method. Theory Coding Scheme developed by Michie and Prestwich was applied to evaluate the extent to which the theories were applied in each study [3]. As reported, use of theory was an inclusion criterion for our selection of studies; all studies selected satisfied the measurement item "theory/model of behavior mentioned", and thus, did not require any further assessment. Another two items, "quality of measures" and "randomization of participants to condition", were seen as irrelevant to the purpose of this review and were thus excluded. Therefore, the extent of theory application to each study was measured in two steps. First, we used the 21-item scheme to evaluate each study. If a study conformed to the statement of a measurement item, "yes" was recorded, and the study was scored 1, otherwise "no" was recorded, and the study was scored 0. Second, these items were further grouped into six categories of the Theory Coding Scheme (see Table 1), the number of items in each of these categories were 2, 7, 2, 2, 9, 2, respectively. In each category, every item was worth one score. Therefore, the total scores for each category were 2, 7, 2, 2, 9, 2, respectively. Afterwards, we calculated the mean score and the proportion of the record "yes" to evaluate the extent to which these theories were reported.

Results

Ten papers were eligible for inclusion into the review (see Figure 1). They all met at least three criteria of the Mixed Methods Appraisal Tool, indicating high methodological quality of these studies.

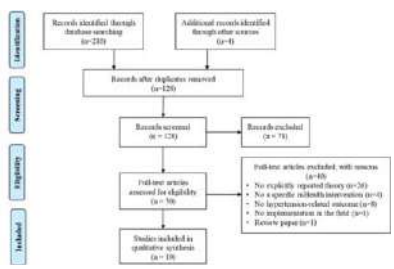


Figure 1– Literature Search and Screening Process

Q1: What are the theories commonly used to guide mHealth interventions that support patient self-management of hypertension? How are these theories used?

Six commonly used theories to guide mHealth interventions for patient self-management of hypertension are: Health Belief Model (HBM), Social Cognitive Theory (SCT), Self-Determination Theory (SDT), Theory of Planned Behavior (TPB), Transtheoretical Model (TM), and Technology Acceptance Model (TAM). Related constructs or processes to which these theories applied are mapping to the design, implementation, and evaluation of the mHealth interventions (see Figure 2).

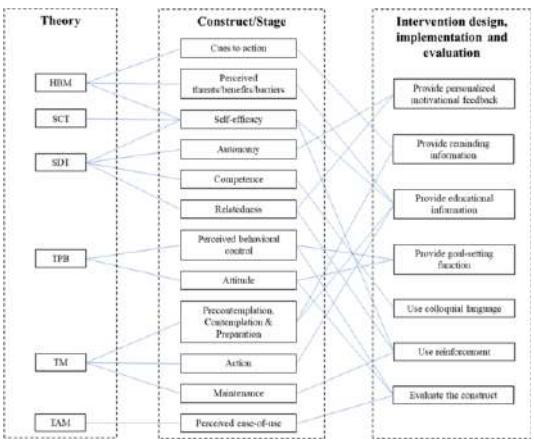


Figure 2– The Relations Among the Six Theories, Their Constructs, and the Corresponding mHealth Interventions

Q2: To what extent are these theories reported?

The mean scores for the six categories were shown in Table 1. Studies reported "Reference to underpinning theory" to a large extent. "Targeting of relevant theoretical constructs" was reported to a medium extent. "Using theory to select recipients or tailor interventions", "Measurement of constructs", and "Testing of mediation effects" were reported to a less extent. None of the included studies reported refinement of a theory.

Table 1– Frequency of Studies that Cover a Particular Category in the Theory Coding Scheme

Theory coding scheme category [3]	Item [3]	Total score	Mean score (%)
Reference to underpinning theory	1, 2	2	1.7 (85)
Targeting of relevant theoretical constructs	1, 4, 6-10	7	4.2 (60)
Using theory to select recipients or tailor interventions	4, 6	2	0.6 (30)
Measurement of constructs	11, 12	2	0.7 (35)
Testing of mediation effects	11-19	9	2.3 (26)
Refining theory	20-21	2	0 (0)

Conclusions

This review identified six commonly used theories to guide mHealth interventions for patient self-management of hypertension: HBM, SCT, TPB, TDM, TTM, and TAM. Most studies only mentioned the use of theory. Several studies described the relevant theoretical constructs. Only a few studies used theory to select recipients, tailor interventions, measure constructs, or test mediation effects. No study reported refinement of the theories. The results are useful for improving the design of future interventions for hypertension management as well as other chronic diseases.

References

[1] J.E. Painter, C.P.C. Borba, M. Hynes, D. Mays, K. Glanz, The use of theory in health behavior research from 2000 to 2005: A systematic review, *Ann Behav Med* **35** (2008), 358-362.

[2] A.J. Rothman, "Is there nothing more practical than a good theory?": Why innovations and advances in health behavior change will arise if interventions are used to test and refine theory, *Int J Behav Nutr Phys Act* **1** (2004), 11.

[3] S. Michie, A. Prestwich, Are interventions theory-based? Development of a theory coding scheme, *Health psychology* **29** (2010), 1.

[4] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement, *Ann Intern Med* **151** (2009), 264-269.

[5] P. Pluye, E. Robert, M. Cargo, G. Bartlett, A. O'cathain, F. Griffiths, F. Boardman, M.-P. Gagnon, M. Rousseau, Proposal: A mixed methods appraisal tool for systematic mixed studies reviews, *Montréal: McGill University* **2** (2011), 1-8.

Address for correspondence
Associate Professor Ping Yu
Centre for IT-enabled Transformation
School of Computing and Information Technology
Faculty of Engineering and Information Sciences
University of Wollongong
Wollongong, New South Wales, 2522, Australia
Phone: +61 2 4221 5412
Email: ping@uow.edu.au

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190746

Three Methods for Engaging Patients and Care Partners in Patient Portal Research

Gillian Strudwick[a], Craig Kuziemsky[b], Richard Booth[c], Sarah Rossetti[d], Anna Chyjek[a], Moshe Sakal[a], Alexandra Harris[e], John Strauss[a]

[a]*Information Management Group, Centre for Addiction and Mental Health, Toronto, Ontario, Canada*
[b]*Telfer School of Management, University of Ottawa, Ottawa, Canada*
[c]*Arthur Labatt Family School of Nursing, Western University, London, Ontario, Canada*
[d]*Department of Biomedical Informatics and School of Nursing, Columbia University, New York, USA*
[e]*Trillium Health Partners, Mississauga, Ontario, Canada*

Abstract

There has been a significant increase in the volume of research conducted on patient portals in recent years. Similarly, there has been a number of benefits described in the literature when patients are engaged in research at all stages. This poster will provide an overview of three ways that patients can be engaged in patient portal research. The methods are based on those used in mental health patient portal research in Canada.

Keywords:

Mental health, Patient engagement, Patient portal

Introduction

Patient portals are increasingly being used in health settings globally as a way of providing patients with easier access to their health information, typically in the form of their health professional documentation [1]. Some of these portals also provide patients with features that allow them to book appointments, message a health professional, and obtain educational resources among other uses [2]. As portals become more commonplace, it is expected that there will be an increasing amount of research into the benefits and drawbacks of the technology, impact of the technology on health service utilization, perceptions of professionals and beyond. Given the literature that describes great benefits when patients are engaged in research, it is important that future patient portal research do so moving forward. Thus, this poster will review methods in which patients and care partners can be engaged in research on this topic.

Methods

The three methods described in this poster were used in research aimed at understanding how best to implement a patient portal in a mental health setting, and how to evaluate it [3,4]. The methods were selected based on those deemed feasible and potentially high impact from the literature on patient and care partner engagement in research [5]. They were not meant to be an exhaustive list. This research is on-going at Canada's largest mental health and addiction teaching hospital. Patients and care partners were engaged in all stages of the research. Benefits and drawbacks of the methods will be discussed in the poster.

Results

The three methods of engagement are:

Patient and Care Partner Representatives as a part of the research team: Patients and care partners are often open to contribute to research by being a member of the investigator team. By engaging these important stakeholders, the most important research questions can be identified and feasible data collection methods that involve patients can be uncovered. Other benefits and considerations will be described in the poster.

Engaging with a Patient and Care Partner Advisory Committee: Once research on the topic has been funded, it may be valuable to establish a committee to ensure the research stays on course, to support the many decisions that the researchers have to make, to support the creation of many patient-facing documents (e.g. recruitment materials), and to hold researchers accountable.

User-centered design and usability testing the patient portal with those who will use it: Before launching a portal for use by patients, it will be important to engage patients in the iterative design and development cycle as well as 'testing' it with patients to identify usability challenges. There are numerous simple, fast and cheap ways of doing this.

Conclusions

From the perspective of the researchers, the methods for engaging patients and their care partners in this research has been effective thus far. Findings have specific relevance to mental health patient portal researchers, and could have relevance to numerous studies conducted in the field of health informatics. Next steps may include the development of a toolkit to support others in effectively engaging patients and care partners.

Acknowledgements

This work was funded by the Canadian Institutes of Health Research, and eCampus Ontario.

References

[1] T. Otte-Trojel, et al, What do we know about developing patient portals? A systematic literature review, *JAMIA*, **23** (2015), 162-168.

[2] A. Griffith, A. Skinner, J. Thornhill, M. Weinberger, Patient Portals, *Applied Clinical Informatics* **7** (2016), 489-501.

[3] G. Strudwick, R. Booth, J. Strauss, Identifying how patient portals may be effectively use among mental health

populations to support digital inclusion: A study protocol, *Stud Health Technol Inform*. 250 (2018), 30.

[4] G. Strudwick, et al., Engaging patients and family members in the evaluation of a mental health patient portal: protocol for a mixed-methods study, *BMJ Open* **8** (2018), e025508.

[5] J. Domecq et al. Patient engagement in research: a systematic review, *BMC Health Services Res*. **14** (2014), 89.

Address for correspondence

Gillian Strudwick RN, PhD,1001 Queen St W, Toronto, Ontario, Canada, M6J 1H4, +1-416 535-8501 ext 39333; gillian.strudwick@camh.ca

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190747

Effectiveness of Cisual Messages to Promote Help-Seeking for Depression

Machi Suka, Takashi Yamauchi, Hiroyuki Yanagisawa

Department of Public Health and Environmental Medicine, the Jikei University School of Medicine, Tokyo, Japan

Abstract

In order to develop more persuasive messaging to promote help-seeking for depression, we created six, differently designed visual messages and tested their effectiveness with possible audience members. Different visual designs brought differences in emotional responses to the message and willingness to use the message. Help-seeking intentions for depression increased significantly after exposure to either the plain text message or the visual messages.

Keywords:
Depression, Emotional Responses, Persuasive

Introduction

Failure and delay in initial treatment contact for mental disorders has been recognized as an important public health problem. Communicating persuasive messages may be useful in solving this problem. However, further evidence is needed to identify strategies for successful, public health messaging with the aim of promoting access to mental health care.

We have undertaken a research project to develop persuasive messages encouraging help-seeking for depression. At the first step, we developed rating scales for measuring the audience's perceptions of effectiveness of health messages [1]. At the second step, we created and pretested depression help-seeking messages, and then found out that the message of 'depression needs treatment' may be most effective in increasing people's help-seeking intentions for depression [2,3]. As the final step, this study attempted to enhance the effectiveness of the message by applying visually appealing design in accordance with the CDC Clear Communication Index (https://www.cdc.gov/ccindex/).

Methods

Development of visual messages

Our depression, help-seeking message consisted of three parts [2,3]. The first part was the main message statement: depression needs treatment - about 80% of untreated patients will not recover. The second part provided information on early signs of depression: depression can be recognized early by mental symptoms such as depressed mood, loss of interest, etc. and physical symptoms such as disturbed sleep, increased fatigue, etc. The last part was the call to action: if you think you might be depressed, speak with your familiar primary care doctor.

Six kinds of visual messages were created by professional designers based on different design concepts (Figure 1). We confirmed that all messages were ideally designed in accordance with the CDC Clear Communication Index.

Assessment of visual messages

Study 1: A web-based survey was conducted in August 2018 among Japanese adults aged 20-34 years (young group) and 60-74 years (old group). Eligible participants (young group n=2,944, old group n=3,237) excluding those who had received treatment for their mental disorders were randomly assigned to one of the six visual messages. After they read the message for at least 15 seconds, they were asked to rate it in terms of persuasiveness, future use, design quality, and emotional response. All items were scored on a 1-to-5 point scale [1-3].

Study 2: A web-based survey was conducted in August 2018 among Japanese adults aged 35-45 years. Those who had not received treatment for their mental disorders and did not report a positive help-seeking intention for depression were selected through a prescreening process. Eligible participants (n=7,017) were randomly divided into four groups: assigned no message, plain text message, and selected visual messages ('Percent' and 'Heart'), respectively. Intension, subjective norm, and attitude toward help-seeking for depression were measured using the vignette methodology after exposure to the message [2,3].

The mean scores were compared using two-way analysis of variance. The prevalence of intention, subjective norm, and attitude toward help-seeking for depression were compared using chi-square test. Significant levels were set at $p<0.05$.

Results

Study 1: The ratings of messages did not differ significantly between the two age groups. Significant differences between the messages were found in the future use, four of the five design quality items, and four of the seven emotional response items. Overall, 'Percent' and 'Heart' were rated relatively high, while 'Greens' and 'Photo' were rated relatively low (Table 1).

Study 2: Compared with the no message group, the other three groups showed significantly greater increases in the prevalence of intention, subjective norm, and attitude toward help-seeking for depression after exposure to the messages. However, no significant differences were observed between the plain text message group and the visual message groups (Table 2).

Conclusions

The application of visually appealing design brought about increasing emotional responses to the messages and willingness to use the messages. Help-seeking intentions for depression increased significantly after exposure to the messages, however the magnitude of this effect was not significanly different between the plain text message and the visual messages.

Simple *Illust* *Percent*

Greens *Heart* *Photo*

Figure 1 Visual messages - 'Depression Needs Treatment'

Table 1 Assessment of the Visual Messages in Study 1

		Simple	Greens	Heart	Illust	Percent	Photo	p		
								Age(A)	Mes(B)	A×B
Persuasiveness		3.04	3.02	3.05	3.02	2.98	2.98	0.582	0.079	0.664
Future use		2.66	2.55	2.67	2.60	2.63	2.57	0.027	0.005	0.092
Design quality	Readability	3.47	3.30	3.38	3.40	3.47	3.25	<0.001	<0.001	0.001
	Organization	3.51	3.40	3.40	3.41	3.42	3.30	<0.001	<0.001	0.268
	Attractiveness	3.09	3.05	3.11	3.06	3.20	3.02	<0.001	<0.001	0.039
	Tone	3.00	2.97	3.94	2.98	3.00	2.90	<0.001	0.048	0.717
	Helpfulness	3.22	3.18	3.21	3.20	3.21	3.17	0.011	0.567	0.531
Emotional response	Surprise	2.63	2.57	2.67	2.62	2.80	2.63	<0.001	<0.001	0.009
	Anger	2.18	2.12	2.26	2.15	2.30	2.22	<0.001	<0.001	0.823
	Fear	2.66	2.55	2.63	2.58	2.72	2.51	<0.001	<0.001	0.692
	Happiness	2.10	2.15	2.23	2.10	2.18	2.11	<0.001	0.008	0.538
	Sadness	2.65	2.59	2.63	2.58	2.69	2.59	0.425	0.063	0.661
	Quilt	2.30	2.25	2.33	2.27	2.32	2.26	0.047	0.416	0.596
	Anxiety	2.68	2.62	2.70	2.64	2.70	2.58	<0.001	0.058	0.224

Table 2 Prevalence of Intention, Subjective Norm, and Attitude Toward Help-Seeking for Depression in Study 2

	None	(n=878)	Plain text	(n=2634)	Percent	(n=1751)	Heart	(n=1754)	p
Intention	27	3.1%	293	11.1%	212	12.1%	189	10.8%	<0.001
Subjective norm	166	18.9%	703	26.7%	468	26.7%	449	25.6%	<0.001
Attitude	399	45.4%	1429	54.3%	955	54.5%	948	54.0%	<0.001

Acknowledgements

This work was supported by the JSPS KAKENHI Grant Number 16K09147 and the Uehara Memorial Foundation Research Grant.

References

[1] Suka M, Yamauchi T, Yanagisawa H. Perceived effectiveness rating scales applied to insomnia help-seeking messages for middle-aged Japanese people: a validity and reliability study. *Environ Health Prev Med* **22(1)** (2017), 69.

[2] Suka M, Yamauchi T, Yanagisawa H. Comparing responses to differently framed and formatted persuasive messages to encourage help-seeking for depression in Japanese adults: a cross-sectional study with 2-month follow-up. *BMJ Open* **8(11)** (2018), e020823.

[3] Suka M, Yamauchi T, Yanagisawa H. Comparing responses to three different messages encouraging professional help-seeking for depression: a cross-sectional web-based survey. *Environ Health Prev Med* 2019 (to be appeared)

Address for correspondence
Machi Suka, Department of Public Health and Environmental Medicine, the Jikei Universitye School of Medicine,
3-25-8 Nishi-Shimbashi Minato-ku, Tokyo 105-8461, Japan
TEL +81-3-3433-1111 FAX +81-3-5472-7526
E-mail suka@jikei.ac.jp.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190748

A Review on Health Information Behavior Study Between 1998 and 2018 in China

Xinglong Tang[a] Shijuan Li[a]

[a] Department of Information Management, Peking University, Beijing, China

Abstract

As a part of our research project (Beijing Social Science Foundation Project, No. 18XCB007), this study aims to provide an overview of the state of art of health information behavior study in China for the past decades. There were 43 studies that met our selection criteria, and they were reviewed regarding to their research objects, methods, and frequent research topics respectively, which provides guidance for future research in this area.

Keywords:
Consumer Health Information; Information Seeking Behavior; Publications

Introduction

Study in health information behavior has been developed rapidly for the past decades, not only in introducing studies abroad to China, but also in bibliometrics analysis [1], model building study [2] and reivews of abroad study such as in elderly health information searching [3], etc.. However, a holistic view of the current state of art of this topic in China is lack.

Methods

Literatures were extracted from three Chinese primary academic resources, including China Academic Journal Network Publishing Database, China Science Periodical Database, China Science and Technology Journal Database. A complex search strategy was utilized, and the search string included "user health informatics", "consumer health informatics", "health information technology", "user health information behavior", "consumer health information behavior" and "health information bahavior". The literature selection process is shown in Figure 1. 309 full text journal papers published between 1998 and 2018 in the Library Information Science (LIS) field were found.

The papers were included if they are: (1) journals papers; (2) written in Chinese; (3) Full-text; (4) in the LIS field. After removing duplication automatically by Endnote and selecting manually, there were 43 papers included finally according to the above selection criteria.

Results

The literatures were reviewed regarding to research objects, research methods and frequently occurred themes.

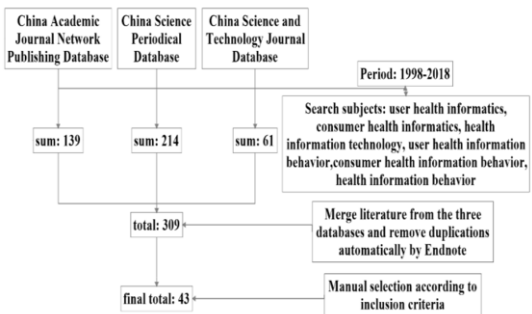


Figure 1– literature search and selection process diagram

Research Objects

The study objects cover both the general consumer [4-9] and specific groups. Especially, the elderly [10-12] and the youths [13; 14] such as undergraduates [15-25] have been paid more attention. Wang's research explored the elderly's health information avoidance behavior, trying to investigate related factors [26]. Other work stated that college students were capable of screening health information in social media [25]. Moreover, specific groups [27], such as public library users [28], sports online community users [29], even Chinese living in the U.S. [30], were paid attention.

Research Methods

Both qualitative and quantitative [31-33] methods were used in the research. Semi-structured interview was utilized to explore the urban and rural young people's health information seeking behavior [14]. Wang etc. used questionnairs to investigate the relationship between two behaviors: browsing sports online community, and purchasing and using smart wrist strap [29]. Moreover, mixed methods were utilized to address the research purpose. Zheng etc. used questionnaires to investigate the participants' related personal characteristics and then interviewed them to have a greater understanding of their attitudes [13].

Frequently Occurred Research Themes

Most frequently occurred research themes in the health information behavior domain fall into three categories: health inforamtion needs, health information seeking and health information adoption.

Health Information Needs

According to Wilson's information behavior model in 1996 [34], health information needs is the precondition of seeking behavior. Cao etc. surveyed public library users' health

information needs with focus on promoting library service [28]. mHealth users' needs were studied, too, by Xu and Zhao. They found mobile health APP users pay more attention to health care, health improvement, medicine and health tracking service information [35]. Limited work has been found that emphasized the information needs of consumers' family members, or carers, or guadians, which is regarded as a research gap found.

Health Information Seeking

This is one of the most prosperous area reviewed. The elderly's health information seeking behavior [10-12], the youths' searching behavior [15; 20; 23; 24] and the public consumers' seeking behavior [7; 9; 19; 20; 36] were all explored. Wu and Yi's experiment for the old adults found out that they have limited capabibliy in searching, and they strongly depend on the quality and credibility of the website [10]. Li etc. compared information seeking behavior between urban and rural youths and led to the concept of "digital gap", however, it's expected to be illustrated in more details [14].

Health Information Adoption

Health information technology acceptance and adoption is another rich publication area in recent years. Li etc. found the factors (credit level of hospital, friends' recommendation, etc.) contribute to the consumers' trust formation, therefore, they have influence on people's acceptance and adoption for health information [18]. As is shown in Hou's work, most users in China are lack of the capability of screening information, and even trust the harmful information [8]. However, the university students were doing better in this [25]. The research objects were restricted to undergraduates, which couldn't lead to a generalization for wider population.

Conclusions

There were 43 studies reviewed, which tried to describe the current research status in China with respect to research objects, research methods and frequently occurred research themes in the field of user health information behavior. Progress has been achieved during the past decades, especially in the Health Information Seeking, and Health Information Adoption area. However, still there are research gaps found, such as the information needs of consumers' family members, or carers, or guadians, etc. Moreover, it has been desired to propose solutions to help disadvantaged groups, such as the elderly, to acquire health information, and suggest proper mechanism especially with the development of innovative information technology.

Acknowledgements

This work is sponsored by the Beijing Social Science Fund (China, project No.18XCB007).

References

[1] Z. Shuai, W. Wentao, L. Jing, and X. Yangqun, Research Review on Online Health Information Behavior of Foreign Users, Library Tribune 38 (2018), 138-147.

[2] W. Wentao, L. Jing, J. Jing, and X. Yangqun, E-health Users' Trust Research Motivation, Path Formation and Conceptual Framework, Library (2018), 85-93.

[3] Y. Jinghan, A Review of Overseas Research on Online Health Information Searching Behavior of Elderly People, Journal of Zhoukou Normal University (2015), 151-154.

[4] D. Shengli, Study of the Factors Influencing User Intention of Access to Health Information Based on Social Q&A, Information Science 34 (2016).

[5] D. Shengli and F. Shaoxiong, Research on the Influence of Individual Differences in Health Information Seeking, Journal of Information Resources Management 6 (2016), 5-11.

[6] D. Shengli, F. Shaoxiong, and L. Jin, Group Differences of Online Health Information Resources Selection Among Young People in Different Task Contexts, Library and Information Service 61 (2017), 98-106.

[7] D. Shengli, F. Shaoxiong, and C. Xiaoyu, The Role of Information Media on User Health Information Seeking —— Based on Health Literacy and Information Retrieval Capability Perspectives, Information Science 35 (2017), 126-132.

[8] H. Xiaorong, F. Yang, and C. Juan, The Research of User Perception and Utility of Health Information Based on Wechat Platform, Journal of Modern Information 36 (2016), 89-93.

[9] Z. Xin and W. Dan, An Analysis of User's Online Health Information Search Task, Information and Documentation Services (2017), 74-83.

[10] W. Dan and L. Yizhe, An Experiment Study on Older Adults' Online Health Information Seeking Behavior, Library and Information Service 58 (2014), 102-108.

[11] W. Dan and L. Yizhe, Older Adults Online Health Information Searching Behavior and Cognition under Different Contexts, Library Tribune 35 (2015), 38-43.

[12] Z. Shubei and D. Xiaozhao, Study on Influence Factors of Older Adults' Online Health Information Seeking, Library and Information Service 59 (2015), 60-67+93.

[13] Z. Ce, K. Jun, and F. Shaoxiong, Research on the Influence of Personality Traits in Health Information Seeking Between Young People from the Perspective of Information Platform, Library and Information Service 61 (2017), 77-85.

[14] L. Ying, Y. Weina, and L. Yuan, Health Information Seeking Behavior of Urban and Rural Youth in Digital Environment, Library and Information Service 60 (2016), 115-123.

[15] Z. Ce, F. Shaoxiong, and K. Jun, Analysis on the Influence of Undergraduate Students' Physical Health on Online Health Information Seeking, Journal of Information Resources Management (2017), 81-87.

[16] L. Guiling, The Characteristics of Health Information Behavior in Different Stages of the Individual, Journal of Modern Information 36 (2016), 92-95+104.

Due to pages' limitations, the full references can be accessible via https://unknowndreaming.github.io/references/references.html

Address for correspondence
Corresponding author: Shijuan Li
Email: shijuan.li@pku.edu.cn

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190749

Should the Picture Archiving and Communication System (PACS) Settings Be Standardized? Questionnaire Survey for Safe Medical Image Management

Takumi Tanikawa [a], Ayako Yagahara[a], Akihisa Fukuda[b], Daisuke Ando[c], Tatsuya Suzuki[d], Kouhei Harada[e], Shuichi Karata[f], Masahito Uesugi[g]

[a] Faculty of Health Sciences, Hokkaido University of Science, Sapporo, Hokkaido, Japan
[b] Department of Radiology, Hokkaido Medical Center, Sapporo, Hokkaido, Japan
[c] Department of Radiological Technology, Hokkaido University Hospital, Sapporo, Hokkaido, Japan
[d] Department of Radiological Technology, Asahikawa Medical University Hospital, Asahikawa, Hokkaido, Japan
[e] Department of Radiological Technology, Sapporo Medical University Hospital, Sapporo, Hokkaido, Japan
[f] J-MAC SYSTEM Inc, Sapporo, Hokkaido, Japan
[g] Department of Medical Management and Informatics, Hokkaido Information University, Ebetsu, Hokkaido, Japan

Abstract

The purpose of this research was to record the differences in picture archiving and communication system (PACS) settings in hospitals. A questionnaire survey was conducted among 261 hospitals that used the PACS system in Japan. As a result, the image deletion method in the PACS server, and other rules were different among hospitals. Furthermore, these variations were dependent on settings of each medical institution. It seemed necessary to define standardized rules for safe image management.

Keywords:
Picture Archiving and Communication System, Data Curation, Surveys and Questionnaires

Introduction

The Digital Imaging and Communications in Medicine (DICOM) standard contributes seamless information exchange such as image diagnostic equipment and picture archiving and communication systems (PACS) in medical institutions. Based on DICOM standard, commercial products such as image diagnostic equipment and PACS have already been supplied to many countries around the world, and are mainly operated in medical institutions. Over the decades, based on requirements from customer hospitals or competition among companies, many common functions between these products have developed besides the DICOM standard functions.

The image management function of PACS plays a very important role in maintaining confidentiality, integrity, and availability while electronically preserving medical records. Furthermore, patient history is recorded for trace management, and a mechanism, in accordance with the regulation for electronic preservation of medical records in Japan, is also implemented. However, details of these mechanisms vary from manufacturer to manufacturer, which could be a problem if system specifications or image management functions are not confirmed by user during system installation or operation.

In this study, we conducted a questionnaire survey on the PACS image management targeting the radiological department of hospitals and evaluating variations in management manners and system settings.

Methods

A questionnaire survey was conducted between February and March 2018 in all 261 hospitals having the PACS system in Hokkaido prefecture, Japan [1]. Questionnaires were sent to the system personnel and administrators in the radiology departments of all these hospitals.

The contents consisted of a face sheet and the following two items: 1) rules regarding the PACS receiving image of data that was already registered (whether the PACS server rejected the resent image data or overwrote it?), 2) PACS image deletion method (whether the PACS storage server deleted the image logically or physically). Additionally, if the answer was to reject the image data, indicators that decided the rejection, such as “study instance UID”, “series instance UID”, “SOP instance UID”, “accession no.”, and “AE title”, were subsequently asked. These questions were decided upon as image management functions were different among the different vendors and/or products in PACS. Candidates for each answer are provided by several options, which can be freely described if necessary.

In the face sheet, following items were asked regarding hospital information: number of hospital beds, number of staffs in the radiological department, whether medical information staff were internally assigned in the radiological department, vendor and product names of the current PACS, year in which the current PACS was introduced, and storage size of the PACS.

The survey sheets were sent by mail and collected by web base response sheets (Google Form). This study was conducted with the approval from the Hokkaido University of Science Ethics Committee.

Results

Responses were received from 93 hospitals (response rate: 36%). In computed tomography (CT) examination, 54.8% hospitals always overwrote data if existent image data was resent to PACS, while 40.9% refused to receive data based on certain rules such as “study instance UID” and/or “SOP instance UID”. This trend was similar in computed radiography (CR) and magnetic resonance imaging (MRI) examination. In

67.7% of hospitals, the authority of deleting images from PACS was limited to specific staff, whereas 32.3% had no such limitation. In 50.5% of hospitals, logical deletion was adopted for image deletion from PACS, but 33.3% adopted physical deletion. This trend did not depend on the particular vendor or manufacturer, but on the individual settings of each hospital.

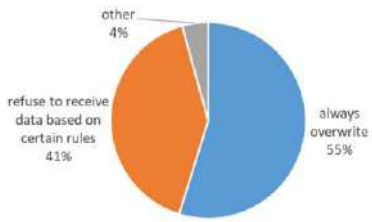


Figure 1. Response in cases of image resent to picture archiving and communication system (PACS) and computed tomography (CT) (n=38)

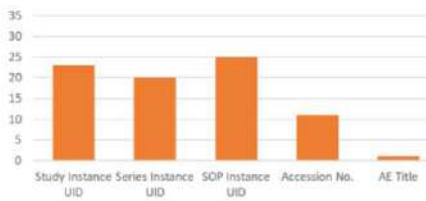


Figure 2. Indicators for refusing to receive data computed tomography (CT) (n=38)

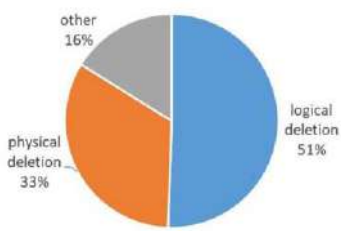


Figure 3. Deletion policy in PACS (n=93)

Discussion

The two methods of image deletion are physical deletion and logical deletion. Physical deletion is a method of deleting images from the storage completely; therefore, once deleted, restoration is extremely difficult. Logical deletion is a method of adding an identifier for hiding an image, and the image is held on the storage in practice. Physical deletion poses the problem of difficulty in restoring the data from the past if the operator mistakenly deletes the image. Although logical deletion is inferior in terms of storage cost, retaining deleted images as history has a larger merit in maintaining important medical records. From the viewpoint of information security and medical safety, it is necessary to standardize the image deletion method.

Images are resent to PACS in case of post-processing the images or modifying the DICOM tag information. Although resending images caused by post-processing has a different frequency among each modality, distribution of overwriting and rejection were almost the same in every modality. SOP instance UID is uniquely numbered to identify the image; however, whether the new number should be registered to the post-processed image has not yet been defined in the DICOM standards. The meaning of "unique" and its reference point are left to the judgement of the operation side, and this fact could be a reason for this variation. With respect to preserving evidence, it is also necessary to standardize the definition of uniqueness of an image, and the handling rule of the image in PACS.

The DICOM standard is a standard for the interface between equipment, so internal functions of the equipment are not covered by the standards. The results of this study indicated that PACS image management method was different among hospitals, and the difference did not depend on vendors or manufacturers but on the setting of each equipment. For the safety management of medical information systems, it is necessary to observe each nation's laws and regulations. Indeed, if this regulation is applied to management of medical images, every vendor and manufacturer would also conform to them. This study focused on the legal gray area. Simple operation of image management using PACS is necessary for the hospital staff. The results of this study indicated the need for standards and regulations in the PACS image deletion method and rules for image data received by PACS of already registered data.

Conclusion

The aim of this study was to record the problems related to the PACS image management function from the view point of medical institutions. Results of the questionnaire survey showed that the image deletion method in the PACS server, and the overwriting rule in case of existent images being re-transmitted to the server were different between different hospitals. Furthermore, these variations were not dependent on the vendors or manufacturers, but on the settings of each hospital. Since the image management method of PACS is not standardized, it was proposed that appropriate regulation was needed for safe management of electric medical record as per the law.

Acknowledgements

This study was supported by the Hokkaido Radiological Technology research Grant (Japanese Society of Radiological Technology). The authors would like to express their deep gratitude to the members of the Hokkaido branch of Japanese Society for Radiological Technology.

References

[1] New Medicine in Japan Data Book Series - Medical Equipment System White Paper 2018, ME promotion association cooperation.2018.

Address for correspondence
Corresponding author: Takumi Tanikawa
Mailing address: 7-15-4-1 Maeda, Teine, Sapporo, Hokkaido, Japan 006-8585
Email: tanikawa-t@hus.ac.jp
Phone: +81-11-676-8523

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190750

Navigating the Search for Patient Generated Health Data

Victoria L. Tiase[a,b]; Katherine A. Sward[b,c]; Mollie R. Cummins[b,c]

[a] Department of Information Services, NewYork-Presbyterian Hospital, New York, NY, USA; [b] College of Nursing, University of Utah, Salt Lake City, UT, USA; [c]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA

Abstract

With massive amounts of mobile health data generated by patients, there is a growing amount of research conducted to understand their impact on patient care. The MeSH heading for patient generated health data was established in early 2018, complicating searches for PGHD research prior to 2018. In conducting a search of scientific databases, keywords are presented along with their degree of representation in the literature to help inform future searches.

Keywords:

Patient generated health data, patient portals, patient reported outcome measures

Introduction

With the proliferation of mobile health tools and technologies, such as smartphone apps, activity trackers and sensors, patients are generating more data than ever before. As defined by the U.S. Office of the National Coordinator for Health Information Technology, patient generated health data (PGHD) is 'health-related data created, recorded, or gathered by or from patients (or family members or other caregivers) to help address a health concern [1].

To facilitate optimal use of PGHD, methodological research is needed for feasibility, evaluation, metric selection and data modeling [2]. However, the MeSH term for patient generated data was initiated in January 2018, making the search for scientific works prior to that time difficult for researchers. We seek to support research in this area with proactive identification of optimal keywords to utilize in searches.

Methods

An iterative process of multiple searches was conducted using the PubMed database using various keywords related to patient generated health data. A search filter was applied to restrict publication dates to the last 10 years. Documentation of the inclusion and exclusion search terms along with the number of results generated from each keyword search term was recorded in Microsoft Excel. Histograms were created to depict the distribution of search terms relating to patient generated health data.

Results

Approximately twenty-seven search terms were identified. One key finding was the importance of using hyphens with keywords. Some examples of this are patient generated and patient-generated; patient reported and patient-reported; and self tracking and self-tracking. This is an example of one of the graphs that will be reported in the results section of the poster.

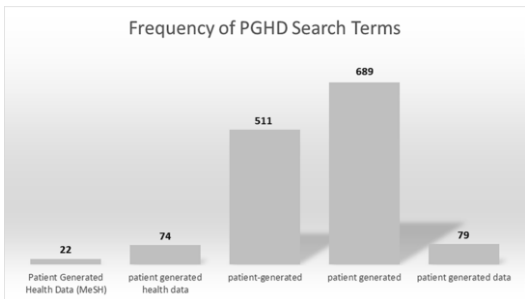


Figure 1 - patient generated health data search terms

Conclusions

Exploring research related to patient-generated health data is important for an understanding of the state of the science in this area. Due to a lack of MeSh terms and the hyphenation of words associated with PGHD, understanding the frequency of keywords can be helpful to researchers in creating search strategies and expediting searches in this area.

Acknowledgements

Ms. Tiase would like to acknowledge the support of Jonas Philanthropies and the Jonas Nurse Leaders Scholar Program.

References

[1] V. Tiase, Navigating the patient-generated health data deluge, *Nurs Manage* **48** (2017), 7-8.

[2] W.A. Wood, A.V. Bennett, and E. Basch, Emerging uses of patient generated health data in clinical research, *Mol Oncol* **9** (2015), 1018-1024.

Address for correspondence

Victoria L. Tiase
vtiase@nyp.org

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190751

A Framework for Applied AI in Healthcare

Tran Truong[a,b], Paige Gilbank[b], Kaleigh Johnson-Cover[b], Adriana Ieraci[b]

[a] Techna Institute, University Health Network, Toronto, Ontario, Canada,
[b] Translational Research Program, University of Toronto, Toronto, Ontario, Canada

Abstract

Significant efforts are being made to develop artificial intelligence technologies for health settings. In a health system that has been notoriously slow to adopt innovative technologies, it is important to consider the implementation of a new technology early in the development stage, especially one that will have added challenges of trust and transparency. To facilitate this process, an implementation framework for artificial intelligence technologies in clinical settings has been created.

Keywords:

Artificial Intelligence, Health Information Systems, Data Science

Introduction

As technology advances and the use of Aritifical Intelligence (AI) technology is adopted in various fields, there are increasing efforts to develop AI technology for healthcare applications. Within the class of AI technology, Machine Learning (ML) systems are being developed that draw from statistics, mathematics, rule-based systems and biological systems to create solutions that can adapt and learn, thus reducing manual burden. Private companies are building ML into medical decision-making, pursuing tools that support physicians [1]. Physician-researchers are predicting that familiarity with ML tools that analyze big data will be a fundamental competency for the next generation of physicians, leading to improved care in areas such as radiology and anatomical pathology [2]. Millions of dollars are being invested in AI, most implementations are still proofs of concept and there remain few examples of successful widespread implementation in health settings [3].

Due to the "black box" nature of AI and the high dependency on accurate patient data for model training, there are unique challenges to successful implementation of AI in healthcare that are not encompassed in common implementation frameworks, such as the Ottawa Model of Research Use [4]. Creating an implementation framework to help healthcare orgainzations understand the key considerations and guide implementation efforts for AI will accelerate adoption and ensure that these technologies are optimally integrated to improve both patient care and patient outcomes.

Methods

An environmental scan was conducted utlizing informal meetings with eight Subject Matter Experts (SMEs) from four hospitals, one national homecare organization, and one academic institution. The SMEs provided examples of health AI technologies and talked about the considerations and challenges of implementation. The learnings from these meetings were then brought to a brainstorming session where the Affinity Diagram grouping method, common in human-centered design [5, 6], was used to help identify key themes that were recurrent in the experiences of implementing AI technologies in the health setting. A literature review was then conducted to further explore the identified themes. A visualization of the framework was created utilizing a nested hierarchy to demonstrate thematic importance. To further understand the critical factors to clinician buy-in, a qualitative research study is underway. This study aims to uncover clinician perceptions, acceptance levels, and professional standards around AI for clinical settings.

Results

The key themes that were identified from the experiences of implementing AI technologies in health settings include data, trust, ethics, readiness for change, expertise, buy-in, regulatory strategy, and scalability.

- Data: AI technology can only be as good as the data used to create it. Thus, factors including data quality, quantity, and collection need to be controlled for.
- Trust: Trust is generated by understanding, transparency and overall explainability of AI technologies, and is needed by both patients and physicians for effective adoption.
- Ethics: Ethical challenges arise with the collection and use of patient data, with implementation and dissemination of the technology developed.
- Readiness: The readiness for change of a given clinic or healthcare institution greatly influences the successful implementation of any new practice. Key factors include appropriate infrastucture, sufficient understanding of AI, and effective change management procedures at the institution.
- Expertise: The input of leaders in the field, including technologists, front-line hospital staff, and clinicians, is required to ensure effective development and implementation of user-friendly technology.
- Buy-in: Successful implementation is facilitated by generating buy-in from all levels of staff from the very beginning of any given project. All members/stakeholders must understand the value and need of the chosen technology.
- Regulatory Strategy: AI technologies that impact patient care will be subject to regulatory oversight in the jurisdiction of commercial use. It is important to have a regulatory strategy during the product development, to ensure a streamlined process.

- Scalability: Challenges for dissemination and adoption by other clinics or hospitals. ML models trained with local data may have biases that prevent them to be generalizable.
- Evaluation: It is essential to have a plan to evaluate the success of the implementation by predefined metrics, including impact on care, medical outcomes or patient experience, integration into workflow, sustainability, and economic considerations.

Utilizing a nested hierarchy of thematic significance, a visualization of the implementation framework was compiled (Figure 1). At the core of the framework is the crux of all ML projects: data. Data in healthcare includes large volumes of heterogeneous data from various systems, with different levels of veracity. These qualities of health data create a challenge for AI technologies [7] that require large amounts of high quality data to achieve high levels of accuracy. Availability, quantity and quality of health data are key considerations in the framework. Without these assurances, a ML initiative cannot gain enough traction to be considered at the next level of implementation. The second level includes the ethics around privacy and secondary use of data [8] and the trust of AI "black box" technology [9] by healthcare providers and patient users. These themes are fundamental to be able to achieve the next level in the framework, including buy-in, readiness and expertise [10]. These involve many levels of stakeholder engagement and organizational strategy alignment, making them vital to successful implementation. Healthcare organizations will need to recognize the value in AI technology and support the required development of infrastruture and expertise. Finally, in the outermost level of the implementation framework are the themes: regulatory, scalability and evaluation. These themes focus on ensuring the value and the long term sustainability of the AI technologies. It is worth noting that all themes should be considered throughout the design and implementation process.

Figure 1– Preliminary Framework for Applied AI in Healthcare

Conclusions

With the rapid emergence of AI technologies for health, organizations need to better understand the key considerations for successful implementation of such technologies. The implementation framework presented identifies these considerations in a nested hierarchy to demonstrate the core dependencies, with data at the center. This framework can be used by healthcare organizations to inform strategic initiatives to ensure readiness for AI technologies or inform individual technology implementation. Companies developing AI technologies can use this framework to understand the implementation requirements when they work with healthcare organizations.

Discussion

The implementation framework presented is a preliminary framework and requires further development. A qualitative study is underway to explore the core themes of data, ethics and trust from the perspective of the health care providers. Results from the study will be used to iterate on the framework.

Acknowledgements

We would like to acknowledge the Translation Research Program for their support on this project.

References

[1] D.S. Char, N.H. Shah, and D. Magnus, Implementing machine learning in healthcare – Addressing ethical challenges, N Engl J Med **378** (2018), 981-983.

[2] Z. Obermeyer, and E.J. Emanuel, Predicting the future – Big data, machine learning, and clinical medicine. N Engl J Med **375** (2016), 1216-1219.

[3] T. Lancet. Artificial intelligence in health care: within touching distance. **390** (2017), 2739.

[4] National Collaborating Centre for Methods and Tools. Ottawa model of research use: A framework for adopting innovations. Hamilton, ON: McMaster University. (Updated 03 October, 2017), (2010). Retrieved from http://www.nccmt.ca/resources/search/65 (accessed March 31, 2019).

[5] K.N. Otto, & K.L. Wood, Product Design: Techniques in Reverse Engineering and New Product Development. Prentice-Hall Inc, Upper Saddle River, 2001.

[6] R. King, Better Design in Half the Time: Implementing Quality Function Deployment (QFD) in America. Metheun, MA: GOAL/QPC, 1987.

[7] I.E. Naga, M.R. Kosorok, J. Jin, M. Mierzwa, R.K. Ten Haken, Prospects and challenges for clinical decision support in the era of big data. JCO Clin Cancer Inform **2** (2018), 1-12.

[8] M. Ienca, A. Ferretty, S. Hurst, M. Puhan, C. Lovis, E. Vayena, Considerations for ethics review of big data health research: A scoping review. PLOS ONE **13** (2018).

[9] A. J. London, Artificial Intelligence and Black-Box Medical Decisions: *Accuracy versus explainability.* The Hastings Centre Report (2019), 15-21.

[10] V.N. Patel,E.H. Shortliffe, M. Stefanelli, P. Szolovits, M.R. Berthold, R. Bellazzi, A. Abu-Hanna, The Coming Of Age of Artificial Intelligence In Medicine. **46** (2009) 5-17. doi:10.1016/j.artmed.2008.07.01

Address for correspondence

Tran Truong, BMath, MHSc Candidate, Techna, University Health Network, 79-100 College St., Toronto, Ontario, Canada, M5G 1L5. Email: Tran.Truong@uhn.ca, Phone: 416-581-7653

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190752

Personal Health Information Management Practices of Older Adults: One Size Does Not Fit All

Anne M. Turner[a, b], Katie Osterhage[a], Andrea Hartzler[b], Jean O. Taylor[a], George Demiris[c]

[a] *Department of Health Services, University of Washington, Seattle, WA, USA,*
[b] *Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, USA,*
[c] *Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA, USA*

Abstract

Older adults are the largest consumers of health care, have the greatest number of chronic conditions, and generate the greatest amount of health data. Yet, information systems designed to aid health information management do not align with their needs and practices. We describe a process of identifying the personal health information management (PHIM) activities and objectives of older adults (60 years and older) from different residential settings.

Keywords:

Aging; Information Technology; Consumer Health Information

Introduction

Aging often brings increased utlilzation of health care services, resulting in personal health information (PHI) which must be recalled, recorded, organized, and utilized through various personal health information management (PHIM) strategies. PHIM practices, which enable a person to retreive and utilize PHI for maintenance of their health, have been observed in select populations [1; 2]. Research examining PHIM practices of older adults calls attention to barriers to PHIM technologies [3] and facilitating roles of healthcare providers [4] and caregivers [5; 6]. Less is known about PHIM practices of older adults themselves [7].

We sought to better understand the PHIM practices of older adults. This study is part of the larger SOARING project (Studying Older Adults and Researching their Information Needs and Goals) at the University of Washington, a longitudinal study investigating PHIM needs and practices among older adults aged 60 and above in a variety of residential settings (i.e., independent living, retirement communities, assisted living). The goal of the SOARING project (soaringstudy.org) is to inform the design of health information technologies that better meet the PHIM needs of older adults.

Methods

We conducted home visits with 88 older adults. We used purposive sampling to recruit participants from a variety of living situations from a metropolitan area in the US Pacific Northwest. Ethnographic tools were used including interviews, observation via photographs of information artifacts and living arrangements, along with surveys about burden of disease, health literacy, and decision-making. Interviews were audio-recorded, transcribed, and uploaded into an online qualitative analysis software (Dedoose) for coding and analysis. Two researchers applied codes to the interview transcripts. To check for inter-coder reliability 10% of the transcripts were double coded by the two researchers. Inter-coder agreement was 90%. Through thematic analysis of interview transcripts, survey results, and photos, we identified distinct PHIM activities of older adults.

Results

Participants: The 88 older adult participants had an average age of 76 years (sd 10), were predominantly female (69%), white (72%), and had at least a bachelor's degree (61%). Participants resided in diverse living situations: independent residences (25%), independent-shared dwellings (27%), retirement communities (27%), assisted living facilities (19%), and homeless (2%). Seventy percent of participants lived alone. Most participants reported having access to a computer (89%) and Internet (83%) where they lived.

PHIM activities: We identified five key PHIM activities of older adults through thematic analysis of interview transcripts. For each activity, we identified the "objective", or goal, and the "approach" used to meet that objective (see Table 1). Older adults varied in how they sought, tracked and organized information as well as the degree to which they involved others in PHIM activities.

Discussion

Given that older adults' approaches to PHIM activities vary, a uniform health information management technology will not meet the needs of all older adults. Our findings demonstrate the need for informatics tools that can be tailored and adjusted to the needs and preferences of individual older adults. Tools should be designed to consider the experiences of individual users (e.g., cognitive and functional ability, experience with technology) and their distinct approaches to carrying out PHIM activities.

Table 1– PHIM Activities

PHIM Objective	PHIM Approach	Distribution (n=88)
Organizing: Strategy used for handling health-related print materials	• Filing • Piling • Tossing	48.9% 38.6% 12.5%
Tracking: Generating and/or logging of health-related measures	• Never • Sometimes • Consistently	44.3% 28.4% 27.3%
Seeking: Type of engagement with obtaining health-related information	• Active • Combined • Passive	51.1% 29.5% 19.3%
Sharing: Including others (family/friends) in communication and management of health-related information	• Independent but shares • Collaborative-team/partnership • No sharing • Proxy	44.3% 36.4% 11.4% 8.0%
Emergency Planning: Preparing / maintaining information for health-related emergency	• Yes-by self • Yes-by others • No planning	40.5% 29.8% 29.8%

Conclusions

Older adults displayed a variety of approaches to meeting the objectives of specific PHIM activities. Technology designed to support older adults should take into consideration this variety as well as personal preferences with respect to PHIM, remembering that one size does not fit all.

Acknowledgements

This research was supported through funding from the Agency for Healthcare Research and Quality R01HS022106. The findings and conclusions expressed here are those of the authors and do not necessarily represent the views of AHRQ.

References

[1] W. Pratt, K. Unruh, A. Civan, and M.M. Skeels, Personal health information management, *Communications of the ACM* **49** (2006), 51-55.

[2] A. Moen and P.F. Brennan, Health@ Home: the work of health information management in the household (HIMH): implications for consumer health informatics (CHI) innovations, *Journal of the American Medical Informatics Association* **12** (2005), 648-656.

[3] H. Jimison, P. Gorman, S. Woods, P. Nygren, M. Walker, S. Norris, and W. Hersh, Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved, *Evid Rep Technol Assess (Full Rep)* **175** (2008), 1-1422.

[4] J.S. Ancker, H.O. Witteman, B. Hafeez, T. Provencher, M. Van de Graaf, and E. Wei, The invisible work of personal health information management among people with multiple chronic conditions: qualitative interview study among patients and providers, *Journal of medical Internet research* **17** (2015), e137.

[5] R.J. Holden, Y.L.P. Karanam, L.H. Cavalcanti, T. Parmar, P. Kodthala, N.R. Fowler, and D.R. Bateman, Health information management practices in informal caregiving: An artifacts analysis and implications for IT design, *International journal of medical informatics* **120** (2018), 31-41.

[6] J.O. Taylor, A.L. Hartzler, K.P. Osterhage, G. Demiris, and A.M. Turner, Monitoring for change: the role of family and friends in helping older adults manage personal health information, *Journal of the American Medical Informatics Association* **25** (2018), 989-999.

[7] A.M. Turner, K. Osterhage, J. Loughran, I. Painter, G. Demiris, A.L. Hartzler, and E.A. Phelan, Emergency information management needs and practices of older adults: A descriptive study, *International journal of medical informatics* **111** (2018), 149-158.

Address for correspondence

Anne M. Turner, MD, MLIS, MPH, FACMI, amturner@uw.edu,
Department of Health Services, Box 357660,
School of Public Health,
University of Washington, Seattle, WA 98105

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190753

SIM-CIG: A Serious Game to Practice and Improve Clinical Guidelines Adoption Based on Computer-Interpretable Guidelines

Sebastian Uribe-Ocampo[a], Ever A. Torres[a], Ivan F. Luna[a], Jose F. Florez-Arango[a,b], Jack W. Smith[a,b]

[a] *Department of Simulation in Healthcare, Universidad Pontificia Bolivariana, Medellin, Antioquia, Colombia,*
[b] *College of Medicine, Texas A&M University, United States of America*

Abstract

Serious games have been used to increase the accuracy and usege of clinical guidelines during routine clinical practice. This document presents the development of a serious game called SIM-GIC, a video game designed to simulate virtual patients and evaluate the decision making of players based on computer-interpretable clinical guidelines. The system is currently being developed with a content focus on antenatal care guidelines, where a number of obstetric guidelines were coded in XML files.

Keywords:

Game Development, Medical Education, Serious Gaming, Virtual Simulation, Role Play

Introduction

Clinical practice guidelines are used by students and professionals as aids during medical encounters [1], to standardize care protocols based on clinical features of the patients [2]. Clinical guidelines containsets of recommendations to make the best decision possible based on available evidence [3; 4]. For health-care institutions, a high adoption level of clinical guidelines would in an increase the frequency of the application of the best clinical practices, avoiding the use of ineffective treatments and optimizing the use of scarce resources [2; 5].

Regardless of these efforts, adherence to guidelines often remains low, causing omission of therapies recommended in the guidelines and contributing to preventable harm [6]. The low adoption levelsare mainly due to the difficulty in translating the evidence into practice, particularly where guidelineadoption would require changes in routine clinical practice[7; 8]. Several initiatives to target this issue of increasing adherence to guidelines have proposed educational resources implementated in video-games [6], where virtual patients can be used to promote guideline uses in a effective and fun way [9]. The following poster presents SIM-CIG, a video game designed to simulate virtual patients and evaluatethe decision making of players based on computer-interpretable clinical guidelines.

Clinical guidelines are often written as large documents in textual and diagrammatic format, making it difficult to integrate and apply in a standardize patient care process as it is [10]. Computer- Interpretables-Guidelines (CIGs) appeared as translations of guidelines from their published formats into computer interpretable algorithms translating written guidelines into computing tasks that unfold over time [11] using an inference engine to interprets the computer statements corresponding to guideline text [12].

SIM-CIG attempts to adopt CIGs in the design of game objectives, where the information in the CIGs would control the actions required to be performed by the studentto "win the game". With the implementation of CIGs as a game objective , we aim to develop a video-game with editable game objectives similar to how clinical decision support systems works, avoiding source code edition as much as possible and allowing for institutions or individuals modify the CIGs according to their needs.

Methods

Games Objectives: SIM-CIG is limited to simulated an antenatal health-care environment, evaluating decision making of professionals relative to obstetric guidelines based on the pregnancy conditions of virtual patients. Clinical functionalities were extracted through a systematic review of guidelines related to antenatal care. The selected guides can be found in references [13]. Each functionality was classified and grouped into: physical actions and interaction with the patient, equipment required, record of information about in the Electronic Health Record (EHR), prescription, management and cessation of medication, clinical orders carried out for screening, diagnostic or with therapeutic purposes, instructions and information provided to the patient.

CIG Syntax Implementation: There are many ways in which the terminology can be defined, it was decided to use terms related to the planning theory in artificial intelligence [14]. Where the knowledge is structured in terms of actions and their consequences, and the syntax is structured around a condition called goal, with a sequence of actions need it to achieve it. This terminology was used to designed the game planning syntax structure where the medical knowledge will be codified. It is decided that our CIG will be in XML file format. In simple terms, information in clinical guidelines is generally found in terms of actions suggested based on specific clinical features presented in the patient, and this will be the core idea of how the game will evaluated player's actions, where every rules will at least have a goal, a precondition and an action. For each condition, a boolean expression is declared with logical operators, inside each boolean expression a semantic path reference is described to relate the information with a specific ID in a database.

Pregnancy Model: The pregnancy simulation is based on Synthetic Patient Population Simulator pregnancy model [15], which can model some pregnancy data and conditions that contribute to the medical history of the synthetic patients.

Synthea provides a JSON file format for each disease and medical conditions. To implement the pregnancy model for the game environment the JSON files were transformed into XML files and limited to simulating a 9 months' normal pregnancy and one with preeclampsia complications.

Database data model and Electronic Health Record: Records and logs of activities are stored in a Microsoft SQL Server database with a data model similar to the openMRS [16], selected due to how drugs, orders and information is handle, making it possible to employ ICD-10, SNOMED-CT or RXNORM coding. Our CIG references a specific ID of the db Concept table and evaluates the game objectives achievements based on the values registered in the OBS database.

Results

Clinical knowledge was coded in the XML file that used the games to stablished the rules. For example, one of the goals is to treats a patient whose Rh type is negative, and according to clinical guidelines after 28 weeks of pregnancy a patient with Rh type negative should be treat with Prophylactic Anti-D.

In Figure 1. is presented how SIM-CIG looks alongside the EHR designed to register and evaluate the information acquired in the encounter.

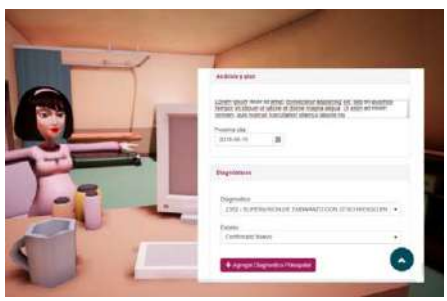


Fig. 1: Medical Encounter SIM-CIG .

Discussion

In this first approach the knowledge code from the clinical guidelines is mainly for obtetric care, after finishig the test stage next step involves covering additional guidelines and evaluated more knowledge. Synthea is the main core of the patient data generation, but still don't cover majors aspect of a real pregnancy. Future versions will attend this issue to bring more reality.

Conclusions

CIGs demonstrated to be integratable into the source code of a serious game, creating a customizable objective game for evaluating medical knowledge. The Synthea Patient Generator can be used to generate data with sufficient clinical information to simulate a pregnant virtual patient in order to provide experience with antenatal care guidelines in a serious game environment.

Acknowledgements

The authors would like to thank Universidad Pontificia Bolivariana Simulation Lab for their collaboration

References

[1] Z. Philips, L. Ginnelly, M. Sculpher, K. Claxton, S. Golder,R. Riemsma, N. Woolacoot, and J. Glanville, "Review of guidelines for good practice in decision-analytic modelling in health technology assessment," pp. iii–iv, ix–xi, 1–158, Sep 2004.

[2] S. A. Bini and J. Mahajan, "Achieving 90% Adoption of Clinical Practice Guidelines Using the Delphi Consensus Method in a Large Orthopedic Group," Journal of Arthroplasty, vol. 31, no. 11, pp. 2380–2384, Nov 2016.

[3] Committee on Patient Safety and Quality Improvement, "Committee Opinion No. 629," Obstetrics & Gynecology, vol. 125, no. 4, pp. 1027–1029, Apr 2015.

[4] G. Penney and R. Foy, "Do clinical guidelines enhance safe practice in obstetrics and gynaecology?" pp. 657–673, Aug 2007.

[5] S. Duckett and P. Breadon, "Questionable care: Avoiding ineffective treatment," 2015.

[6] P.J. Pronovost,"Enhancing physicians' use of clinical guidelines," pp. 2501–2506, Dec 2013.

[7] M. R. Keiffer, "Utilization of Clinical Practice Guidelines Barriers and Facilitators." pp. 327–345, Jun 2015.

[8] I. del Cura-Gonza´lez, J. A. Lo´pez-Rodr´ıguez, T. Sanz-Cuesta, R. Rodr´ıguez-Barrientos, "Effectiveness of a strategy that uses educational games to implement clinical practice guidelines among Spanish residents of family and community medicine (e-EDUCAGUIA project): A clinical trial by clusters," Implementation Science, vol. 11, no. 1, p. 71, Dec 2016.

[9] A. E. Olszewski and T. A. Wolbrink, "Serious Gaming in Medical Education," Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare, vol. 00, no. 00, p. 1, 2016.

[10] R. Goud, A. Hasman, and N. Peek, "Development of a guideline- based decision support system with explanation facilities for outpatient therapy," Computer Methods and Programs in Biomedicine, vol. 91, no. 2, pp. 145–153, 2008.

[11] M. Peleg, "Computer-interpretable clinical guidelines: A method- ological review," pp. 744–763, Aug 2013.

[12] S. Tang and M. Hanneghan, "Game content model: An ontology for documenting serious game design," Proceedings - 4th International Conference on Developments in eSystems Engineering, DeSE 2011, pp. 431–436, 2011.

[13] Ministerio de Salud y Proteccio´n Social, Gu´ıas de Pra´ctica Cl´ınica para la prevencio´n, deteccio´n temprana y tratamiento de las complicaciones del embarazo, parto o puerperio 2013, 2013, no. 11.

[14] R. J. Brachman and H. J. Levesque, "Chapter 15 - Planning," Knowledge Representation and Reasoning, pp. 305–325, 2004. [Online].

[15] J. Walonoski, M. Kramer, J. Nichols, A. Quina, C. Moesel, D. Hall,

C. Duffett, K. Dube, T. Gallagher, and S. McLachlan, "Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record," Journal of the American Medical Informatics Association, vol. 25, no. February, pp. 230–238, 2017.

[16] OpenMRS, "Data Model - Documentation - OpenMRS Wiki." [Online]. Available: https://wiki.openmrs.org/display/docs/Data+Model

Address for correspondence

sebastian.uribeo@upb.edu.co.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190754

Change Management in Healthcare Organizations: *Soft Skills* Training Strategies Through Blended Learning Environments

Giussi Bordoni MV[a], Baum A[a], García G[a], Moriñigo P[a], Luna D[a], Otero P[a], Otero C[a], González Bernaldo de Quirós F[a]

[a] Department of HealthInformatics, Hospital Italiano de Buenos Aires, Argentina.

Abstract

Teaching soft skills for change management in healthcare organizations is becoming increasingly necessary, even more, when implementing health information systems (HIS). There is little evidence that these skills can be learned through online teaching environments. This paper describes the experience of having taught soft skills to health informatics master's degree students, through blended learning environments.

Keywords:

Environment, Health information systems, Students

Introduction

Implementing change management in health information systems (HIS) requires leaders with highly developed soft skills. [1; 2] According to the literature, competencies such as empathy, flexibility, active listening skills, clear communication and negotiation strategies can be developed or strengthened through training programs. The development of pedagogical strategies to strengthen these competences through e-learning has been, so far, hardly explored. [3]

The International Medical Informatics Association (IMIA) cites several competencies that health informatics professionals (HIP) should have. Among them, HIP should be able to handle the resolution of socio-organizational issues, so as to achieve a successful change management. [4] Including information and communication technologies (ICT) in the education field can result in a new teaching-learning process, overcoming the barriers of space and time presented by face-to-face training modality. Distance learning has become an alternative at postgraduate levels. Advantages of e-learning have been widely reported, but there is scarce literature describing pedagogical strategies to teach soft skills through it. [5]

The aim of this paper is to describe the design and development of a course to teach soft skills for managing change in HIS projects, to professionals of different disciplines, through a master's degree in health informatics within a blended learning environment.

Methods

Setting

Hospital Italiano de Buenos Aires (HIBA) is a third-level university hospital created in 1853. In 2016, the professional master's degree in Health Informatics was created as a blended learning program. The graduate profile that this master's degree proposes is a professional capable of leading digital transformation projects within the healthcare field.

"Organizational Issues and Change Management" Class Design

The "Organizational issues and Change Management" class takes place within the second year of the master's degree. It delves into content related to soft skills and lasts a total of 12 weeks in virtual mode and 4 hours of face to face activities.

The teaching team is responsible for selecting the subject's contents, designing pedagogical strategies, and guiding students along the course.

The contents were divided into three different sections along the 12 weeks course. In section 1, the students reviewed concepts related to organizational behavior and theories, paradigms' changes (from the industrial to the digital era) and people's behavior. Concepts related to healthcare organization's particularities, organizational culture, diversity of actors, hidden organizational charts, power relationships and types of change, were addressed in section 2. Finally, in section 3, students were provided with tools to better handle resistance to change and the topics reviewed were Information System Implementation Change Management-barriers and facilitators-, sociotechnical approach, communication strategies, support and training, conflict negotiation tools, and evaluation and management monitoring. Contents were approached through individual activities in the first 2 sections and through group activities in the 3rd section.

Results

The first cohort participated between March and June 2018. 53 students from different professions participated (Medicine, Nursing, Engineering, others). 64% of the students were male.

To reach the goal of consider resistance to change as a natural attitude of human beings, the students were called to reflect on their own resistance to the changes and build a collaborative board (Padlet) with representative images. Other goal of section 1 was to identify organizational properties, people's behavior and tensions generated when introducing changes. To reach these, students watched a film and completed a chart. The aim of section 2 was to understand the complexity of healthcare organizations, recognize the type of changes (Lorenzi), list restraining and driving forces (Lewin´s Force Field Analysis Model), and identify types of technology adopters (Rogers). To reach this, students watched Grey's Anatomy's "Bad blood" chapter and completed the evaluation chart. Finally, in section 3, students were asked to analyze a HIS implementation case in order to position themselves as leading agents of change and develop skills to handle difficult conversations. Training in active listening and acquiring negotiation skills were goals of section 3 too. All sections

were accompanied by forum discussions to enrich the knowledge of all the participants.

Discussion

One of the challenges of teaching through virtual environments is to design pedagogical strategies to simulate usual scenarios of health information systems (HIS) implementation, allowing students to learn key skills for their graduate profile. Along their virtual training, students were asked to experience first-hand how it felt to identify themselves as protagonists when managing change. We identified that it was easier for students to solve activities based on the analysis of cases reported in third person. For example, it was easier to answer questions such as "Why do people resist change?" than the question "Why do I resist change?" Students were gradually able to embrace the role of leaders and respond to assignments with a greater degree of personal involvement.

Face-to-Face training was key to empathize with conflict scenarios and integrate what was learned through virtual training. Since virtual and face-to-face training generate completely different feelings and experiences, teachers have been challenged to come up with strategies to transform the way virtual learning is perceived.

Limitations

Although this paper describes a local experience, the reported teaching strategies could be considered for similar projects.

Conclusion

There is a clear need to include soft skills training in health informatics' education programs. Teaching these skills through virtual environments can be challenging.

References

[1] S.H. Fenton, M.J. Gongora-Ferraez, and E. Joost, Health information technology knowledge and skills needed by HIT employers, *Appl. Clin. Inform.* **3** (2012), 448–61. doi:10.4338/ACI-2012-09-RA-0035.

[2] W. Hersh, The Health Information Technology Workforce, *Appl. Clin. Inform.* **01** (2010), 197–212. doi:10.4338/ACI-2009-11-R-0011.

[3] G.S. Britton, and G.S. Britton, Using Online Project-Based Capstone Experiences to Enhance Soft Skills Development, 2013. https://www.learntechlib.org/p/120006/ (accessed November 5, 2018).

[4] J. Mantas, E. Ammenwerth, G. Demiris, A. Hasman, R. Haux, W. Hersh, E. Hovenga, K.C. Lun, H. Marin, F. Martin-Sanchez, G. Wright, and IMIA Recommendations on Education Task Force, Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics, *Methods Inf. Med.* **49** (2010), 105–120. doi:10.3414/ME5119.

[5] T. Myers, A. Blackman, T. Andersen, R. Hay, I. Lee, and H. Gray, Cultivating ICT Students' Interpersonal Soft Skills In Online Learning Environments Using Traditional Active Learning Techniques, *J. Learn. Des.* **7** (2014), 39–53. https://www.jld.edu.au/article/view/194/196.html.

Address for correspondence

María Victoria Giussi Bordoni MD
Hospital Italiano de Buenos Aires
Gascón 450, Buenos Aires city. Argentine
Phone: +549 2317415348
Mail: maria.giussi@hiba.org.ar

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190755

Adoption Factors Related to Electronic Vaccine Record in the Public Primary Care Network of Buenos Aires City

Vazquez MV[a], Palermo C[a], Islas M B[a], Zapata M[a], Giussi Bordoni MV[a], Esteban S[a], Baum A[a]

[a] *Health Information Systems Office, Ministry of Health, Buenos Aires City, Argentina*

Abstract

The Electronic Health Record used in the public primary care network of Buenos Aires City has a specific module for the vaccines registration. The present study explores the factors of EHR adoption by nurses. We found 5 barriers and one facilitator for adoption. Barriers are related with organize the flow of patients especially during vaccine campaigns, adapt the work stations and integrate the records with the vaccine central program.

Keywords:

Nursing Records, Vaccines, Electronic Health Records

Introduction

In accordance with CDC recommendations, several public health agencies have implemented Immunization Information Systems (IIS) for the collection of data about the immunization status of the population and the individual [1].

Many Electronic Health Record (EHR) can help track planned vaccinations [2]. There are three main immunization record issues an EHR can help solve: the human error in the administration of a vaccine not according to the age of the person, the visualization of the calendar and vaccination status of a person in a flexible way, and the tracking of the stock of vaccines, this last is important since they are the most expensive input in primary care.[3] Another important factor is that the vaccination electronic recording may people reduce revaccination due to the loss of paper-based record [4].

In Buenos Aires public healthcare network the vaccination is carried out by nurses who are responsible for the record in order to comply with the Ministry of Health vaccination program. In the last 2 years an EHR was implemented but still electronic vaccination record coexists with the paper record. This paper analyzes the facilitators and barriers for the vaccination record adoption in the Primary Care Centers where vaccination is available.

The Buenos Aires city healthcare network is conformed by a Ministry of Health, 35 Hospitals, 74 Primary Care Centers (44 where vaccination is available), 1 Ambulatory Reference Medical Facility and its structured into 12 geographical areas in order to organize health care delivery. Since June 2016 an Electronic Health Record (EHR) is being gradually implemented in outpatient [5].

In Argentina, the National Vaccination Calendar includes 19 vaccines whose administration is mandatory and free. It is accessed in public health centers [6]. Each healthcare facility registers the application of vaccines in a paper form. These are compiled by the Vaccination Program of the Ministry of Health of the Buenos Aires City. This information is value for vaccine replacement and control of the immunization status of the population of the city.

Methods

We developed an exploratory design with mixed methodology that combines quantitative and qualitative tools.

To determine the level of electronic vaccine record adoption in each primary care center (n = 44), the number of vaccine records in EHR was compared with the amount reported to the vaccine program in paper format for the same period of time. Three ranges were defined, being considered 'high' the cases in which the registry in the HCE exceeded 70%, 'medium' between 40 and less than 70% and 'low' less than 40%.

To determine the factors of adoption, between May and July of 2018, 16 semi-structured, 13 group and 3 individual interviews were carried out with nurses from primary care centers, accomplishing 44 interviewees. In the same period of time, participant observations were carried out.

The analysis of the results was made based on the principles and procedures of the grounded theory [7] and the common themes found are listed in the Result section.

Results

The level of electronic vaccine record adoption in each primary care center (n = 44) is shown in Table 1.

Table 1 – Electronic vaccine record adoption by primary care facility.

High Adoption Level	14
Middle Adoption Level	16
Low Adoption Level	14

The sample consisted of 16 primary care facilities with different levels of use -high level=6, middle level=5 and Low level=5- in order to understand the factors that could influence in

adoption level. The factors found in the interviews were grouped into 5 dimensions:

Transition from paper to electronic record: the vaccine module is not yet integrated with the centralized vaccine report. This results in double registration by the nurses, longer time in the registry and possible delays in care. Given this situation, nurses prefer to comply with the paper record for the centralized report.

Patient flow: the procedure through which EHR is opened requires an administrative process [8]. It was observed that most of the time, patient's access to the immunization sector without going through the administrative post. This fact determines an obstacle to electronic registration. Other factors perceived as barriers to adoption are: the large influx of people during vaccination campaigns, compulsory health checks on children at school entrance and the overload of tasks of administrative staff if all patients went through their post.

Work overload: another obstacle is the large number of tasks performed by the nursing staff, in addition to vaccination. The nurses state that they prioritize care and paper registration over the electronic, a fact that leads to the accumulation of paper records that are sometimes transcribed to the EHR at the end of the day overloading their work.

History of autonomy from the rest of health professionals: the nurses pointed out that the EHR is a tool aimed at other healthcare professionals: *"the EHR is oriented to the medical practice".* Before EHR implementation in Primary Care, the staff of nurses had a fragmented record of their own practices, disintegrated form the whole clinical notes. Some professionals asked for a specific module within the EHR for the record of nurses practices: "*There should be an exclusive module for nursing*".

Nurses Workflow*:* two issues related to the nursing workflow were noticed in the participant observations. The first one is that usually while a nurse vaccinates the patient, the other registers in the EHR. The second issue is that there is not one computer per nurse, but per office. This result in many records made on paper and then transcribed to the system.

Perceptions of advantages and disadvantages of the electronic vaccine record: the main perceived advantage is that the electronic vaccine record avoids people revaccinations in case of paper-based record loss, improving continuity of care. Among the disadvantages are the time that the registration takes and the impossibility of completely replacing the paper.

Conclusion

The use of the EHR by the nurses of Buenos Aires city primary care network is related with multiple factors descripted in literature and not exclusively related to software issues [9]. Among the most important are the need to organize the flow of patients, adapt the work stations and integrate the records with the vaccine central program. Thus, it is important to emphasize that workflow considerations are decisive factor regarding the adoption of an IS. The perception of lack of representation of nursing workflow in the EHR was a recurring concern in the interviews conducted. Different papers report the lack of articulation of nursing records with physicians ones [10]. The most important limitation of this work is related to the impossibility of the methodological tools used, to be able to account for the non-explicit factors that are related to the adoption of the EHR. It is hoped more data-driven studies can be developed to strengthen the validity of the factors listed, and may suggest accurate interventions to improve workflow and system integration.

A complete and appropriate electronic immunization record guarantees a higher quality in patient care. It also allows the generation of statistical information related to the vaccination status of the population, improving decision-making and healthcare management. Exploring the barriers and facilitators of the EHR adoption is a key to improve immunization records.

References

[1] CDC, and Ncird, National Center for Immunization and Respiratory Diseases (NCIRD) - IIS Strategic Plan 2018-2020, 2017. https://www.cdc.gov/vaccines/programs/iis/strategic-plan/downloads/iis-strategic-plan-2018-2020-h.pdf (accessed November 21, 2018).

[2] T. Burkhardt, et al, *Ther. Umschau*. 73 (2016) 297–300. doi:10.1024/0040-5930/a000795.

[3] Immunization Records Management and How an EHR Can Help (2017). https://blueehr.com/blogs/ehr-can-help-immunization-records-management/ (accessed November 21, 2018).

[4] L. Rodewald, et al, Immunization performance measurement in a changing immunization environment., *Pediatrics*. 103 (1999) 889–97. doi:10.1542/peds.107.1.91.

[5] M.V. Giussi, et al, Change Management Strategies: Transforming a Difficult Implementation into a Successful One., *Stud. Health Technol. Inform.* 245 (2017) 813–817. http://www.ncbi.nlm.nih.gov/pubmed/29295211 (accessed November 6, 2018).

[6] Argentinian Ministry of Health, Calendario Nacional de Vacunación y Certificados (2018). https://www.argentina.gob.ar/salud/inmunoprevenibles/calendario (accessed November 6, 2018).

[7] A.L. Strauss, and J.M. Corbin, Basics of qualitative research : grounded theory procedures and techniques, Sage Publications, 1990.

[8] C. Burgo, et al, Estrategias en la implementación de estándares para la identificación de pacientes sobre un software legacy en el Sistema de Salud de la Ciudad de Buenos Aires, n.d. http://47jaiio.sadio.org.ar/sites/default/files/cais-4.pdf (accessed November 21, 2018).

[9] C.S. Kruse, et al., Adoption Factors of the Electronic Health Record: A Systematic Review, *JMIR Med. Informatics*. 4 (2016) e19. doi:10.2196/medinform.5525.

[10] J.L. Molina, et al, Enfermería global, Servicio de Publicaciones, Universidad de Murcia, 2010. http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1695-61412010000200021 (accessed November 22, 2018).

Address for correspondence

Maria Victoria Vazquez - Ministry of Health, Buenos Aires City, Argentina - E-mail: mariavictoriavazquez@buenosaires.gob.ar

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190756

A Comparison of In-Person and Online Training in a Statewide Clinical Education Program for Dissemination of HIV, HCV and STD Clinical Evidence

Dongwen Wang

Department of Biomedical Informatics, Arizona State University, Scottsdale, AZ, USA

Abstract

We compared in-person and online training in a statewide education program to disseminate HIV, HCV, and STD clinical evidence. In a study period of three months, 250 clinicians completed four training courses delivered in dual formats. Course evaluation was positive regarding useful information, easy comprehension, knowledgeable trainer, appropriate format, knowledge increase, intention to use knowledge, and plan to change practice. Online training became a preferred format by clinicians when compared to in-person training (p=0.01).

Keywords:

Internet training, teacher training, knowledge

Introduction

The pandemic of HIV infection has been a major public health challenge for decades. Hepatitis C virus (HCV) infection has also increased rapidly in recent years. Co-infection of HIV, HCV, and other sexually transmitted diseases (STDs) such as syphilis and gonorrhea is frequently reported. With the frequent updates of clinical evidence on HIV, HCV, and STD, timely and effective dissemination of the latest research to community healthcare providers is crucial for translating medical knowledge into routine patient care [1-2].

Since the early 1990s', the New York State Department of Health AIDS Institute has been sponsoring the Clinical Education Initiative (CEI) program for community healthcare providers serving HIV, HCV, and STD patients [3]. In the early years, the CEI program utilized a traditional, in-person approach for training by sending domain experts across the state to give lectures or workshops. This in-person education program successfully trained thousands of clinicians. With the development of information and communication technologies, CEI initiated its online training program in 2008, aiming to complement and to integrate with in-person education. Since then, CEI developed hundreds of multimedia learning modules, online continuing medical education (CME) and continuing nursing education (CNE) courses, interactive case simulation tools, and other digital resources [4-5]. These resources were disseminated through the web, mobile apps, email newsletters, and online social networks to thousands of clinicians from 170+ countries over the world [6]. The participating clinicians had very positive evaluations of the CEI online training program [7-9].

Compared to traditional in-person continuing professional education, online training is advocated for wide geographical availability, rapid outreach to the target audience, flexibility in resource access and use, as well as cost-efficiency [10-11]. Yet studies to compare online and in-person training directly are scarce and inconclusive.

Here we report a study to compare online and in-person training within the New York State CEI program. The results from this study can help to compare online and in-person cross-disciplinary continuing professional education for HIV, HCV, and STD clinicians, and to guide the future design of education programs for more effective dissemination of clinical evidence to community healthcare providers.

Methods

We conducted the study in a period from April 1, 2015, to June 30, 2015. Among the hundreds of courses and training sessions offered by the CEI program, we selected for analysis only those delivered during the study period in both online and in-person formats to clinicians. For a traditional in-person training session, a clinician in the community first participated in a lecture or workshop delivered by a domain expert, and then logged on to the CEI Student Portal online system to complete a course evaluation. For an online course, a clinician logged on to the CEI Student Portal system, selected a specific course, reviewed the multimedia course materials, and then completed the same course evaluation as for in-person training [9].

For study measures, we focused on the training content, format, participants' improvement in knowledge, and impact on their clinical practice. Specifically, we collected the following self-reported data through course evaluation: (1) usefulness and relevance of the information in a training course; (2) easiness of comprehension; (3) expertise/knowledge of trainer/speaker; (4) appropriateness of training format; (5) participant's knowledge of the clinical topic before and after the training; (6) intention to use the knowledge learned; and (7) plan to change clinical practice after training. We used a five-point Likert-scale measure (*strongly agree*, *agree*, *neutral*, *disagree*, and *strongly disagree*), which were further grouped into positive (*strongly agree* and *agree*) and non-positive (*neutral*, *disagree*, and *strongly disagree*) responses for data analyses on items (1), (2), (3), and (6). For item (5), we used five discrete levels to indicate a clinician's knowledge on a specific training topic (*novice*, *not very knowledgeable*, *knowledgeable*, *very knowledgeable*, and *expert*) before and after the training, and then calculated the difference (*at least one level increase* vs. *no increase*) for data analysis. Additional details of CEI program evaluation can be found in our previous publications [7-9].

We used the SPSS package for statistical analysis [12]. We compared the proportions of positive responses between in-person and online training. We used the chi-square test to examine the statistical significance of differences.

Results

During the study period, four training courses were delivered in dual formats to 250 clinicians (online 131, in-person 119). These four courses were: (1) *STD-HIV Inter-Relationship* (81 completions); (2) *Treatment for Hepatitis C: New Tests, New Drugs & New Recommendations* (91 completions); (3) *Vaginitis* (32 completions); and (4) *The Clinical Diagnosis and Treatment of Syphilis* (46 completions).

Clinicians' course evaluation was positive for both in-person and online training regarding useful information (in-person 96%, online 92%), easiness of comprehension (in-person 91%, online 93%), knowledgeable trainer (in-person 97%, online 90%), and appropriate format (in-person 74%, online 87%). In terms of the impact, we found: (1) 56% clinicians participated in in-person training and 43% in online training reported at least one level increase in knowledge; (2) 87% clinicians in in-person training and 86% in online training intended to use the learned knowledge; and (3) 39% in in-person training and 22% in online training planned to change practice. In-person training recorded better evaluation in terms of knowledgeable trainer ($p=0.04$), improvement of knowledge ($p=0.03$), and change of practice ($p<0.01$). Online training recorded better evaluation in terms of appropriate format ($p=0.01$). No significant difference was found in terms of useful information ($p=0.25$), easy comprehension ($p=0.49$), and intention to use knowledge ($p=0.79$). The detailed evaluation data are shown in Table 1.

Table 1 – Evaluation of In-Person and Online Training

Measures	In-Person (n=119)	Online (n=131)	Total (n=250)
Information Useful and Relevant			
Positive	114 (96%)	121 (92%)	235 (94%)
Non-Positive	5 (4%)	10 (8%)	15 (6%)
Easy to Comprehend			
Positive	108 (91%)	122 (93%)	230 (92%)
Non-Positive	11 (9%)	9 (7%)	20 (8%)
Knowledgeable Trainer*			
Positive	115 (97%)	118 (90%)	233 (93%)
Non-Positive	4 (3%)	13 (10%)	17 (7%)
Format Appropriate*			
Positive	88 (74%)	114 (87%)	202 (81%)
Non-Positive	31 (26%)	17 (13%)	48 (19%)
Improvement of Knowledge*			
Positive	67 (56%)	56 (43%)	123 (49%)
Non-Positive	52 (44%)	75 (57%)	127 (51%)
Intend to Use Knowledge			
Positive	104 (87%)	113 (86%)	217 (87%)
Non-Positive	15 (13%)	18 (14%)	33 (13%)
Will Change Practice*			
Positive	46 (39%)	29 (22%)	75 (30%)
Non-Positive	73 (61%)	102 (78%)	175 (70%)

* Significant difference ($p<0.05$) between in-person and online training detected

Conclusions

Both online and in-person training can effectively disseminate HIV, HCV, and STD clinical evidence to community healthcare providers. Although online training is still not equal to in-person training in certain aspects, it has become a preferred format by clinicians to update their knowledge. Future research is required: (1) to verify the generalizability of the findings from this study; and (2) to combine the features of online and in-person training to better address clinicians' needs and preferences for more effective knowledge dissemination.

Acknowledgements

This work was supported by the Agency for Healthcare Research and Quality (AHRQ) through grant R24 HS022057, and by New York State Department of Health AIDS Institute through contracts C023557, C024882, and C029086. The content is solely the responsibility of the author and does not necessarily represent the official views of the sponsors. The author would like to thank: 1) CEI staff Amneris Luque, Xuan Hung Le, Monica Barbosu, Terry Doll, Matthew Bernhardt, and Tom DellaPorta for their contributions to the program; and 2) AHRQ and New York State program officers Marian James, Beatrice Aladin, Cheryl Smith, Howard Lavigne, Lyn Stevens, and Bruce Agins for their support.

References

[1] E.H. Morrato, T.W. Concannon, P. Meissner, et al, Dissemination and implementation of comparative effectiveness evidence: key informant interviews with Clinical and Translational Science Award institutions. J Comp Eff Res 2 (2013), 185-194.

[2] E.A. Zerhouni, Translational and clinical science--time for a new vision. N Engl J Med 353 (2005), 1621-1623.

[3] The New York State HIV-HCV-STD Clinical Education Initiative. http://www.ceitraining.org. Accessed 18 October 2017.

[4] D. Wang, X.H. Le, A. Luque, Development of digital repositories of multimedia learning modules and interactive case simulation tools for a statewide clinical education program. Proc 6th International Workshop on Knowledge Representation for Health-Care (2014), 145-151.

[5] X.H. Le, A.E. Luque, D. Wang, Assessing the usage of a guideline-driven interactive case simulation tool for insomnia screening and treatment in an HIV clinical education program. Stud Health Technol Inform 192 (2013), 323-327.

[6] D. Wang D, X.H. Le, A.E. Luque, Identifying effective approaches for dissemination of clinical evidence – correlation analyses on promotional activities and usage of a guideline-driven interactive case simulation tool in a statewide HIV-HCV-STD clinical education program. Stud Health Technol Inform 216 (2015), 515-519.

[7] D. Wang, A.E. Luque, Evaluation of a statewide HIV-HCV-STD online clinical education program by healthcare providers – a comparison of nursing and other disciplines. Stud Health Technol Inform 225 (2016), 267-271.

[8] D. Wang, Development and evaluation of a statewide HIV-HCV-STD online clinical education program for primary care providers. In: Shachak A, Borycki E, Reis S (ed) Health Professionals' Education in the Age of Clinical Information Systems, Mobile Computing and Social Networks. Academic Press (2017), 371-397.

[9] D. Wang, M. Abrams, Health care providers' profiles and evaluations of a statewide online education program for dissemination of clinical evidence on HIV, hepatitis C virus, and sexually transmitted disease: cross-sectional study. JMIR Med Educ. 2019;5(1):e10722.

[10] E.C. Buxton, Pharmacists' perception of synchronous versus asynchronous distance learning for continuing education programs. Am J Pharm Educ 78 (2014), 8.

[11] M. Bell, K. MacDougall, Adapting online learning for Canada's Northern public health workforce. Int J Circumpolar Health (2013), 72.

[12] IBM SPSS Software. https://www.ibm.com/analytics/ spss-statistics-software. Accessed 18 October 2018.

Address for correspondence

Dongwen Wang, PhD; Professor of Biomedical Informatics, Arizona State University; 13212 E. Shea Blvd., Scottsdale, AZ 85259; dongwen.wang@asu.edu; 1-480-884-0264.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190757

Comparing Usability of User Interfaces to Collect Family Health History

Amal Ponathil[a], Necmettin Firat Ozkan[a], Jeffrey Bertrand[a], Brandon Welch[b] & Kapil Chalil Madathil[a]

[a]Human-Systems Integration Laboratory, Clemson University, Clemson, SC, USA,
[b]Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA

Abstract

Family health history (FHx) is vital in early detection of genetic diseases. This research studied two different FHx collection interface and compared them based on the IBM CSUQ usability metrics. We found the conversational interface to be significantly better in terms of overall satisfaction, system usefulness, interface quality and information quality than the traditional interface.

Keywords:
Family health history collection, usability, user experience.

Introduction

Family Health History (FHx) play an important role in understanding the risk of individual illnesses, especially those with genetic causes [1–2]. Diagnosis of these diseases at an earlier stage is vital for initiating preventive measures [3]. For efficient and accurate diagnosis, having accurate FHx information is critical. Various methods like emails, phone conversations have been employed by the medical service providers to collect these FHX information. However, Qureshi et al., (2009) portrayed that self-collection FHx tools have certain advantages like improved accuracy and quicker data collection than the traditional methods [4].

Although there are different types of self collection tools, electronic FHx tools have been at the forefront in the recent years. Electronic FHx tools have the advantage of easy editability and the ability to share the information among their family members. This helps in getting accurate information and is a quicker way to collect the information.

Despite all the advantages of the electronic FHx, its usability is crucial since a very diverse user population from different age and education groups will interact with the system [5]. Previous studies have focused on the usefulness and potential difficulties of the FHx tool [6,7]. Wang et al. (2015) developed a virtual assistant tool and compared if this new tool could identify more health conditions compared to the Surgeon General's My Family Health Portrait [8]. However, none of the studies focused on including the diverse age group and/or studied the usability of such interfaces.

In order to bridge this knowledge gap, we tried to understand the effect of the different age group on the usability and interface design for collection of FHx data.

Methods

A total of 54 participants (27 males and 27 females) with a mean age of 44.91 years (SD=22 years) were recruited for this study via email and word-of-mouth. On completion of the study, the participants were compensated with a $20 Walmart gift card.

Apparatus

The study was conducted in a quiet and contolled room on a desktop with a 17.5 inch monitor. Two FHx interfaces analyzed in the study were:

Conversational interface

This interface consisted of a conversational dialog box in the bottom right as the data entry method as shown in the figure 1. In addition, the family pedigree provides a real time update as the family history is entered using the dialog box.

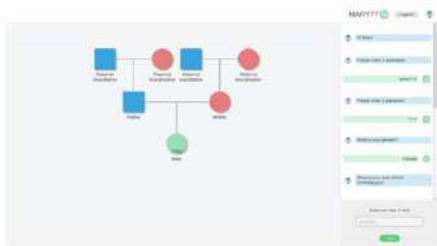


Figure 1.– Conversational interface

Traditional interface

This interface consisted of the traditional method of data entry, i.e. by clicking on the pencil icon (update history column) to enter the family health information as shown in figure 2.

Figure 2.– Traditional interface

Procedure

A mixed experimental design was used with the type of interface being the within subject variable and age group (young adults: 18-30 years, older adults: above 60 years) being the between subject variable. Participants were randomly

assigned to the FHx interfaces and were counterbalanced to avoid order effects. Participants were provided with a fictional family health history information that included the family member's information and past cancer history in the family. Overall, the participants had to complete 5 tasks for each interface: i) create a user profile, ii) add the family health history, iii) re-access the platform, iv) edit the information and v) share the information with a family member. On completion of the tasks using either of the interfaces, the IBM Computer System Usability Questionnaire (CSUQ) was given to the participants to evaluate the usability of the interfaces [9].

Hypotheses

- **H1**: Age group moderates the relationship between the interface design and the overall satisfaction, with an increase in the satisfaction as the interface design changes from traditional to conversational and age group changes from younger to older adults.
- **H2**: Age group moderates the relationship between the interface design and the interface quality, with an increase in the perceived interface quality as the interface design changes from traditional to conversational and age group changes from younger to older adults.

Results

Usability Evaluation

The participants had a higher overall satisfaction on using the conversational interface (M= 5.65, SE=0.12) than the traditional interface (M= 4.77, SE=0.18) and system usefulness score on using the conversational interface (M= 5.79, SE=0.12) than the traditional interface (M=4.82, SE=0.21).

There was a statistically significant interaction between the age group and the interface type on information quality and on interface quality. For the information quality, in the younger age group, the participants reported a higher information quailty score on the conversational interface (M = 5.65, SD = 0.85) than the traditional interface (M = 4.51, SD = 1.30). For the interface quality, in the younger age group, the participants reported a higher interface quality score on the conversational interface (M = 5.79, SD = 0.96) than the traditional interface (M = 4.38, SD = 1.41).

Conclusions

This study looked at the effect of aging on the usability and interface design of the FHx data collection process. Although we did not find an interaction with the overall satisfaction as hypothesized, we did find a significant main effect of the interface design. We found the conversational interface was rated higher on all the usability metrics on the IBM CSUQ scale. Participants were highly satisfied using the interface as they only had to answer the questions asked by the conversational virtual agent to complete the task. Unlike in the traditional interface, where they had to manually search for all the data entry boxes to enter the family health information.

Acknowledgements

This research is supported in part by funding from the National Cancer Institute (5K07CA211786) and the Hollings Cancer Center's Cancer Center Support Grant (P30CA138313) at the Medical University of South Carolina.

References

[1] C.D. Ford, R.N. Rooks, and M. Montgomery, Family health history and future nursing practice: Implications for undergraduate nursing students, *Nurse Educ. Pract.* **21** (2016) 100–103.

[2] C. Hughes Halbert, B. Welch, C. Lynch, G. Magwood, L. Rice, M. Jefferson, and J. Riley, Social determinants of family health history collection, *J. Community Genet.* **7** (2016) 57–64.

[4] E.J. Rutherford, J. Kelly, E.A. Lehane, V. Livingstone, B. Cotter, A. Butt, M.J. O'Sullivan, F. O Connell, H.P. Redmond, and M.A. Corrigan, Health literacy and the perception of risk in a breast cancer family history clinic, *Surgeon*. (2016). doi:10.1016/j.surge.2016.06.003.

[5] N. Qureshi, B. Wilson, P. Santaguida, J. Little, J. Carroll, J. Allanson, and P. Raina, Family history and improving health, *Evid. Rep. Technol. Assess.* . (2009) 1–135.

[6] W.B. Lober, B. Zierler, A. Herbaugh, S.E. Shinstrom, A. Stolyar, E.H. Kim, and Y. Kim, Barriers to the use of a personal health record by an elderly population, *AMIA Annu. Symp. Proc.* (2006) 514–518.

[7] A.A. Ozok, H. Wu, M. Garrido, P.J. Pronovost, and A.P. Gurses, Usability and perceived usefulness of personal health records for preventive health care: A case study focusing on patients' and primary care providers' perspectives, *Appl. Ergon.* **45** (2014) 613–628.

[8] S.J. Czaja, C. Zarcadoolas, W.L. Vaughon, C.C. Lee, M.L. Rockoff, and J. Levy, The usability of electronic personal health record systems for an underserved adult population, *Hum. Factors*. **57** (2015) 491–506.

[9] C. Wang, T. Bickmore, D.J. Bowen, T. Norkunas, M. Campion, H. Cabral, M. Winter, and M. Paasche-Orlow, Acceptability and feasibility of a virtual counselor (VICKY) to collect family health histories, *Genet. Med.* **17** (2015) 822–830.

[10] J.R. Lewis, IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use, *International Journal of Human–Computer Interaction*. **7** (1995) 57–78.

Address for correspondence

Brandon M Welch, MS, PhD
Medical University of South Carolina
Biomedical Informatics Center
135 Cannon Street, 405L
Charleston, SC, USA 29425
welchbm@musc.edu
(843) 518- 3769

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190758

Use of in-Hospital Geomagnetic Fingerprinting Localization

Keiko Yamashita[a], Shintaro Oyama[a], Tomohiro Otani[b], Satoshi Yamashita[a], Taiki Furukawa[a], Daisuke Kobayashi[a,c], Kikue Sato[a], Aki Sugano[a], Chiaki Funada[a], Kensaku Mori[d], Yoshimune Shiratori[a]

[a] Medical IT Center, Nagoya University Hospital, Nagoya, Aichi, Japan,
[b] NTT DATA Corporation Technology Innovation Headquarters, Tokyo, Japan,
[c] Medical System Field of Study, Kobe University Graduate School of Medicine, Kobe, Hyogo, Japan,
[d] Department of Intelligent Systems, Graduate School of Informatics, Nagoya University, Nagoya, Aichi, Japan,

Abstract

Recently, visualizing location of people and things in a hospital has become an issue particularly for improving work efficiency and incident prevention. Although radio frequency waves such as Wi-Fi and Bluetooth are commonly used in indoor positioning, they have several limitations owing to their physical characteristics. We proposed in-hospital hybrid positioning technique, involving a combination of radio waves and geomagnetic fingerprinting techniques. We compared accuracy of proposed technique with that of Wi-Fi- and BLE-based techniques.

Keywords:

Geomagnetic fingerprinting localization, hybrid in-hospital positioning, Wi-Fi and Bluetooth low energy beacon

Introduction

Recently, IoT technologies have influenced various fields to improve operation efficiency and prevent accidents by using smart devices; particularly, in logistics and factory automation fields by visualizing people and objects for 24 hours on a massive scale, they have succeeded in reducing costs, workers' burden, and workplace hazards drastically [1]. Moreover, in the medical field, issues regarding increase in medical care expense and shortage of medical staff, accompanied by rapid progression of the aging problem increased urgency to reduce labor costs and burden on medical staff without increasing medical incidents. Visualization of location is a key technology in this cross-sectoral innovation. For example, in factories, reducing movement of workers by a single step can improve efficiency due to reduced time wastage and mistakes. Furthermore, by introducing smart devices such as smartwatches, some companies have succeeded in reducing workers' heatstroke accidents. Generally, localization methods are divided into two ways: outdoor and indoor. Most outdoor positioning systems are represented using a global positioning system (GPS). However, indoor positioning systems do not have a *de facto* standard method. Generally, radio frequency (RF) waves such as Wi-Fi, bluetooth low energy (BLE), indoor messaging system (IMES), and indoor GPS are used as reference points for position estimation. However, these RF based techniques have accuracy limitation due to signal attenuation, radio interference, and hardware variability (standard compliance) problem.

In Nagoya university hospital, we have been considering ways to localize smartphones that are used in nursing service. Use of Wi-Fi access points and BLE beacons were rejected due to saturation of 2GHz band radio waves in the ward. Hence, IMES was incompatible with the smartphones. Thus, we decided to introduce geomagnetic field pattern matching developed by GiPStech for device localization. Their indoor positioning technology won GeoIoT World for Indoor Localization & Proximity Services, 2016 award [2]. The position estimation strategy based on geomagnetism uses pattern matching of a 3D magnetic vector field map. Current position can be calculated by comparing continuous magnetic vector measurements using particle filter algorithm. The advantage of this method is that although it requires certain number of cornerstones such as BLE beacons to stabilize accuracy, it can be used in places without a good number of beacons. It can cover a wide area seamlessly and at less maintenance cost. However, there were no preceding reports on use of geomagnetic localization in the hospital. Thus, in this study, we compared accuracy of in-hospital geomagnetic field fingerprinting with Wi-Fi and BLE beacon techniques.

Methods

Comparison of position measurement accuracy experiment using geomagnetic field (GM) only, a hybrid of GM/Wi-Fi, and GM/BLE was performed in general inpatient ward (8th floor, 52 beds). As shown in Fig.1, we set up seven test-paths that cover the ward. Node was used when setting up test path on a system for creation of correct answer data of position information set pillars, which helped to easily determine accurate position.

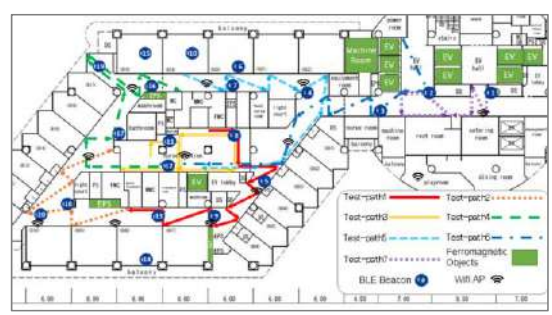


Figure 1-Seven test-paths settled in the 8th floor ward.

Before examination, magnetic field information was pre-logged and stored as geomagnetic field map in a database server in advance. For Wi-Fi positioning, we used 12 existing wireless access points in the ward as a cornerstone. With transmitter power control turned off, we obtained maximum Wi-Fi radio field intensity (RSSI) and MAC address of each access points. RSSI and multilateration techniques were used for location measurement [3]. For BLE positioning, we installed 20 beacons

in the ward and registered their absolute positions. Beacons were installed in the corridor and nurses' station on a wall at 2.00m high. Based on bluetooth radio wave intensity received from multiple beacons, its position was measured using three-point positioning method.

Next, we verified measurement using the most accurate positioning method in the experiment. We set three routes, assuming medical staffs' movement, and made positioning. To prepare correct answer data of position information, we attached the barcode to the wall of the hospital rooms' doorway. We obtained barcode image using smartphone camera and matched that information with actual position and timing. We chose the hospital room selected during verification as a general hospital room in a quadruple room.

Results

The overall average positioning errors in GM only, GM/Wi-Fi, and GM/BLE were 7.62m, 3.19m, and 2.60m respectively. The most accurate method was GM/BLE used in three routes assuming medical staffs' actual movement.

Table 1-GM only, GM/Wi-Fi, and GM/BLE positioning errors (Meter, Average ± Standard Deviation) of 7 Routes

Route	GM only	GM/Wi-Fi	GM/BLE
#1	5.74 ± 7.03	2.20 ± 1.82	2.72 ± 2.55
#2	5.58 ± 6.43	1.93 ± 1.03	1.76 ± 1.18
#3	6.28 ± 9.90	2.25 ± 2.83	2.21 ± 2.93
#4	0.98 ± 0.53	2.75 ± 2.28	1.27 ± 0.98
#5	3.88 ± 7.20	6.99 ± 2.81	1.59 ± 1.41
#6	18.10 ± 15.48	1.18 ± 0.78	2.61 ± 2.64
#7	12.41 ± 10.45	4.84 ± 3.13	6.24 ± 4.39

Hospital rooms without barcode during verification were excluded from accuracy calculation of the experiment. The hospital room selected by this route used common room of bay for verification. The average positioning error was 1.20m.

Table 2-Result of GM and BLE beacon hybrid method, of path 1-3

	Absolute Error (m)		
#Node	Route1	Route2	Route3
#1	1.76	0.51	0.94
#2	1.31	1.00	0.68
#3	0.28	0.73	1.52
#4	0.74	1.39	0.52
#5	2.10	1.57	0.80
#6	0.72	0.70	1.24
#7	1.04	0.68	2.53
#8	2.25	0.87	0.91
#9	0.68	0.80	0.66
#10	0	1.69	1.41
#11	1.46	0.30	1.40
#12	0	1.68	1.55
#13	-	1.46	0.57
#14	-	0	0.90
#15	-	2.21	1.40
#16	-	0	0
#17	-	-	2.20
#18	-	-	0
Average	1.24	1.11	1.20
SD	0.73	0.62	0.65

Discussion

Regarding positioning accuracy in Nagoya university hospital, we believe that positioning accuracy of about 3.00m is sufficient to grasp identity of person being treated because the area per patient of a four-person room is 3.00m^2. For radio frequency beacons placement, most of the positioning method places beacons in polygons at short intervals every few meters. For example, it requires to place 5 beacons in 5.00 m^2 area to obtain an average of 1.20m positioning accuracy inside a hospital [4]. Notably, the positioning of beacons is very important when measuring positioning accuracy. Thus, it is often common practice to examine placement position several times. However, in the verification performed, we only examined beacons placement once. The beacons were arranged at average intervals of 10.00m. We did not verify accuracy of Wi-Fi only nor BLE only methods, these would be assumed to be very low. According to the relationship of distance of mean error and number of beacons by Li, estimated mean error of BLE only method would be much worse than 2.33m with beacon density of this experiment [4]. Concerning the result, any magnetic substances such as pipes and communication cables in wall of the hospital building helped form a characteristic magnetic field space, which is easy to extract. Conversely, the magnetic field can be influenced by ferromagnetic objects such as an elevator (EV in Fig.1) and electric power space (EPS in Fig.1) resulting in loss of precision as seen in route #5/6/7. We recommend using several beacons when using GM method in-hospital and believe that adequate beacon settings in high magnetic field fluctuated area needs to be further verified when using GM/Beacon hybrid method.

Conclusions

Hybrid localization using geomagnetic (GM) fingerprinting and Bluetooth low energy (BLE) beacon were found to be accurate in conducting in-hospital tracking in this study. The experimental results show that although a small number of BLE beacons were placed and arranged in position without careful planning, a good position accuracy value was obtained. Therefore, it can be said that the position measurement method using geomagnetism is an effective method for indoor position measurement in a hospital.

Acknowledgements

I am deeply grateful to all the staffs in Nagoya university hospital for their assistance with our study.

References

[1] H. Liu, H. Darabi, P. Banerjee, and J. Liu. Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews* **37** (2007), 1067-1080.

[2] GiPStech. Indoor localization that works. http://www.gipstech.com/ (cited 2019-March-31).

[3] J. Yang and Y. Chen. Indoor localization using improved RSS-based lateration methods. *GLOBECOM 2009 - 2009 IEEE Global Telecommunications Conference* (2009), 1-6.

[4] Z. Li, Y. Yang, and K. Pahlavan. Using iBeacon for newborns localization in hospitals. In: *2016 10th International Symposium on Medical Information and Communication Technology*, *IEEE* (2016), 1-5.

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190759

Making Sense of Clinical Laboratory Results: An Analysis of Questions and Replies in a Social Q&A Community

Zhan Zhang[a], Yu Lu[a], Caleb Wilson[b], Zhe He[b]

[a] *Department of Information Technology, Pace University, New York, New York, USA,*
[b] *School of Information, Florida State University, Tallahassee, Florida, USA,*

Abstract

Previous literature has revealed that patients have difficulty making sense of their clinical data. To address this barrier, we first need to understand what kinds of support that patients may seek. In this study, we analyzed question posts and their replies in a social Q&A community to understand what types of support people are providing to and receiving from the community and what contextual information they provide in order to elicit relevant answers.

Keywords:
Consumer Health Information; Information Seeking Behavior

Introduction

Growing evidence suggests that patients are increasingly interested in timely and easy access to their clinical data, such as laboratory test results [1]. With the wide adoption of personal health record technology, such as online patient portals, healthcare organizations are enabling patients' easy access to their clinical data. However, patients' current use of clinical data is significantly limited due to the techical nature of the clinical report, preventing patients from making informed health decisions based upon it [2]. Our ultimate research goal is to address this barrier by designing and developing novel informatics tools to help patients better understand their clinical data and make personalized decisions. To achieve this goal, we first need to understand what kinds of support that patients may seek.

There is an ongoing trend for patients turning to online resources, such as social Q&A sites and health forums, to fill their knowledge gaps [3]. These online communities provide us an opportunity to examine consumer information searching and types of information that people need concerning a particular disease and their querying behavior in online health communities [4]. In our recent study, we analyzed the questions that people posted on a major social Q&A site, Yahoo! Answers, to understand lay people's information needs in making sense of their laboratory test results, one type of clinical data. In this paper, we take our analyses a step further to answer two research questions: 1) what kinds of *contextual information* people provide in the question in order to elicit relevant answers, and 2) what types of support people receive from the online healthcare community. We believe that this could provide a foundation for our future work to design and develop informatics tools to provide more personalized information that better meets patients' needs. Towards this end, we further analyzed the question posts identified through our previous work and report the methodologies and results below.

Methods

In our previous work, we collected a total of 58,422 questions and their threaded replies in the diabetes category of Yahoo! Answers between 2009 and 2014. We retrieved potentially relevant posts in this dataset using keywords suggested by the literature [5], such as "blood sugar", "glucose", and "HbA1c". Then, we took a random sample of 1619 posts containing keywords. Two researchers independently reviewed posts for relevance. Duplicate and irrelevant posts were discarded. This screening resulted in 967 posts eligible for further analysis.

Two researchers independently performed content analysis on the relevant 967 posts. We first iteratively developed a codebook for the kinds of contexual information provided in the posts, using the open coding technique. Two coders, C1 and C2, independently analyzed 100 randomly sampled posts and created the initial list of codes. Researchers then discussed the codes in a group session to determine which codes to keep, merge, or remove. After the list of codes was set, we created a data dictionary, defining each code to standardize the coding process. Next, C1 and C2 independently coded another set of 100 posts to test inter-rater agreement (90%). Once resolving all disagreements, C1 and C2 coded the rest of the posts to conclude the analysis. When reviewing the replies to relevant question posts, we focused on the "best answer", which was selected by the person posting the original question. C1 and C2 followed the same procedure to review the replies.

Results

Types of Information Provided by Askers

Askers provided a variety of information in the question to effectively communicate their current conditions and informational needs to the community members. Table 1 shows the types of information that askers provided in their questions and the frequency of being provided in the posts (if a particular type of information was provided more than once in a post, it was considered as appearing one time in that post). As shown in the table, askers provided mainly two types of information, demographic and medical information, to contextualize their questions.

Major demographic information provided in the questions include age, gender, height, and weight. Less frequently, askers provided pregnancy information to help others better

interpret their problems: "*What does it mean if I am pregnant and my homocysteine level is low, particularly 4.9.*"

Furthermore, askers provided many different types of medical information in their questions, including lab tests, symptoms, medication, medical and family history, doctor's diagnosis, and lifestyle. More specifically, to help others interpret their lab test results, askers provided readings from lab tests. Askers also provided information about their symptoms to help others interpret their problems. Some askers went further to provide information about their personal medical hisotry or family medical history to ask the community for opinions.

Some askers also mentioned medications that they are currently using: "*My dosage now is 112mcg […] My doctor suggested that I need to lower my dosage. Need help understanding, I am very confused*." Lastly, a great number of askers also listed information about their lifestyle, including diet, exercise routines, drink, and smoke.

Table 1– Types of Information Provided in the Questions

	Provided Information	Frequency
Demographic	Age	148
	Gender	105
	Height	28
	Weight	67
	Pregnancy	16
Medical	Lab tests	678
	Symptoms	167
	Medication	84
	Medical history	68
	Family history	47
	Doctor's diagnosis	70
	Lifestyle	175

Supports Received from the Online Community

The contextual information provided by askers in their questions helped community members to provide relevant answers. The answers addressed the original question by providing suggestions, opinions, information, personal experience, and emotional support. They also requested more information from the question poster.

Suggestion: The community mainly provided suggestions about the next step, medication, and lifestyle. Many replies suggested that askers should see a doctor, get a second opinion, and discuss their questions/concerns with their physician. Some answers also recommended additional tests that the asker may need to take. Furthermore, based on the provided information, answers also suggested treatment and medication (e.g., "*Metformin to start. Never ever take Actos or Avandia. They may kill you*"). Finally, the community emphasized the importance of having a healthier lifestyle.

Information: The community provided information on a number of topics, including tests, symptoms, treatments, and resources. For example, answers explained genernal information about about a particular lab test, such as normal value range, test options for evaluating a disease, and lab procedure. Answers also provided information on common symptoms associated with a particular disease, and potential risks and common causes of these symptoms. Furthermore, community members provided information on treatment and medication, including general treatment approach for a given condition, treatment risk, medication side-effects, and general knowledge about a particular treatment/medication. Finally, it is interesting to see that answers would point the asker to external websites for further information.

Opinion: Furthermore, community memebers provided their personal opinions and interpretations regarding test results (e.g., "*…indicate that you are at the pre-diabetes stage of developing type 2 diabetes*"), prognosis (e.g., "*The high glucose is likely due to the pneumonia*"), and treatment (e.g., "*That is a terrible drug*").

Personal Experience: Sometimes, answers shared personal experiences to offer emotional support or to provide "assured" information to support their opinions or suggestions.

Request More Information: Finally, answers often asked follow-up questions and requested more information so that they would be able to help more.

Conclusions

This study is part of a large research effort supportting patients to better understand and act on their clinical data. This preliminary work provides a foundation for our future work. For example, we will use these results to design surveys and interview protocols to further understand the challenges that lay individuals have in comprehending clinical data and to guide the design of consumer-facing informatics tools to better present data of varying technical complexity to lay individuals. Furthermore, our findings revealed that members provide both objective and subjective information to the community. This observation highlights the importance of developing mechanisms to address quality of online health information. Lastly, our future work will expand to other health conditions (e.g., cancer) and health forums and include other types of health information (e.g., radiology report) to assess the generalizability of these findings.

References

[1] F. Pillemer, R.A. Price, S. Paone, G.D. Martich, S. Albert, L. Haidari, G. Updike, R. Rudin, D. Liu, and A. Mehrotra, Direct release of test results to patients increases patient engagement and utilization of care. *PLoS One* 11 (2016), 1–9.

[2] M.K. Hong, C. Feustel, M. Agnihotri, M. Silverman, S.F. Simoneaux, and L. Wilcox, *Supporting families in reviewing and communicating about radiology imaging studies*. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, ACM, New York (2017), 5245-5256.

[3] K. Zickuhr, Generations 2010. Pew Internet and American Life Project (2010).

[4] Y. Zhang, Contextualizing consumer health information searching: an analysis of questions in a social Q&A community. In Proceedings of the 1st ACM International Health Informatics Symposium, ACM, New York (2010), 210-219.

[5] D.B. Sacks, M. Arnold, G.L. Bakris, et al. Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. *Diabetes Care* 34 (2011): e61-99.

Address for correspondence

Zhe He, PhD. School of Information, Florida State University, 142 Collegiate Loop, Tallahassee, Florida, USA. Email: zhe@fsu.edu. Phone: 001(850)644-5775

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190760

Exploring Lung Cancer Screening Discussions on Twitter

Yunpeng Zhao[a], Jinhai Huo[b], Mattia Prosperi[c], Yi Guo[a], Yongqiu Li[d], Jiang Bian[a]

[a] Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, Florida, USA
[b]Department of Health Services Research, Management and Policy, University of Florida, Gainesville, Florida, USA
[c]Department of Epidemiology, University of Florida, Gainesville, Florida, USA
[d]Operations and Decision Technologies, Kelley School of Business, Indiana University Bloomington, Bloomington, Indiana, USA

Abstract

Lung cancer is the leading cause of cancer-related death in the United States. Low-dose computed tomography (LDCT) for lung cancer screening (LCS) can reduce lung cancer deaths by 20% compared to chest x-rays. Nevertheless, the uptake of LDCT is low for high-risk smokers. In this study we used Twitter data to understand laypeople's emotions towards LCS, find topics on LCS-related tweets, and assess the impact of promotional information on laypeople's discussions.

Keywords:

Social media, early detection of cancer, lung neoplasms

Introduction

Lung cancer is the leading cause of cancer-related death in the United States. In 2018 it was estimated that 234,030 new cases of carcinoma of the lung were diagnosed and that 154,050 patients died from this disease [1]. Research has shown that screening for lung cancer with low-dose computed tomography (LDCT) reduces deaths from lung cancer by 20% compared to chest x-rays [2]. However, our recent study showed the uptake of LDCT for lung cancer screening (LCS) remains low and there was inappropriate use of LCS among individuals who were ineligible in 2015 [3].

Social media is more than a network to connect or build virtual relations with friends. It reignited research into studying public health-related topics using user generated health data. In our previous study, we used Twitter data to understand the impact of promotional information on laypeople's discussions in the case of Lynch syndrome [4]. In this study, we used a similar approach to answer 3 research questions (RQs): 1) what are laypeople's emotions towards LCS; 2) what are common topics associated with LCS-related tweets; and 3) what is the impact of promotional information on laypeople's discussions.

Methods

Our approach, as shown in Figure 1, started with collecting tweets from two different data sources by using a list of LCS-related keywords. We then classified tweets into promotional information vs. laypeople's discussions, assessed laypeople's emotions and attitudes towards LCS when they discuss LCS on Twitter, and compared the topics addressed in promotional information and laypeople's discussions.

Figure 1– The study workflow.

Step 1: Data Collection

Our data were queried from two different sources. First, we collected tweets related to LCS from December 9, 2017 to October 26, 2018 using a Twitter crawler based on a set of keywords related to LCS (e.g., "low-dose ct"). The list of LCS-related keywords was developed through a snowball sampling process, where we iteratively queried our collection of Twitter data and randomly selected sample tweets to discover new LCS-related keywords until no new keywords were found. We also used the same list of keywords to extract tweets from our random sample database, which was collected using the Twitter streaming application programming interface (API) from January 1, 2013 to December 30, 2017.

Step 2: Data Preprocessing

We preprocessed the collected data to eliminate tweets that were duplicates or not in English. We also geocoded the tweets using a tool we developed in our previous study [5].

Step 3: Tweet Classification

We annotated 1,121 tweets to create training data, and then experimented with two deep learning methods, convolutional neural networks (CNN) and long-short term memory (LSTM) networks, to automatically classify the tweets into promotional information and laypeople's discussions groups. This method classified 621 of the training tweets as promotional information and 400 as laypeople's discussions.

Step 4: Sentiment Analysis

The Linguistic Inquiry and Word Count (LIWC) [6] is a validated text analysis tool which can assess indivdiuals' emotions through counting the percentage of emotional words used in a given text. We applied the LIWC tool on all of the tweets in the laypeople's discussion cateogry to assess their emotion scores (i.e., positive emotion, negative emotion, anxiety, anger, and sadness).

Step 5: Topic Modeling

Topic modeling is a statistical, unsupervised model that can discover abstract topics in a collection of documents. We used the Biterm algorithm [7], which has shown superior performance compared with the commonly used Latent Dirichlet allocation (LDA) model on short texts, to find the abstract latent topics presented in the overall Twitter data. We then used the trained model to infer topics for each tweet.

Results

Data Collection

Overall we collected 29,775 tweets and retained 25,576 for analysis after excluding duplicates and non-English tweets.

Tweet Classification

The CNN model outperformed the LSTM classifer on the training data. Using the trained CNN model, 25,576 tweets were classified as promotional information and 16,360 tweets as laypeople's discussions (from 9,216 unique individuals).

Sentiment Analysis

For RQ1, we applied LIWC analysis on tweets of laypeople's discussions. As shown in Figure 2, laypeople's emotion towards LCS fluctuated across the years.

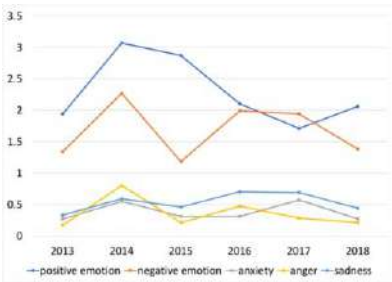


Figure 2– Laypeople's emotion changes towards LCS by year.

Topic modeling

The Biterm topic model was trained on all LCS-related tweets to extract 100 topics. The topic with the highest probability produced by the Biterm model was selected for each tweet. The top 6 topics are shown in Figure 3.

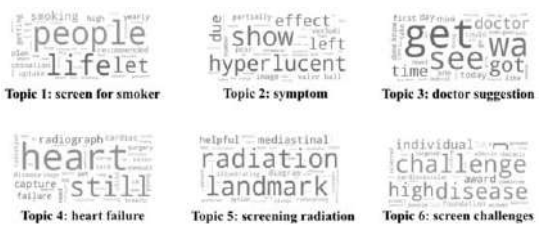


Figure 3– Top topics among lung cancer screening-related tweets in both promotional information and laypeople's discussions.

For RQ2, we calculated the top 5 topic distributions on promotional information compared to laypeople's discussions. As shown in Figure 4, topics in laypeopole's discussions are similar to the topics in promotional information, confirming the impact of promtional information on laypeople's discussions in Twitter.

To answer RQ3, we calculated the Pearson correlations of different topics between promotional information and laypeople's discussions based on the monthly tweet volume as shown in Table 1. There was a high correlation on the topic of "screen challenges" and moderate correlations on the topics of "heart failure" and "screen radiation". In other words, promotional information has an impact on laypeople's discussions related to "screen challenges", "heart failure", and "screen radiation".

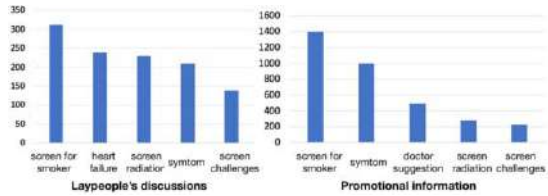


Figure 4– Top 5 topics between promotional information and laypeople's discussions related to lung cancer screening.

Table 1– Pearson correlation coefficients of different topics between promotional information and laypeople's discussions based on their monthly tweet volumes.

Topic	Correlation coefficient	p-value
heart failure	0.67	0.03
screen challenges	0.70	0.02
screen radiation	0.66	0.03

Conclusions

Our results demonstrated that using natural language processing tools and machine learning models can be useful to assess people's emotions towards a health-related topic, discover themes in their discussions, and understand the correlation between promotional information and laypeople's discussions.

Acknowledgements

The study was supported by the University of Florida Health Cancer Center (UFHCC) Research Pilot Grant through the Florida Consortium of National Cancer Institute Centers Program at the UF (Dr. Huo and Dr. Bian) and in part by NSF Award #1734134 and NIH UL1TR001427. The content is solely the responsibility of the authors and does not necessarily represent the official views of the sponsors.

References

[1] R.L. Siegel, K.D. Miller, and A. Jemal, Cancer statistics, 2018, *CA. Cancer J. Clin.* 68 (2018) 7–30. doi:10.3322/caac.21442.

[2] The National Lung Screening Trial Research Team, Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening, *N. Engl. J. Med.* 365 (2011) 395–409. doi:10.1056/NEJMoa1102873.

[3] J. Huo, C. Shen, R.J. Volk, and Y.-C.T. Shih, Use of CT and Chest Radiography for Lung Cancer Screening Before and After Publication of Screening Guidelines: Intended and Unintended Uptake, *JAMA Intern. Med.* 177 (2017) 439. doi:10.1001/jamainternmed.2016.9016.

[4] J. Bian, Y. Zhao, R.G. Salloum, Y. Guo, M. Wang, M. Prosperi, H. Zhang, X. Du, L.J. Ramirez-Diaz, Z. He, and Y. Sun, Using Social Media Data to Understand the Impact of Promotional Information on Laypeople's Discussions: A Case Study of Lynch Syndrome, *J. Med. Internet Res.* 19 (2017) e414. doi:10.2196/jmir.9266.

[5] A. Hicks, W.R. Hogan, M. Rutherford, B. Malin, M. Xie, C. Fellbaum, Z. Yin, D. Fabbri, J. Hanna, and J. Bian, Mining Twitter as a First Step toward Assessing the Adequacy of Gender Identification Terms on Intake Forms, *AMIA Annu. Symp. Proc. AMIA Symp.* 2015 (2015) 611–620.

[6] J.W. Pennebaker, R.L. Boyd, K. Jordan, and K. Blackburn, The Development and Psychometric Properties of LIWC2015, (2015). https://repositories.lib.utexas.edu/handle/2152/31333 (accessed November 9, 2018).

[7] X. Yan, J. Guo, Y. Lan, and X. Cheng, A biterm topic model for short texts, *Proceedings of the 22nd international conference on World Wide Web*, 2013 (2013) 1445-1456.

Address for correspondence
Jiang Bian
Email: bianjiang@ufl.edu

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

Subject Index

This page intentionally left blank

MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

Author Index

C

D

H

M

T

This page intentionally left blank